Improving the performance of Density Functional Theory self-consistent cycle
From matrix analysis to eigenvectors evolution

Dr. Edoardo Di Napoli
Aachen Institute for Advanced Study in Computational Engineering Science
Schinkelstrasse 2, 52062 Aachen Germany

19th of July 2010
Synopsis

- Density Functional Theory (DFT) and its self-consistent cycle
  - The Hohenberg-Kohn theorem and the Kohn-Sham scheme
  - Full-potential linearized augmented plane-wave method (FLAPW)

- The new frontier: improving the performance of FLAPW and DFT in general
  - FLAPW self-consistent cycle: where does the time go?
  - Plan of action: the new computational philosophy
  - Eigenvectors evolution
  - Matrix generation
  - Operations optimization
  - High performance direct eigensolvers

- Research milestones and conclusions
  - Five years from now
  - Conclusions and acknowledgments
Synopsis

- Density Functional Theory (DFT) and its self-consistent cycle
  - The Hohenberg-Kohn theorem and the Kohn-Sham scheme
  - Full-potential linearized augmented plane-wave method (FLAPW)

- The new frontier: improving the performance of FLAPW and DFT in general
  - FLAPW self-consistent cycle: where does the time go?
  - Plan of action: the new computational philosophy
  - Eigenvectors evolution
  - Matrix generation
  - Operations optimization
  - High performance direct eigensolvers

- Research milestones and conclusions
  - Five years from now
  - Conclusions and acknowledgments
Synopsis

- Density Functional Theory (DFT) and its self-consistent cycle
  - The Hohenberg-Kohn theorem and the Kohn-Sham scheme
  - Full-potential linearized augmented plane-wave method (FLAPW)

- The new frontier: improving the performance of FLAPW and DFT in general
  - FLAPW self-consistent cycle: where does the time go?
  - Plan of action: the new computational philosophy
  - Eigenvectors evolution
  - Matrix generation
  - Operations optimization
  - High performance direct eigensolvers

- Research milestones and conclusions
  - Five years from now
  - Conclusions and acknowledgments
Teaser

DON'T PANIC
Hohenberg-Kohn theorem

Definitions

- $\hat{H}_v = \hat{T} + \hat{W} + \hat{V}$
  - Hamiltonian operator in second quantization form
- $\hat{H}_v |\Psi\rangle = E_v |\Psi\rangle$
  - Unknown eigenstates of Hamiltonian
- $\hat{V} = \sum_a \int d\mathbf{r} \hat{\psi}_a^\dagger(\mathbf{r}) v(\mathbf{r}) \hat{\psi}_a(\mathbf{r})$
  - External potential (second quantized)
- $n(\mathbf{r}) = N \sum_a \int dx_2 \cdots dx_N |\Psi(\mathbf{r}; a, x_2, \ldots, x_N)|^2$
  - One-particle electron density

Theorem statement (1964)

1. $n(\mathbf{r}) \iff v(\mathbf{r})$
   - $n(\mathbf{r})$ is called $v$-representable
2. $E_v = E[n]$ and $E_0 = \min_n E[n]$
   - Unique functional
3. $\langle \Psi[n] | \hat{T} + \hat{W} | \Psi[n] \rangle = F_{HK}[n]$
   - Universal ($v$-independent)

Instead of $|\Psi\rangle$ and $E_v$ we now work with $n(\mathbf{r})$ and $E[n]$. 
The Kohn-Sham scheme allows for a practical application of the Hohenberg-Kohn theorem.

