Multigrid DFT calculations, optimized localized orbitals and nearly *O(N)* calculations of quantum transport

J. Bernholc,[1] E. L. Briggs,[1] M. Buongiorno Nardelli,[1]
J.-L. Fattebert,[2] V. Meunier,[1] S. Nakhmanson,[1]
W. G. Schmidt,[1] and Q. Zhao[1]

1. North Carolina State University, Raleigh, USA
2. CASC, Lawrence Livermore National Laboratory, CA 94551

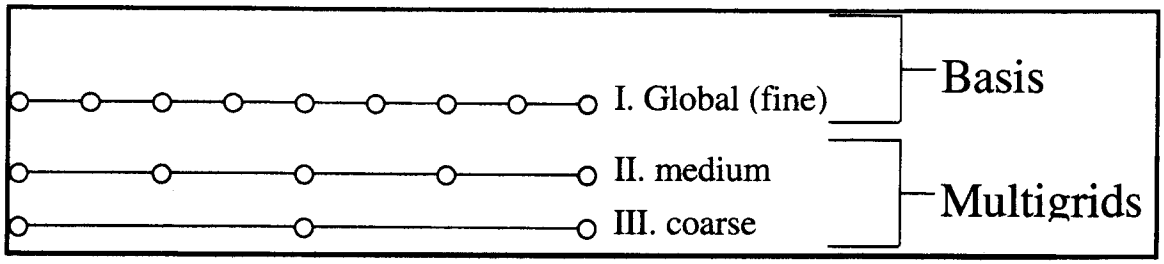
I. DFT on a grid
   - Discretization, multigrid, tests
   - Massively parallel implementation
   - Tests and example large-scale applications
     - strength, pyro- and piezo-electricity in nanotubes
     - reflectance anisotropy for real-time feedback control of the growth of semiconductors

II. Non-orthogonal, optimally-localized orbitals on a grid
   - Massively parallel
   - Nearly *O(N)* calculations
   - Test results and scaling

III. *O(N)* quantum transport
   - Matrix Green's function formalism
   - Massively parallel
   - Application to "real" systems
     - nanotube-metal contacts
     - nanotube/cluster molecular sensors

IV. Summary

Real-space multigrid method for quantum simulations



density functional equations solved directly on the grid

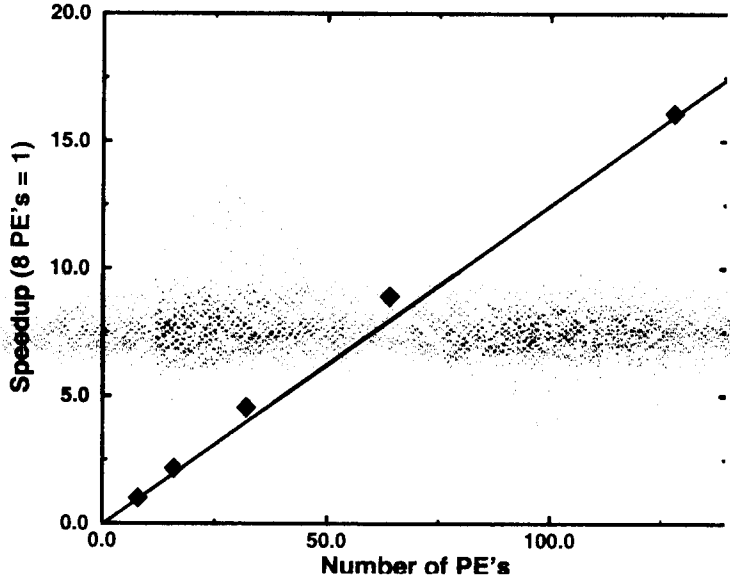
Multigrid techniques remove instabilities by working on one length scale at a time

Automatic preconditioning and convergence acceleration on all length scales

Non-periodic boundary conditions are as easy as periodic

Compact, non-diagonal "Mehrstellen" discretization

Allows for efficient massively parallel implementation



Speedup nearly *linear* in the number of processors (up to 1024)

A. Speed on Cray T3E

Double precision code

77 Mflops/processor

Mixed precision code

177 Mflops/processor

42 Gflops on 256 T3E processors

See E. L. Briggs, D. J. Sullivan and J. Bernholc Phys. Rev. B 54, 14362 (96).

Compact Real-Space Discretization

Higher accuracy achieved by using more local information.
***Local* nature also important for MPP implementations.**

Both $\nabla^2\psi$ and V are discretized along several grid points in each coordinate direction (we use 3 points per direction)

leads to a generalized eigenvalue problem

$$A[\psi_i] + B[V\psi_i] = \varepsilon_i B[\psi_i] + O(h^4)$$

A* : kinetic energy operator to second order in *h
B* : smoothing operator, *I* to second order in *h
***A* and *B* are components of the *Mehrstellen* discretization.**

Two dimensional stencil form: $12h^2 A$

-2	-8	-2
-8	40	-8
-2	-8	-2

$12B$

	1	
1	8	1
	1	

We derived and used a 3D Mehrstellen for orthorhombic grids: suitable for a wide range of problems but not all.

