Electronic energy optimisation in ONETEP

Chris-Kriton Skylaris

c.skylaris@soton.ac.uk

UNIVERSITY OF Southampton
School of Chemistry

Outline

1. Kohn-Sham DFT calculations

   - Direct energy minimisation versus density mixing

   - Using the density matrix to achieve linear-scaling

2. ONETEP scheme: density kernel and NGWFs

   - Density matrix idempotency and normalisation conditions

3. Linear-scaling functionals for density matrix optimisation

   - Penalty, LNV, PM

4. Optimisation of density kernel

   - Density kernel gradients, tensor properties

5. Optimisation of NGWFs

   - NGWF gradients, preconditioning schemes

6. Overall ONETEP calculation scheme

   - Initialisation of density kernel and NGWFs

   - Flowchart, input keywords
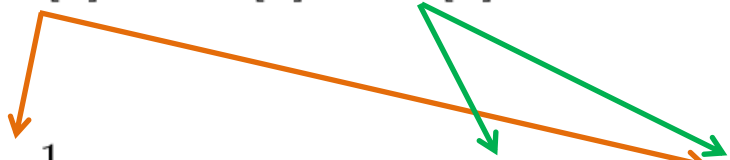
Density Functional Theory (DFT)

- Electronic density

$$n(\mathbf{r_1}) = N \int \cdots \int \Psi(\mathbf{r}_1 s_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) \Psi^*(\mathbf{r}_1 s_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) ds_1 d\mathbf{x}_2 \cdots \mathbf{x}_N$$

- Hohenberg-Kohn

$$E[n] = E_{\mathrm{Kin}}[n] + E_{\mathrm{ext}}[n] + E_{ee}[n]$$

- Kohn-Sham

$$E[n] = \sum_i \langle \psi_i | -\frac{1}{2}\nabla^2 | \psi_i \rangle + E_{\mathrm{ext}}[n] + E_{\mathrm{Coul}}[n] + E_{xc}[n]$$

Non-interacting electrons

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_{\mathrm{eff}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r})$$

Density of interacting electrons

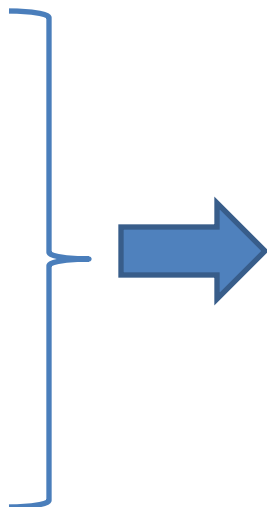$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$$

Self-consistency

- Variational principle

$$\frac{\delta E[\{\psi_i\}]}{\delta \psi_j} = 0$$

- Constraint

$$\langle \psi_k | \psi_n \rangle = \delta_{kn}$$

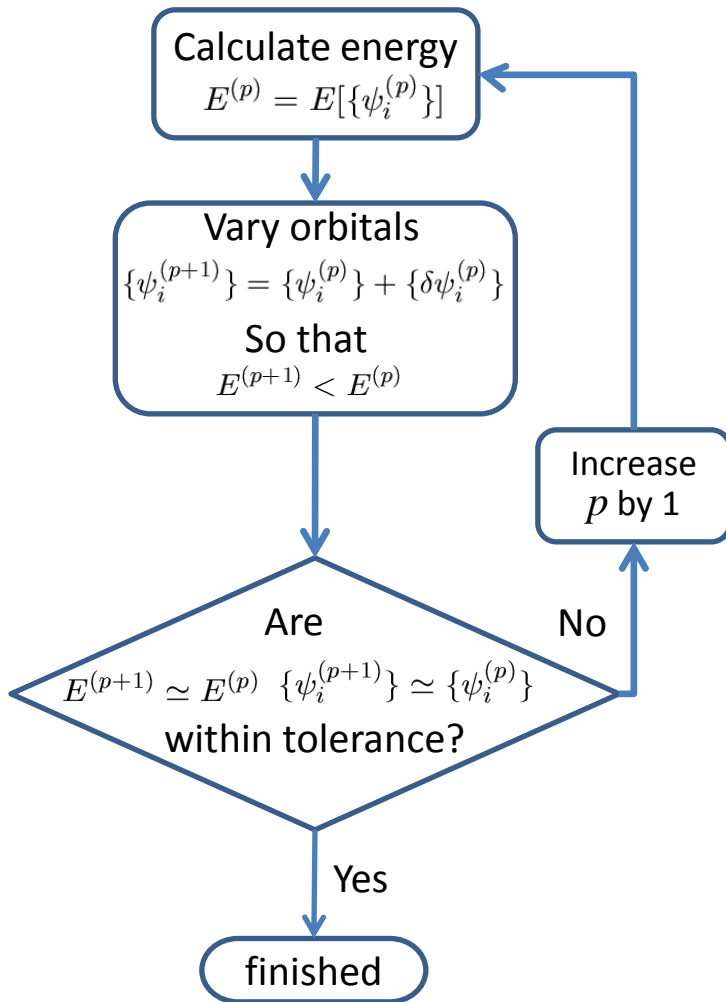- Eigenvalue equation for molecular orbitals (bands)

$$\hat{H} | \psi_n \rangle = \varepsilon_n | \psi_n \rangle$$

- But the Hamiltonian operator depends on the molecular orbitals

$$\hat{H}[\{\psi_i^{(p)}\}] \, |\psi_n^{(p+1)}\rangle = \varepsilon_n^{(p+1)} |\psi_n^{(p+1)}\rangle$$
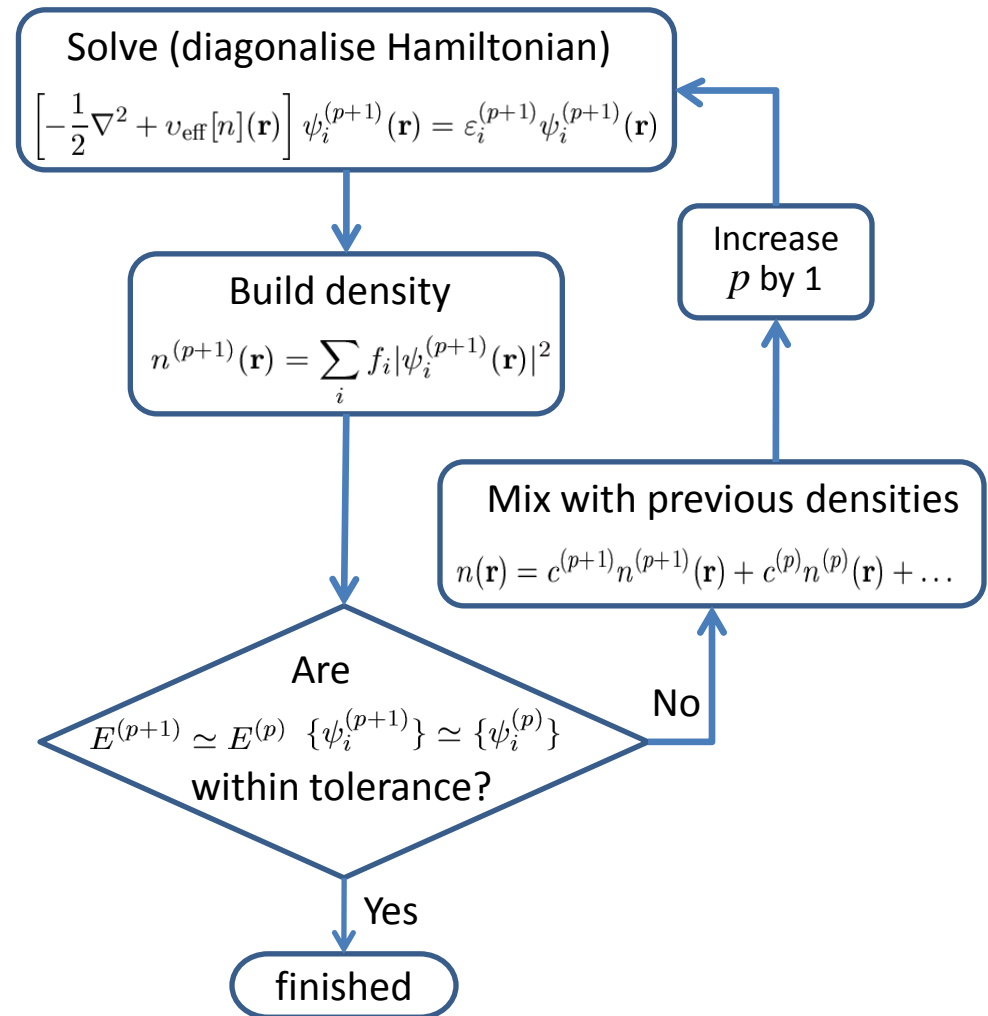
Achieving self-consistency

Direct energy minimisation

Calculate energy
$$E^{(p)} = E[\{\psi_i^{(p)}\}]$$

Vary orbitals
$$\{\psi_i^{(p+1)}\} = \{\psi_i^{(p)}\} + \{\delta\psi_i^{(p)}\}$$
So that
$$E^{(p+1)} < E^{(p)}$$

Increase p **by 1**

Are $E^{(p+1)} \simeq E^{(p)}$ $\{\psi_i^{(p+1)}\} \simeq \{\psi_i^{(p)}\}$ **within tolerance?** — **No**

Yes

finished

- Direct search for well-defined minimum

Diagonalisation and density mixing

Solve (diagonalise Hamiltonian)
$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}[n](\mathbf{r}) \right] \psi_i^{(p+1)}(\mathbf{r}) = \varepsilon_i^{(p+1)} \psi_i^{(p+1)}(\mathbf{r})$$

Build density
$$n^{(p+1)}(\mathbf{r}) = \sum_i f_i |\psi_i^{(p+1)}(\mathbf{r})|^2$$

Increase p **by 1**

Mix with previous densities
$$n(\mathbf{r}) = c^{(p+1)}n^{(p+1)}(\mathbf{r}) + c^{(p)}n^{(p)}(\mathbf{r}) + \dots$$

