Linear-scaling DFT using localized spherical-waves and real-space grids

Peter Haynes

Theory of Condensed Matter, Cavendish Laboratory, Cambridge, UK

pdh1001@cam.ac.uk
http://www.tcm.phy.cam.ac.uk/∼pdh1001/

Acknowledgements

- Dr Chee Kwan Gan
- Dr Chris-Kriton Skylaris
- Arash Mostofi
- Dr Oswaldo Diéguez
- Prof Mike Payne

Outline

- Optimisation of localized functions for density-matrix minimization
- Localized spherical-wave basis sets
- Real-space grids and FFTs

Density-matrix representation

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

Exploit the short range of the DM:

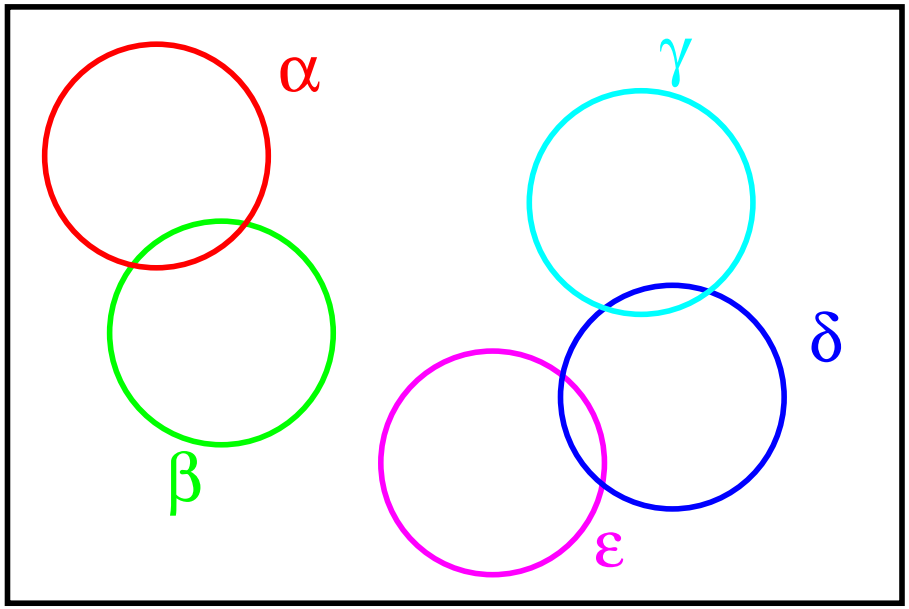
$$\rho(\mathbf{r},\mathbf{r}') \to 0 \text{ as } \left|\mathbf{r}-\mathbf{r}'\right| \to \infty$$

by making:

- $P^{\alpha\beta}$ sparse
- $\{\phi_\alpha(\mathbf{r})\}$ localized

i.e. imposing:

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



Density-matrix minimization

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

- Write the energy as a functional of the DM: $E[\rho]$
- Minimize the energy w.r.t. $\rho(\mathbf{r}, \mathbf{r}')$:
  - Optimize $P^{\alpha\beta}$ and $\{\phi_\alpha(\mathbf{r})\}$
  - Subject to the normalization and idempotency constraints

Physical interpretation:

- $\{\phi_\alpha(\mathbf{r})\}$ define a small subspace (relative to basis set) which optimally contains the occupied Kohn-Sham eigenfunctions
- Optimizing $P^{\alpha\beta}$ for a given set $\{\phi_\alpha(\mathbf{r})\}$ (subject to the constraints) is equivalent to filling up the states according to the Pauli exclusion principle: the density-matrix commutes with the Hamiltonian
- Optimizing $\{\phi_\alpha(\mathbf{r})\}$ corresponds to solving the Kohn-Sham equations

Localized spherical-wave basis sets

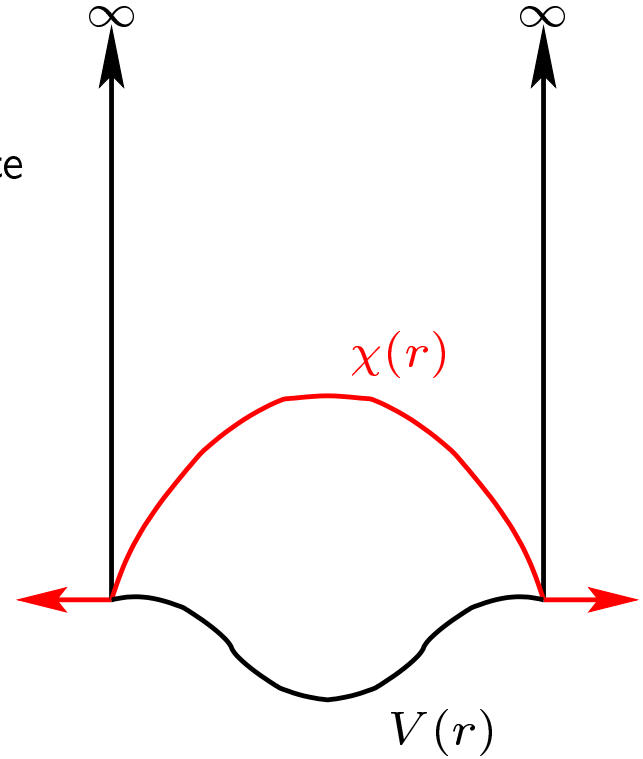
Confine $\{\phi_\alpha(\mathbf{r})\}$ to spherical regions of space $\Rightarrow$

- sphere radius $R_{\rm reg}$ is a variational parameter
- boundary condition on basis functions: vanish on surface

Pseudopotential approximation and perturbation theory $\Rightarrow$

- solve free-particle Schrödinger equation
- need relatively few basis functions for a weak potential

$$\Rightarrow \quad \left(\tfrac{1}{2}\nabla^2 + E\right)\chi(\mathbf{r}) = 0; \qquad \chi\left(|\mathbf{r}| = R_{\rm reg}\right) = 0$$