We have now derived 3D Mehrstellen for additional symmetries
- **Hexagonal**
- **Body-centered**
- **Face-centered**

(To be published: E. L. Briggs, J. L. Fattebert, J. Bernholc)

1D example for Poisson's equation $\Phi''(x) = f(x)$

- In central finite differences

Taylor expansion of $\Phi(x)$ gives:

$$\Phi''(x_i) = h^{-2}|\Phi(x_i-h) - 2\,\Phi(x_i) + \Phi(x_i+h)| + O(h^2)$$

$$h^{-2}|\Phi(x_i-h) - 2\,\Phi(x_i) + \Phi(x_i+h)| = f(x_i) + O(h^2)$$

- In Mehrstellen approach

Taylor expansion of $\Phi(x)$ *and* $f(x)$, and $\Phi''(x) = f(x)$ leads to

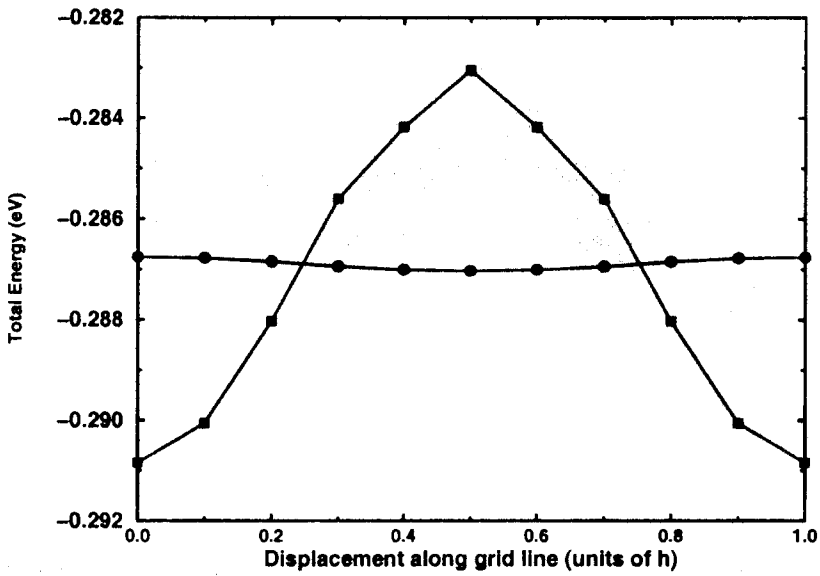
$$h^{-2}|12\Phi(x_i-h) - 24\,\Phi(x_i) + 12\Phi(x_i+h)| = f(x_i-h) + 10f(x_i) + f(x_i-h) + O(h^4)$$

Pseudopotential representation

- In *plane wave* calculations
  V(G) with G > cutoff are automatically neglected

- In real-space calculations standard potentials need to be Fourier-filtered.

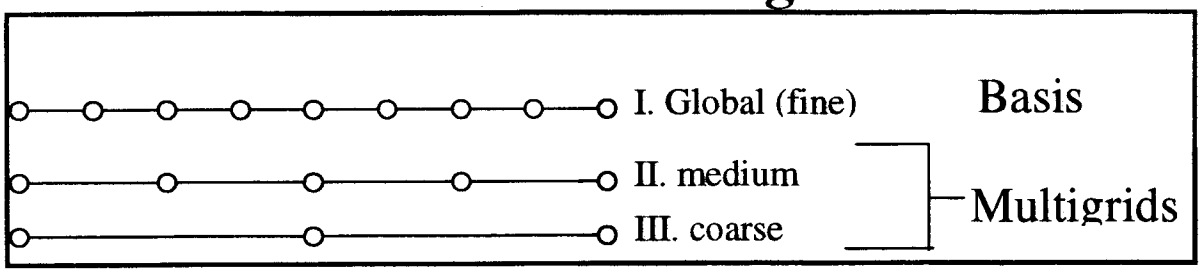
e.g., Cutoff potentials in G-space with a gaussian function Fourier-transform and remove small oscillations at large *r* with R-space gaussian cutoff



Fourier-filtered potentials retain the *accuracy* of plane-waves.

- General rule:
  Avoid high-frequency variation on real-space meshes.
  Slower convergence
  Lower accuracy

Outline of multigrid



1. Select trial ψ_n and compute initial estimate of ε_n

$$\varepsilon_n = \langle\psi_n|H_{mehr}|\psi_n\rangle / \langle\psi_n|B_{mehr}|\psi_n\rangle$$

2. Compute residual vector: $r_h = \varepsilon_n B_{mehr}|\psi_n\rangle - H_{mehr}|\psi_n\rangle$

3. Do several steepest descents updates on the global grid

$$|\psi_n^{new}\rangle = |\psi_n\rangle + \Delta t\, r_h$$

high frequency errors are efficiently removed

4. Transfer the *smooth* residual to a coarse grid using a weighted average

5. Solve $\nabla^2 b_{2h} = r_{2h}$ on the coarse grid.

Note that the potential terms are represented only by r_{2h}.