**Assumption:** interactive \( \nu \)-representable \( n(\mathbf{r}) \) are also non-interactive \( \nu \)-representable

\[
E_I[n] = \mathcal{F}_{HK}[n] + \int v_I(\mathbf{r})n(\mathbf{r}) \quad \iff \quad n(\mathbf{r}) \quad \iff \quad E_0[n] = T[n] + \int v_0(\mathbf{r})n(\mathbf{r})
\]

---

### The self-consistent scheme

1. \( v_0([n], \mathbf{r}) = v_I(\mathbf{r}) + \int w(\mathbf{r}, \mathbf{r}')n(\mathbf{r}) + \nu_{xc}([n], \mathbf{r}) \)
2. \( -\frac{\hbar^2}{2m} \nabla^2 + v_0(\mathbf{r}) \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad ; \quad \varepsilon_1 \leq \varepsilon_2 \leq \ldots \)
3. \( n(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2 \)
4. \[
\begin{align*}
\nu_{xc}([n], \mathbf{r}) &= \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} & \text{exchange pot.} \\
E_{xc}[n] &= \mathcal{F}_{HK}[n] - \frac{1}{2} \int n(\mathbf{r}')w(\mathbf{r}, \mathbf{r}')n(\mathbf{r}) - T[n]
\end{align*}
\]
Full-Potential Linearized Augmented Plane-Wave Method (FLAPW)

Linearized augmented plane wave in a Muffin Tin (MT) geometry

\[ \phi_{k,v}(r) = \sum_{|G+k| \leq K_{\text{max}}} c_{k,v}^G \phi_G(k, r) \]

\[ k \quad \text{Bloch vector} \]

\[ v \quad \text{band index} \]

\[ e^{i(k+G)r} \]

\[ \sum_{l,m} \left[ a_{lm}^{\alpha,G}(k) u_{l}^{\alpha}(r) + b_{lm}^{\alpha,G}(k) \dot{u}_{l}^{\alpha}(r) \right] Y_{lm}(\hat{r}_{\alpha}) \]

Interstitial

Muffin Tin

Boundary conditions

Continuity of wavefunction and its derivative at MT boundary

\[ a_{lm}^{\alpha,G}(k) \quad \text{and} \quad b_{lm}^{\alpha,G}(k) \]
Full-Potential Linearized Augmented Plane-Wave Method (FLAPW)

No shape approximations

\[ V(\mathbf{r}) = \begin{cases} \sum G \, V_I^G e^{iG \mathbf{r}} & \text{Interstitial warped potential} \\ \sum_{lm} V_{MT}^{lm}(r) Y_{lm}(\hat{\mathbf{r}}_{\alpha}) & \text{MT (includ. non spherical terms)} \end{cases} \]

Pseudocharge method (Weinert 1981)

- \( n_{MT}(\mathbf{r}_{\alpha}) \implies q_{l,m} \) (multiple moments) & \( n_I(\mathbf{r}) \implies V_I(\mathbf{r}) \) (interstitial potential)

- \( n_{MT}(\mathbf{r}_{\alpha}) \text{ subs with } \tilde{n}_{MT}(\mathbf{r}_{\alpha}) \) rapidly convergent same \( q_{l,m} \) \( \implies V_I(\mathbf{r}) = \sum_{G \neq 0} \frac{4\pi n(G)}{G^2} e^{iG \mathbf{r}} \)

- \( V_I(\mathbf{r}) \big|_{\partial_{MT} @ \alpha} \implies V_{MT}(\hat{\mathbf{r}}_{\alpha}) \) Green’s function method
Generalized eigenvalue problems

In the FLAPW framework (as in many other implementations of DFT) the Kohn-Sham equations translate to generalized eigenvalue equations

\[ \phi_{k,v}(r) = \sum_{|G+k| \leq K_{\text{max}}} c_{k,v}^G \phi_G(k,r) \quad \rightarrow \quad \hat{H} \phi_{k,v}(r) = \lambda_{k,v} \phi_{k,v}(r) \]

\[ \sum_{G'} [H_{GG'}(k) - \lambda_{k,v} S_{GG'}(k)] c_{k,v}^{G'} = 0 \]

The matrices

The Hamiltonian \( H(k) \) and overlap matrices \( S(k) \) entries are now given by multiple integrals and sums

\[ \{H(k), S(k)\} = \sum \int \int \phi_G^*(k,r) \{\hat{H}, \hat{I}\} \phi_{G'}(k,r) \]
**FLAPW self-consistent cycle**

- **Spherical \( V_0(r) \)**
  - \( n_0(r) \) init.