Are $E^{(p+1)} \simeq E^{(p)}$ $\{\psi_i^{(p+1)}\} \simeq \{\psi_i^{(p)}\}$ **within tolerance?** — **No**
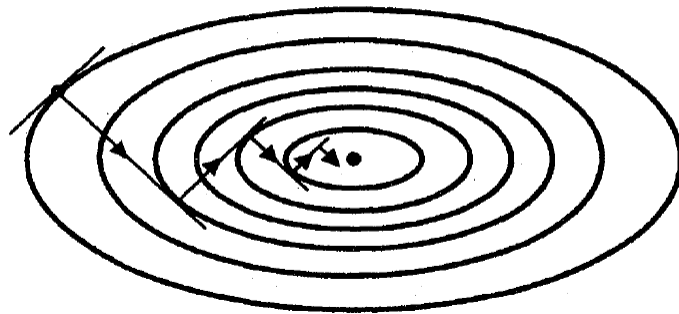
Yes

finished

- Discontinuous changes
- Hamiltonian search (indirect)

Function minimisation
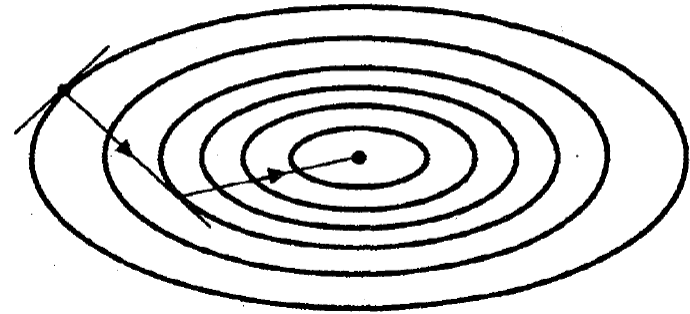
- Iterative procedures
- Need the value of the function and its gradient at each step
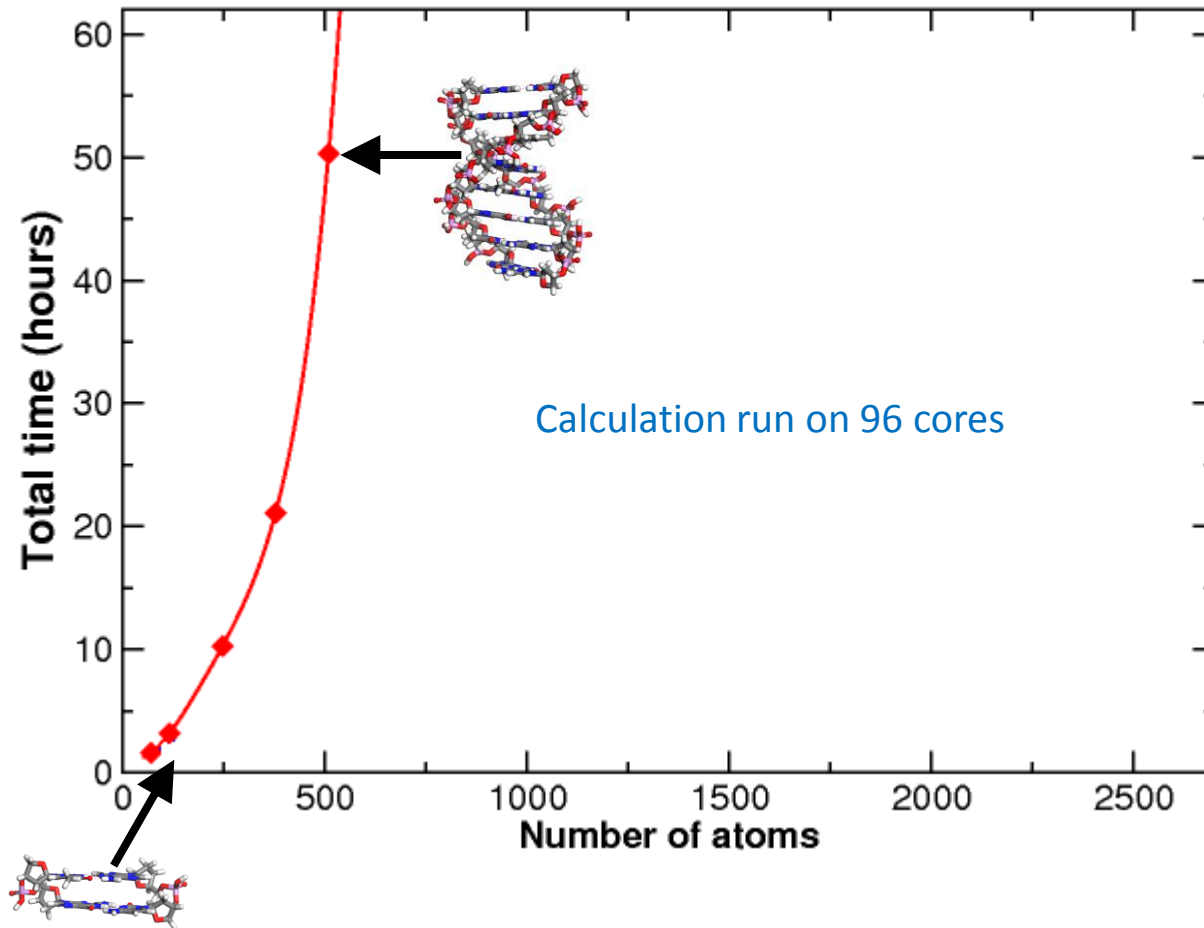- Converge to local minima

STEEPEST DESCENTS

CONJUGATE GRADIENT

- Moves always downhill
- Robust but may need very large number of iterations
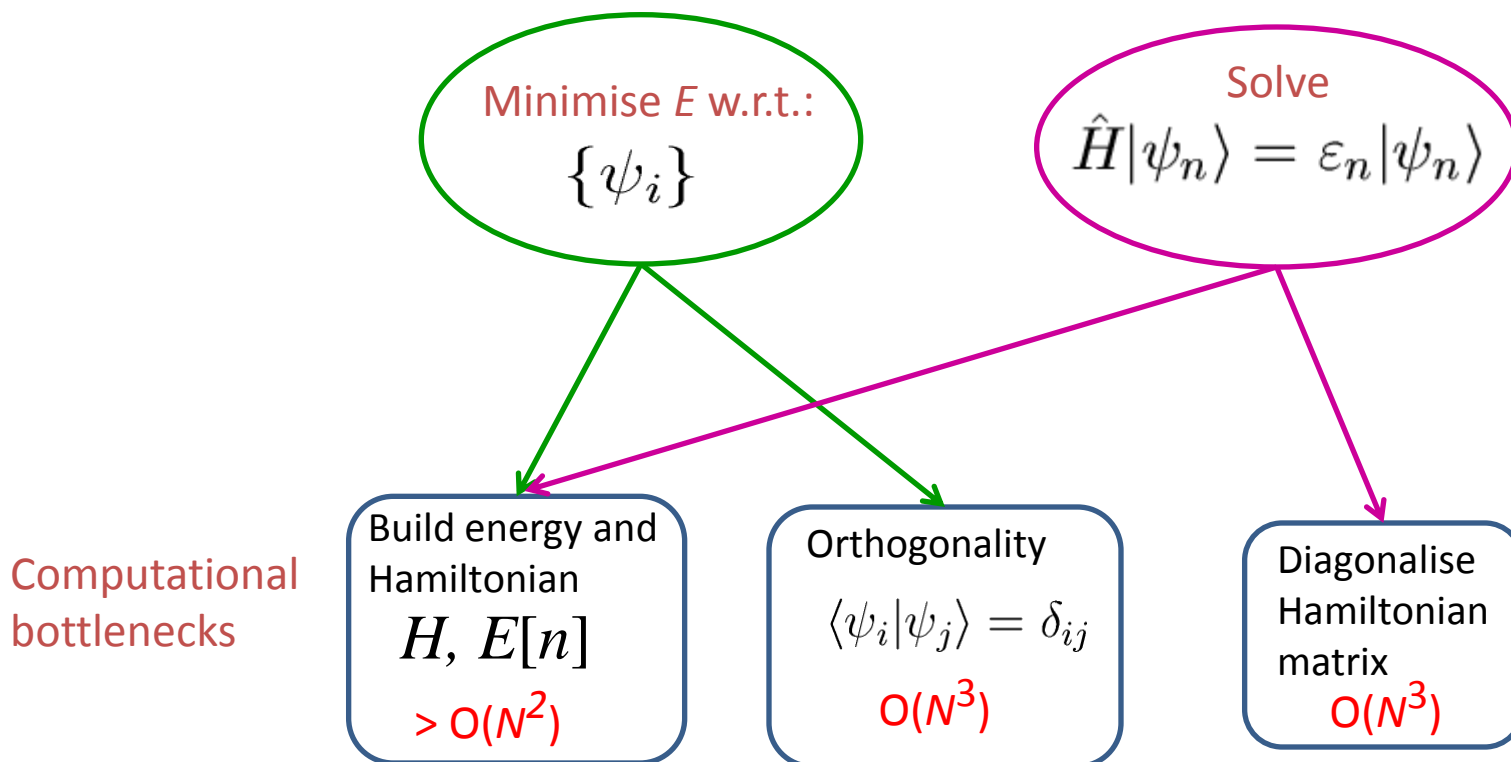
- Converges in N steps for N-dimensional quadratic function
- In practice very efficient even for non-quadratic functions