6. Interpolate b_{2h} to global grid and correct $|\psi_n\rangle$

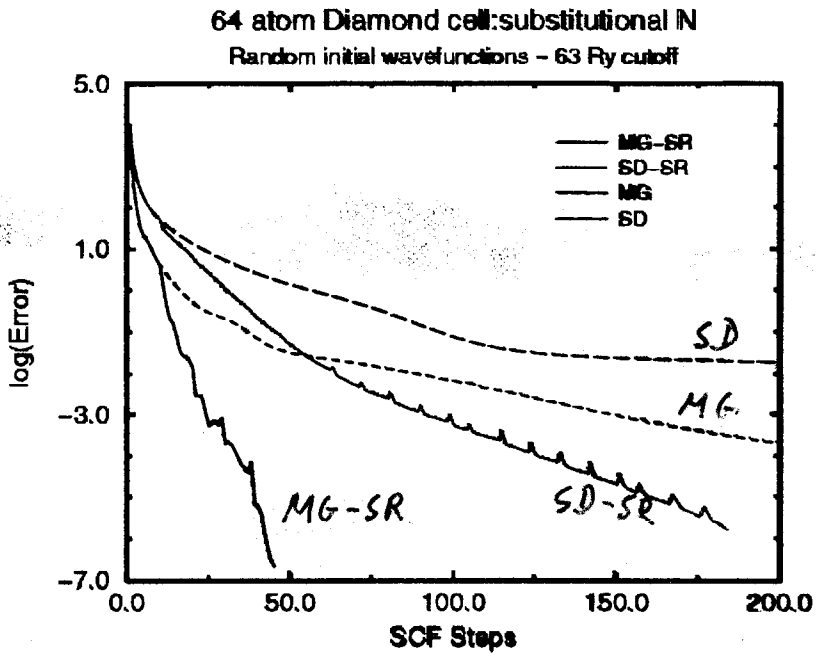
$$|\psi_n^{new}\rangle = |\psi_n\rangle + \Delta t\, b_{2h}$$

Repeat for all $|\psi_n\rangle$, orthogonalize and mix density until self-consistent.

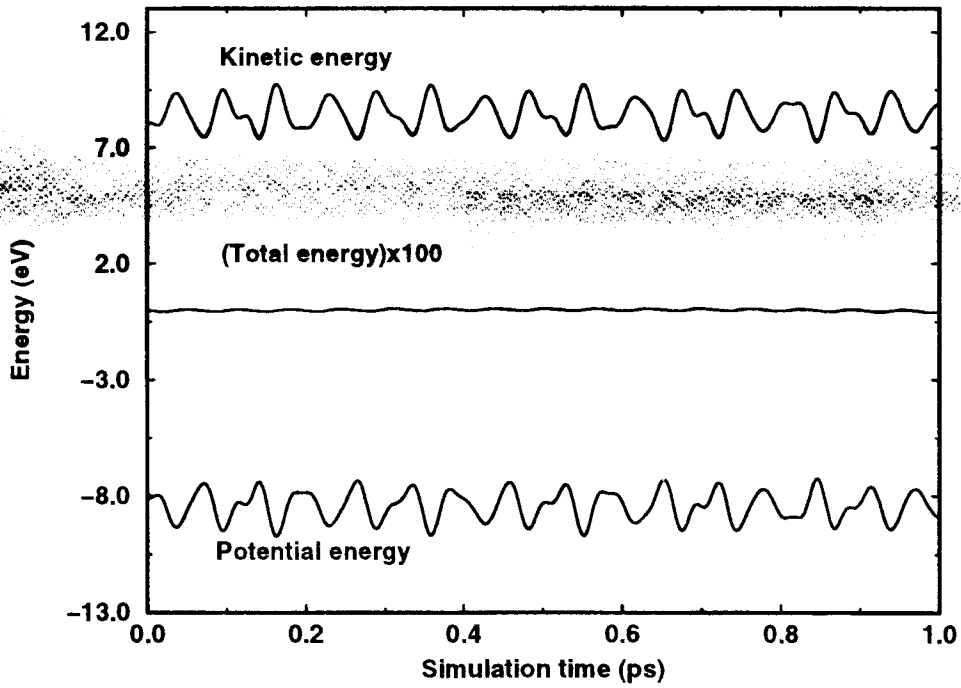
Subspace diagonalization every 8-10 steps significantly improves convergence.

Convergence and accuracy

- Convergence rates for *steepest descents* and *multigrid* with and without *subspace diagonalization*



- Energy conservation during a molecular dynamics run 64 Si atoms at 1100K, $\Delta t = 80$ a.u., $\Delta E_{MAX} = 27$ μV/atom



Tests for uniform grids

Effective cutoff **for comparison to plane-wave calculations**

for plane waves **cutoff = PW kinetic energy**

for multigrid **cutoff = $\pi^2/2h^2$**

results in the same FFT grid as the real-space grid

Time per scf/step *comparable* to Car-Parrinello, but *fewer* SCF steps needed for convergence in the multigrid method

- **Perfect crystal and vacancy in diamond (35 Ry cutoff)**

64-atom cells	Car-Parrinello	Multigrid	Experiment
Perfect Crystal			
Band gap	4.53 eV	4.53 eV	5.50 eV
Cohesive energy	8.49 eV	8.54 eV	7.37 eV
Vacancy			
Formation energy	6.98 eV	7.07 eV	
level splitting	0.32 eV	0.32 eV	

max $\Delta(\varepsilon) = 0.06$ eV **max $\Delta(r_{atomic}) = 0.009$ bohr**

- **C60 isolated molecule and solid**

C60, 35 Ry cutoff	d(C=C)	d(C-C)
molecule (multigrid)	1.39 Å	1.44 Å
solid (Car-Parrinello)	1.39 Å	1.45 Å

- **Scaling with respect to system size**

Diamond, 35 Ry	8-atom cell	64-atom cell
# SCF steps	17	20

- **Scaling with respect to cutoff**

8-atom Diamond	25 Ry	35 Ry	60 Ry	110 Ry
# SCF steps	22	17	21	26

see E. Briggs, D. Sullivan, and J. Bernholc, Phys. Rev. B 52, R5471 (1995).

