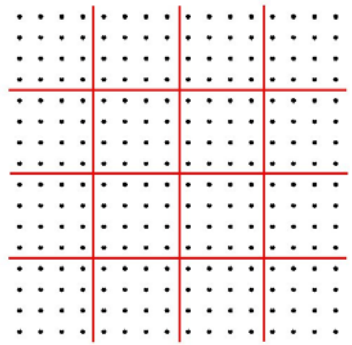
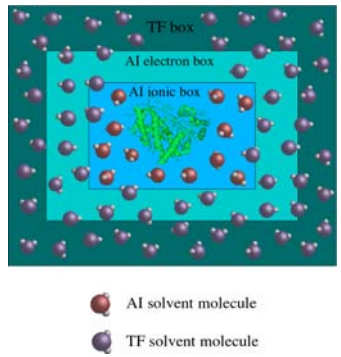
Real-space method with multigrid acceleration (RMG)

parallelization, scaling and applications to electronic structure and quantum transport problems

Wenchang Lu
North Carolina State University

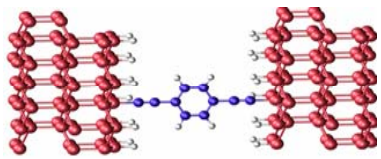


- **Real space method in DFT**
  - **Multigrid acceleration**
  - **Reflectance anisotropy spectroscopy to identify the surface structures**
- **Hybrid simulations in biology: DFT + TF(Thomas-Fermi)**
- **Localized orbitals and nearly O(N) method**
- **Non-equilibrium Green functions in quantum transport**



Examples of applications:

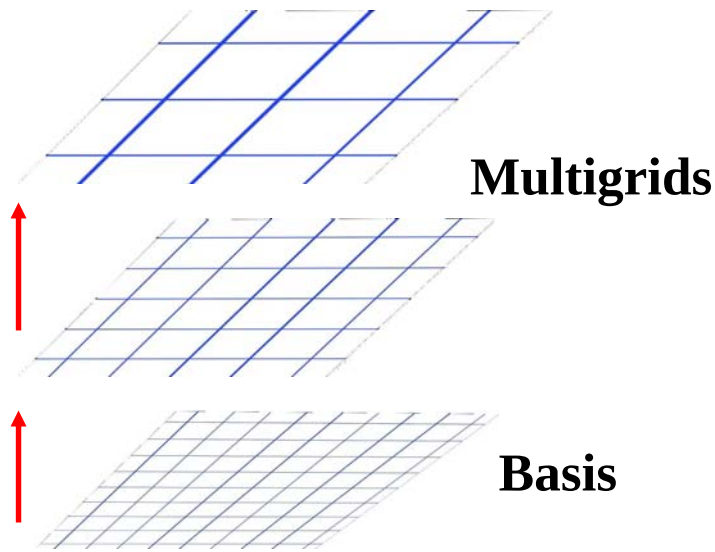
Negative Differential Resistance (NDR) in semiconductor-molecules-semiconductor and metal-molecules-metal juctions.



Collaborators: M. Hodak, V. Meunier, F. Ribeiro, S. Wang, Q. Zhao, G. Schmidt, and J. Bernholc

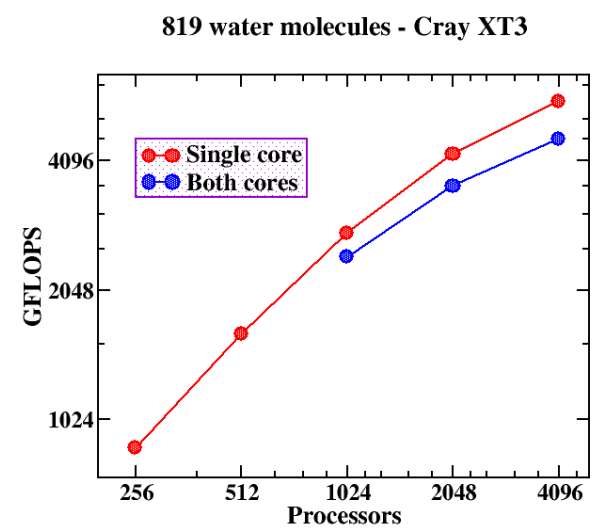
Multigrid method for quantum simulations

- Density functional equations solved directly on the grid
- Multigrid techniques remove instabilities by working on one length scale at a time
- Convergence acceleration and automatic preconditioning on all length scales
- Non-periodic boundary conditions are as easy as periodic
- Compact "Mehrstellen" discretization
- Allows for efficient massively parallel implementation



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

Ultrasoft pseudopotentials:
S. Wang, M. Hodak, W. Lu and J. Bernholc, to be published 2007



Speedup on Cray XT3 with the number of processors 819 water molecules, 41 Ry ultrasoft pseudopotentials

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

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

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	

3D Mehrstellen discretizations for many symmetries of crystal lattices (sc, bcc, fcc, orthorhombic and hexagonal) PRB 54, 14362 (1996).

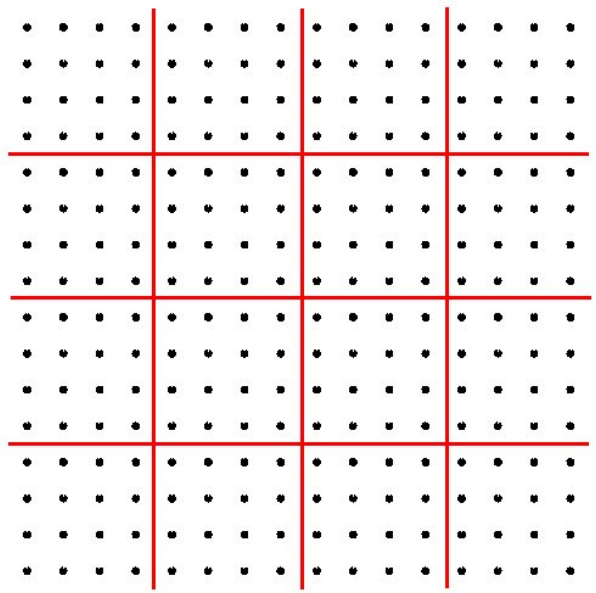
High order Mehrstellen, Wang et al, to be published 2007.

Alternate discretization: Central Finite difference operators
Chelikowsky et. al. Phys. Rev. Lett. 72, 1240 (1994)

Storage and Domain Decomposition

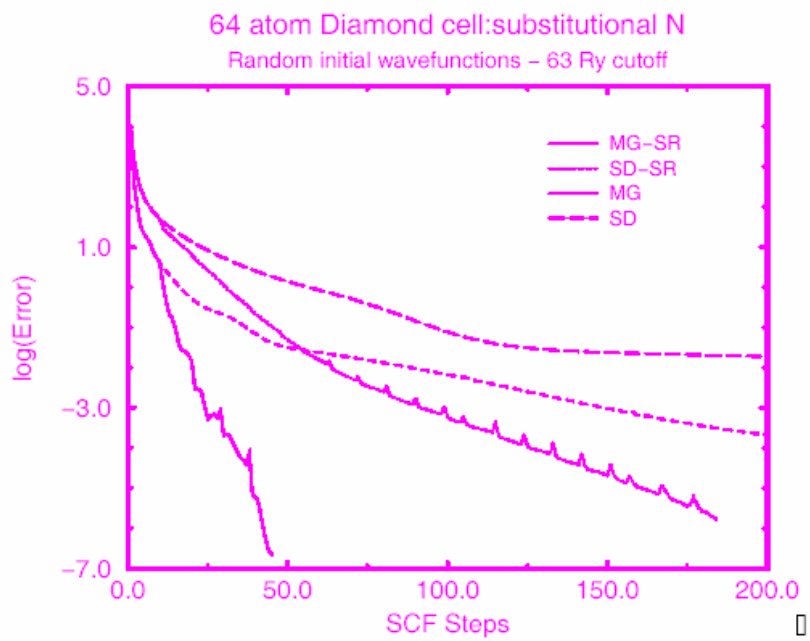
- Wave function values on the real-space grid: $N_x * N_y * N_z$
- Total storage for all wave functions: $N_x * N_y * N_z * N_{wf}$
- Each processor stores N_{wf}/N_{pe} wavefunctions
  - Orthogonalization and subspace diagonalization would require a huge amount of data transfer between processors. It increases quadratically with the system size.

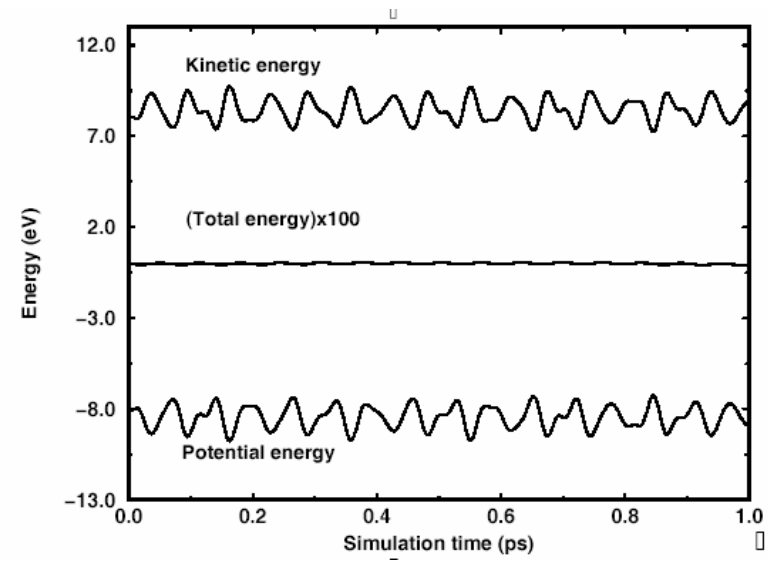
- Domain decomposition: Wave functions, as well as the charge density and potentials are distributed by regions over the processors.
- Communication only between 6 neighboring processors



- Memory: N^2
- Communication: N^2
- Subspace diagonalization: N^3

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., $\Delta E_{MAX} = 27\ \mu V/atom$

A constant of 6832 eV was added to the potential energy

Ultrasoft PP and multiple grid representations

- **Double-grid technique**

  [Ono and Hirose, PRL 82, 5016 (1999)]

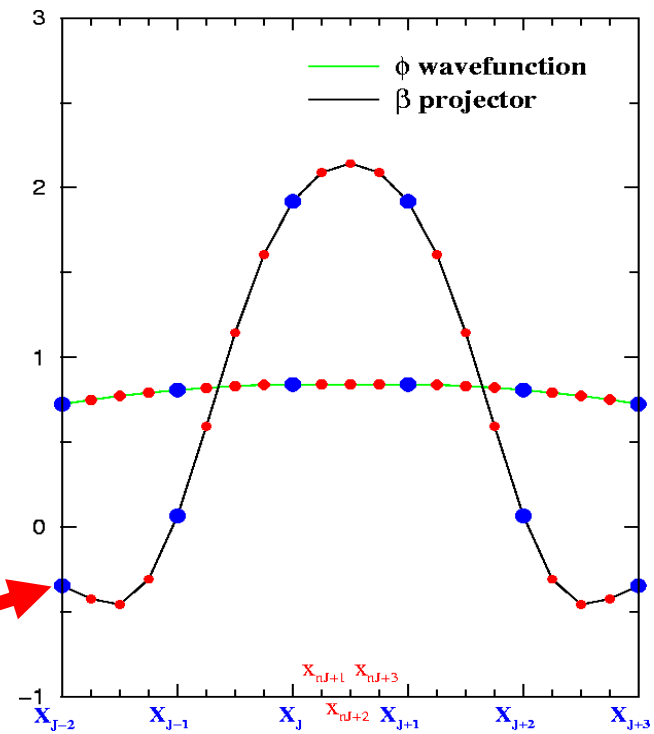
  Wave functions $|\phi\rangle$: coarse grid only

  Projectors $|\beta\rangle$: need to be on a fine grid

  With double grids technique and Fourier interpolation, $|\beta\rangle$ are stored on the coarse grid only with a weight factor which is determined by the values on the fine grid

  $\langle\phi|\beta\rangle$: performed on the coarse grid only

  A non-local projector and a wave function for diamond calculation



- **uniform grids:**

  wavefunctions: coarse grid

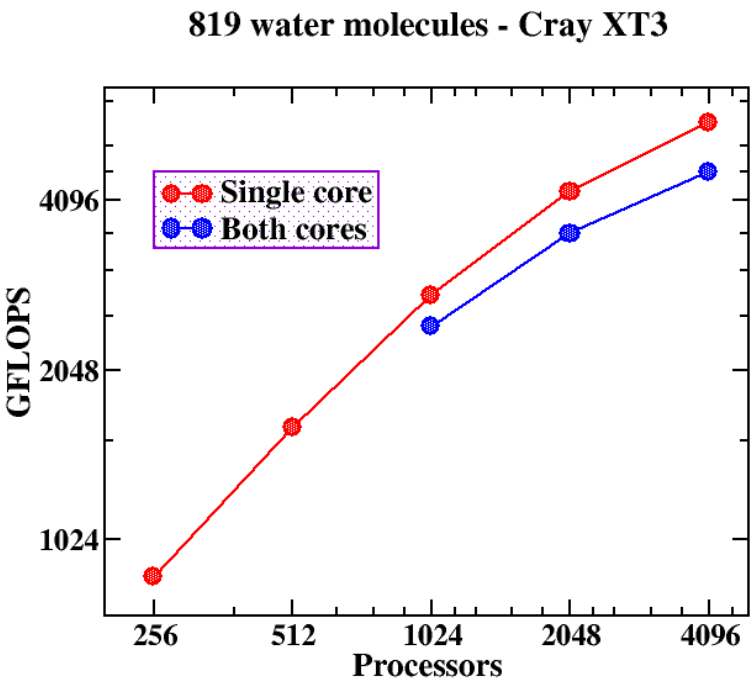
  charge density and potentials: 2x fine grid