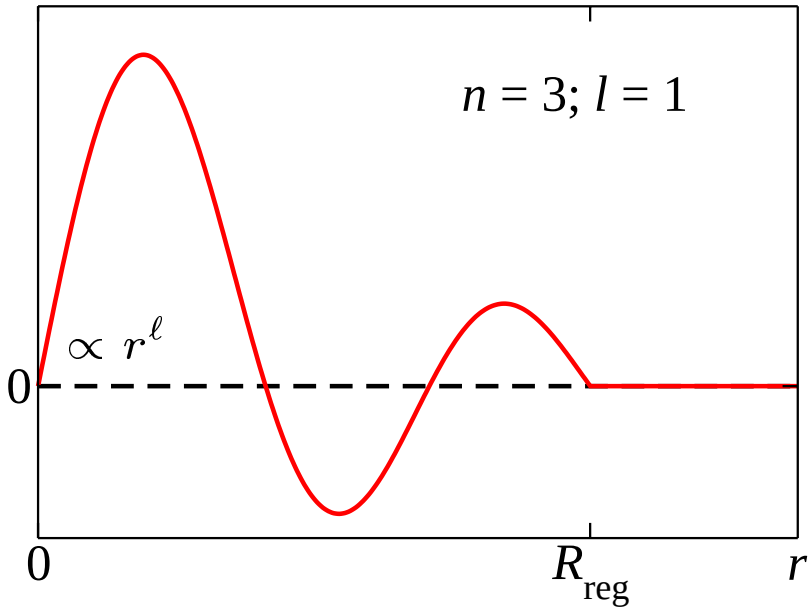
Basis functions

$$\left(\frac{1}{2}\nabla^2 + E\right)\chi(\mathbf{r}) = 0; \qquad \chi\left(|\mathbf{r}| = R_{\rm reg}\right) = 0$$

$$\Rightarrow \chi(\mathbf{r}) = \begin{cases} j_\ell(q_{n\ell}r)Y_{\ell m}(\vartheta,\varphi), & r < R_{\rm reg} \\ 0, & r \geq R_{\rm reg} \end{cases}$$

where:

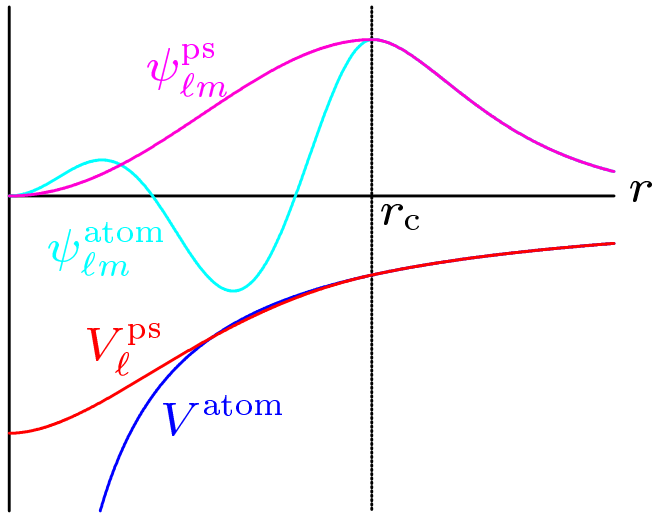
- n is a positive integer
- ℓ is a non-negative integer
- m is an integer: $-\ell \leq m \leq \ell$
- $q_{n\ell}$ is defined by $j_\ell(q_{n\ell}R_{\rm reg}) = 0$
- $E = \frac{1}{2}q_{n\ell}^2$



Properties of basis set

- Single parameter $E_{\rm cut}$ can be used to control the completeness
- Basis functions within a sphere are mutually orthogonal
- For basis functions in different spheres, analytic results can be obtained[1] for:
  - overlap matrix elements
  - kinetic energy matrix elements
  - non-local pseudopotential (in Kleinman-Bylander form) matrix elements:

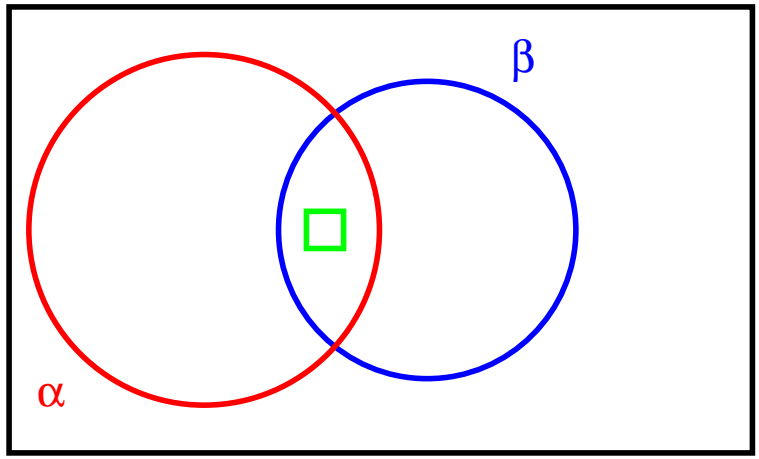
$$\hat{V}_{\rm NL} = \sum_{\ell m} \frac{|\delta\hat{V}_\ell \psi^{\rm ps}_{\ell m}\rangle\langle\psi^{\rm ps}_{\ell m}\delta\hat{V}_\ell|}{\langle\psi^{\rm ps}_{\ell m}|\delta\hat{V}_\ell|\psi^{\rm ps}_{\ell m}\rangle}$$



[1] P. D. Haynes and M. C. Payne, *Comput. Phys. Commun.* **102**, 17 (1997)

Truncating the basis set

- In principle, $E_{\rm cut}$ determines the maximum angular momentum component $\ell_{\rm cut}$
- In practice, we truncate the basis set by introducing a second parameter $\ell_{\rm max}$:
  - to reduce the basis set size $\propto (\ell_{\rm max}+1)^2$
  - to reduce the computational cost
  - as in Gaussian and LCAO type basis sets
  - to avoid over-completeness



Convergence tests

Used a direct diagonalization method to test the basis set *only* and vary the:

- Number of spherical regions
- Locations of the spherical regions
- Radii of the spherical regions $R_{\rm reg}$
- Energy cutoff $E_{\rm cut}$
- Maximum angular momentum component $\ell_{\rm max}$

For more details see:
C. K. Gan, P. D. Haynes and M. C. Payne, *Phys. Rev. B* **63** 205109 (2001).