Grid-optimized orbitals for nearly $O(N)$ DFT

Large-scale electronic structure calculations scale as $O(N^3)$

- A lot of current research focuses on $O(N)$ methods, see S. Goedecker, Rev. Mod. Phys. (1999) for a review.
- Our approach is described in J.-L. Fattebert and J. Bernholc, Phys. Rev. B 62, 1713 (2000).
- Work most related to our approach: Galli and Parrinello (1992), Nunes and Vanderbilt (1994), Hernandez and Gillan (1995)

We want to keep:

- Accurate *ab initio* results
- Efficient iterative algorithm using a good preconditioning to relax the wavefunctions
- Inclusion of unoccupied states may increase the convergence rate

Representation of the electrons

- Eigenfunctions:

$$\Psi = \left(\vec{\Psi}_1, \ldots . \vec{\Psi}_N\right)$$

- Basis of orthogonal functions:

$$\chi = (\vec{\chi}_1, \ldots, \vec{\chi}_N) = \Psi U^T$$

where U is an $N \times N$ orthogonal matrix

- Basis of non-orthogonal functions:

$$\Phi = \left(\vec{\phi}_1, \ldots, \vec{\phi}_N\right) = \chi G^T$$

where $G \in \mathcal{M}_N$ is a lower triangular matrix, $S = GG^T$ (Cholesky decomposition), $S = \Phi^T \Phi$

These 3 representations are mathematically completely equivalent, but:

- The evaluation of the electronic density is cheaper in the basis of the eigenfunctions.

- Using non-orthogonal functions allows to impose localization constraints $\rightarrow$ lower cost.

Nonorthogonal basis

Trial non-orthogonal basis: $\Phi = \left(\vec{\phi}_1, \ldots, \vec{\phi}_N\right)$

$\Rightarrow$ Iterative corrections:

$$\delta\vec{\phi}_i = K\left(\sum_{j=1}^{N} B\vec{\phi}_j\Theta_{ij} - H\vec{\phi}_i\right)$$

where $\Theta = (\Phi^T B\Phi)^{-1}(\Phi^T H\Phi)$

K: linear preconditioning multigrid operator

B: "Mehrstellen" operator

NO ORTHOGONALIZATION OR SUBDIAGONALIZATION REQUIRED!

Density:

$$\rho(x) = 2\sum_{i,j=1}^{N} (S^{-1})_{ij}\phi_i(x)\phi_j(x),\ S = \Phi^T\Phi$$

Energy:

$$E = 2Tr(\Theta) + F[\rho] + E_{ion-ion}$$

Multigrid preconditioning

K: iterative multigrid Poisson solver (V-cycle) for the Poisson problem

$$-\nabla^2 u = d$$

with initial solution

$$\mathbf{u}_0 = \alpha \mathbf{d}$$

and limited to the grids with a number of nodes larger than N.

Smoothing: Jacobi

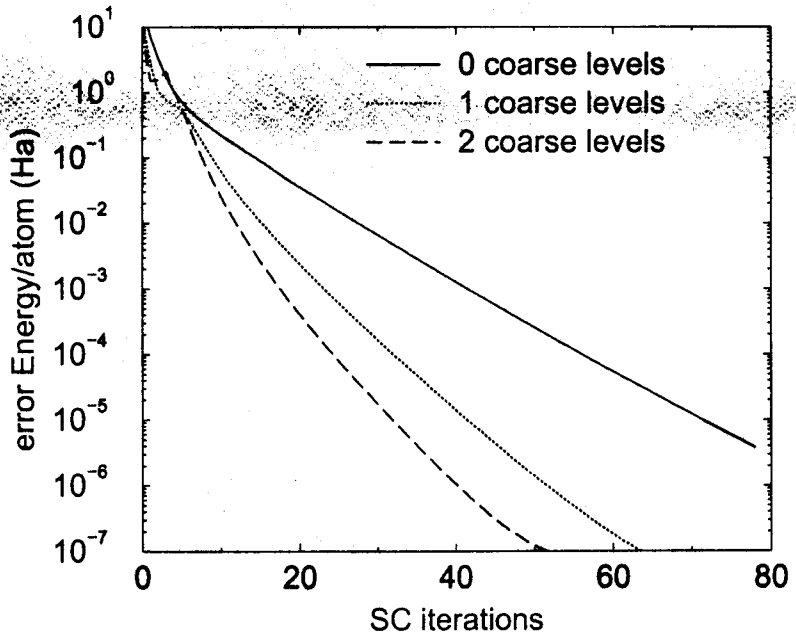
α: approximation of the inverse of the largest eigenvalue of H on the finest non-visited grid

Multigrid preconditioning

Main features:

- Ability to deal with components of different wavelength
- Linear preconditioner $\Rightarrow$ independent of the basis
- parallel
- Low cost ($\sim 15\%$ of total cost)
- works well for $N >$ # of occupied orbitals

Example: C_{60}, grid $64 \times 64 \times 64$, $N = 180$, $N_{occ} = 120$



Minimization with unoccupied states

$N >$ # of occupied orbitals

- Iterative relaxations of the orbitals Ψ as if they were all fully occupied
- Density matrix for Ritz vectors Ψ

$$\rho(x) = 2 \sum_{i=1}^{N} \psi_i(x)\psi_i(x)\theta(\mu - \epsilon_i)$$

$\Rightarrow$ Non-diagonal density matrix $\bar{\rho}_\phi$ in the basis Φ $(N \times N)$

$$\rho(x) = 2 \sum_{i,j=1}^{N} \phi_i(x)\phi_j(x)(\bar{\rho}_\Phi)_{ij}$$

$$\begin{aligned} E &= 2Tr(S\Theta\bar{\rho}_\Phi) \\ &+ F[\rho] + E_{ion-ion} \end{aligned}$$