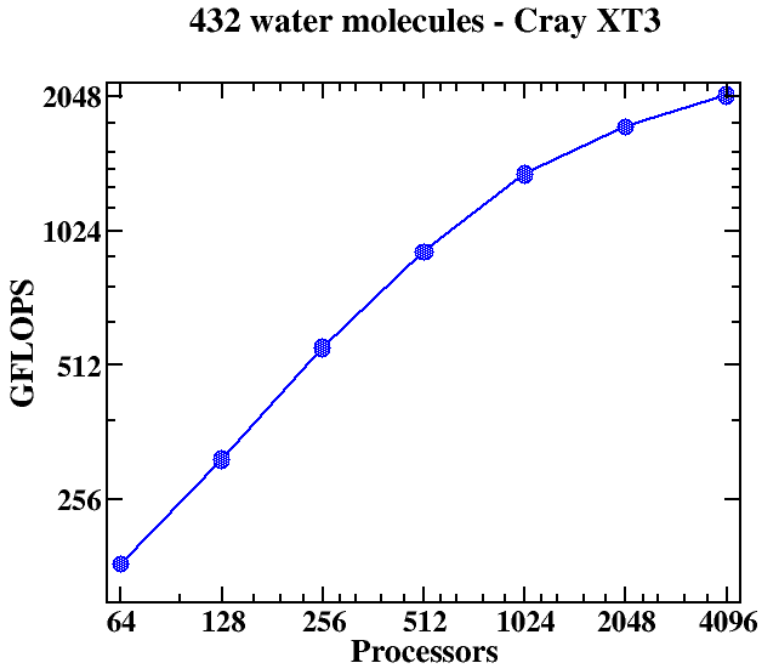
  projectors: 3x or 4x fine grid

  Wang, Lu, Bernholc (2004)

  similar to grid implementation of PAW (Mortensen, Hansen, and Jacobsen, PRB 2005)

Parallel efficiency over large number of processors



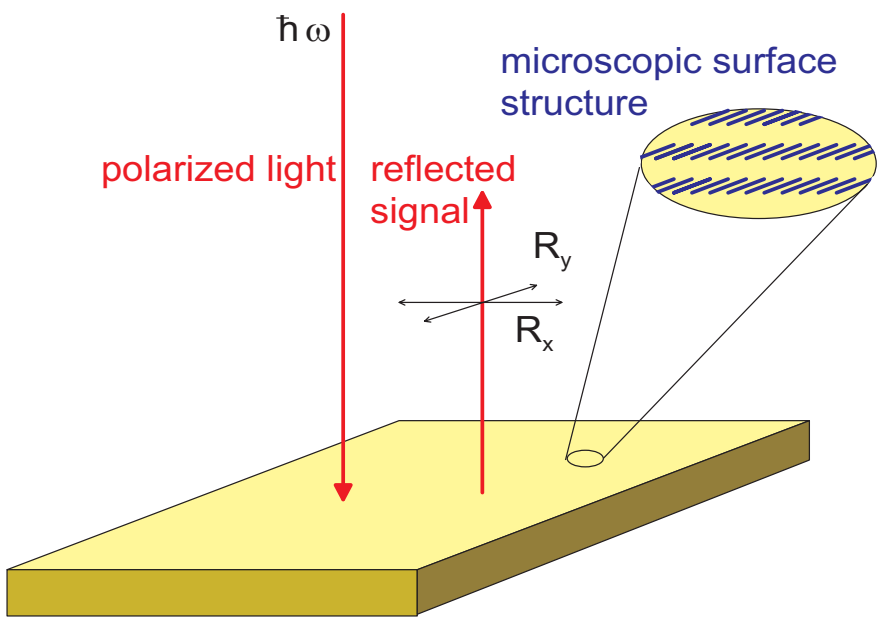


Latest optimizations by Miro Hodak on ORNL-Cray XT3/4

36.5% peak performance with 4096 Cray XT4 cores, 7.78 TFlops

Reflectance Difference Spectroscopy (RDS/RAS)

- Miniaturization of electronic devices requires *in situ* monitoring and feedback control of crystal growth
- "Standard" techniques to monitor surfaces use electron scattering and are restricted to ultrahigh vacuum

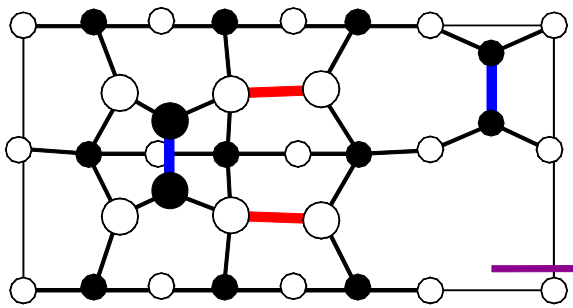


- Difference in reflectance between two polarizations
  bulk contribution cancels out
- *Very* sensitive to atomic structure at the surface
- Can unambiguously determine surface structure, even when other techniques fail
- Optical probe: can be used for feedback control during growth in gaseous environments

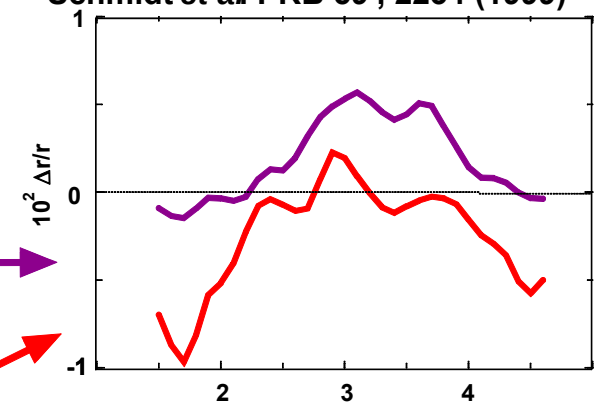
Interpretation of spectra requires *accurate* calculations. We studied many complex surfaces, incl. steps, complex reconstructions and organics.

InP(001)(2x4): surface structure & RDS

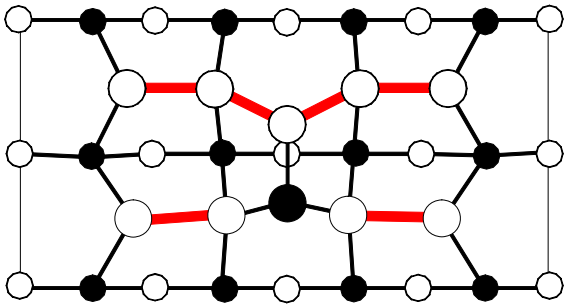
InP(001)α2(2x4)



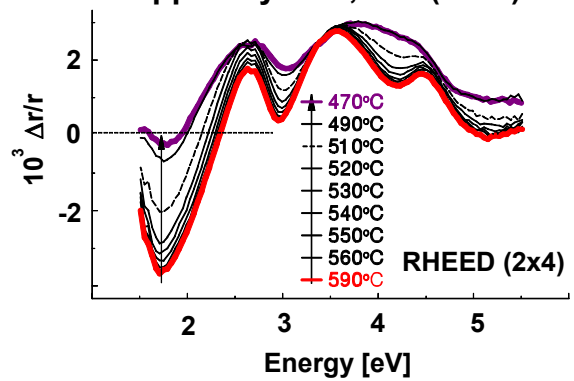
calculated reflectance anisotropy
Schmidt *et al.* PRB 59 , 2234 (1999)



(2x4) mixed-dimer model



measured RAS, KB Ozanyan *et al.*
J. Appl. Phys. 82, 474 (1997)



- **Reflectance**

$$\frac{\Delta R_\alpha(\omega)}{R_0(\omega)} = \frac{4\omega}{c}\,\mathrm{Im}\left[\frac{4\pi\alpha^s_{\alpha\alpha}(\omega)}{\varepsilon_b(\omega)-1}\right]$$

DFT-LDA
qualitative agreement

$\alpha^s_{\alpha\alpha}$ ... αα component of slab polarizability

ε_b ... bulk dielectric function

Many-electron "GW" corrections

- Correction to DFT eigenvalues computed by GW many-body theory. The quasiparticle energies E_i become

[Hedin Phys. Rev. 1965; Hybertsen, Louie, PRB, 1988]

$$E_i \approx \varepsilon_i^{DFT-LDA} + \frac{\left\langle \psi_i \middle| \Sigma\left(\varepsilon_i^{DFT-LDA}\right) - V_{XC}^{LDA} \middle| \psi_i \right\rangle}{1 - \frac{\partial}{\partial E}\left\langle \psi_i \middle| \Sigma(E) - V_{XC}^{LDA} \middle| \psi_i \right\rangle \Big|_{E=\varepsilon_i^{DFT-LDA}}}$$

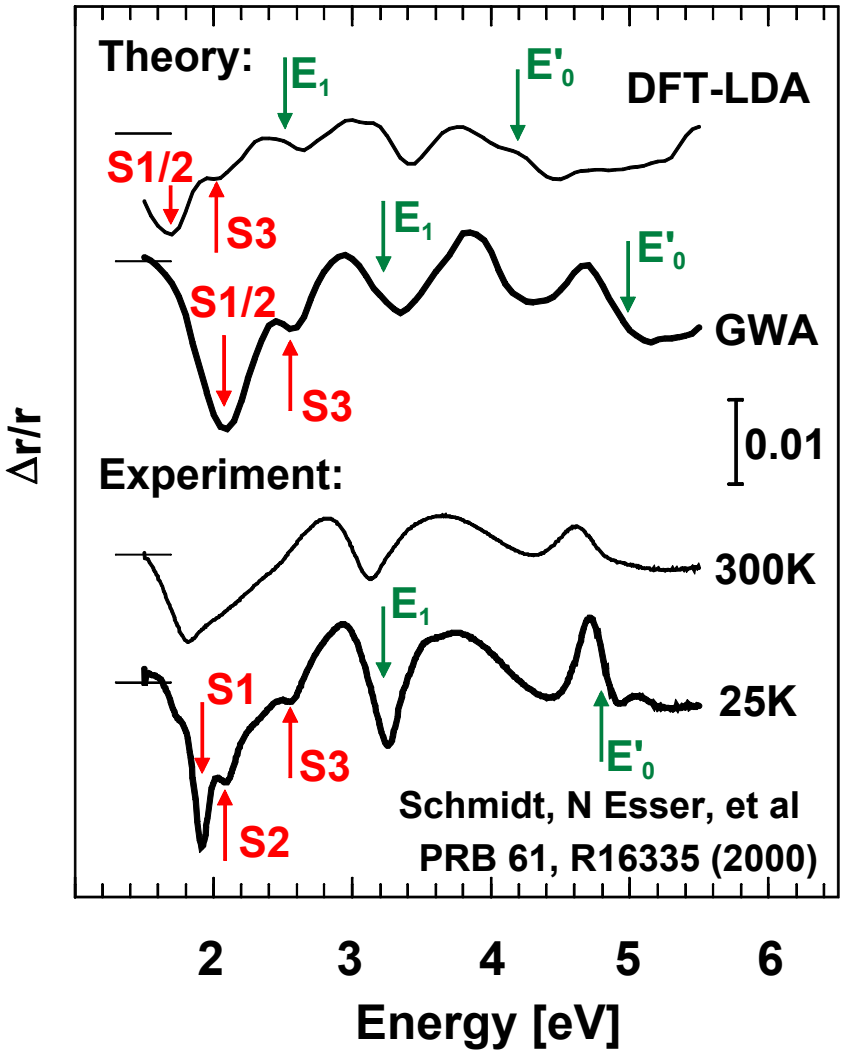
$\Sigma = iG(1,2)W(1,2)$; $1 \equiv (r,t)$; $W(1,2) = \int d3\, v(3-2)/\varepsilon(1,3)$

$G(1,2)$ ≡ one-electron Green's function

$v(1-2)$ ≡ Coulomb potential = $1/|r-r'|$

- In calculating the screened Coulomb potential W a model dielectric function is used Bechstedt, SSC (1992)

In-rich InP(001)(2x4):
self-energy effects on optical anisotropy

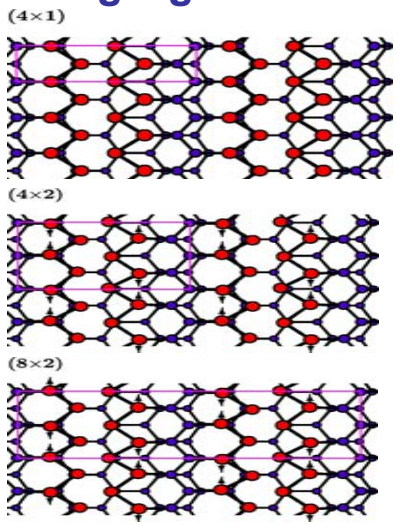


- non-uniform shift of bulk and surface related peaks with GW corrections
- line shape changes
- k-space resolution insufficient to resolve S1/S2

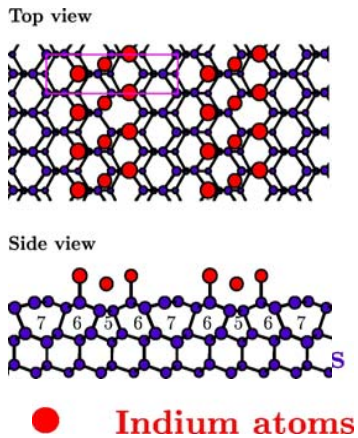
Optical signature of Si(111)-In nanowires