Cubic-scaling computational cost of DFT

Calculation run on 96 cores

Kohn-Sham DFT calculation bottlenecks

Minimise E w.r.t.:
$$\{\psi_i\}$$

Solve
$$\hat{H}|\psi_n\rangle = \varepsilon_n|\psi_n\rangle$$

Computational bottlenecks

Build energy and Hamiltonian
$$H, E[n]$$
$> O(N^2)$

Orthogonality
$$\langle\psi_i|\psi_j\rangle = \delta_{ij}$$
$O(N^3)$

Diagonalise Hamiltonian matrix
$O(N^3)$

Density matrix localisation

Nearsightedness of electronic matter
W. Kohn, *Phys. Rev. Lett*. **76**, 3168 (1996); E. Prodan and W. Kohn, *P.N.A.S.* **102** 11635 (2005)

In systems with a band gap:

$$\rho(\mathbf{r}, \mathbf{r}') \sim e^{-\gamma|\mathbf{r}-\mathbf{r}'|} \to 0 \quad \text{as} \quad |\mathbf{r} - \mathbf{r}'| \to \infty$$

Take advantage of this locality condition to make the density matrix contain a **linear** amount of data

Truncate exponential "tail", impose:

$$\rho(\mathbf{r}, \mathbf{r}') = 0 \quad \text{when} \quad |\mathbf{r} - \mathbf{r}'| > r_{\text{cut}}$$

One-particle density matrix in DFT

- Operator representation

$$\hat{\rho} = \sum_n f_n |\psi_n\rangle \langle \psi_n|$$

- Position representation

$$\langle \mathbf{r}|\hat{\rho}|\mathbf{r}'\rangle = \rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}')$$

- Conditions
  - Idempotency (from **orbital orthonormality and occupancies 1 or 0**)

$$\rho^2(\mathbf{r}, \mathbf{r}') = \int \rho(\mathbf{r}, \mathbf{r}'')\rho(\mathbf{r}'', \mathbf{r}')d\mathbf{r}'' = \sum_n f_n^2 \psi_n(\mathbf{r})\psi_n^*(\mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r})\psi_n^*(\mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}')$$

  - Normalisation (preserving the **number of electrons**)

$$\text{tr}[\rho] = \int \rho(\mathbf{r}, \mathbf{r})d\mathbf{r} = \int \sum_n f_n \psi_n(\mathbf{r})\psi_n^*(\mathbf{r})\,d\mathbf{r} = \sum_n f_n \int |\psi_n(\mathbf{r})|^2 d\mathbf{r} = \sum_n f_n = N_e$$

Energy expressions

- With orbitals

$$E[\{\psi_j\}] = \sum_i \int \psi_i(\mathbf{r}) \left(-\frac{1}{2}\nabla^2\right) \psi_i(\mathbf{r}) d\mathbf{r} + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[n]$$

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$$

- With density matrix

$$E[\rho] = \int \left[-\frac{1}{2}\nabla^2_{\mathbf{r}'}\rho(\mathbf{r}',\mathbf{r})\right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[n]$$

$$\rho(\mathbf{r}',\mathbf{r}) = \sum_i f_i\, \psi_i(\mathbf{r}')\psi_i^*(\mathbf{r}) \qquad n(\mathbf{r}) = \rho(\mathbf{r},\mathbf{r})$$

Self-consistency

$$\hat{H}[\{\psi_i^{(p)}\}] \, |\psi_n^{(p+1)}\rangle = \varepsilon_n^{(p+1)}|\psi_n^{(p+1)}\rangle \quad \text{therefore} \quad \hat{H}[\{\psi_i^{(p)}\}] = \sum_n \varepsilon_n^{(p+1)}|\psi_n^{(p+1)}\rangle\langle\psi_n^{(p+1)}|$$

$$\hat{\rho}[\{\psi_i^{(p)}\}] = \sum_n f_n \, |\psi_n^{(p)}\rangle\langle\psi_n^{(p)}|$$

When $|\psi_n^{(p)}\rangle = |\psi_n^{(p+1)}\rangle$ the **H** and ρ commute $\quad \left[\, \hat{H}[\{\psi_i^{(p)}\}], \, \hat{\rho}[\{\psi_i^{(p)}\}]\,\right] = 0$

Conditions
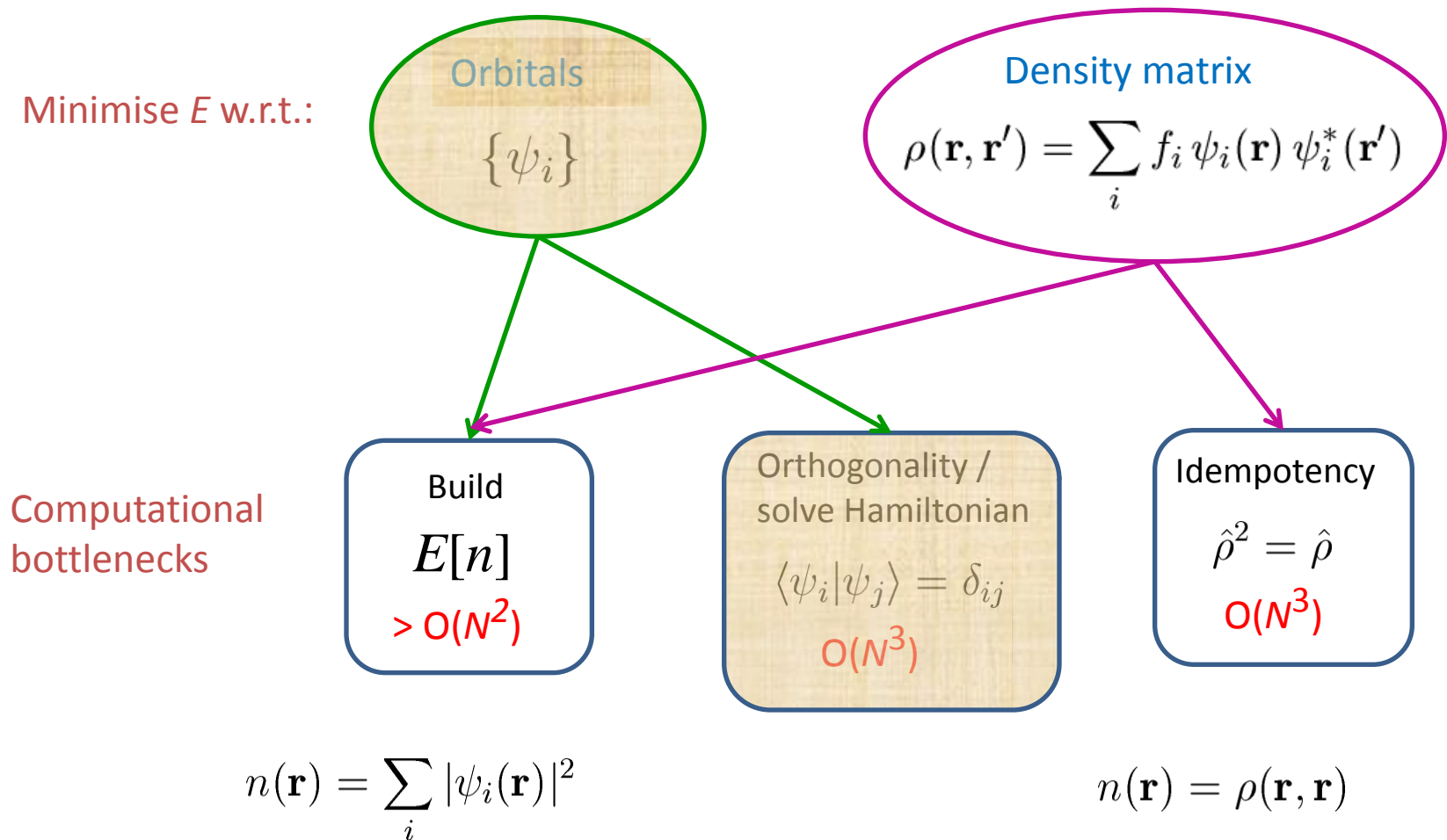
- Idempotency

$$\hat{\rho}^2 = \hat{\rho} \iff f_n = 0 \text{ or } 1$$

- Normalisation

$$N_e = 2\,\mathrm{tr}[\rho] = 2\sum_n f_n \qquad \text{(spin-unpolarised case)}$$

Calculation bottlenecks with the density matrix

Minimise E w.r.t.:

Orbitals

$$\{\psi_i\}$$

Density matrix

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \, \psi_i(\mathbf{r}) \, \psi_i^*(\mathbf{r}')$$

Computational bottlenecks

Build

$$E[n]$$

$> O(N^2)$

Orthogonality / solve Hamiltonian

$$\langle \psi_i | \psi_j \rangle = \delta_{ij}$$

$O(N^3)$

Idempotency

$$\hat{\rho}^2 = \hat{\rho}$$

$O(N^3)$

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$$

$$n(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r})$$

Linear-scaling DFT using the density matrix

Truncate exponential "tail", impose:

$$\rho(\mathbf{r}, \mathbf{r}') = 0 \quad \text{when} \quad |\mathbf{r} - \mathbf{r}'| > r_{\text{cut}}$$

The density matrix

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \, \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}')$$

is expanded in **non-orthogonal localised** functions:

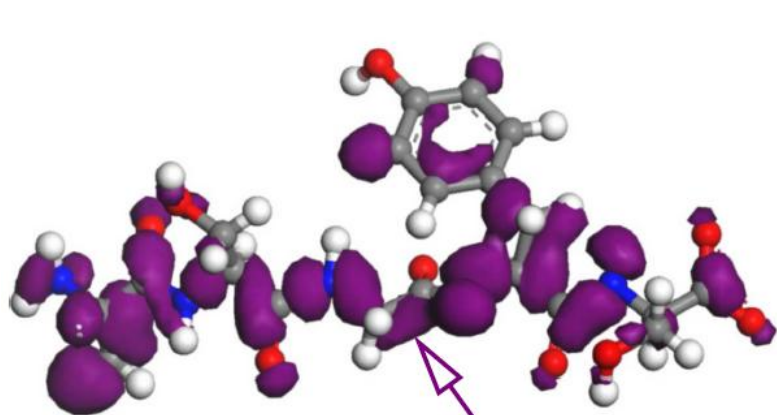$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}')$$

sparse matrix

- **K** matrix made sparse by truncation

- **S** and **H** in terms of $\{\phi_\alpha\}$ also sparse

- Optimise energy with respect to **K** using algorithms that include only **sparse matrix multiplications**