- Calculation of full-potential \( V(r)[n] \)
  - Convergence check
    - \( n(r) \) mixing
      - \( n(r) \) for next cycle

- Determining the Lapw wavefunctions \( \psi_G(k, r) \)
  - Matrices generation
    - \( A_k = \langle \psi(k) | H | \psi'(k) \rangle \)
    - \( B_k = \langle \psi(k) | S | \psi'(k) \rangle \)

- Generalized eigenproblems solution
  - \( A_k x = \lambda B_k x \)
    - Eigenpairs selection
    - Fermi energy calculation
  - Eigenproblems: distribution and setup

---

Dr. Edoardo Di Napoli (AICES)  
Improving Density Functional Theory performance  
19th of July 2010 9 / 29
Did you already panic?

**Part one**
Foundation of DFT and FLAPW method

**Part two**
- FLAPW self-consistent cycle: where does the time go?
- Plan of action: the new computational philosophy
- Eigenvectors evolution
- Matrix generation
FLAPW self-consistent cycle

- **Spherical** $V_0(r)$
  - $n_0(r)$ init.

- Calculation of full-potential $V(r)[n]$
  - Determining the Lapw wavefunctions $\psi_G(k, r)$

- Matrices generation
  \[ A_k = \langle \psi(k) | H | \psi'(k) \rangle \]
  \[ B_k = \langle \psi(k) | S | \psi'(k) \rangle \]

- Generalized eigenproblems solution
  \[ A_k x = \lambda B_k x \]

- Eigenpairs selection
- Fermi energy calculation
- Convergence check
  - $n(r)$ mixing
    - $n(r)$ for next cycle

- Calculation of new charge density $n(r)$
  - $a^G_{lm}$ and $b^G_{lm}$ coeff.
  - 3-D stars, 2-D stars lattice harmonics

- Pseudo-charge method
FLAPW self-consistent cycle

Spherical $V_0(r)$ \(\downarrow\) $n_0(r)$ init.

Calculation of full-potential $V(r)[n]$

Spherical $V_0(r)$ \(\downarrow\) $n_0(r)$ init.

Calculation of full-potential $V(r)[n]$

Convergence check $n(r)$ mixing \(\downarrow\) $n(r)$ for next cycle

Determining the Lapw wavefunctions $\psi_G(k, r)$

Calculation of new charge density $n(r)$

Matrices generation

- $A_k = \langle \psi(k)|H|\psi'(k) \rangle$
- $B_k = \langle \psi(k)|S|\psi'(k) \rangle$

Generalized eigenproblems solution $A_k x = \lambda B_k x$

Fermi energy calculation

Eigenproblems: distribution and setup

Eigenpairs selection

$\sim 40\%$ time usage

$\sim 5-10\%$ time usage

$< 1\%$ time usage
A different philosophy
Performance and Speedup

A series of correlated problems
- DFT set of \textit{generalized eigenproblems} \quad \Rightarrow \quad P_1 \cdots P_i P_{i+1} \cdots P_N \not\Rightarrow (P)^N
- This series \{P_i\} of problems should eventually converge to a stable solution

Current approach:
\begin{itemize}
  \item Calculate everything \quad \Rightarrow \quad \text{Accurate but time consuming}
  \item Libraries as black boxes \quad \Rightarrow \quad \text{Standard but sub-optimal}
\end{itemize}

A counter-intuitive action
\begin{itemize}
  \item Save on calculations \quad \Rightarrow \quad \text{Sloppy but fast!}
  \item Improve performance of iterations \quad \Rightarrow \quad \text{tailored algorithms}
  \item Take advantage of hybrid architectures \quad \Rightarrow \quad \text{scalable algorithms}
  \item Optimize calculation \quad \Rightarrow \quad \text{use optimized sub-libraries}
\end{itemize}
A different philosophy
Performance and Speedup

A series of correlated problems

- DFT set of generalized eigenproblems $\Rightarrow P_1 \ldots P_i P_{i+1} \ldots P_N \nRightarrow (P)^N$
- This series $\{P_i\}$ of problems should eventually converge to a stable solution