- In atoms on Si(111) self-organize into nanowires.
- Undergo a phase transition from (4x1) to (8x2) as temperature is lowered.
- Several models of (4x1) and (8x2) structures exist

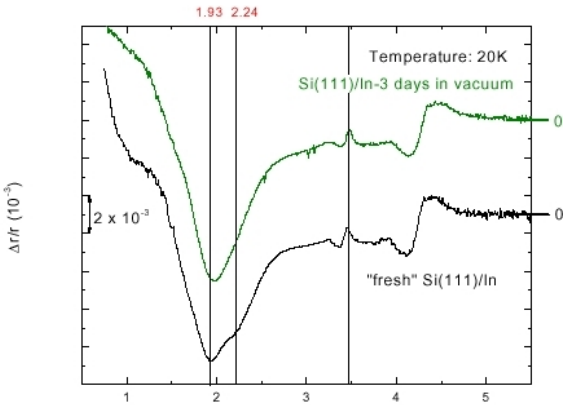
Zigzag model

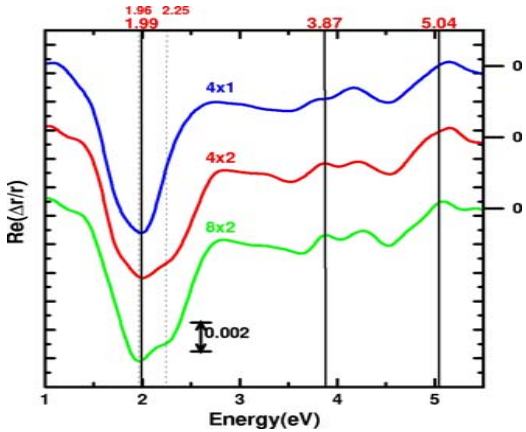


π-bonded-chain-stacking-fault (7656) model

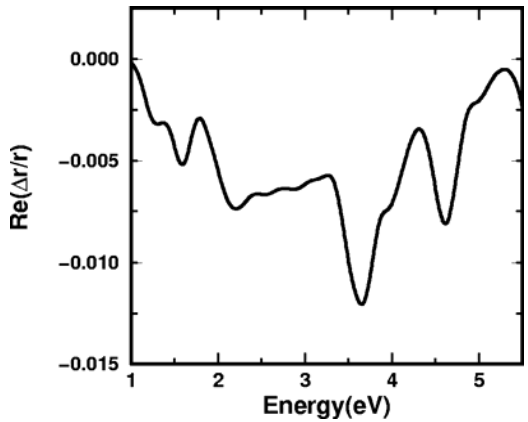


strongly anisotropic signals observed in experiments





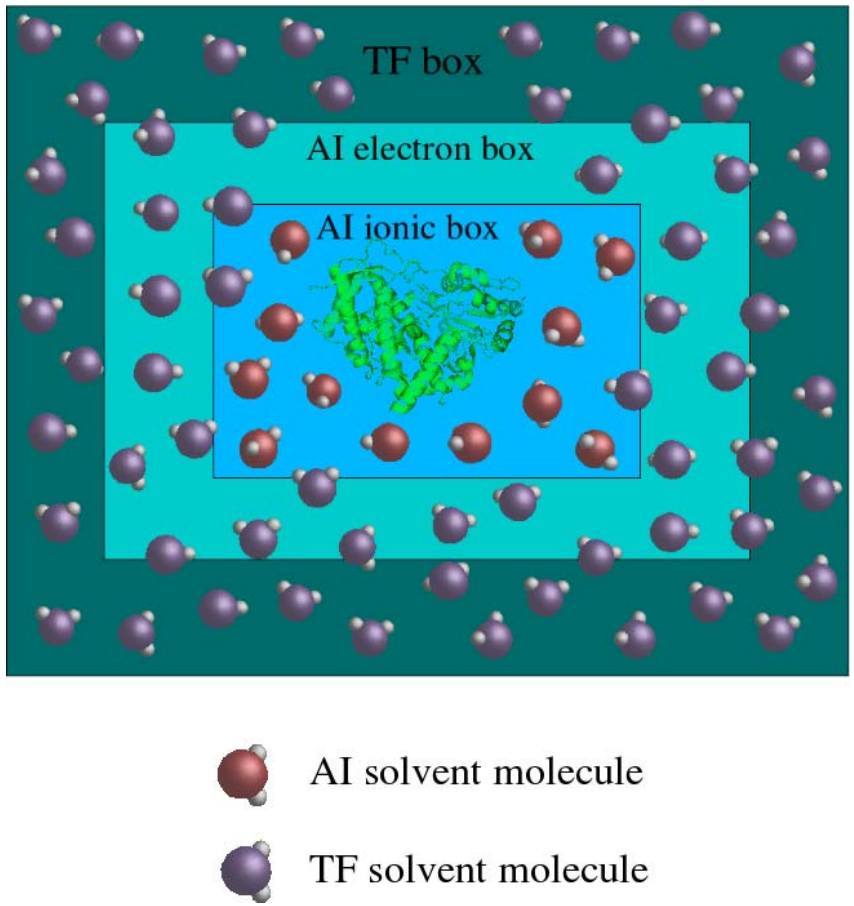
Calculated zigzag



Calculated (7656)

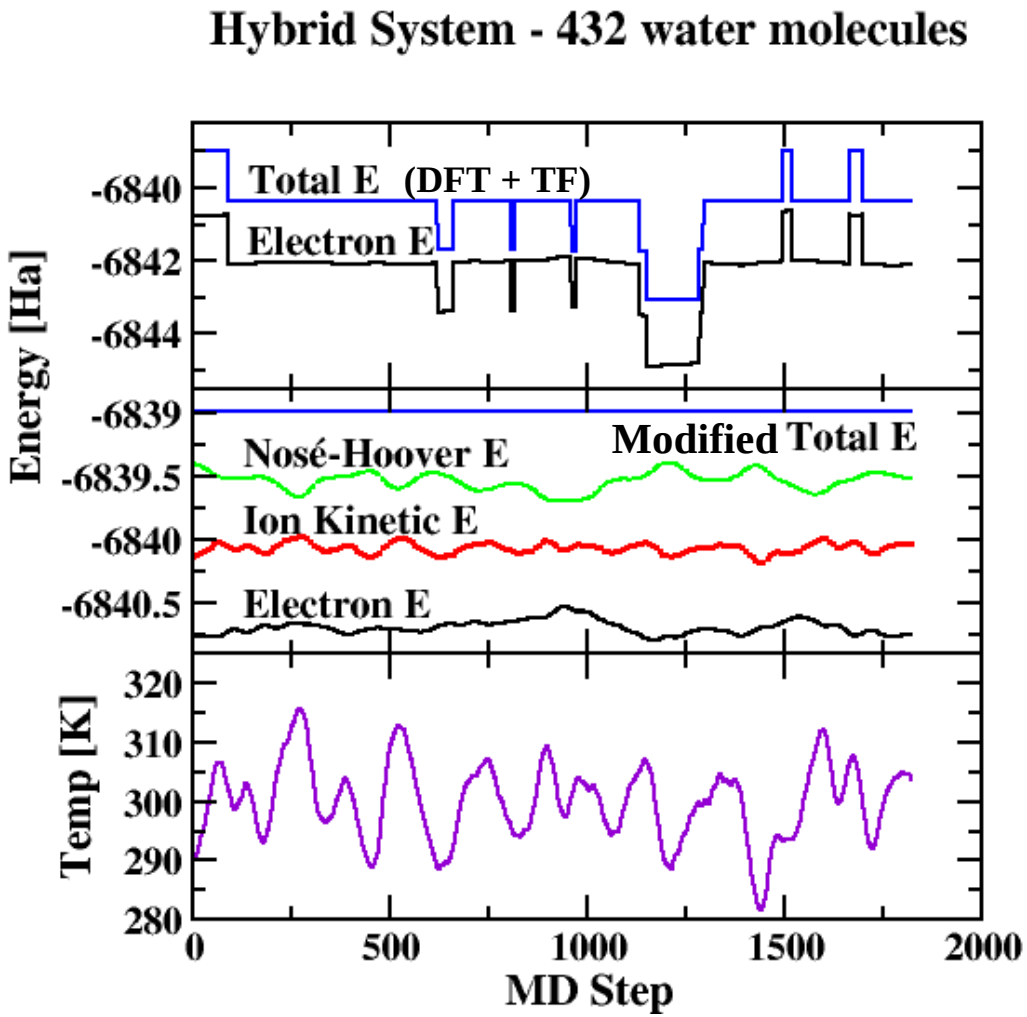
Hybrid simulation method: DFT + TF

- **The simulation system is divided into 3 boxes:**
  - ***ab initio (AI)* ionic box – all molecules solved by *ab initio* have to be completely inside**
  - ***AI* electron box – for solving Kohn-Sham equations with zero boundary conditions**
  - **TF (Thomas-Fermi) box – contains the whole system**
- **TF and *AI* densities can overlap**



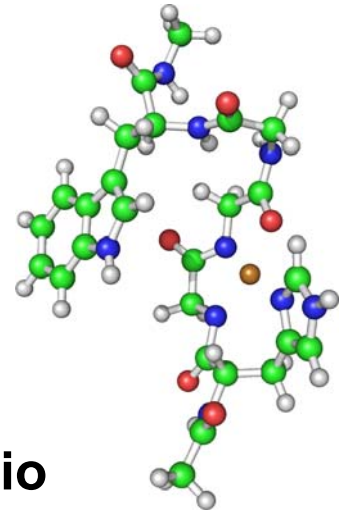
Hybrid Molecular Dynamics

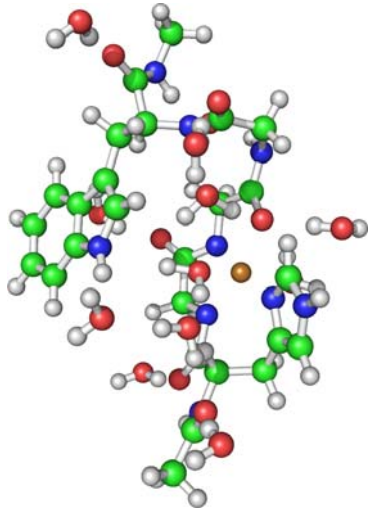
- Total number of water molecules: 432
- Thomas-Fermi box: 44 Bohr
- *ab initio (AI)* electron box: 22 Bohr
- *AI* ion box: 8 Bohr
- Timestep 10 au
- Nose thermostats for temperature control

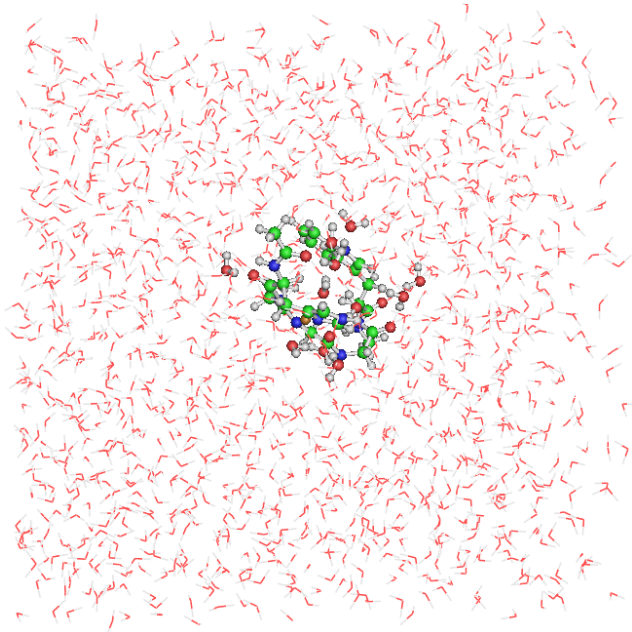


Solvent effect on Cu-PrP^C complex

- Compare three cases:
  - Dry PrPC-Cu
    - 73 atoms – ab initio only
  - PrP^C-Cu with nearby waters
    - 12 water molecules
    - total of 109 atoms - ab initio only
  - Fully solvated PrP^C-Cu
    - hybrid calculation
    - ab initio: PrP^C, 12 waters (109 atoms)
    - Thomas-Fermi: 1469 waters (4407 atoms)
    - ab initio part takes 97% of CPU time



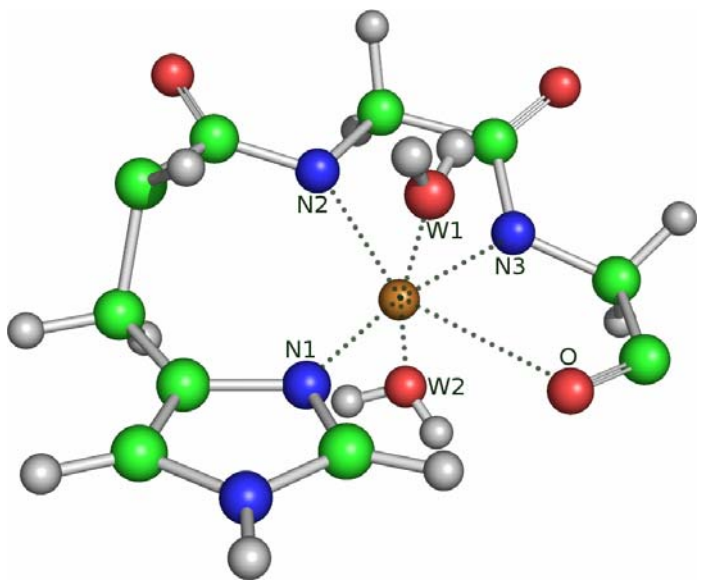




Cu-PrP^{C} binding site geometry

- **Copper-HGGGW bond distances (Å):**

	Dry protein	partial solvation	full solvation
Cu-N1	1.99	1.98	1.98
Cu-N2	2.01	2.00	2.01
Cu-N3	1.91	1.91	1.92
Cu-O	2.23	2.41	2.36
Cu-W1	-	2.78	2.82
Cu-W2	-	2.79	2.7



- **Solvation is important for non-covalent bonds**

Gap between unoccupied and occupied states:

Dry PrPC-Cu: 2.11 eV
Partial solvation: 2.25 eV
Full solvation: 2.38 eV

Solvation is important for electronic properties and cannot be neglected

Also important for reaction energetics

Need for *O(N)* quantum simulations

- **Estimated time for DFT calculations at 200 sustained MFLOPS/CPU**

  10,000 atoms, simple relaxation (100 ionic steps)

  O(N) **6,000 CPU hr, *O(N)* memory 77 GB**

  O(N²) **421,000 CPU hr, *O(N²)* memory 6 TB**

  O(N³) **30,000,000 CPU hr, *O(N²)* memory 6 TB**

  Large systems will require many simple relaxations

  Efficient and reliable *O(N)* methods are a must!