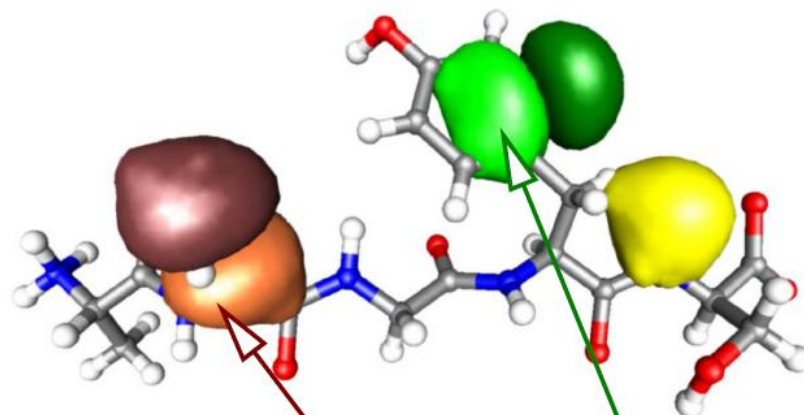- Sparse matrix multiplications can be linear-scaling if the sparsity is suitably exploited
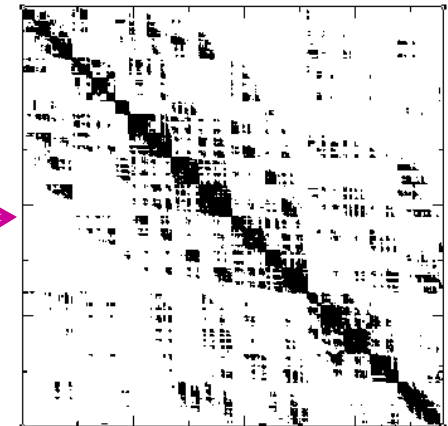
ONETEP aims for linear-scaling with high accuracy

The $\{\phi_\alpha\}$ are **not fixed**, but are optimised *in situ*, (in addition to **K)**

$$\rho(\mathbf{r}, \mathbf{r}') \quad = \quad \sum_n f_n \, \psi_n(\mathbf{r}) \, \psi_n^*(\mathbf{r}') \quad = \quad \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) \, K^{\alpha\beta} \, \phi_\beta^*(\mathbf{r}')$$

Molecular orbitals (MOs)

Non-orthogonal Generalised Wannier Functions (NGWFs)

Density matrix in terms of NGWFs

$$\psi_n(\mathbf{r}) = \phi_\alpha(\mathbf{r}) M^\alpha{}_n \qquad\qquad S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$$

MO   **NGWF**

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \, \psi_n(\mathbf{r}) \, \psi_n^*(\mathbf{r}') = \phi_\alpha(\mathbf{r}) \left(\sum_n M^\alpha{}_n f_n M_n^{\dagger\,\beta} \right) \phi_\beta^*(\mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}')$$

$$K^{\alpha\beta} = \sum_n M^\alpha{}_n f_n M_n^{\dagger\,\beta}$$

Density kernel

$$\rho^2(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\gamma} S_{\gamma\epsilon} K^{\epsilon\beta} \phi_\beta^*(\mathbf{r}')$$
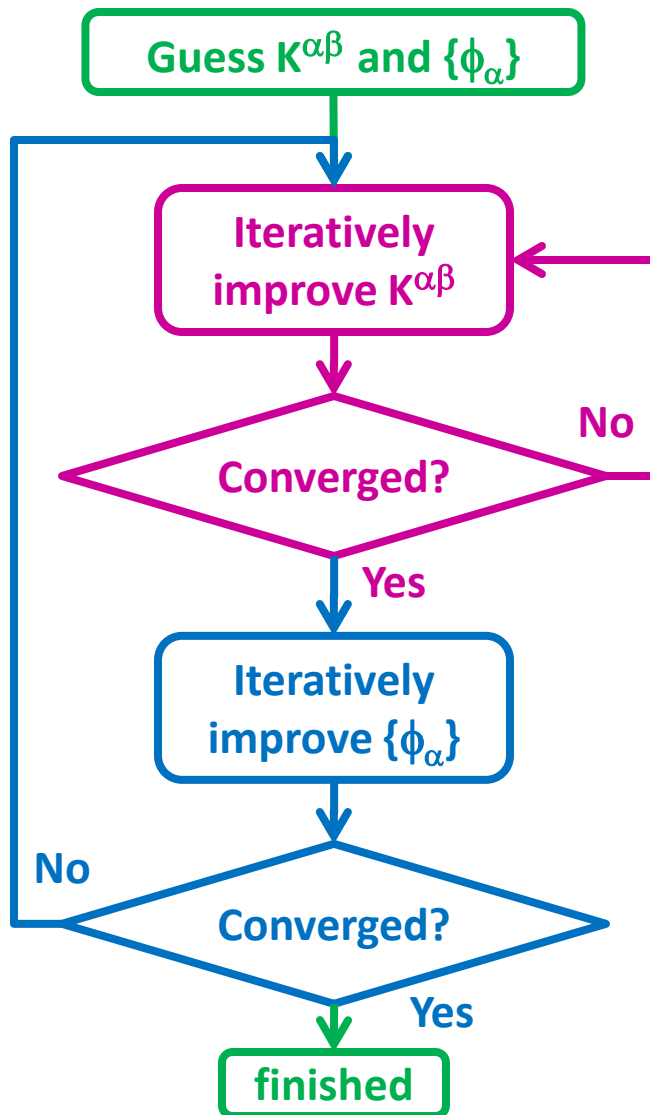
$$\mathrm{tr}[\rho] = K^{\alpha\beta} S_{\beta\alpha} = \mathrm{tr}[\mathbf{KS}]$$

Density matrix optimisation in ONETEP

$$E = E[\mathbf{K}, \{\phi_\alpha\}]$$

Flowchart:

Guess $K^{\alpha\beta}$ and $\{\phi_\alpha\}$

↓

Iteratively improve $K^{\alpha\beta}$

↓

Converged? — **No** (loop back to improve $K^{\alpha\beta}$)

↓ **Yes**

Iteratively improve $\{\phi_\alpha\}$

↓

Converged? — **No** (loop back to top)

↓ **Yes**

finished

- **Outer loop:** Optimise total (interacting) energy E w.r.t. to **K** and $\{\phi_\alpha\}$

- **Inner loop:** Optimise total (interacting) energy E w.r.t **K** for fixed $\{\phi_\alpha\}$ while imposing **idempotency** and **normalisation**

"Linear-scaling" functionals

- These are not DFT exchange and correlation functionals

- These are functionals of the density matrix which have been designed so that when they are minimised:

  - Produce the total energy

  - Implicitly impose the required conditions (idempotency and normalisation) on the density matrix that minimises them

- The aim of using these functionals is to work with expressions which involve only (sparse) matrix multiplications

- We will examine the functionals implemented in ONETEP

  - Penalty functional

  - Li –Nunes-Vanderbilt (LNV) types of functionals

  - Palser-Manolopoulos approach (not a functional)

Penalty functional

$$P[\rho] = \mathrm{tr}[(\rho^2 - \rho)^2] = \sum_n (f_n^2 - f_n)^2$$
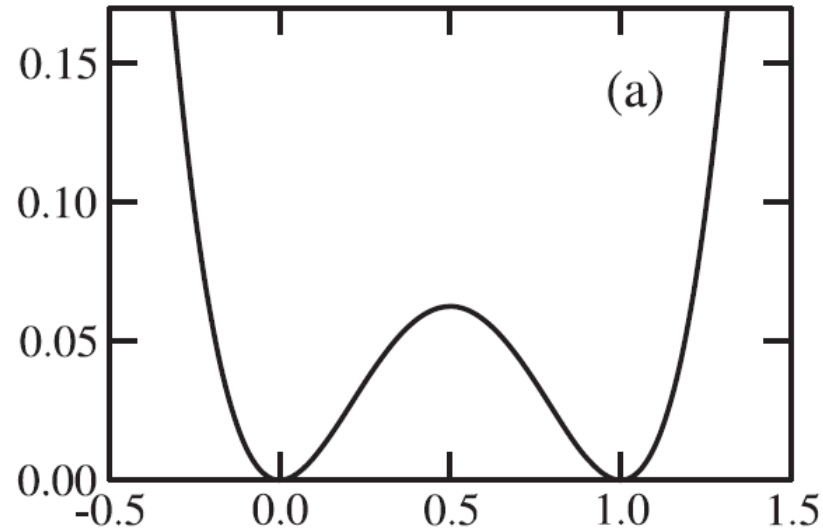
- Minimum (P=0) for idempotent density matrices
- Will always converge, regardless how non-idempotent the initial guess
- Constraints to preserve the total number of electrons are needed

Penalty functional derivatives

$$P[\rho] = \mathrm{tr}[(\rho^2 - \rho)^2] = \sum_n (f_n^2 - f_n)^2$$

- Assume a steepest descents "step" to update f_k

$$\frac{\partial P}{\partial f_k} = 4f_k^3 - 6f_k^2 + 2f_k$$

$$f_k^{(p+1)} = f_k^{(p)} + \lambda d_k^{(p)} \qquad f_k^{(p+1)} = f_k^{(p)} - \lambda \frac{\partial P}{\partial f_k}^{(p)}$$

- If we apply a quadratic approximation (near the minimum):

$$f_{new} = f_{old} - \frac{1}{2}(4f_{old}^3 - 6f_{old}^2 + 2f_{old}) = 3f_{old}^2 - 2f_{old}^3$$