Iterative diagonalization

Linear-scaling $\Rightarrow$ large systems $\Rightarrow$ large basis sets $\Rightarrow$ iterative diagonalization

Want to minimize the energy w.r.t. the wavefunctions $\{\psi_i(\mathbf{r})\}$:

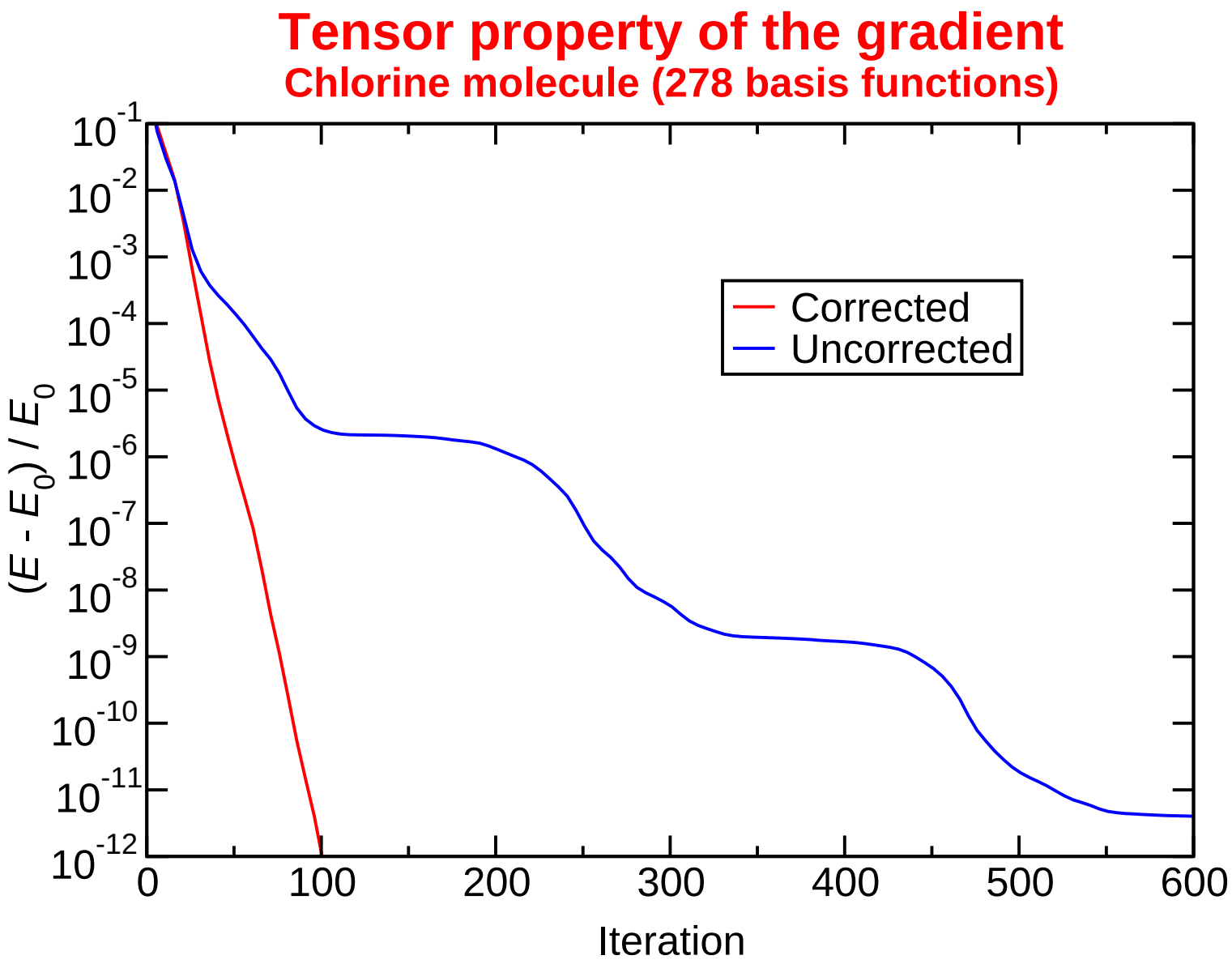
- Expand in a basis set: $\psi_i(\mathbf{r}) = \sum_\mu x_i^\mu \chi_\mu(\mathbf{r})$
- For a non-orthogonal basis set, solve the generalized eigenvalue equation:

$$H_{\mu\nu} x_i^\nu - \varepsilon_i S_{\mu\nu} x_i^\nu = 0$$

- Need to distinguish contravariant and covariant quantities[2]
- Correct gradient is therefore:

$$g_i^\mu = S^{\mu\nu} H_{\nu\xi} x_i^\xi - \varepsilon_i x_i^\mu$$

[2] E. Artacho and L. Miláns 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)


Tensor property of the gradient
Chlorine molecule (278 basis functions)
$(E - E_0) / E_0$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-11}
10^{-12}
0
100
200
300
400
500
600
Iteration
Corrected
Uncorrected

Kinetic energy preconditioning

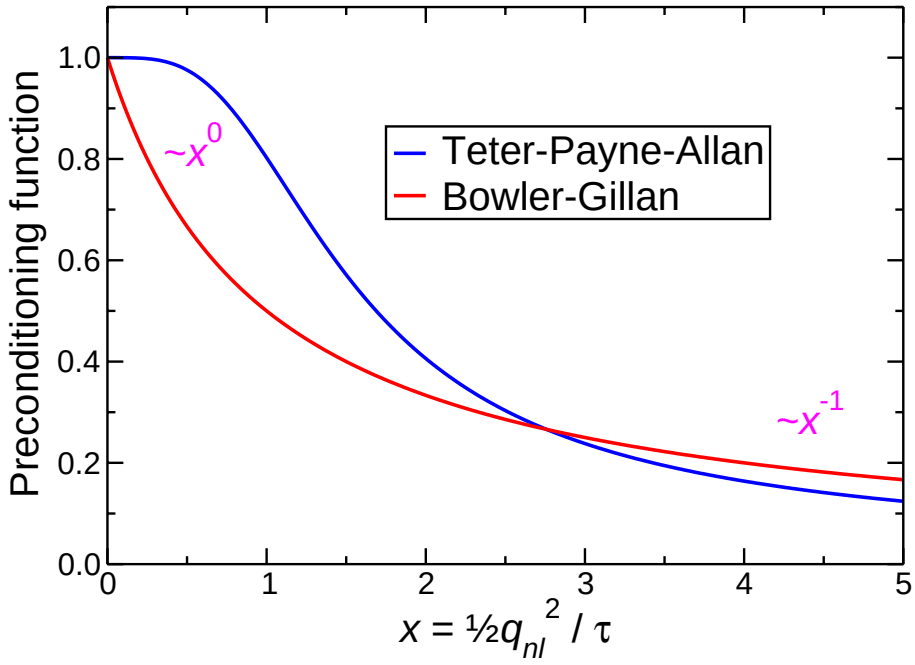
- Gradient is dominated by high kinetic energy basis functions
- Use a preconditioner to compensate
- Preconditioned gradient is:

$$G_i^\mu = (S + T/\tau)^{\mu\nu} H_{\nu\xi} x_i^\xi - \varepsilon_i x_i^\mu$$

- Cannot use "diagonal" approximation

  with spherical-waves
- Adjustable parameter τ

  can be optimized automatically



- For more information see:

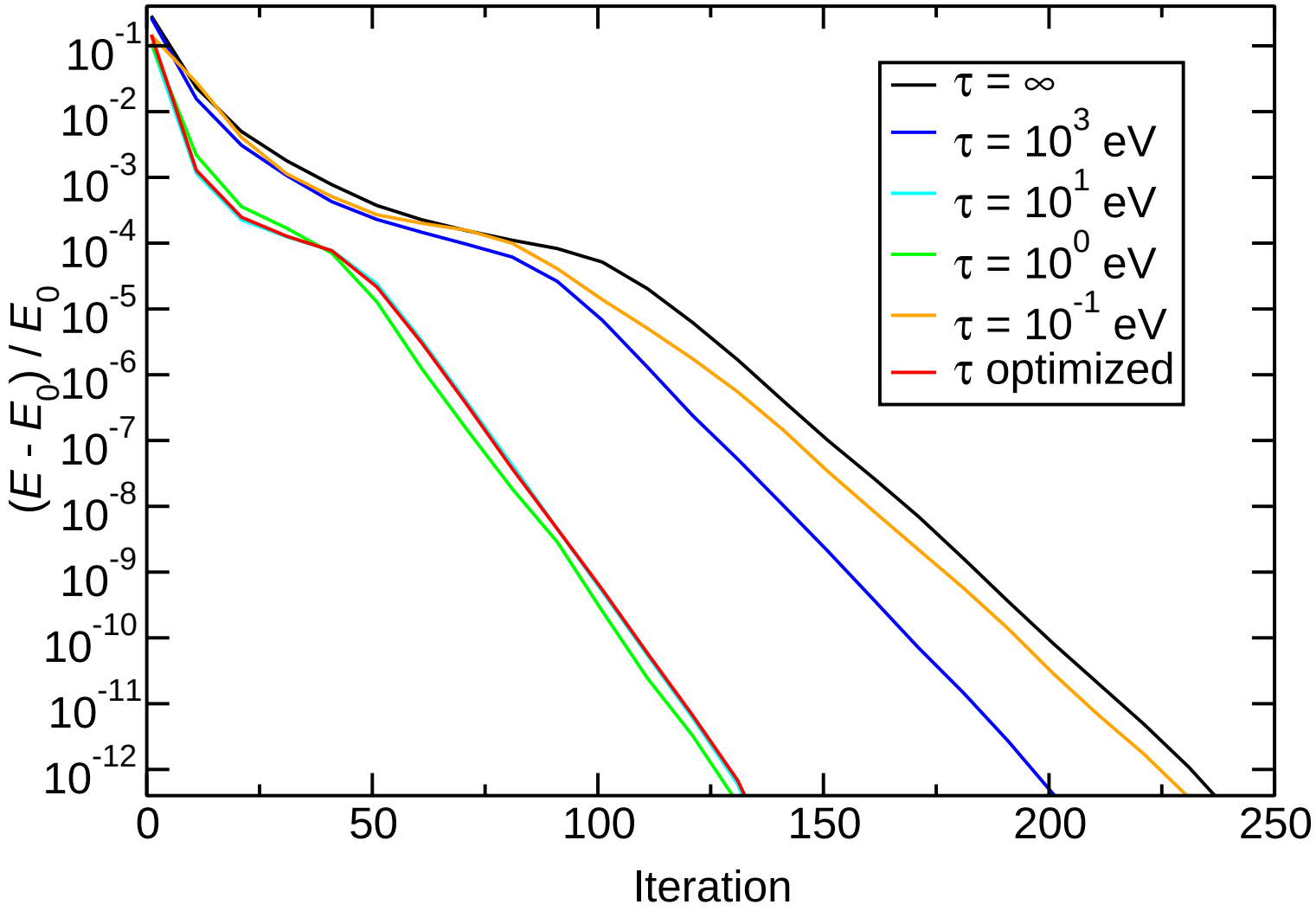
  C. K. Gan, P. D. Haynes and M. C. Payne, *Comput. Phys. Commun.* **134** 33 (2001).