The evaluation of the exact $\bar{\rho}_\Phi$ requires the diagonalization of a matrix $N \times N$

Localizations constraints on Φ

Functions ϕ_i non-zero only inside spheres of radius R_c centered on an atom

- Linear scaling to compute the corrections $\delta\vec{\phi}_i$

$$\delta\vec{\phi}_i = K\left(\sum_{j=1}^{N} B\vec{\phi}_j\Theta_{ij} - H\vec{\phi}_i\right)$$

- Linear scaling to compute the density ρ

$$\rho(x) = 2\sum_{i,j=1}^{N}\phi_i(x)\phi_j(x)\left(\bar{\rho}_\phi\right)_{ij}$$

- Evaluation of the density matrix $\bar{\rho}_\Phi$ (full) remains $O(N^3)$ (diagonalization of a matrix $N \times N$)

PBLAS and ScaLapack for submatrices

Parallel Basic Linear Algebra Subprograms (PBLAS)

Scalable linear algebra package (ScaLapack)

- Computation of density matrix ($O(N^3)$): ScaLapack to diagonalize a matrix $N \times N$

  (eigenvalues + eigenvectors).

- Memory to store submatrices ($O(N^2)$): distribution on the PEs

Timing for ScaLapack diagonalization subroutine (PSSYEV) on T3E (Alpha EV56 processors 450 Mhz, block size=16)

[seconds]	1 PE	4 PEs	16 PEs	64 PEs
$N = 560$	12.8	4.5	2.6	1.8
$N = 1120$	98	27	11	7
$N = 2240$		175	60	29

Interpretation

The method can be seen as a generalization of *ab initio* methods using LCAO (linear combination of atomic orbitals) or Gaussians basis functions:

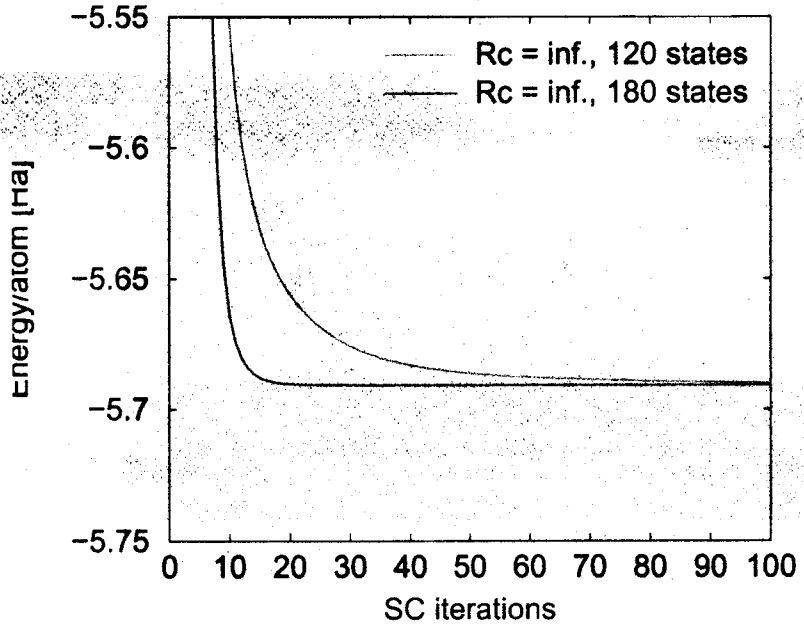
$$\Psi_j = \sum_i c_i \phi_i$$

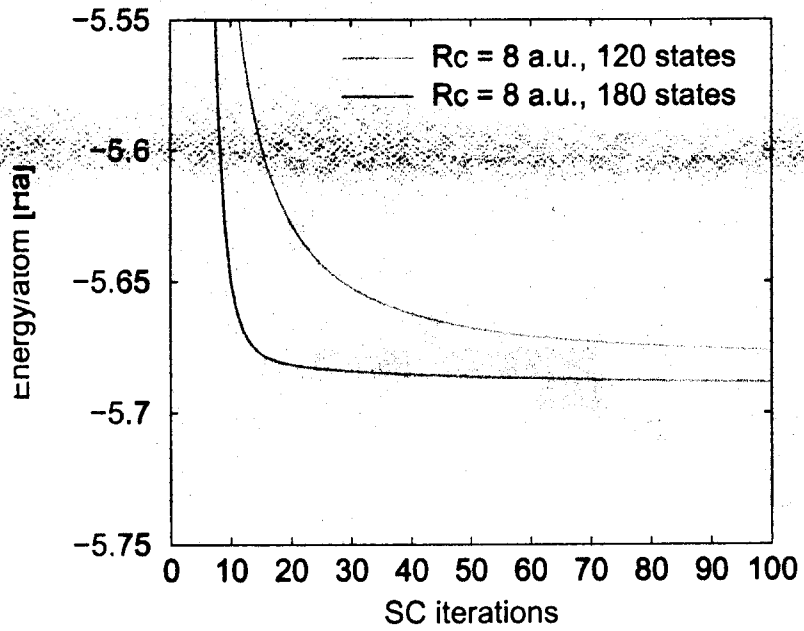
The basis local functions ϕ_i are defined by their values on a grid and are variationally optimized

- More degrees of freedom
- Systematic increase of the accuracy by mesh refinement or expansion of the localization domain
- Numerical integration on a grid