Purification transformation

R. McWeeny, *Rev. Mod. Phys.* **32**(2), 335(1960)

Purification transformation
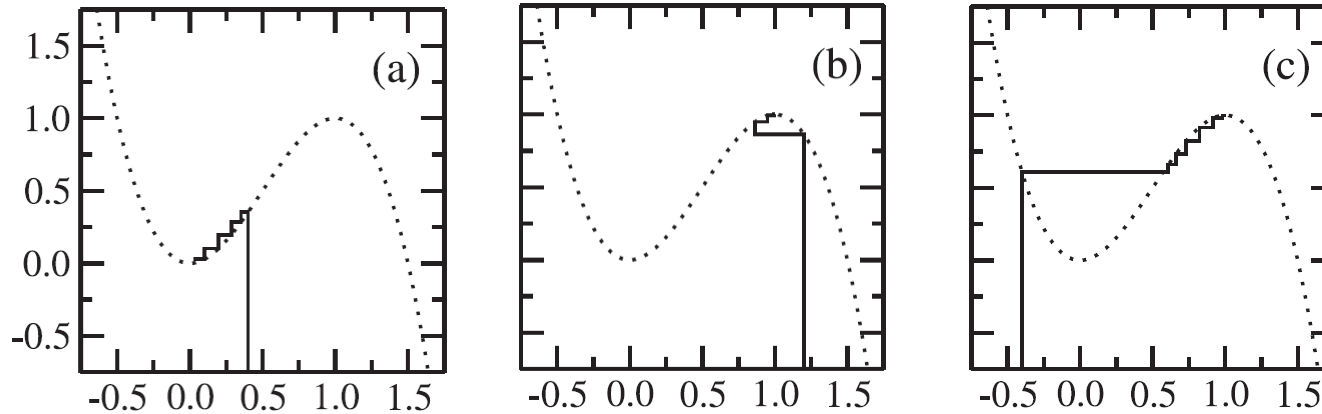
$$f_{new} = 3f_{old}^2 - 2f_{old}^3$$

- Improves the idempotency of a nearly idempotent density matrix
- Forces occupancies in interval [0,1] ("weak" idempotency)
- Quadratic convergence. Example:
  - $f^{(1)} = 1.1$
  - $f^{(2)} = 3 \times 1.1^3 - 2 \times 1.1^2 = 0.968$
  - $f^{(3)} = 3 \times 0.968^3 - 2 \times 0.968^2 = 0.997$
  - Etc..
- But, diverges if the initial occupancies are not within certain bounds

Energy with penalty functional

P. D. Haynes and M. C. Payne, *Phys. Rev. B* **59**, 12173 (1999)

$$Q[\rho] = E[\rho] + \alpha P[\rho]$$
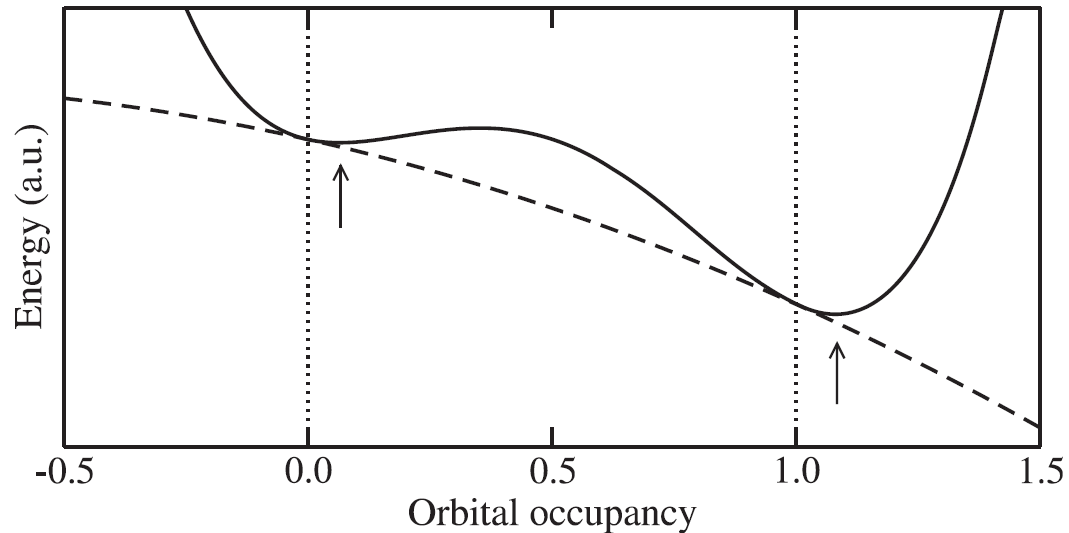
- Balance between minimum energy and minimum penalty
- Near-idempotency, depending on value of α
- Correction expression available for the energy

Li-Nunes-Vanderbilt (LNV) functionals

$$\rho = 3\sigma^2 - 2\sigma^3$$

X.-P. Li., R. W. Nunes and D. Vanderbilt, *Phys. Rev. B* **47**, 10891 (1993),
M. S. Daw, *Phys. Rev. B* **47**,10895 (1993)

$$\Omega_1[\sigma] = \text{tr}[\rho(H - \mu)]$$

J. M. Millam and G. E. Scuseria, *J. Chem. Phys.* **106**, 5569 (1997)

$$\Omega_2[\sigma] = \text{tr}[\rho H] + \mu'(\text{tr}[\sigma] - N_e)$$

- Energy expressions containing a purification transformation of an "auxiliary" density matrix
- Implicitly enforce idempotency
- Break down if purification transformation breaks
- Minimisation of band structure energy (equivalent to diagonalisation)
- But H also depends on ρ: Self-consistency by **density mixing**

LNV functionals in ONETEP

`exact_lnv = F`

$$L_1[\sigma] = E[\rho] + \mu'\left(\text{tr}[\sigma] - N_e\right) = E[3\sigma^2 - 2\sigma^3] + \mu'\left(\text{tr}[\sigma] - N_e\right)$$

`exact_lnv = T (default)`

$$L_2[\sigma] = E\left[\frac{N_e}{\text{tr}[\rho]}\rho\right] = E\left[\frac{N_e}{\text{tr}[3\sigma^2 - 2\sigma^3]}(3\sigma^2 - 2\sigma^3)\right]$$

- No density mixing, just **direct minimisation** of interacting energy
- Purification never allowed to break down
  - Occupancy maxima and minima tracked
  - Restored by penalty functional if out of safe range
- Electron number conserved without need to know the chemical potential

P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, *J. Phys.: Condens. Matter* **20**, 294207 (2008)
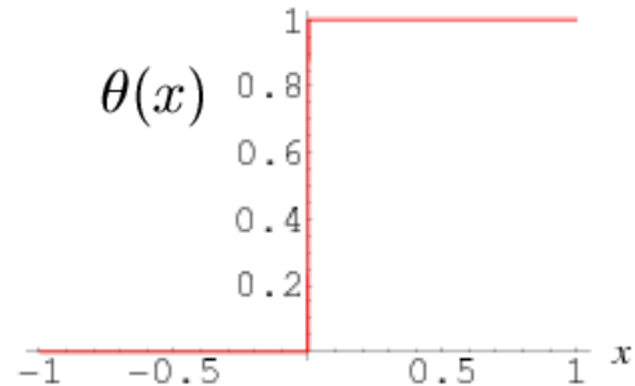
Canonical purification

$$\hat{\rho} = \theta(\mu\hat{I} - \hat{H})$$

$$H(\mathbf{r}, \mathbf{r}') = \sum_i \varepsilon_i \, \psi_i(\mathbf{r})\psi_i(\mathbf{r}')$$

$$\theta(\mu\hat{I} - \hat{H})(\mathbf{r}, \mathbf{r}') = \ldots = \sum_i \theta(\mu - \varepsilon_i)\psi_i(\mathbf{r})\psi_i(\mathbf{r}') = \sum_i f_i \, \psi_i(\mathbf{r})\psi_i(\mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}')$$

$$f_i = \begin{cases} 1, & \text{if } \varepsilon_i < \mu \\ 0, & \text{if } \varepsilon_i > \mu \end{cases}$$
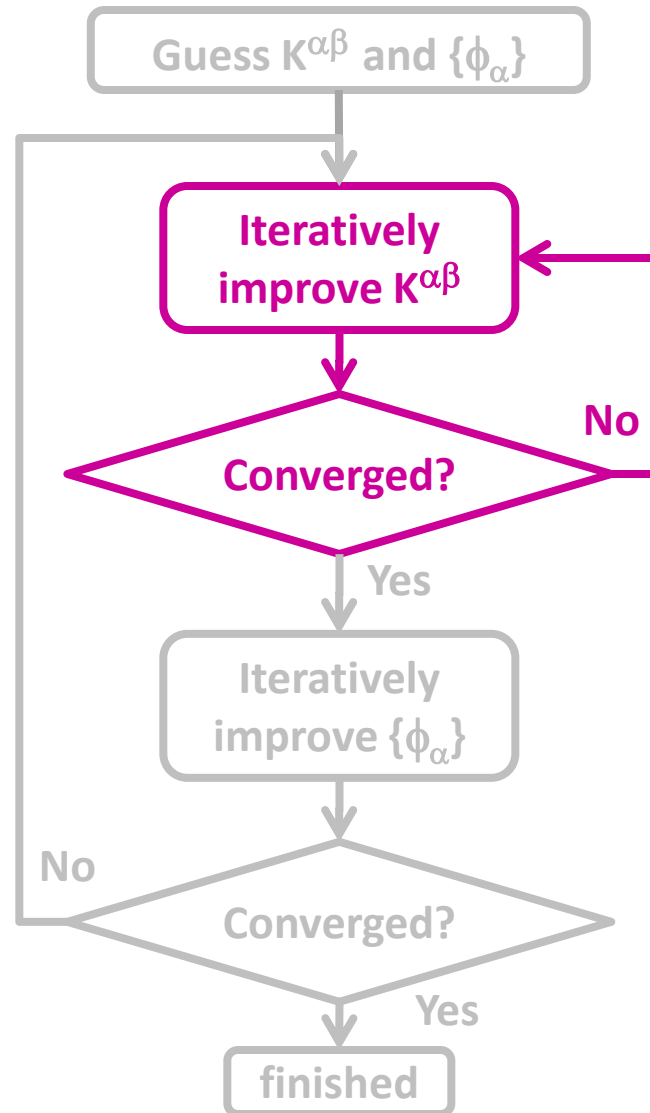
- One can approximate ρ as a polynomial expansion of H
- Can do this iteratively with a formula that resembles the purification transformation:

    A. H. Palser and D. E. Manolopoulos, *Phys. Rev. B* **58**, 12704 (1998)

- Always converges
- No need to know μ

Density kernel optimisation (inner loop)

Guess $K^{\alpha\beta}$ and $\{\phi_\alpha\}$

Iteratively improve $K^{\alpha\beta}$

Converged?