M. P. Teter, M. C. Payne and D. C. Allan, *Phys. Rev. B* **40** 12255 (1989)

D. R. Bowler and M. J. Gillan, *Comput. Phys. Commun.* **112** 103 (1998)

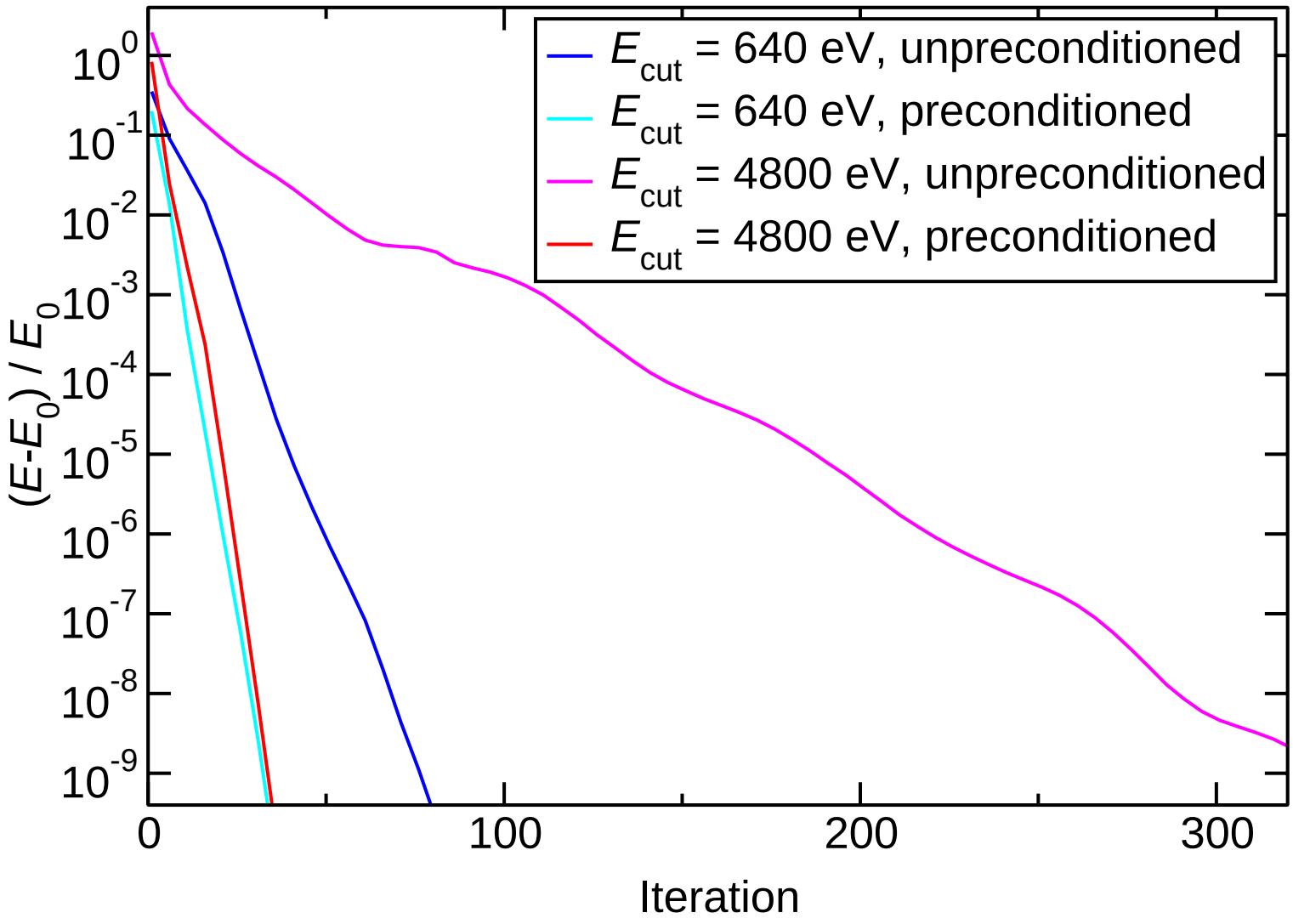
Choice of preconditioning parameter τ

Crystalline silicon (3520 basis functions)



Preconditioning results

Chlorine molecule (278/784 basis functions)



Optimizing localized functions

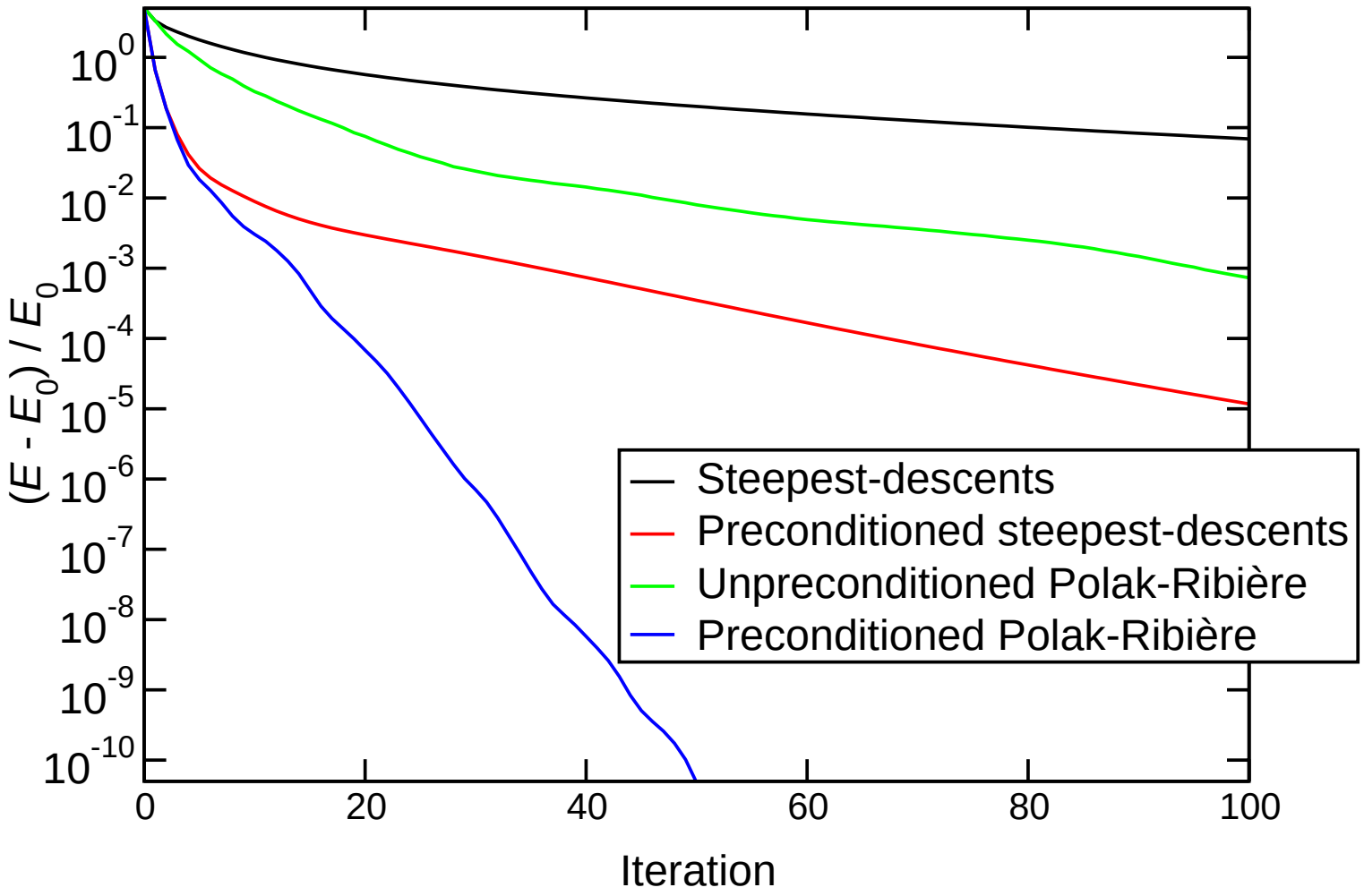
- Optimize a set of localized functions $\{\phi_\alpha(\mathbf{r})\}$ instead of wavefunctions $\{\psi_i(\mathbf{r})\}$
- Can no longer impose orthogonality, so minimize:

$$E[\{\phi_\alpha\}] = \mathrm{Tr}\left(\mathsf{S}^{-1}\mathsf{H}\right)$$

  where $\mathsf{S}_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$ and $\mathsf{H}_{\alpha\beta} = \langle\phi_\alpha|\hat{\mathsf{H}}|\phi_\beta\rangle$
- Need to consider tensor properties of gradient with respect to:
  - basis set
  - localized functions