Importance of unoccupied states

Example: C60



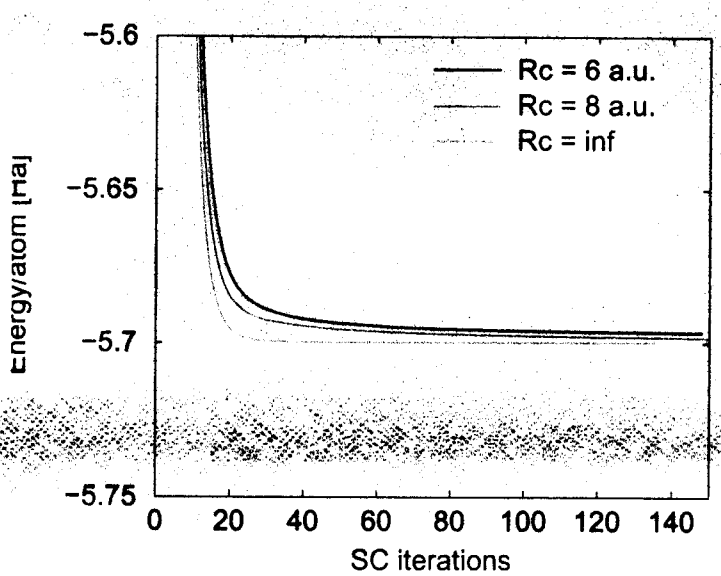


Example: nanotube with 160 carbon atoms

Cell 32.6 × 21.8 × 21.8 a.u., Grid 96 × 64 × 64

Convergence as a function of the localization radius (480 states: 2 occ. + 1 unocc. localized on each atom)

Start: random localized functions



Eigenvalues and band gap

Electronic properties depend on the gap (eigenvalue difference) between the highest occupied orbital (HOMO) and the lowest unoccupied orbital

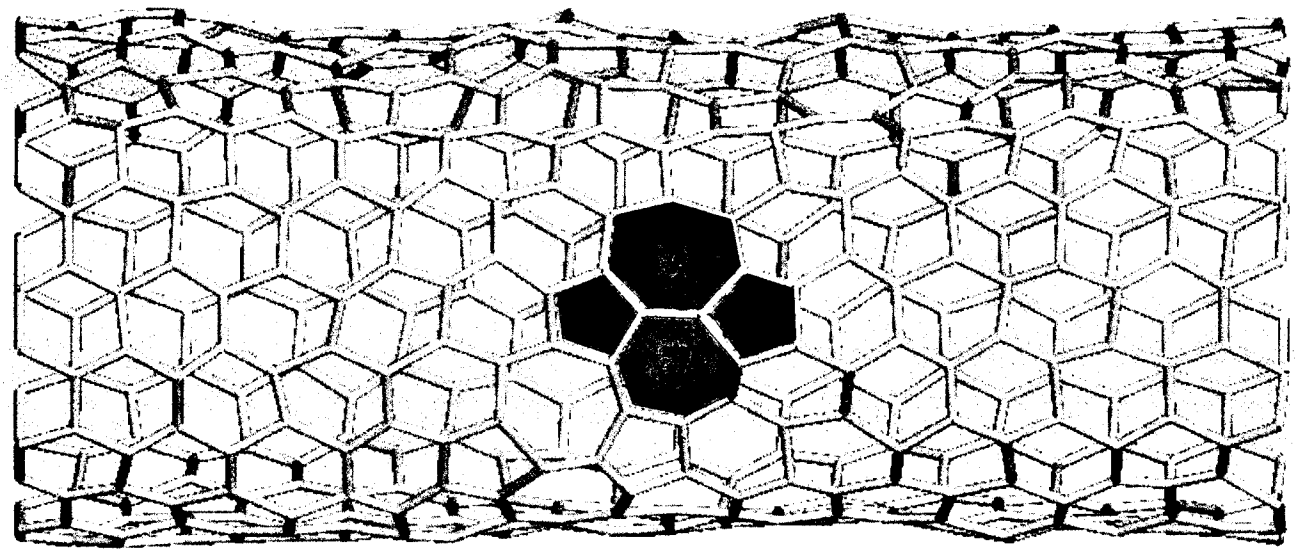
Example: Nanotube with 160 carbon atoms

[eV]	$R_c = 5a.u.$	$R_c = 6a.u.$	$R_c = 8a.u.$	$R_c = \infty$
LUMO	1.53	1.27	1.17	1.17
HOMO	0.64	0.41	0.32	0.32
gap	0.89	0.86	0.85	0.85

Ionic relaxation

Successful ionic relaxation of a nanotube with 320 atoms, $R_c = 6a.u.$ (agrees with full DFT results)