No

Yes

Iteratively improve $\{\phi_\alpha\}$

Converged?

No

Yes

finished

Derivatives with respect to the density kernel

Relationship between canonical Kohn-Sham orbitals and NGWFs

$$\psi_n(\mathbf{r}) = \phi_\alpha(\mathbf{r}) M^\alpha_{\ n} \quad \Leftrightarrow \quad \phi_\alpha(\mathbf{r}) = \sum_n \psi_n(\mathbf{r})(M^\dagger)_{n\alpha}$$

$$\langle \psi_n | \psi_m \rangle = \delta_{nm} \qquad \langle \phi_\alpha | \phi_\beta \rangle = S_{\alpha\beta} \ , \quad (S^{-1})^{\alpha\beta}$$

$$K^{\alpha\beta} = \sum_n M^\alpha_{\ n} f_n (M^\dagger)_n^{\ \beta}$$

When varying K:

- Occupancies change

- Kohn-Sham orbitals change

- NGWFs do not change

Example: Gradient of simple LNV functional

$$\Omega[\sigma] = E[3\sigma^2 - 2\sigma^3] = E[3\mathbf{LSL} - 2\mathbf{LSLSL}]$$

Purified density kernel: $\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL}$

Differentiate with respect to the elements of **L**

$$\frac{\partial E}{\partial L^{\alpha\beta}} = 6(\mathbf{SLH} + \mathbf{HLS})_{\beta\alpha} - 4(\mathbf{SLSLH} + \mathbf{SLHLS} + \mathbf{HLSLS})_{\beta\alpha} = \mathbf{G}_{\beta\alpha}$$
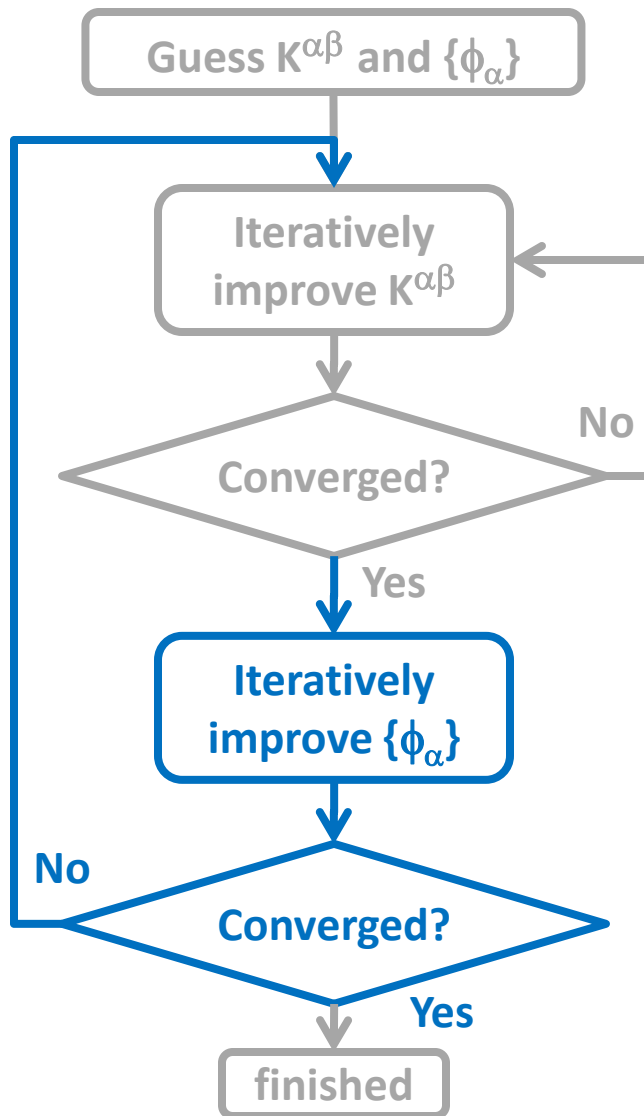
Tensor correction

- To obtain search directions (in steepest descents or conjugate gradients) a **contravariant** gradient is needed

$$G^{\alpha\beta} = S_-^{\alpha\gamma} G_{\gamma\epsilon} S_-^{\epsilon\beta} = 6(\mathbf{LHS_-} + \mathbf{S_-HL})^{\alpha\beta} - 4(\mathbf{LSLHS_-} + \mathbf{LHL} + \mathbf{S_-HLSL})^{\alpha\beta}$$

E. Artacho and L. M. del Bosch, *Phys. Rev. A* **43**, 5770 (1991).

C. A. White, P. Maslen, M. S. Lee and M. Head-Gordon, *Chem. Phys. Lett*. **276**, 133 (1997).

NGWF optimisation (outer loop)

Guess $K^{\alpha\beta}$ and $\{\phi_\alpha\}$

Iteratively improve $K^{\alpha\beta}$

Converged?   No

Yes

Iteratively improve $\{\phi_\alpha\}$

No   Converged?   Yes

finished

Derivatives with respect to NGWFs

$$\phi_\alpha(\mathbf{r}) = \sum_m D_m(\mathbf{r}) C_{m\alpha}$$

Gradient of LNV functional with respect to psinc expansion coefficients of NGWFs

$$g_m{}^\alpha = \frac{\partial L}{\partial C^*_{m\alpha}} = 2w \left[(\hat{H}\phi_\beta)(\mathbf{r})K^{\beta\alpha} + \phi_\beta(\mathbf{r})Q^{\beta\alpha} \right]_{\mathbf{r}=\mathbf{r_m}}$$

Tensor-corrected **covariant** gradient

$$g_{m\alpha} = g_m{}^\beta S_{\beta\alpha} = 2w \left[(\hat{H}\phi_\beta)(\mathbf{r})K^{\beta\gamma}S_{\gamma\alpha} + \phi_\beta(\mathbf{r})Q^{\beta\gamma}S_{\gamma\alpha} \right]_{\mathbf{r}=\mathbf{r_m}}$$

Linear-scaling calculation of gradient with:
- FFT box technique
- Sparse Hamiltonian and overlap matrices
- Sparse density kernel

C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Dieguez and M. C. Payne,
Phys. Rev. B **66**, 035119 (2002)

Maintaining localisation: Advantages of orthogonal basis set

Localisation by truncation of basis set expansion

$$\phi(\mathbf{r}) = \sum_i D_i(\mathbf{r}) c^i$$
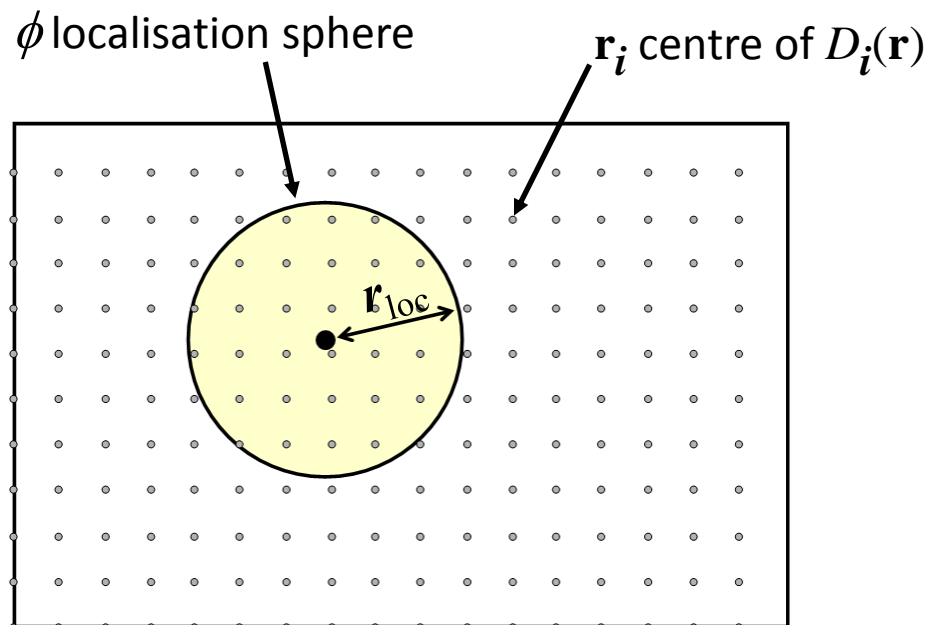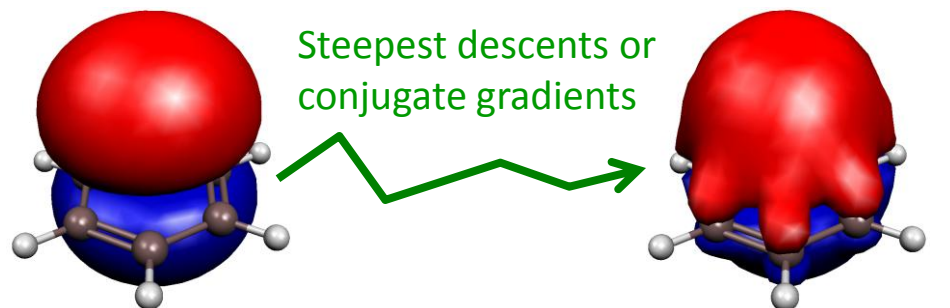
Steepest descents or conjugate gradients