- **Time-dependent *O(N)* simulations for 10,000 atoms**

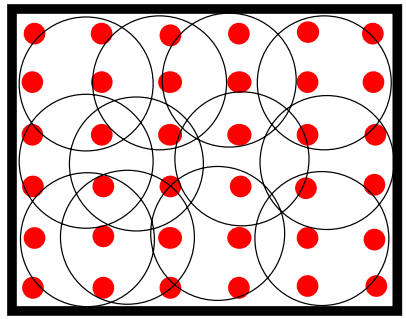
  1 ps ~ 84,000 CPU hr (~ 1400 steps)

  1 ns ~ 84,000,000 CPU hr (~ 1,400,000 steps)

  Multiscale time-acceleration methods are a must!

Introduction to linear-scaling DFT

- **One can achieve *O(N)* scaling by using overlapping spatially localized basis functions.**
  (Several different formulations, see Goedecker, Rev. Mod. Phys. 1999)
- **Non-orthogonal basis φ_i is *variationally* optimized:**
  Sole condition: $\varphi_i(r-R_k), = 0$ *for* $|r-R_k| > R_{Loc}$



Localization regions

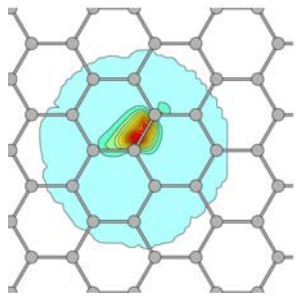
- **The optimized orbitals are dual, non-orthonormal Wannier functions with optimal decay properties $w(r) \approx r^{-3/2} e^{-hr}$, *see He and Vanderbilt PRL 86 (2001)***

- **Overlap $S_{ij} = \langle\varphi_i|\varphi_j\rangle$ and Hamiltonian matrices become sparse in real space**
  ***High* accuracy solutions ⇒ grids or *large* LCAO**

$$E_{tot} = \sum_{occ} \varepsilon_i + F(\rho) = \sum_{i,j}^{occ} S_{ij}^{-1} H_{ji} + F(\rho)$$

Expand $S^{-1} = [I - (I - S)]^{-1} = I + (I - S) = 2I - S$

All matrix operations become sparse and *O(N)*

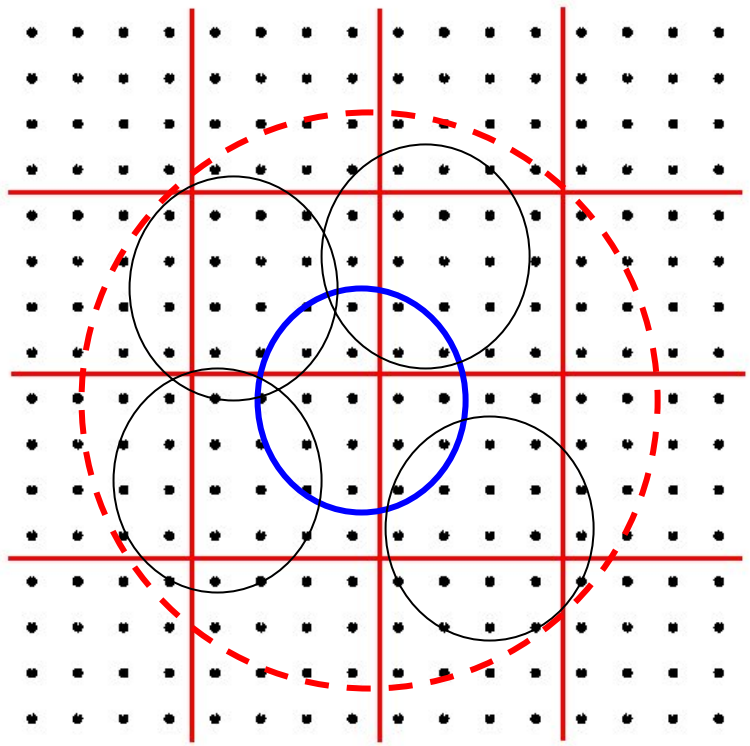


Shape of an optimized orbital: valence bond function

Data storage and parallelization

- **Charge density and potentials**: same as in the extended orbitals method
- **Orbitals** are distributed over processors. **Communication**: scales linearly with the system size.
- **Diagonalization:** ScaLapack up to 20k orbitals. For an even larger system, sparse matrix algebra or iterative methods may be used.

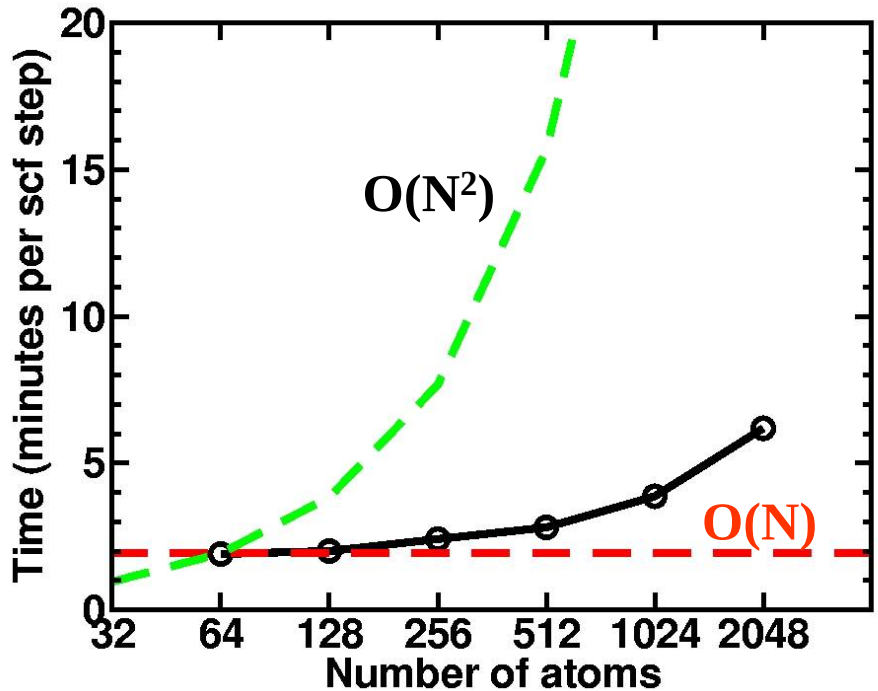
- If the system size per processor is fixed, both the memory and communication between processors are **constant**, which leads to a **linear scaling** with the size of the system.



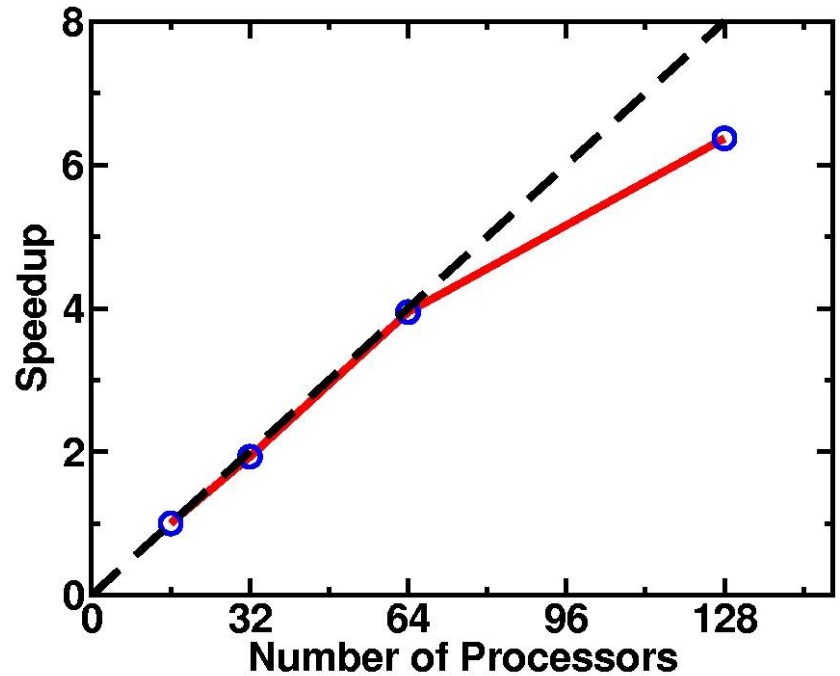
The **blue orbital** only interacts with the orbitals in the **red circle**.

Scaling, Size and Accuracy

32 atoms per processor



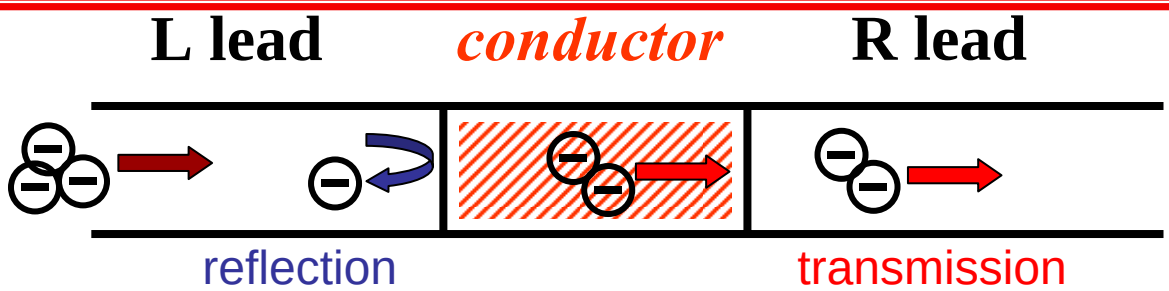
1024 atom Carbon nanotube

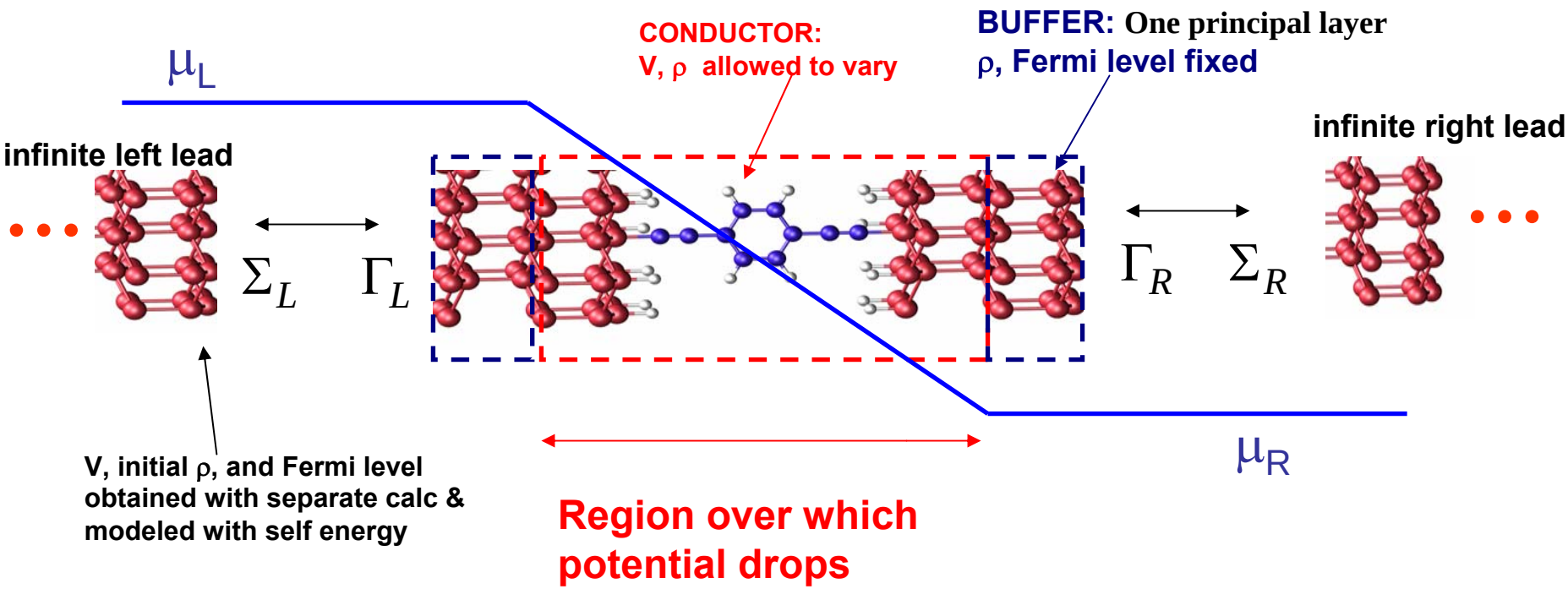


- **Scales nearly linearly (~60-90%) over a large range of system sizes**
- **Scales linearly with the number of processors in a large range**