Current approach:

- Calculate everything $\Rightarrow$ Accurate but time consuming
- Libraries as black boxes $\Rightarrow$ Standard but sub-optimal

A counter-intuitive action

- Save on calculations $\Rightarrow$ Sloppy but fast!
- Improve performance of iterations $\Rightarrow$ tailored algorithms
- Take advantage of hybrid architectures $\Rightarrow$ scalable algorithms
- Optimize calculation $\Rightarrow$ use optimized sub-libraries
The process of convergence

Area of converged solution \( n_f(r) \)

Full-data convergence process

Charge density manifold

Starting density \( n_0(r) \)
The process of convergence

Area of converged solution $n_f(r)$

Charge density manifold

Starting density $n_0(r)$

- Approximate-data convergence process
- Full-data restart
- Full-data convergence process
Four paths, same philosophy

Matrix updating
- A considerable amount of matrix entries do not undergo a significant change from one iteration to the next
  \[\Rightarrow\]
- Saving on computations that update the Hamiltonian matrix A and the overlap matrix B

Eigenvector evolution
- At fixed $k$, subspace angles between eigenvectors from adjacent iterations are considerably small
  \[\Rightarrow\]
- Eigenvectors from previous iteration can be re-used to find lowest set of eigenpairs needed to calculate $n(r)$

HP libraries
- New hybrid architectures require new computational paradigms
  \[\Rightarrow\]
- New scalable algorithms for standard and generalized eigensolvers can optimize access to resources and speed up computations

Optimization
- Low-level sub-routines push standard operations to peak levels
  \[\Rightarrow\]
- Intelligent self-tuning procedure can recast most operations in terms of fast sub-routine calls

Dr. Edoardo Di Napoli (AICES)  Improving Density Functional Theory performance  19th of July 2010  14 / 29
Eigenvectors evolution
Subspace deviation angle (2nd lowest eigenvalue)
$\langle \theta_{sub} \rangle = 3.92\% \quad (100\% = \frac{\pi}{2})$

$\sigma(\theta_{sub}) = 13.50\%$
A mixed solver

Research in progress

1. Initial few iterations $P_1 \ldots P_i \ i \ll N$ using the Direct Solver
2. A series of iterations $P_i \ldots P_{j-1}$ is performed an using Iterative Solver
3. Re-start with one iteration $P_j$ using the Direct Solver
4. Perform a new series of iterations $P_{j+1} \ldots P_{k-1}$ using an Iterative Solver

: 

(m) ... until convergence

Iterative solver: more work to be done

1. Implicit Restarted Arnoldi Method $\rightarrow$ Krylov-Schur process (Stewart)
2. Green function integration method $\rightarrow$ work in progress on FEAST
3. SLEPc $\leftarrow$ Subspace Iteration method $\rightarrow$ development of new algorithm
Re-using Eigenpairs
A simple example using ARPACK

N=50

LAPACK (QR) vs ARPACK (IRAM)

N=400

LAPACK (QR) vs ARPACK (IRAM)

Dr. Edoardo Di Napoli (AICES)  Improving Density Functional Theory performance  19th of July 2010  18 / 29
Matrix updating

\[
\begin{align*}
\text{iter=3} & \uparrow 10\% & \text{iter=4} & \uparrow 10\% & \text{iter=3} & \uparrow 5\% & \text{iter=4} & \uparrow 5\%
\end{align*}
\]
Matrix updating

iter=3 ↑ 10% → iter=4 ↑ 10% → iter=3 ↑ 5% → iter=4 ↑ 5%
Matrix updating

iter=3 ↑ 10% → iter=4 ↑ 10% → iter=3 ↑ 5% → iter=4 ↑ 5%
Matrix updating

iter=3 → 10% → iter=4 → 10% → iter=3 → 5% → iter=4 → 5%
Matrix updating
How to take advantage of it

Question: does partial updating influence distribution of matrix entries among processors?