What happens with a non-orthogonal basis:

$$\frac{\partial E}{\partial c^{*i}} = g_i$$

$$S_{ij} = \langle D_i(\mathbf{r}) | D_j(\mathbf{r}) \rangle$$

$$c^i_{\text{new}} = c^i_{\text{old}} - \lambda \sum_j (S^{-1})^{ij} g_j$$

Iterative update contains gradients from outside of localisation region

ϕ localisation sphere

$\mathbf{r}_i$ centre of $D_i(\mathbf{r})$

$\mathbf{r}_{\text{loc}}$

Occupancy preconditioning

- Derivative of E w.r.t. orbitals

$$\frac{\delta E}{\delta \psi_n^*(\mathbf{r})} = 2f_n[\hat{H}\psi_n(\mathbf{r}) - \epsilon_n \psi_n(\mathbf{r})]$$

- Derivative of E w.r.t. NGWFs

$$\frac{\delta E}{\delta \psi_n^*(\mathbf{r})} = \int \frac{\delta E}{\delta \phi_\alpha^*(\mathbf{r}')} \left(\frac{\delta \phi_\alpha(\mathbf{r}')}{\delta \psi_n(\mathbf{r})}\right)^* d\mathbf{r}' = \frac{\delta E}{\delta \phi_\alpha^*(\mathbf{r})} M_{\alpha n}$$

$$\frac{\delta E}{\delta \phi_\alpha^*(\mathbf{r})} = \frac{\delta E}{\delta \psi_n^*(\mathbf{r})} (M^\dagger)_{n\beta} (S^{-1})^{\beta\alpha}$$

Relation between Kohn-Sham orbitals and NGWFs

$$\phi_\alpha(\mathbf{r}) = \sum_n \psi_n(\mathbf{r})(M^\dagger)_{n\alpha}$$

$$\psi_n(\mathbf{r}) = \phi_\alpha(\mathbf{r}) M^\alpha_n \quad \langle \psi_n | \psi_m \rangle = \delta_{nm}$$

$$\langle \phi_\alpha | \phi_\beta \rangle = S_{\alpha\beta} , \quad (S^{-1})^{\alpha\beta}$$

$$K^{\alpha\beta} = \sum_n M^\alpha_n f_n (M^\dagger)_n^{\beta}$$

Remove ill-conditioning due to $f_n \sim 0$

$$\frac{\delta E}{\delta \phi_\alpha^*(\mathbf{r})}$$

f_n in [0,1]

$$2[\hat{H}\phi_\beta(\mathbf{r})K^{\beta\alpha} - \phi_\beta(\mathbf{r})K^{\beta\gamma}H_{\gamma\eta}(S^{-1})^{\eta\alpha}]$$

make all $f_n = 1$ `occ_mix = 1.0`

$$2[\hat{H}\phi_\alpha(\mathbf{r}) - \phi_\beta(\mathbf{r})(S^{-1})^{\beta\gamma}H_{\gamma\alpha}]$$

Use for search directions

Covariant

$$2[\hat{H}\phi_\beta(\mathbf{r})(S^{-1})^{\beta\alpha} - \phi_\beta(\mathbf{r})(S^{-1})^{\beta\gamma}H_{\gamma\eta}(S^{-1})^{\eta\alpha}]$$

Kinetic energy preconditioning

Length scale ill-conditioning

- Convergence rate falls with increasing ratio of max/min eigenvalues of the Hamiltonian

- High energy eigenstates are dominated by kinetic energy

- Need to reduce the contribution of kinetic energy in the NGWF gradient
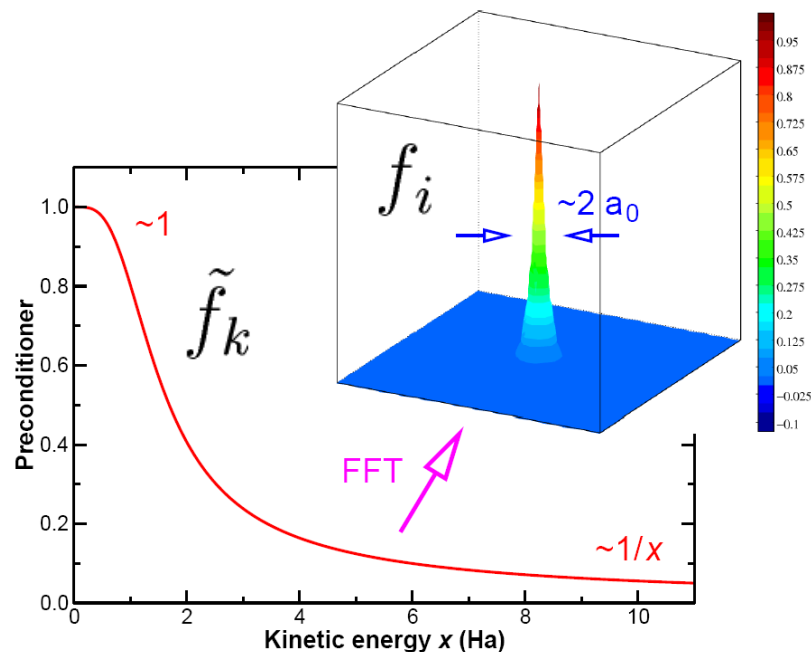
- Pre-multiply NGWF gradient in reciprocal space with a function that behaves as the inverse of the kinetic energy at high wave vectors and approaches 1 at low wavevectors:
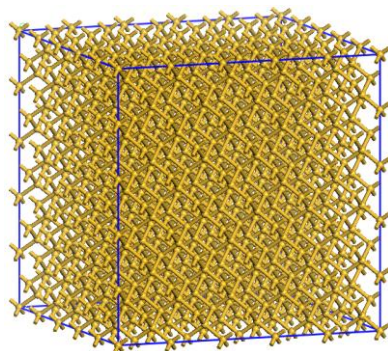
$$\tilde{g}_{k\alpha} \longrightarrow \tilde{f}_k \tilde{g}_{k\alpha} \qquad \text{Example:} \quad \tilde{f}_k = \frac{k_0^2}{k_0^2 + \mathbf{k}^2} \qquad \texttt{k\_zero}$$

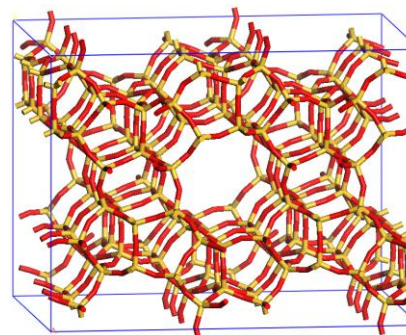A. A. Mostofi, P. D. Haynes, C.-K. Skylaris and M. C. Payne, *J. Chem. Phys.* **119**, 8842 (2003)
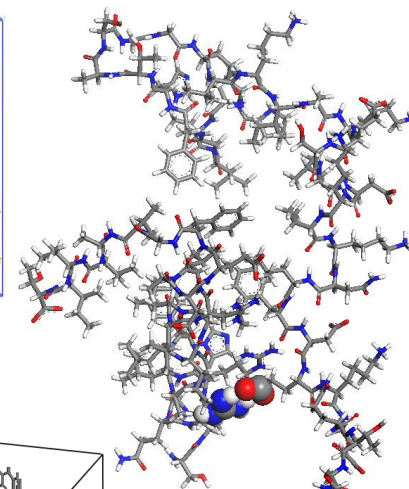
Fast convergence
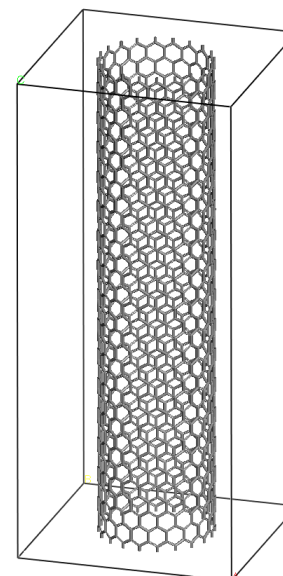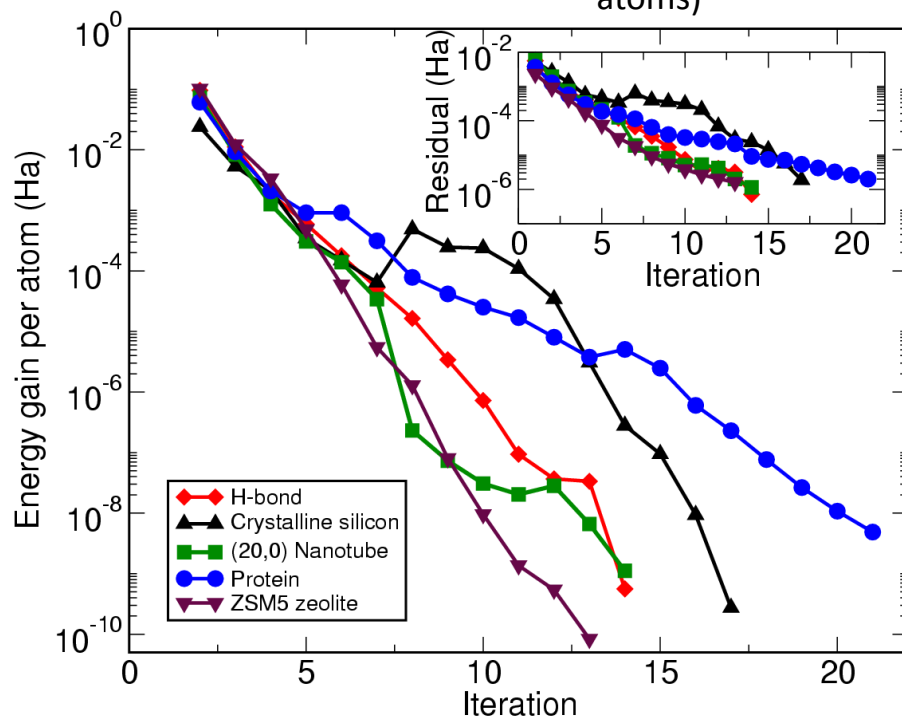**True linear-scaling**
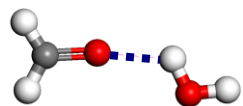
H-bond (7 atoms)