Further improvements possible

Introduction to Quantum Transport

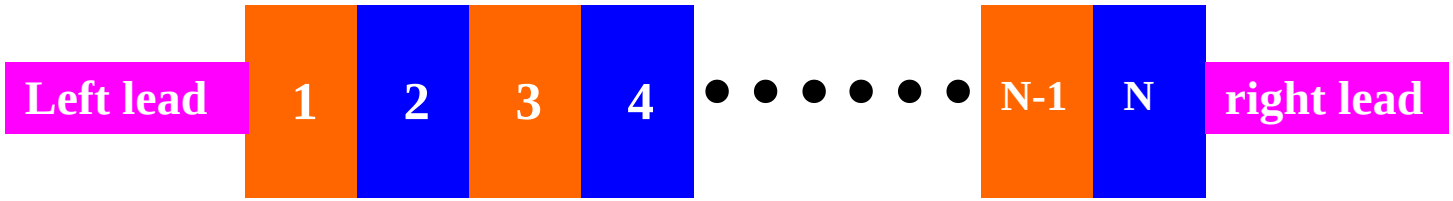


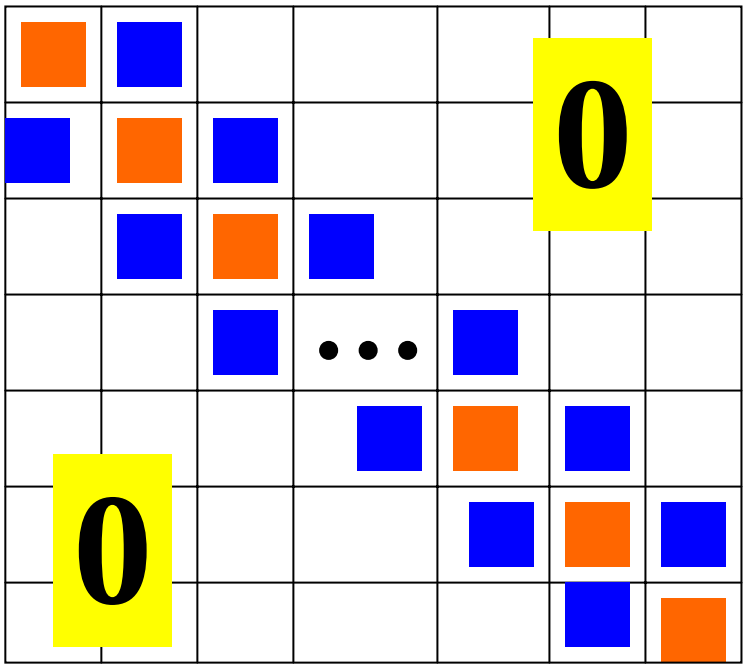


H_L, S_L | $H_{L\,C}, S_{LC}$ | H_C, S_C | $H_{R\,C}, S_{RC}$ | H_R, S_R

Semi-Tridiagonal Matrix in Principal Layers

Conductor part: Nearest neighbor interaction between principal layers





- **Hamiltonian and overlap matrixes are semi-tridiagonal.**
- **Self-energy Σ have values only in first block (left lead) or in the last block (right lead).**

- **Green's function**
  $G = (\varepsilon S - H - \Sigma)^{-1}$
- **Only the semi-tridiagonal parts are needed for self consistent calculations.**

Quantum transport parallelization

I-V calculations typically 20 biases

The processors are divided into 20 bias groups, each group performs the calculations for one bias. parallel scalability: 99%

Sum over k-points 1~20 kpoints

The processors in one bias group are divided into kpoint subgroups, each subgroup does calculations for one k-point.

Integration over energy for each bias ~100 energy points

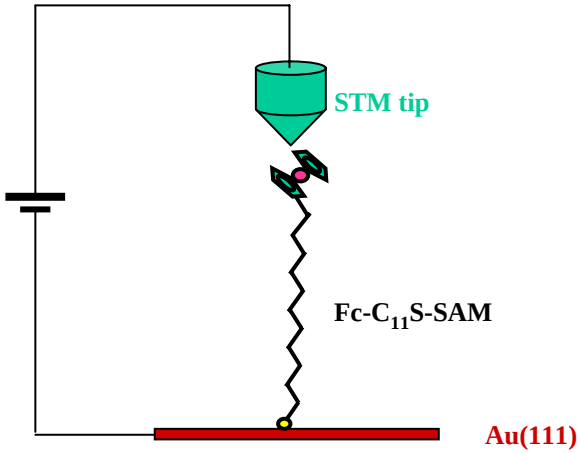
The processors in one bias group are divided into 100 subgroups, **one subgroup performs matrix operations for one energy point**

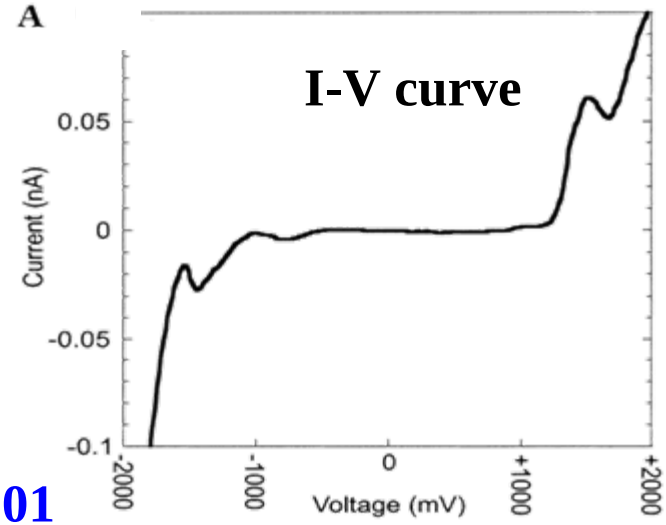
Matrix operations with PBLAS and ScaLapack

PBLAS and ScaLapack are used to perform matrix operation for each subgroup. This subgroup, a 2D processor grid, can be up to 64 processors, depending on the system size.

20 bias groups X 100 energy subgroups X 64 matrix = 128,000 processors

Ferrocenyl-SAM on Au

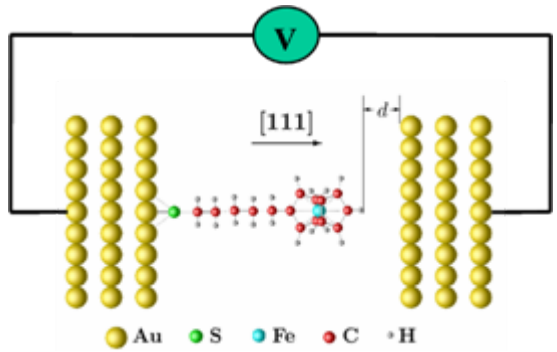




Gorman, Carrol, Fuierer (NCSU), Langmuir 2001

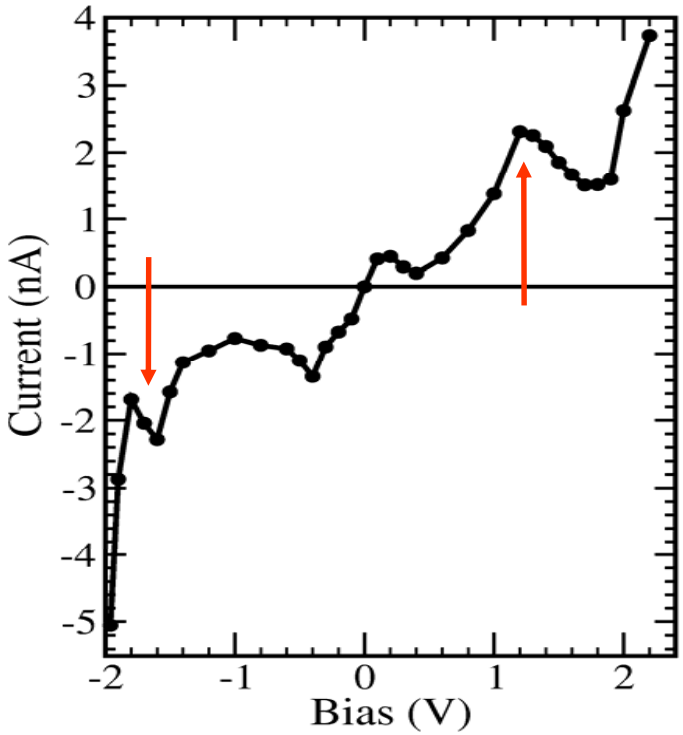
- **Fc-C_nS-Au(111): self-assembled molecules**
- **STM measurement of the I-V curve.**
- **Negative differential resistance around 1.6 V.**

- **Theory: 372 atoms, ~ 4000 electrons.**
- **Fc-C_nS is anchored to the left**
- **The right Au lead acts as STM tip**

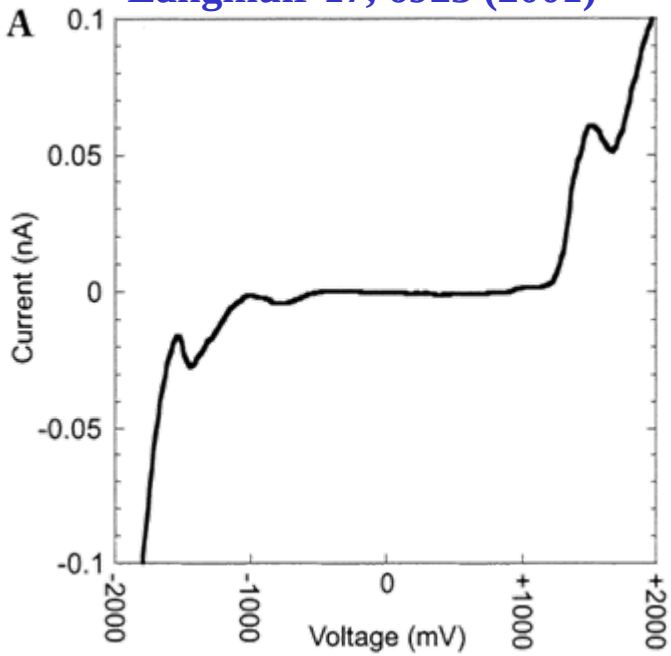


I-V characteristic of the Fc-C_5S-Au(111)

Our calculations

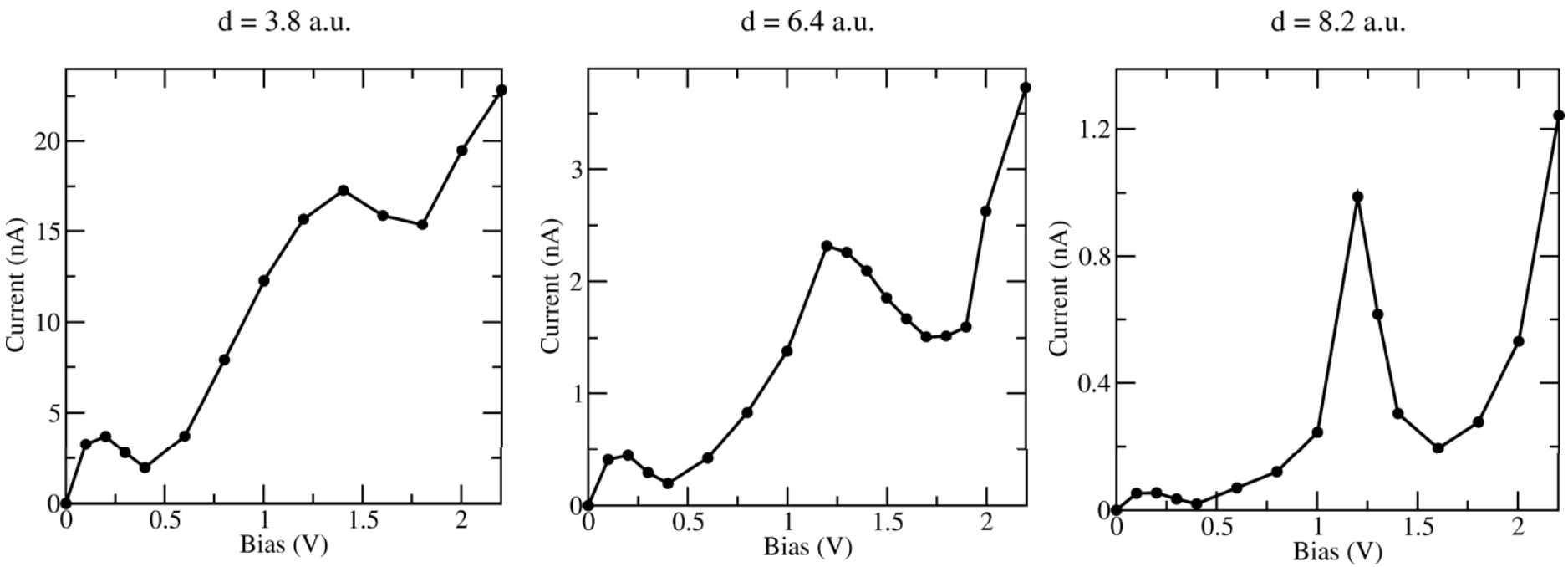


Gorman, Carrol, Fuierer (NCSU), Langmuir 17, 6923 (2001)



- **NDR at -1.6, -0.4, 0.2, and 1.2 V.**
- **NDR at large biases are in good agreement with STM measurement.**
- **No NDR at small biases is observed in experiments.**