Energy of a defect



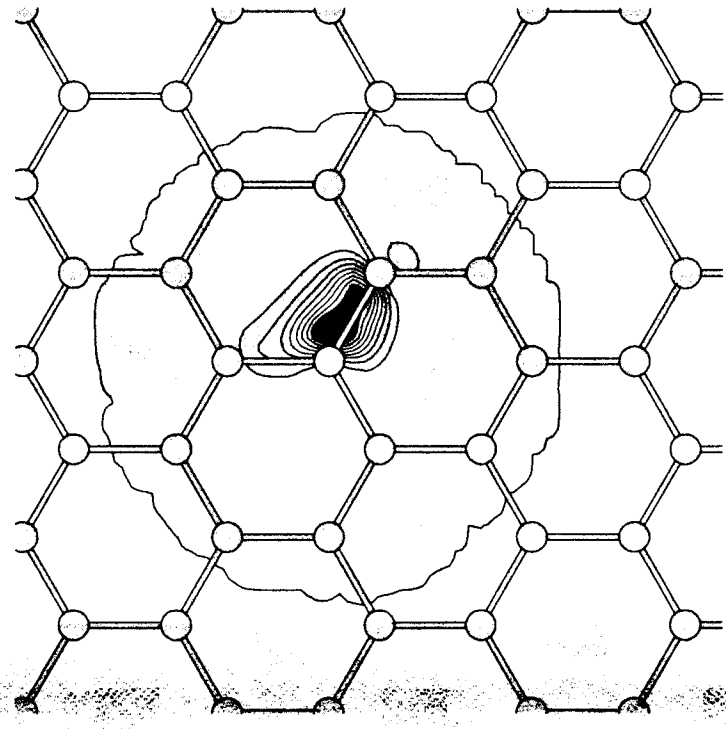
Results for nanotube (10,0) with 160 atoms (with ionic relaxation)

[Ha]	$R_c = 8 a.u.$	$R_c = \infty$
Defect	-911.84	-911.86
No defect	-911.97	-911.99

Defect energy: 0.13 [Ha]

Illustration: carbon nanotube

Isosurface for the square of a localized function in the plane of the nanotube, localization radius = 6 a.u.



- Smoothly decaying functions
- Localized on the bond

Timing

- Implementation in C on CRAY-T3E
- Based on BLAS, Lapack, BLACS, PBLAS, ScaLapack and shmem libraries

Timing for 1 SC step (T3E, DEC alpha 450 MHz Processors, 256 MB RAM)

$R_c = 6.2 a.u.$, 3 orbitals/atom, $h = 0.34 a.u.$: (grid $56 \times 56 \times 96$ for 140 atoms)

# atoms	140	280	560	1120
# orbitals	420	840	1680	3360
# PEs	32	64	128	256
# storage func.	237	252	255	255
CPU time/PE [s]	69	82	104	173
Subdiagonalization [s]	1.4	2.6	9	30

90 Mflops/PE for test on 128 PEs

Optimal localization?

L. He and D. Vanderbilt, PRL 86, 5341 (2001)
Exponential Decay of Wannier Functions and Related Quantities

In 1D, it is shown analytically and numerically that:

$$w(x) \approx x^{-\alpha} e^{-hx}$$

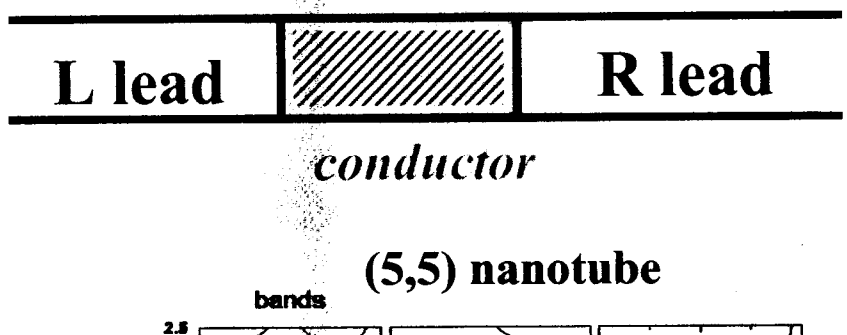
***h* is determined by the potential, but α is universal**

$\alpha =$ *3/4* **for orthonormal Wannier functions (WF)**
1/2 **for** $\rho(x,x')$
1/2 **for projected nonorthonormal WF** ($P = \Sigma_k |\psi_k\rangle\langle\psi_k|$)
3/2 **for dual nonorthonormal WF** $|\overline{\varphi}_i\rangle = \sum_j S^{-1}_{ji} |\varphi_j\rangle$

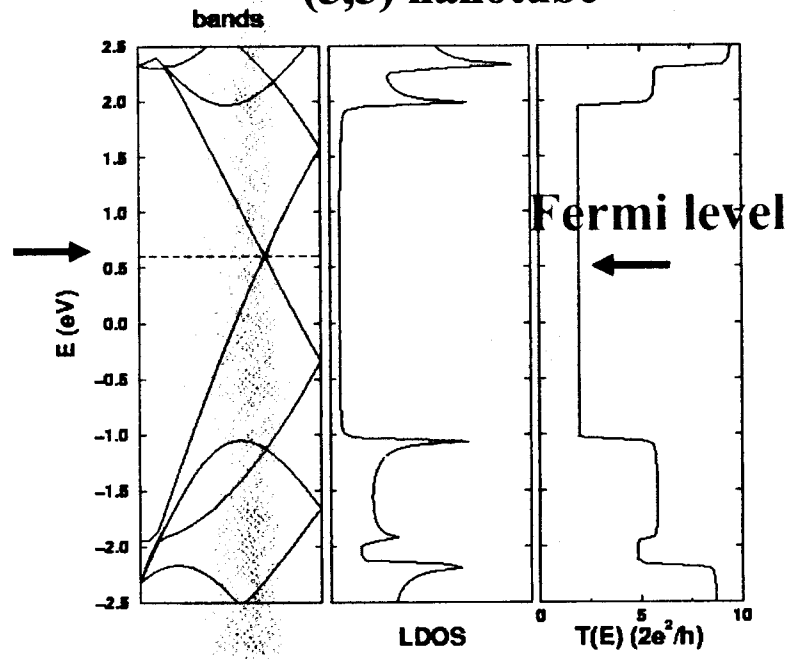
In 3D, the initial slopes are as above, but the authors "suspect that there might be a crossover to larger α values in the far tails."