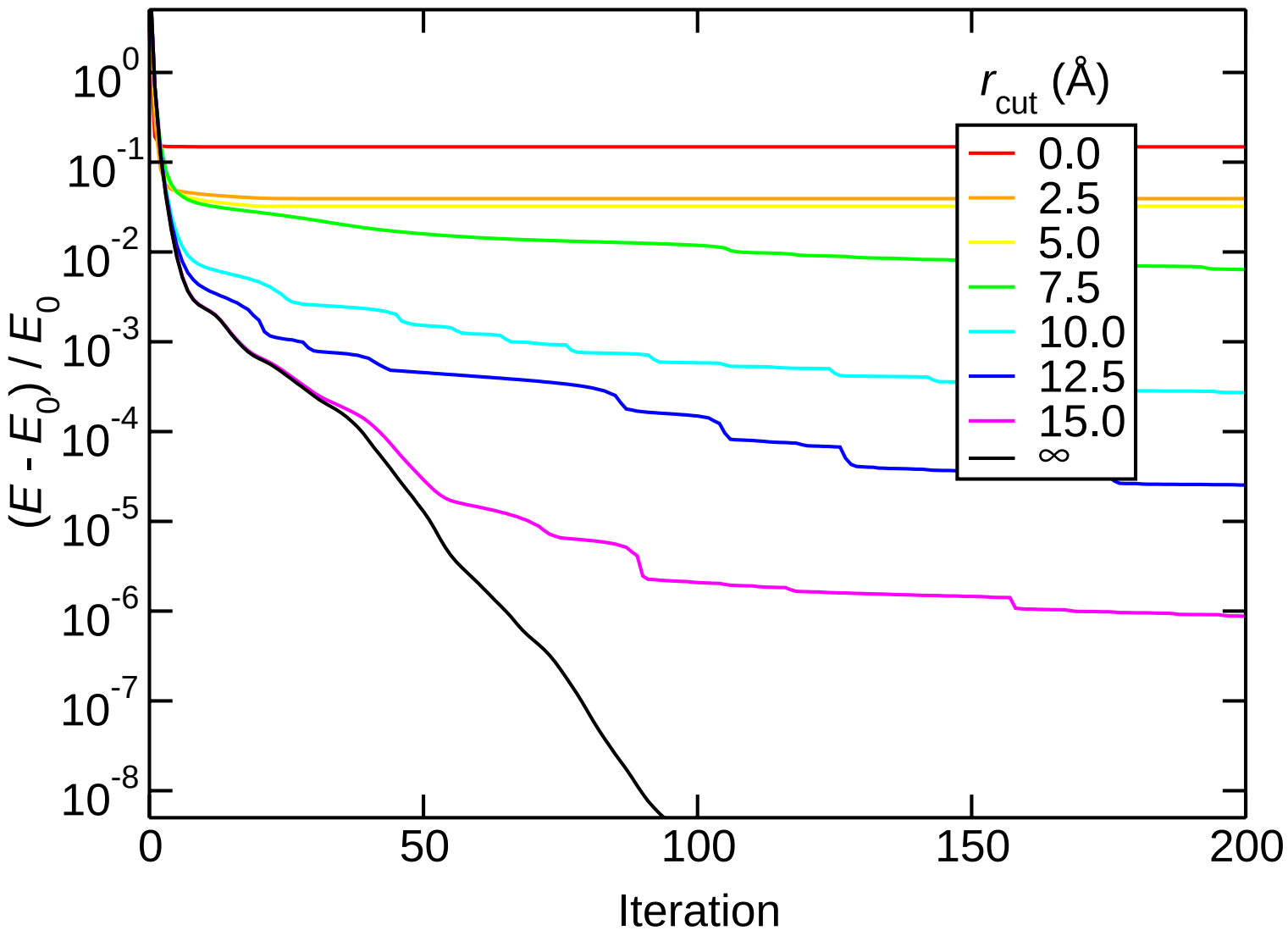
Performance of various algorithms

Crystalline silicon



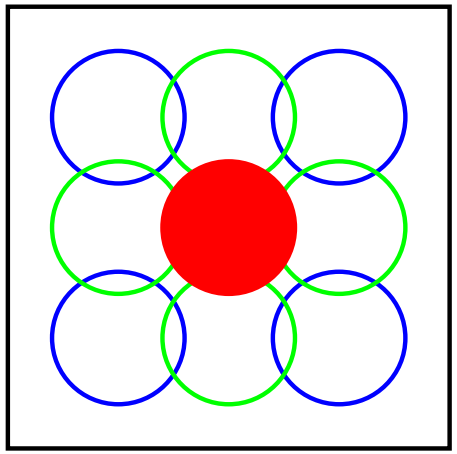
Effect of truncating the search direction

64-atom silicon cell (n-n bond length 2.35 Å)

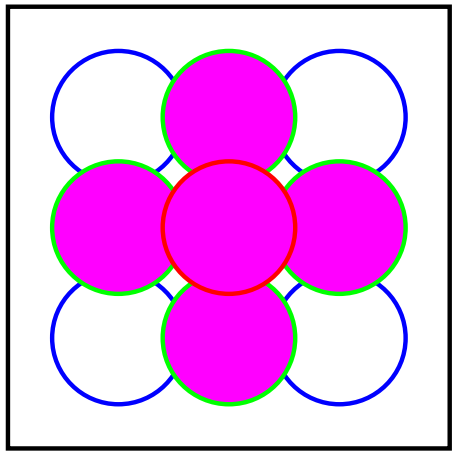


The problem

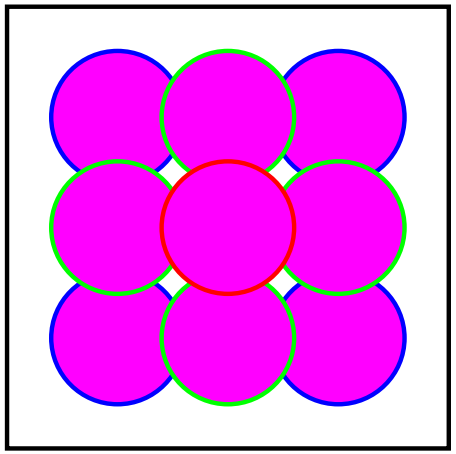
- Variational error
- Slow convergence
- Conjugate gradients breaks down