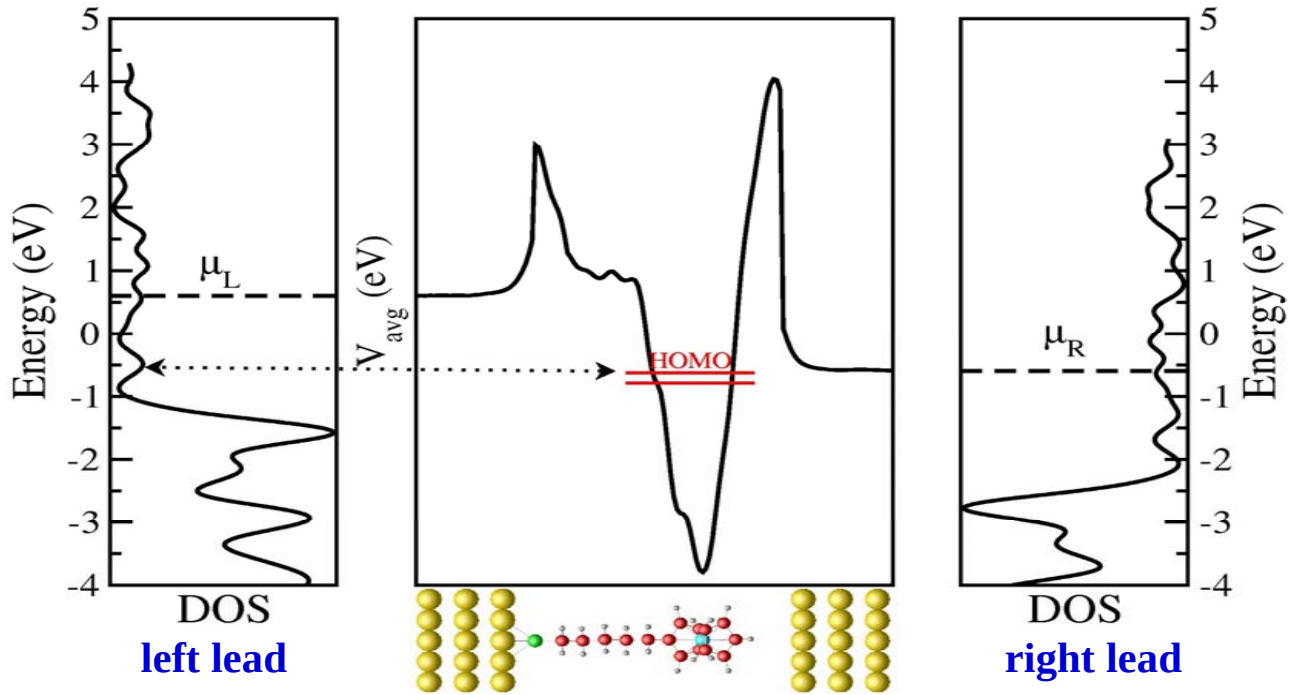
The effect of the metal-molecule distance



- All of I-V curves show the same features, concerning the position of the NDR and the line shape.
- Current decreases with the increase of vacuum distance.
- With an increase in the distance, the NDR at large bias is enhanced, while NDR at small bias is attenuated.

Schematic model of NDR due to resonant tunneling

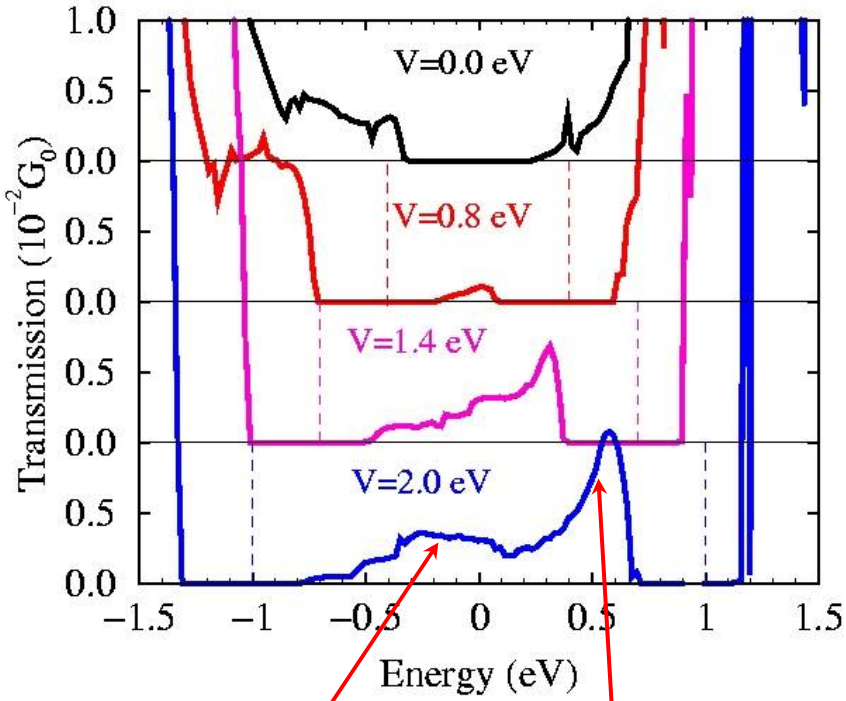
V = 1.2 V

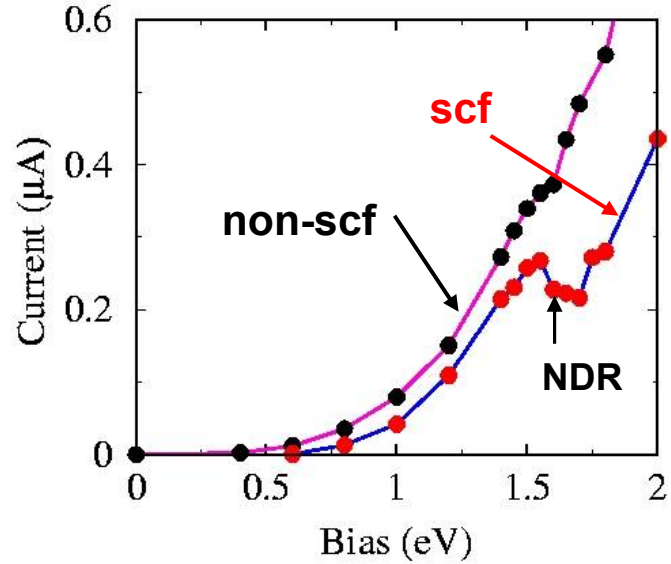


- **Two potential tunnelling barriers: one at Au-S surface and the other at vacuum between the ferrocene and Au surface.**
- **At 1.2 V, the system approaches resonant coupling, resulting in the current peak in the I-V curve.**
- **For negative bias, the resonant coupling occurs at -1.6 V, due to the inner opposite bias.**

Wang, Lu, Bernholc, PRB 2006

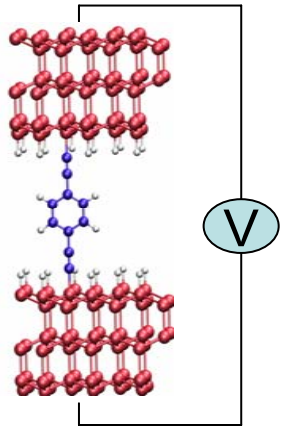
Transmission and I-V Characteristic of Si-molecule-Si juctions





HOMO contribution

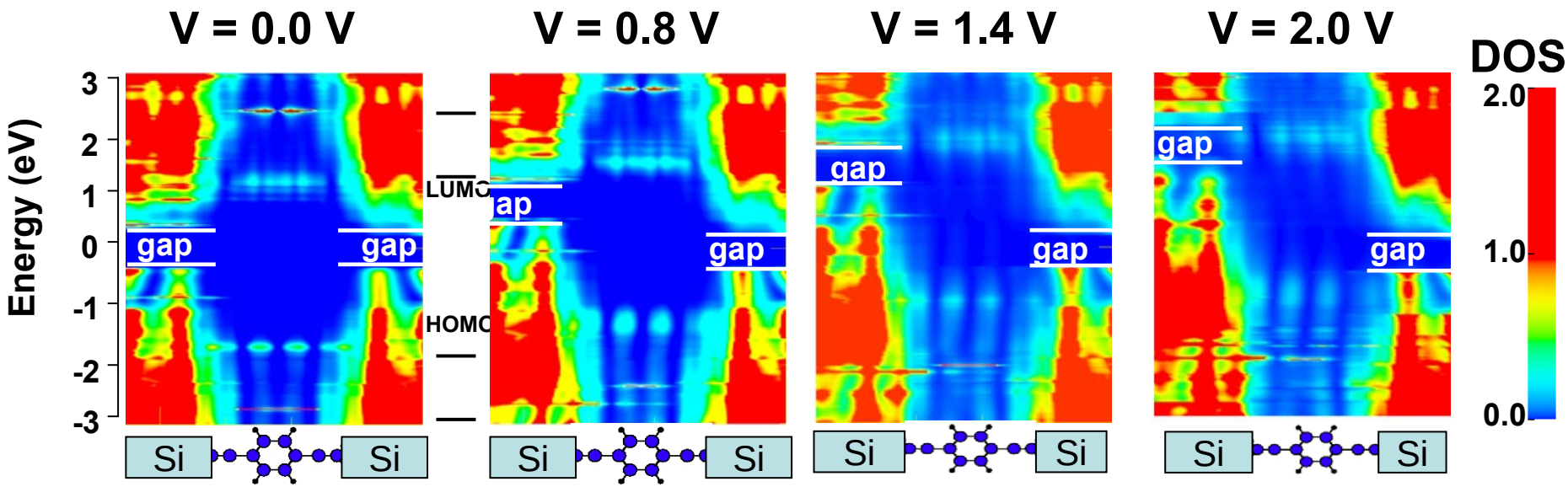
LUMO contribution



- Transmission arises from HOMO and LUMO tails
- NDR appears around 1.6 eV
- Importance of SCF under large bias

Lu, Meunier, Bernholc, PRL95, 206805 (2005).

Position-dependent density of states



- Molecular orbitals broaden with applied bias.
- When molecular orbitals match the band edge, a peak appears in the I-V curve.
- When molecular orbitals fall into the band gap of Si, the current drops, i.e., negative differential resistance (NDR) occurs.

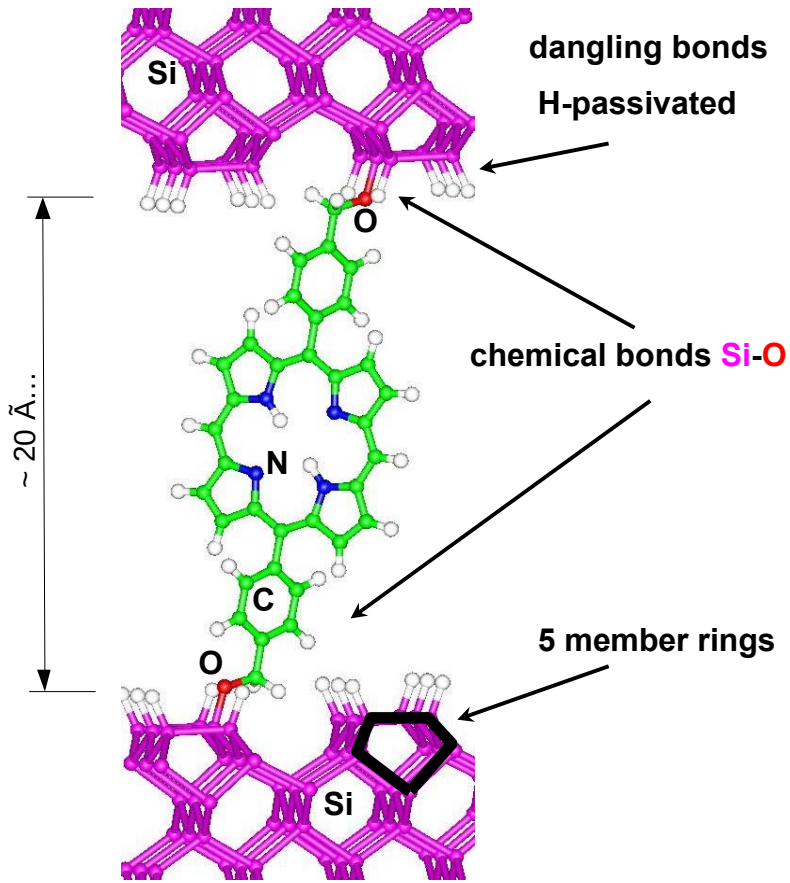
Unambiguous demonstration of shifts of molecular orbitals under bias

Porphyrins on Si(100)

- Experimental demonstrations of self-assembled monolayers (SAMs) of porphyrins on Au(111) and Si(100) substrates
- Oxidation states of porphyrins may be used for information storage (electrochemistry) (Zhiming Liu, *et al.*, Science 302, 1543 (2003))
- Possible applications to molecular-based memories, molecular transistors, logic, etc...