Introduction to quantum conductance

- Electron transmission through a device requires a Green's function treatment of an *open* system
  Left lead -- Conductor -- Right Lead
- In general, the quantum conductance measures the number of electron channels extending through the conductor *and* the leads, each contributing $2e^2/h$.
- For a perfect metallic nanotube and perfect contacts, both bands at the Fermi level contribute equally.
- For a disordered nanotube or for poor contacts, the conductance is much less.
- Conductances computed using a new, very efficient method (Buongiorno Nardelli PRB 1999; Buongiorno Nardelli and Bernholc, PRB RC 1999, Buongiorno Nardelli, Fattebert, Bernholc, PRB 2001)



(5,5) nanotube



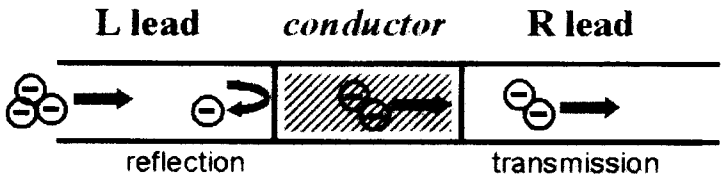
Two bands cross at the Fermi level

$\Rightarrow$ Conductance $\equiv T(E_F) = 2$

Units of $2e^2/h \approx (12.9\ k\Omega)^{-1}$

Green's functions for quantum transport

1. Consider a conductor *C* connected to semi-infinite left (*L*) and right (*R*) leads.



2. Expand in *local* orbitals and divide the system in "layers," so that matrix elements exist only between adjacent layers.

3. In layer "basis," the Green's function equations for the entire *Left lead – Conductor – Right lead* system are:

$$\begin{pmatrix} G_L & G_{LC} & 0 \\ G_{CL} & G_C & G_{CR} \\ 0 & G_{RC} & G_R \end{pmatrix} = \begin{pmatrix} (\varepsilon - H_L) & h_{LC} & 0 \\ h_{CL} & (\varepsilon - H_C) & h_{RC} \\ 0 & h_{CR} & (\varepsilon - H_R) \end{pmatrix}^{-1}$$

4. One can separately solve for the "self-energy" of the left lead:

$$\Sigma_L = h_{CL} g_L h_{LC},$$

where Σ_L has the dimension of *C* and g_L is the Green's function for the semi-infinite lead.

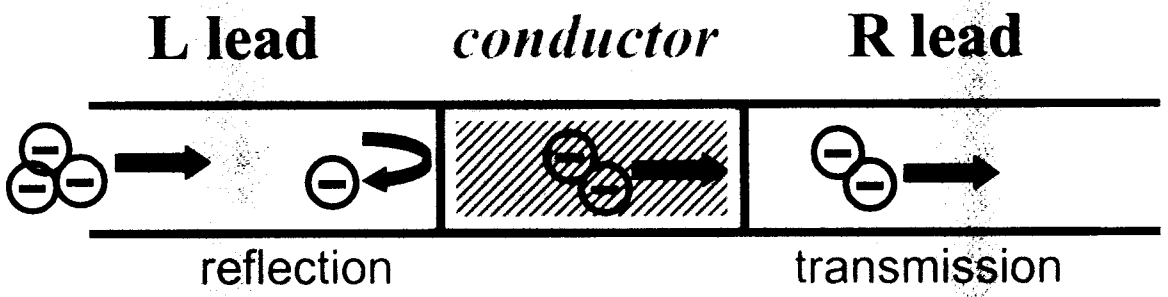
4. $h_{CL} g_L$ interact only across *one* layer, and g_L can be obtained recursively for any periodic lead by recursively doubling the period (see also below).

5. The Green's function for the *Left*-lead-Conductor system becomes

$$G_C = (\varepsilon - H_C - \Sigma_L)^{-1}$$

One can "extend" the lead by "merging" the conductor with the lead, forming a "new"lead. This results in O(N) propagation of G_C in the basis of layer orbitals.

Quantum transport in a local-orbital basis



$$G = \frac{2e^2}{h} T = \frac{2e^2}{h} Tr(\Gamma_L G_C^r \Gamma_R G_C^a)$$

$$G_C = (\varepsilon S_C - H_C - \Sigma_L - \Sigma_R)^{-1}$$

$$\Gamma_{\{L,R\}} = i[\Sigma^r_{\{L,R\}} - \Sigma^a_{\{L,R\}}]$$

- **G_C -- Green's function of the conductor**
- **$\Gamma_{L,R}$ -- coupling functions between the conductor and the leads.**
- **H = Hamiltonian, S = overlap matrix**

- **Self-energies are computed using Surface Green's function matching theory and the concept of layer orbitals:**

$$\Sigma_L = (\varepsilon S_{LC} - H_{LC})^+ (\varepsilon S_{00}^L - H_{00}^L + (\varepsilon S_{01}^L - H_{01}^L)^+ \overline{T_L})^{-1} (\varepsilon S_{LC} - H_{LC})$$

$$\Sigma_R = (\varepsilon S_{RC} - H_{RC}) (\varepsilon S_{00}^R - H_{00}^R + (\varepsilon S_{01}^R - H_{01}^R)^+ T_R)^{-1} (\varepsilon S_{CR} - H_{CR})^+$$