x_α^μ



g_μ^α



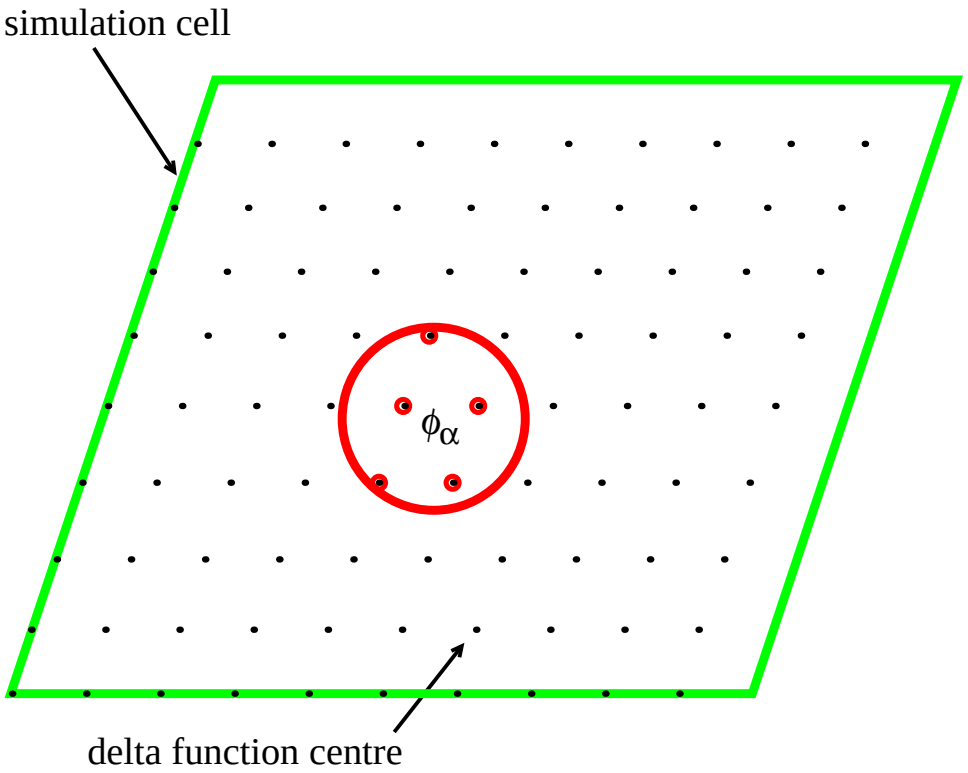
g_α^μ

Real-space grids

Represent the localized functions $\{\phi_\alpha(\mathbf{r})\}$ by their values at points on a regular grid:

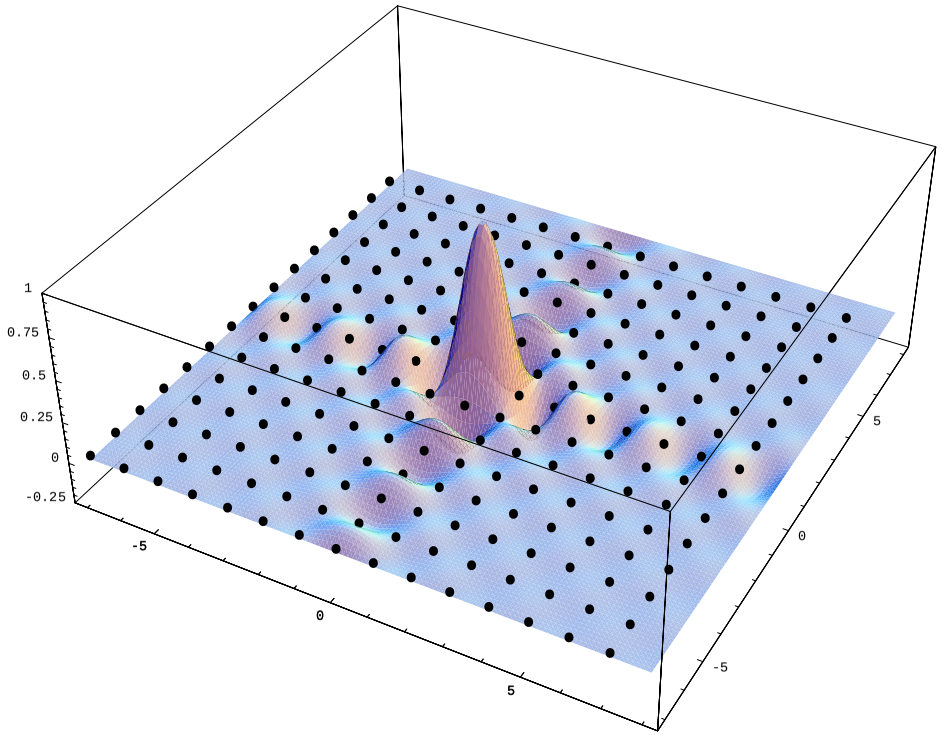
$$\phi_\alpha(\mathbf{r}) = \sum_{KLM} C_{KLM,\alpha} D_{KLM}(\mathbf{r})$$

where $C_{KLM,\alpha} = 0$ if $\mathbf{r}_{KLM}$ does not lie inside the sphere of ϕ_α



Delta function basis set

- Bandwidth Limited Periodic "delta" functions, $D_{KLM}(\mathbf{r})$. Centred on points $\mathbf{r}_{KLM}$ of regular grid with the symmetry of the simulation cell
- Grid spacing determines equivalent plane-wave kinetic energy cutoff



Properties of $D_{KLM}(\mathbf{r})$

- Centred on real-space grid points: $\mathbf{r}_{KLM} = \frac{K}{N_1}\mathbf{A_1} + \frac{L}{N_2}\mathbf{A_2} + \frac{M}{N_3}\mathbf{A_3}$
  - $\mathbf{A}_1$: primitive lattice vector
  - N_1: number of points in direction 1
- An "impulse function", i.e. a plane-wave expansion with constant amplitudes

$$D_{KLM}(\mathbf{r}) = \frac{1}{N_1 N_2 N_3} \sum_{\mathbf{G}} \exp[\mathrm{i}\mathbf{G} \cdot (\mathbf{r} - \mathbf{r}_{KLM})]$$

  - $D_{KLM}(\mathbf{r}_{FGH}) = \delta_{KF}\delta_{LG}\delta_{MH}$
  - Orthogonality
  - Real-valued everywhere
  - Equivalent to a plane-wave basis set
  - The projection of $f(\mathbf{r})$ on to $D_{KLM}(\mathbf{r})$ is equal to the value of the bandwidth limited version of $f(\mathbf{r})$ at the centre of $D_{KLM}(\mathbf{r})$