See www.zettacore.com

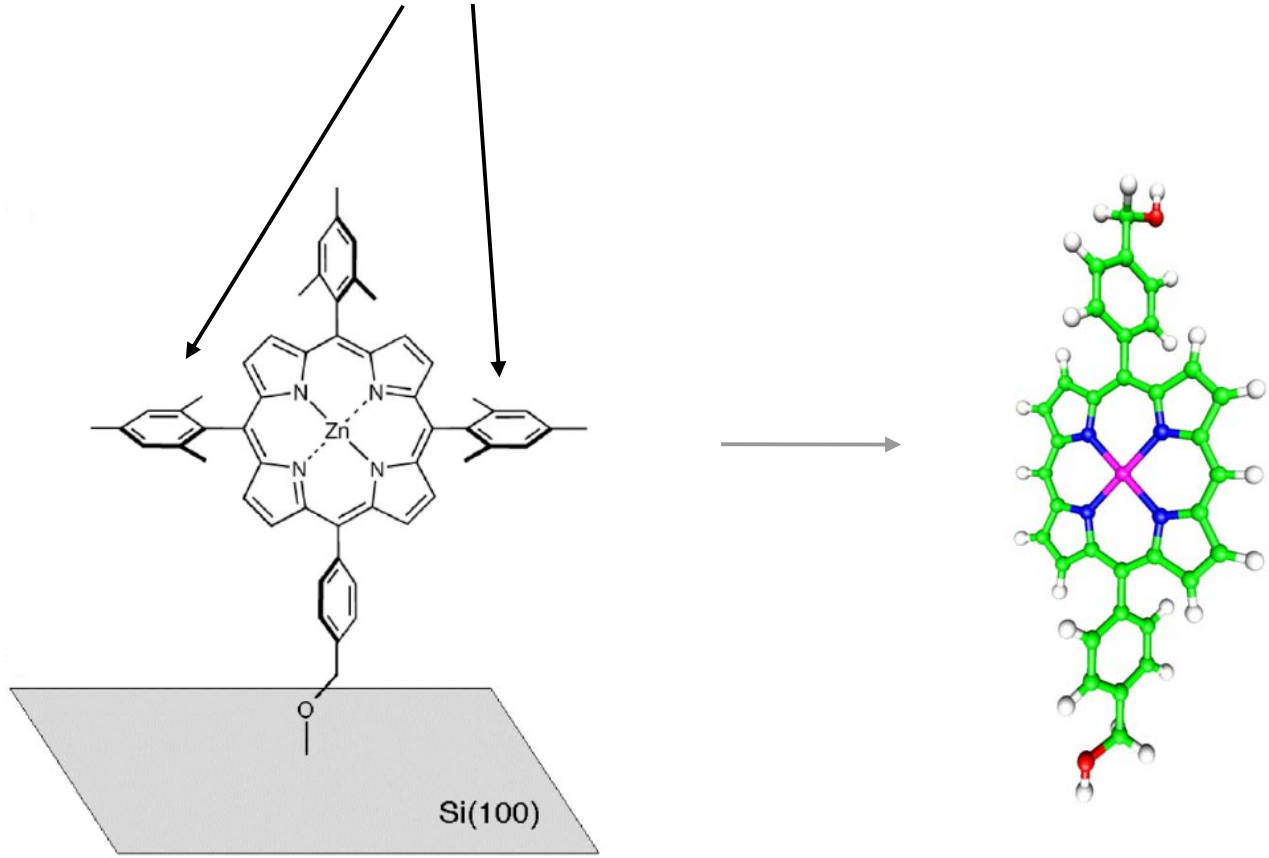
Si (100) surface reconstruction



system has inversion symmetry

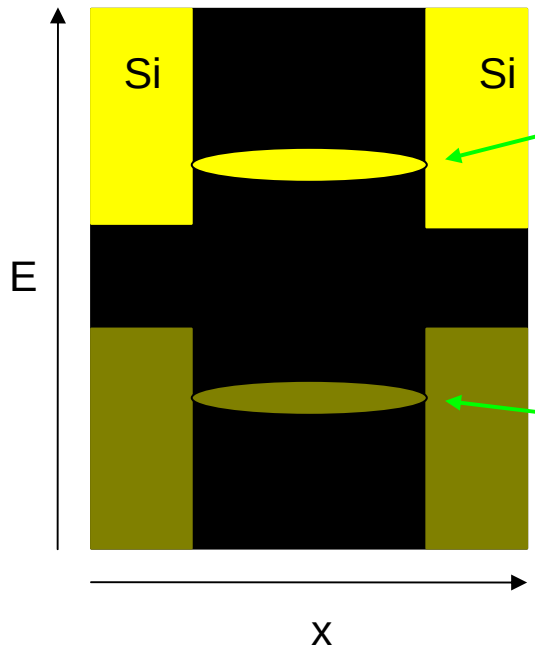
Full and simplified Zn-Porphyrin

For simplicity the side groups are not present in our calculation:



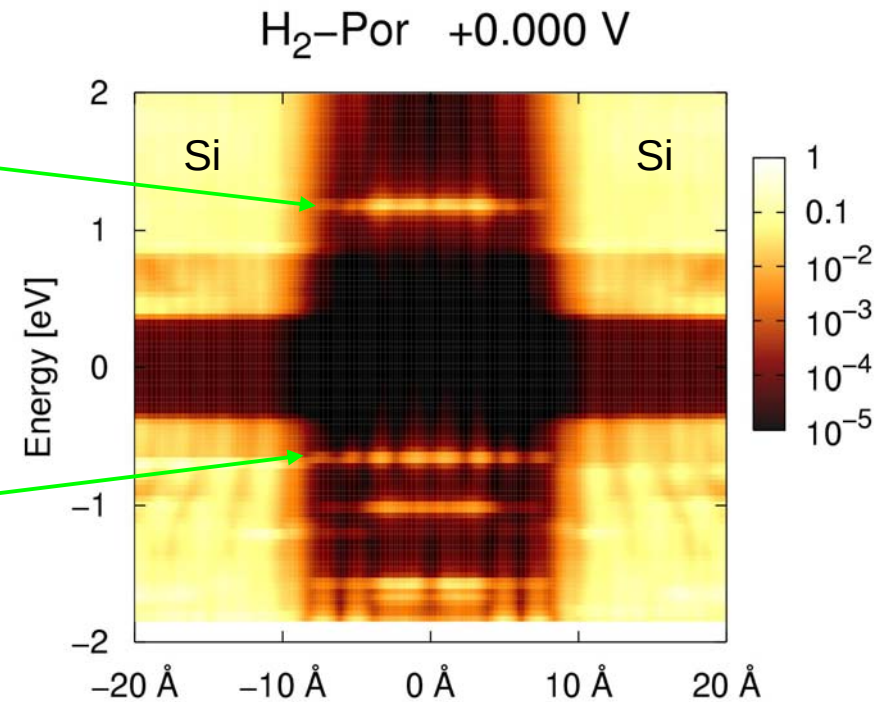
(Fig. from: Liu, et al., Science **302**, 1543 (2003))

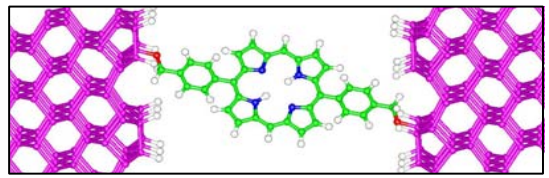
H_2-Porphyrin density of states (DOS)