Crystalline silicon (1000 atoms)

ZSM5 zeolite (576 atoms)

Protein (988 atoms)

Legend:
- H-bond
- Crystalline silicon
- (20,0) Nanotube
- Protein
- ZSM5 zeolite

(20, 0) Nanotube (1280 atoms)

C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *Phys. Stat. Sol. (b)* **243**(5), 973 (2006)

Calculation of S⁻¹

The Inverse of
$$S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$$
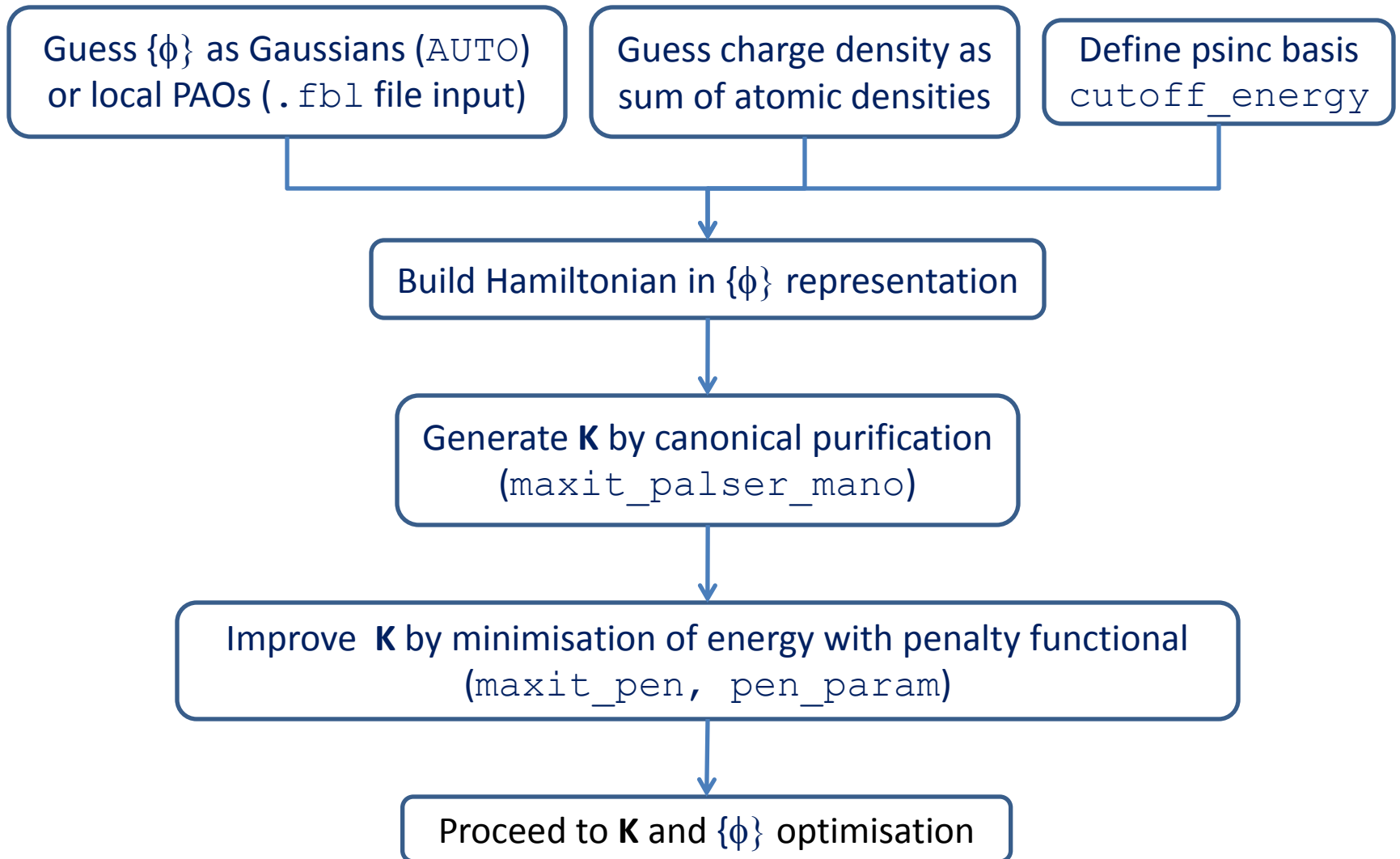Is a **contravariant** tensor, like the density kernel

Can be iteratively generated with the Hotelling formula

$$\mathbf{S}_{\text{new}}^{-1} = 2\mathbf{S}_{\text{old}}^{-1} - \mathbf{S}_{\text{old}}^{-1}\,\mathbf{S}\,\mathbf{S}_{\text{old}}^{-1} \qquad \texttt{maxit\_hotelling}$$
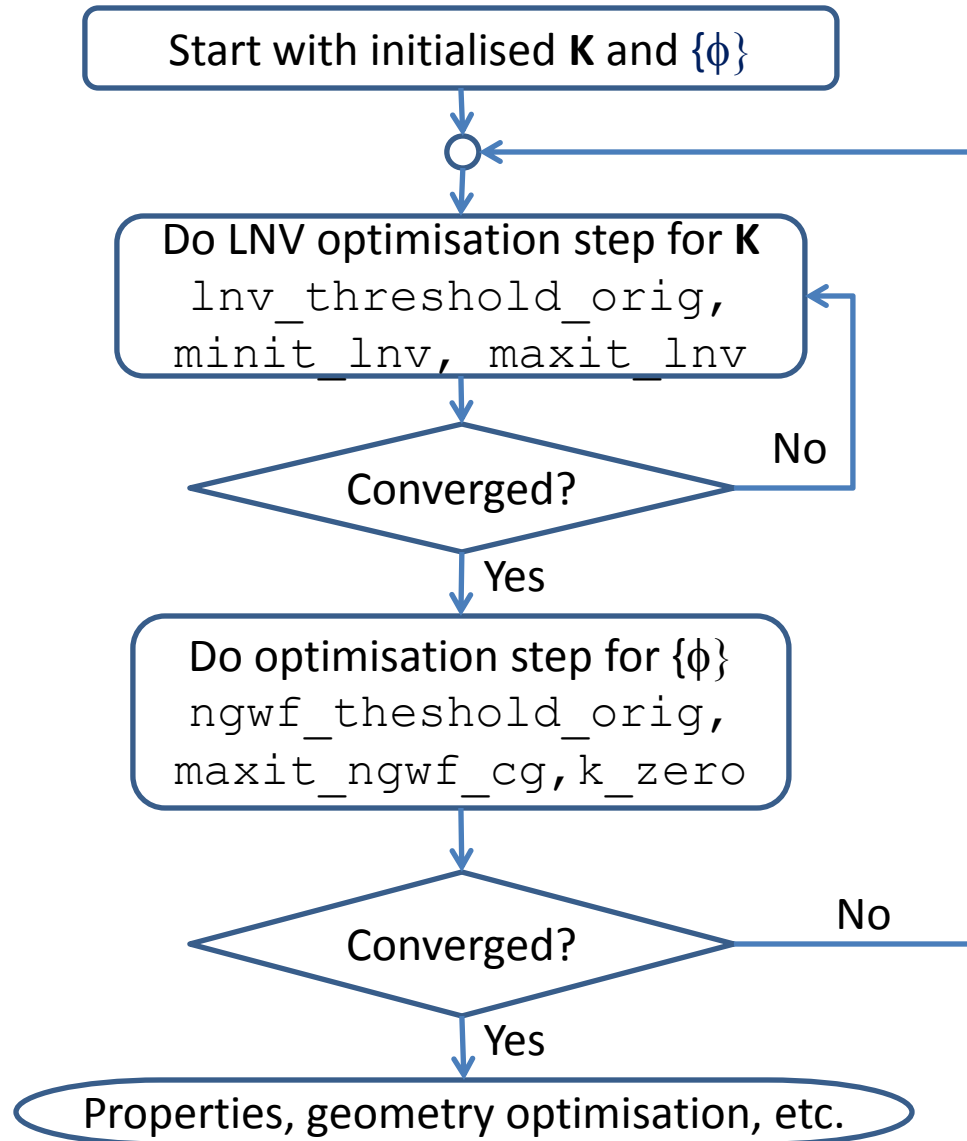
- Similar to purification transformation

- Converges rapidly

- Can generate **S**⁻¹ with linear-scaling cost from the beginning

- Can be used to update **S**⁻¹ when NGWFs are updated with minimal computational effort

- Takes advantage of kernel truncation

Overall ONETEP calculation scheme

Initialisation of density kernel and NGWFs

Guess {ϕ} as Gaussians (`AUTO`) or local PAOs (`.fbl` file input)

Guess charge density as sum of atomic densities

Define psinc basis `cutoff_energy`

Build Hamiltonian in {ϕ} representation

Generate **K** by canonical purification (`maxit_palser_mano`)

Improve **K** by minimisation of energy with penalty functional (`maxit_pen, pen_param`)

Proceed to **K** and {ϕ} optimisation

NGWF and density kernel optimisation

Key points

- Direct energy minimisation, no density mixing
- All quantities (Hamiltonian, energy, gradients) built with linear-scaling cost
  - ➢NGWF localisation in real space, FFT-box technique in reciprocal space
  - ➢Sparse Hamiltonian and overlap matrixes
  - ➢Sparse density kernel (`kernel_cutoff`)
- Linear-scaling iterative algorithms using sparse matrix algebra
- Matrix products preserve sparsity patterns, e.g. **KSK** is less sparse than **K**
- Tensorially correct gradients
- Preconditioned NGWF optimisation