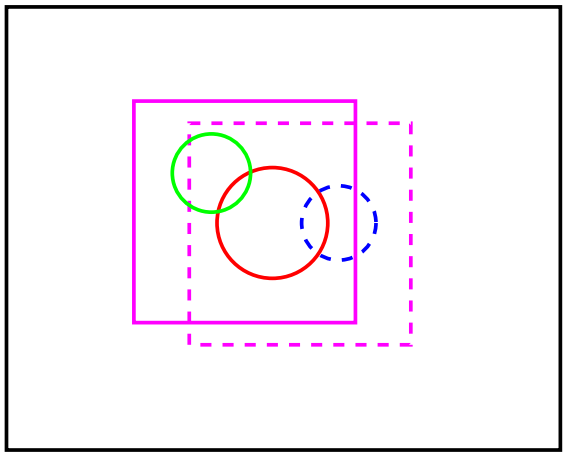
$$\int_V D_{KLM}(\mathbf{r}) f(\mathbf{r}) d\mathbf{r} = \frac{V}{N_1 N_2 N_3} f_{BWL}(\mathbf{r}_{KLM})$$

where $f_{BWL}(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{f}(\mathbf{G}) \exp[\mathrm{i}\mathbf{G} \cdot \mathbf{r}]$

FFT box

Connection with plane-waves suggests the use of FFTs e.g. to calculate the kinetic energy operator:

- More accurate than finite difference methods[3]
- For linear-scaling, cannot use full simulation cell



Must use single size of box for all pairs of functions to ensure:

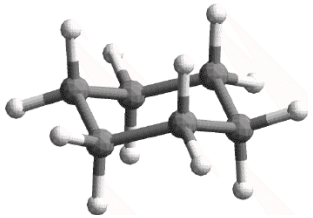
- Matrix representations of operators are Hermitian
- Consistent action of the Hamiltonian on each localized function

Also use the FFT box to construct the charge density via Fourier interpolation.

[3] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard and M. C. Payne, *Comput. Phys. Commun.* **140** 315 (2001).

Test calculations

Compare accuracy of FFT box method:



- Cyclohexane in $(40\ a_0)^3$ cell
- Norm-conserving pseudopotentials
- Grid spacing 0.5 a_0 ($=$ 40 Ry)

Error in total energy: $\Delta E = E(\mathrm{FFT\ box}) - E(\mathrm{simulation\ cell})$

rNGWF (a_0)		ΔE (mRy)	
H	C	total	per atom
5.0	5.0	6.48×10^{-1}	3.60×10^{-2}
5.0	6.0	3.26×10^{-1}	1.80×10^{-2}

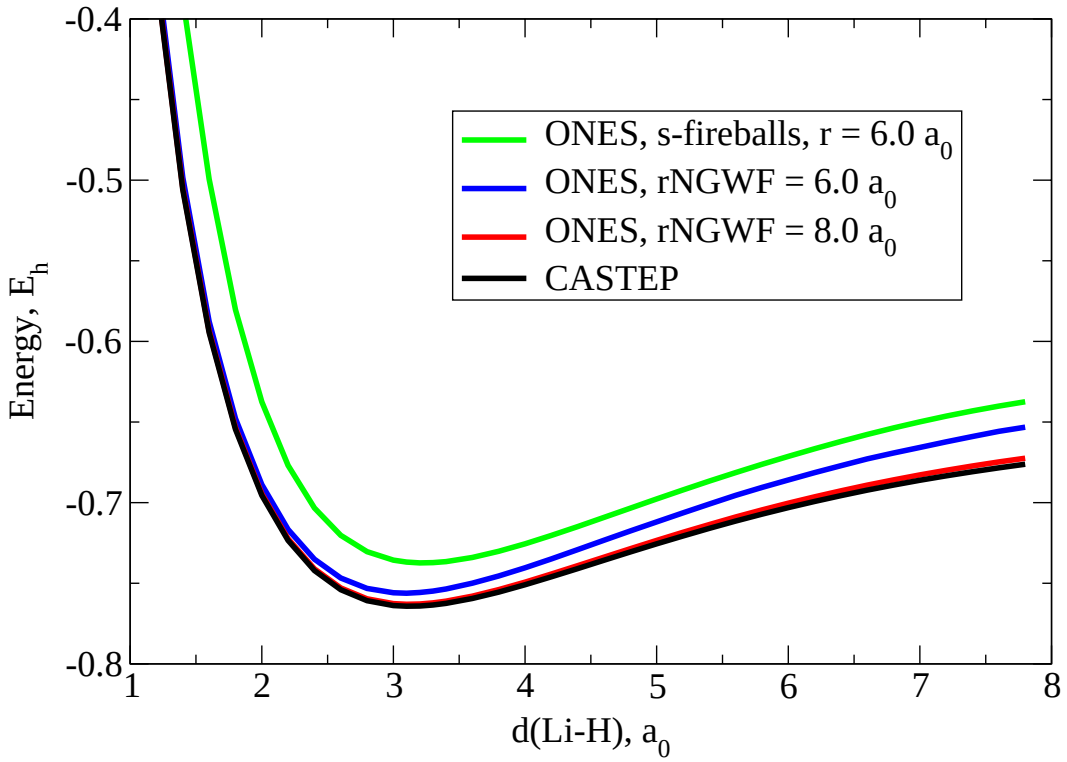
Convergence of energy with NGWF radius

- SiH_4 in $(40\ a_0)^3$ cell
- Energy converged when its rms gradient with respect to the NGWFs is $< 10^{-6}$
- Grid spacing 0.5 a_0 ($=$ 40 Ry)
- Norm-conserving pseudopotentials

rNGWF (a_0)	Energy (Ry)
5.0	-12.419298
6.0	-12.448340
7.0	-12.455436
8.0	-12.457244
9.0	-12.457674

Potential energy curve of LiH

- Norm-conserving pseudopotentials
- Kinetic energy cutoff: 40 Ry
- NGWF radii of 6.0 and 8.0 a_0
- Local Density Approximation



Conclusions

- Two schemes for optimizing localized functions:
  1. Localized spherical-waves
     - Systematic improvement of accuracy
     - Necessity for preconditioning and care over tensor properties
     - Non-orthogonality and localization cause problems with iterative methods
  2. Real-space grids
     - Connection with plane-waves
     - Orthogonality overcomes problems with iterative methods
     - FFT box method enables linear scaling