Answer 1: No, if the new **Elemental** library (in developing phase) is used
Answer 2: No, if updating the entries is already done in a distributed fashion

Research directions

- Study which “metric” most significantly normalizes the changes below threshold
- Conduct a quantitative study on the relation between threshold value, speedup and accuracy of solution
- Develop an automated smart slider that adjusts the parametric threshold for optimized trade-off b/w speed and accuracy
- Last but not least: is there a physical reason behind the unchanging patterns? Can we uncover it symbolically?
Still not panicking?

Part two
The new philosophy: eigenvector evolution and matrix generation

Part three
- Operations optimization
- High performance direct eigensolvers
- Five years from now
- Acknowledgments

**Matrix mult.**

\[
A^\alpha_{lm}(k) = \sum_{G} c_{k,\nu}^G \alpha^\alpha, G_{lm}(k) \\
B^\alpha_{lm}(k) = \sum_{G} c_{k,\nu}^G b_{lm}^\alpha, G_{lm}(k)
\]

**in Muffin-tin**

\[
n(r) = \int \sum_{\varepsilon_{\nu} < E_F} |\phi_{k,\nu}(r)|^2 d^3k = \sum_{k,\nu} |\phi_{k,\nu}(r)|^2 \omega(k,\nu) \\
\phi_{k,\nu}(r) = \sum_{k,\nu} [A^\alpha_{lm}(k)u^\alpha_l(r) + B^\alpha_{lm}(k)\dot{u}^\alpha_l(r)] Y_{lm}(\hat{r}_\alpha)
\]

**MMM Algorithms**

- Usual implementation: Triple Loop
- Goto BLAS implementation (DGEMM)

160 x of Triple Loop
Matrix-Matrix Multiplication (MMM) on quadcore Intel platform

Performance [Gflop/s]

- **Best implementation**

![Graph showing performance improvement]

- **Triple loop**

- **Same (mathematical) operation count (2n³)**

Courtesy of Markus Püschel
Towards a HP direct eigensolver
Elemental vs ScaLAPACK

Complex lower Cholesky factorization (pzpotrf)

```
GFlops(m, t) = \frac{4}{3} \frac{m^3}{10^4 t}
```

- Elemental
- ScaLAPACK (MKL 10.2.2.025)

Speedup

Dimension $\cdot 10^4$

- 0
- 0.5
- 1
- 1.5
- 2
- 2.5
- 3

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
Towards a HP direct eigensolver
Elemental vs ScaLAPACK

Complex $A := L^{-1} AL^{-H}$ (pzhegst)

GFlops($m,t) = \frac{4m^3}{10^9}$

- Elemental
- ScaLAPACK (MKL 10.2.2.025)

Speedup

Dimension

19th of July 2010
Towards a HP direct eigensolver
A fast and scalable tridiagonal eigensolver

![Graph](image)

- **MR³-SMP**
- **MRRR (MKL)**
- **MRRR (LAPACK)**
- **DC (MKL)**

Dr. Edoardo Di Napoli (AICES)

Improving Density Functional Theory performance

19th of July 2010 26 / 29
In the next five years

**HP self-consistent cycle**
- Matrix generation
- Matrix structure analysis
- New iterative solvers
- Mixed direct-iterative numerical methods
- Generalization to DFT methods à la Car-Parrinello

**Parallel libraries for eigensolvers (coll. with AICES)**
- Tridiagonal eigensolvers
- Reduction to tridiagonal
- Transform to standard
- Backtransformation

**HP Numerical tensors methods**
- Algebraic operations
- Parametric dependence
- Differential operations
Acknowledgments

Thanks to...

- JARA-SIM for supporting this research
- Prof. P. Bientinesi and his group at AICES (M. Petschow and D. Fabregat) for continuous collaborative support and encouragement
- Prof. S. Blügel and Dr. D. Wortmann for their collaboration efforts and patience
- Prof. C. Honerkamp and Prof. S. Andergassen for their hospitality and friendship
- The RWTH Physics department for hosting this seminar and their support, together with the Forschungszentrum Jülich, for my Helmholtz application
Did you finally panic?

Always Know Where Your Towel Is