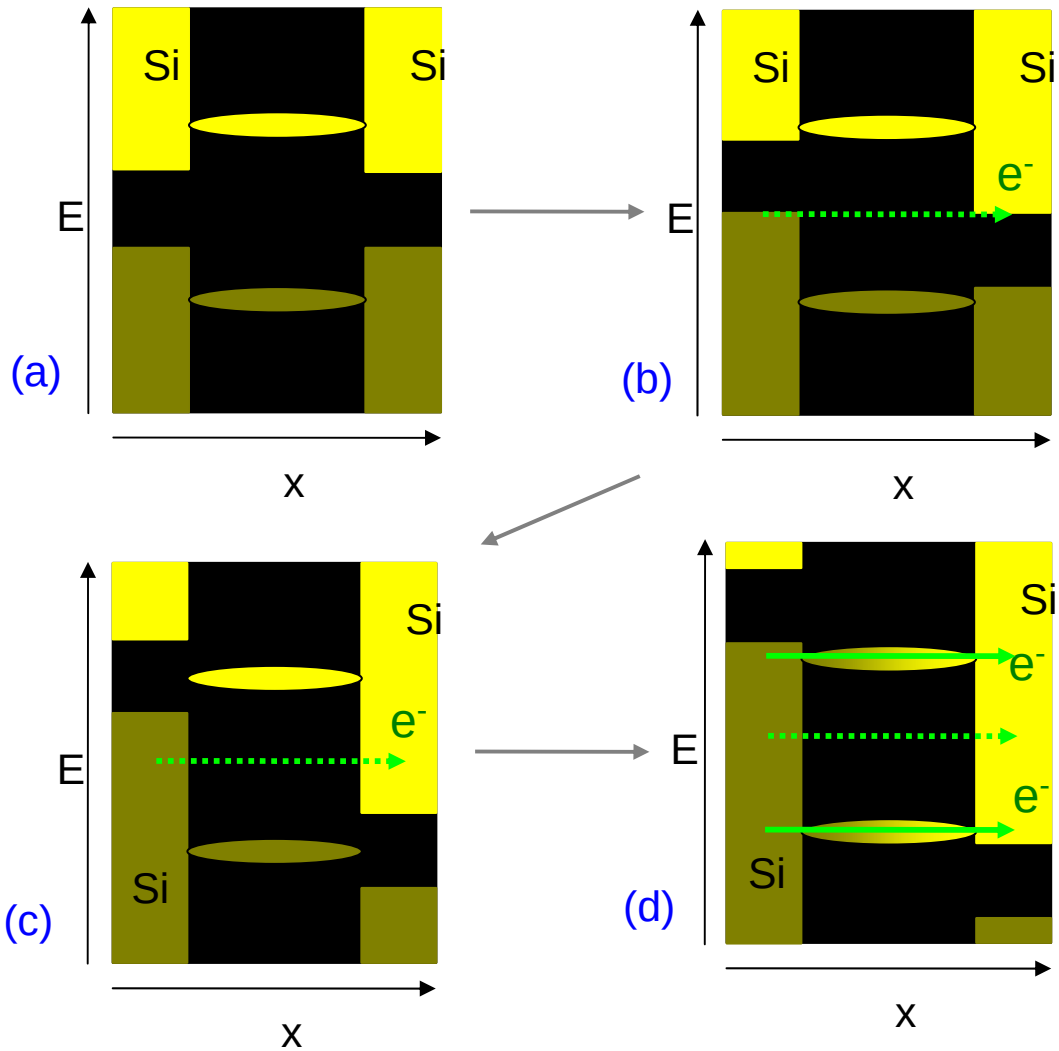
LUMO

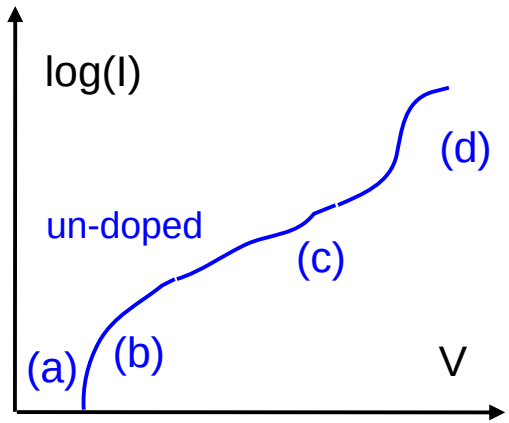
HOMO





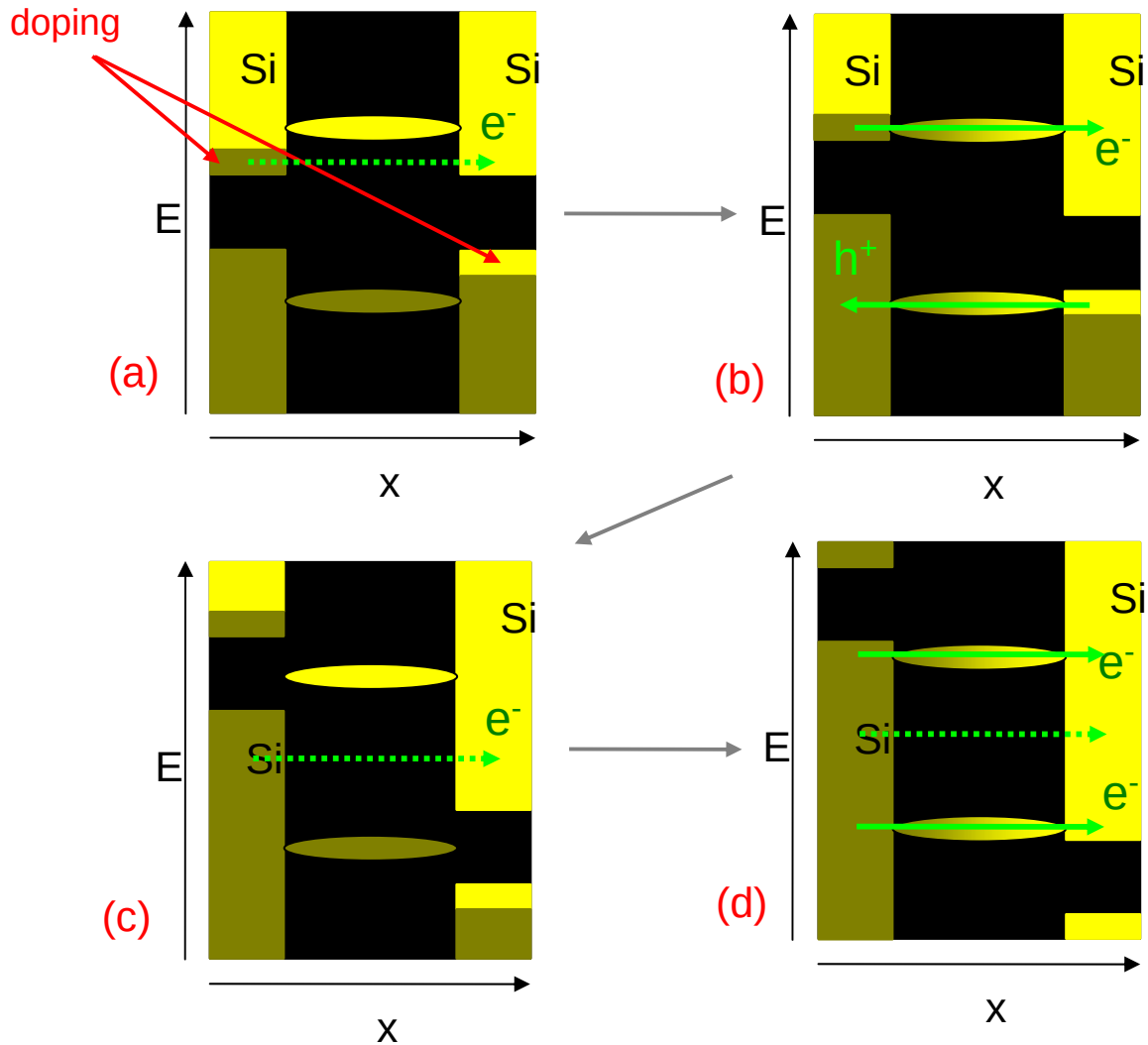
DOS and Current vs. Bias

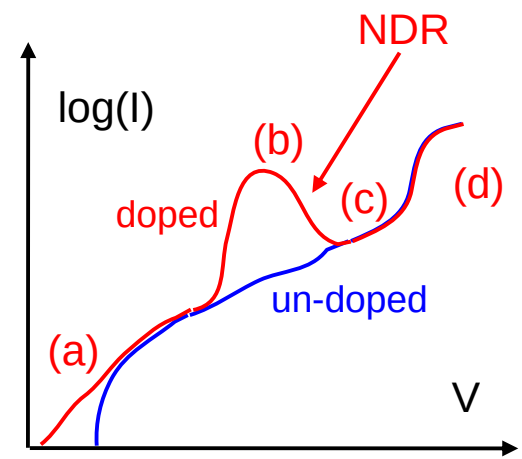




- **(a) Bias = 0; no current**
- **(b) Bias ≈ E_{gap}: very low Si-Si tunnelling current**
- **(c) Bias > E_{gap}: tunneling current increases**
- **(d) HOMO or LUMO become open channels: resonant tunneling through molecular channels**

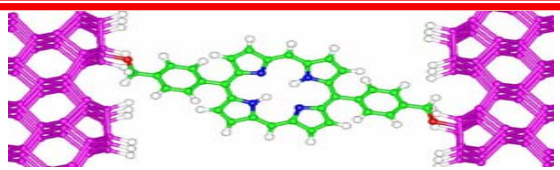
Doped DOS and Current vs. Bias

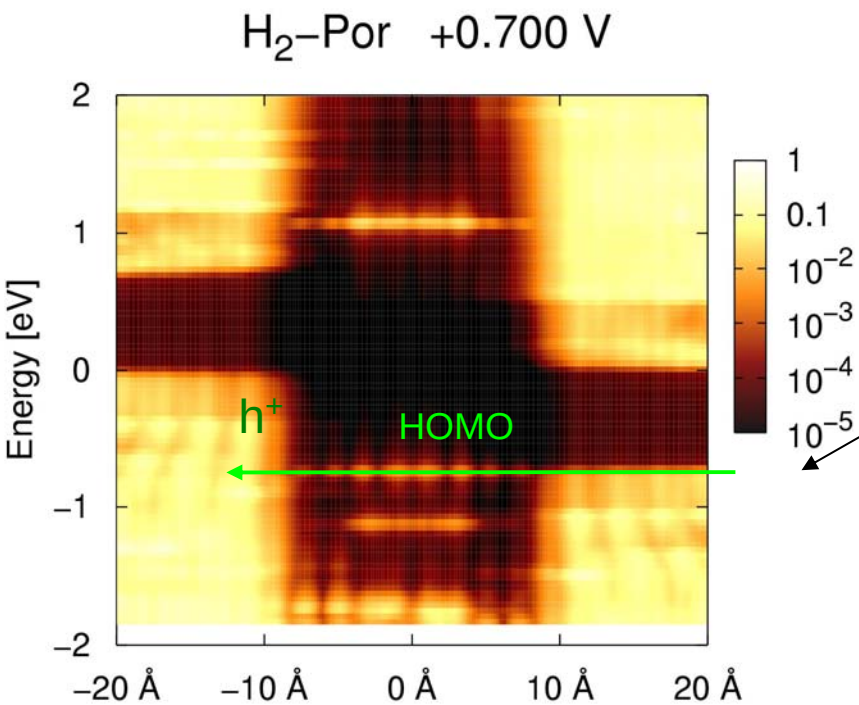


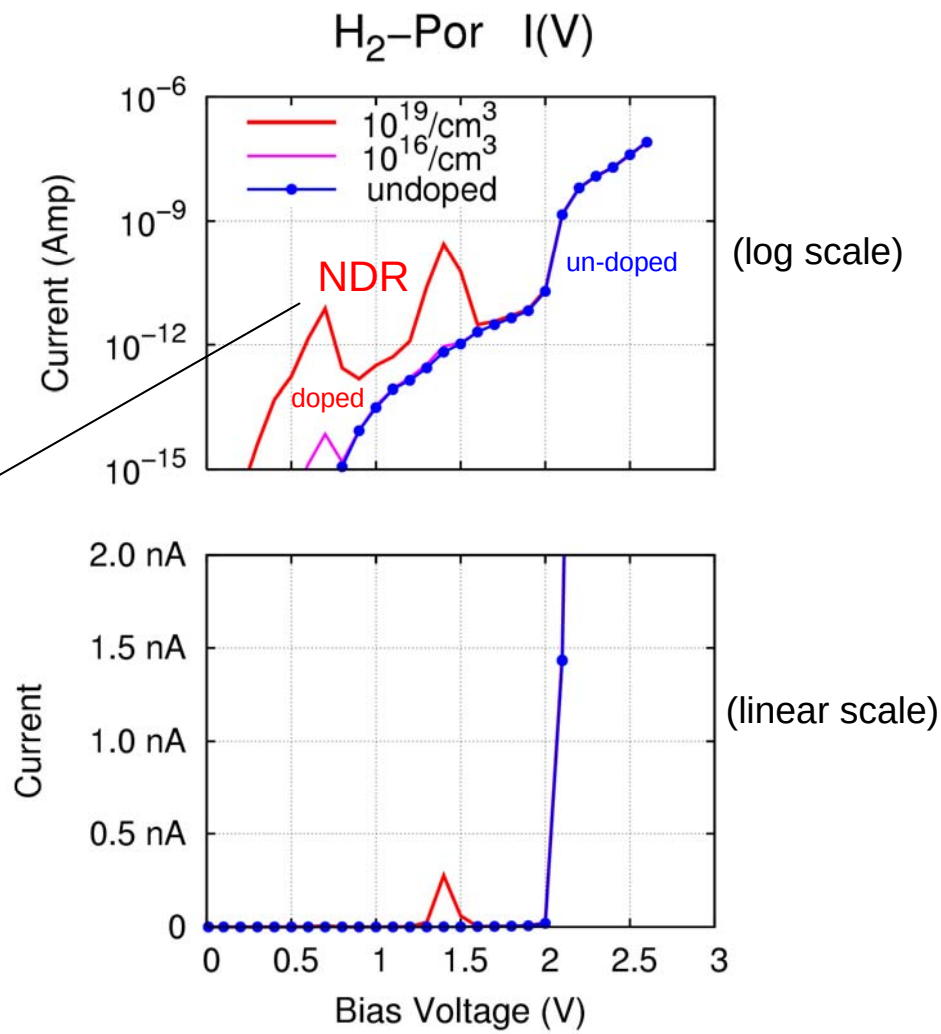


- **Doped case: n-p junction; V is relative to the bias at which the gaps are aligned.**
- **(a) Only weak Si-Si tunneling current**
- **(b) HOMO aligned with top of right valence band and/or LUMO aligned with bottom of left conduction band: strong current through open molecular channels**
- **(c) HOMO and/or LUMO in the gap: low current from Si-Si tunneling**
- **(d) HOMO and/or LUMO become open channels: resonant tunneling**

H_2-Porphyrin NDR

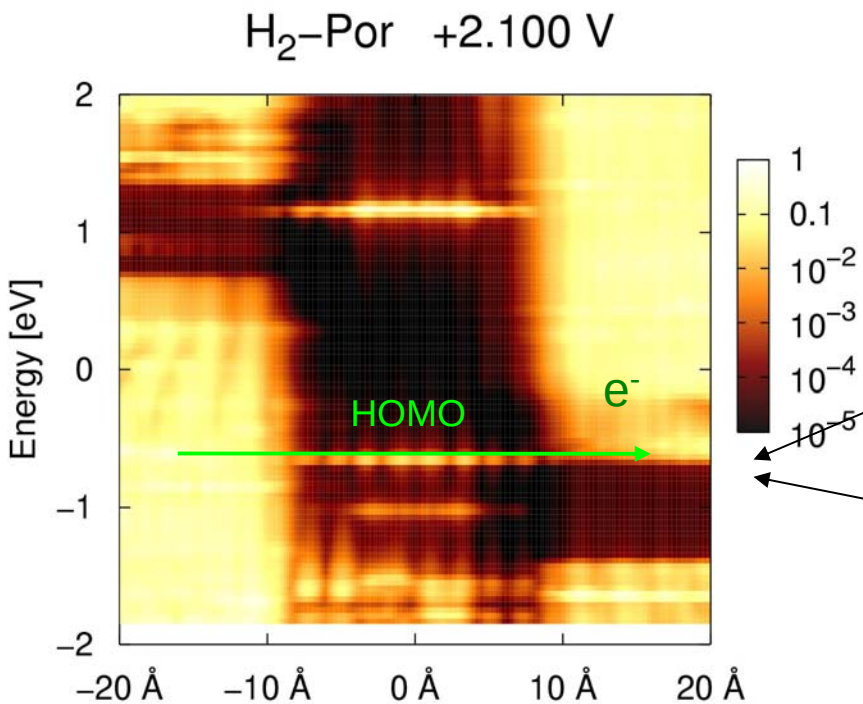


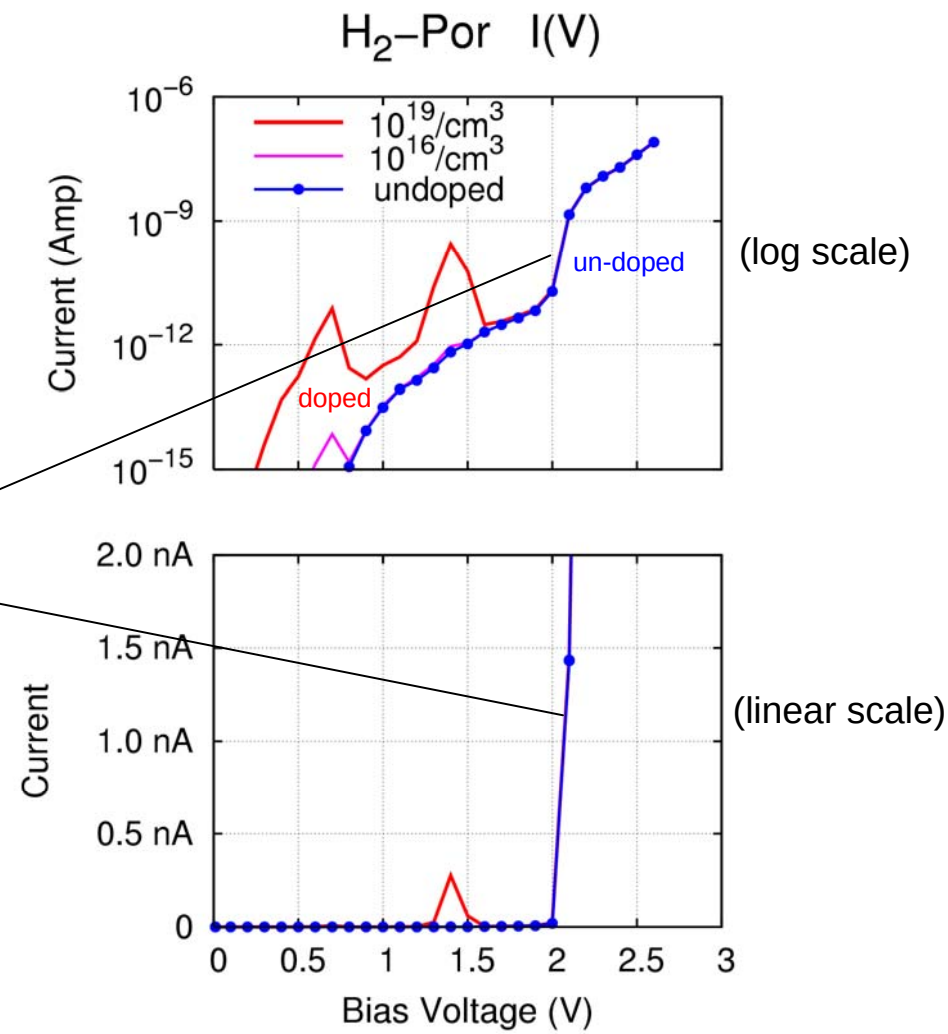




- in the doped p-n case excess holes on the right lead are subject to resonant transmission to the left lead
- for bias > 0.7 V HOMO goes into the Si gap and current decreases

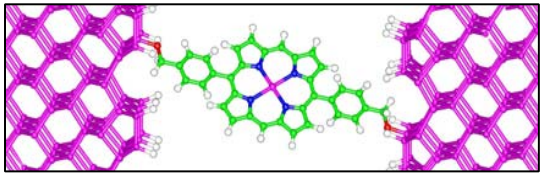
H_2-Porphyrin turn-on BIAS

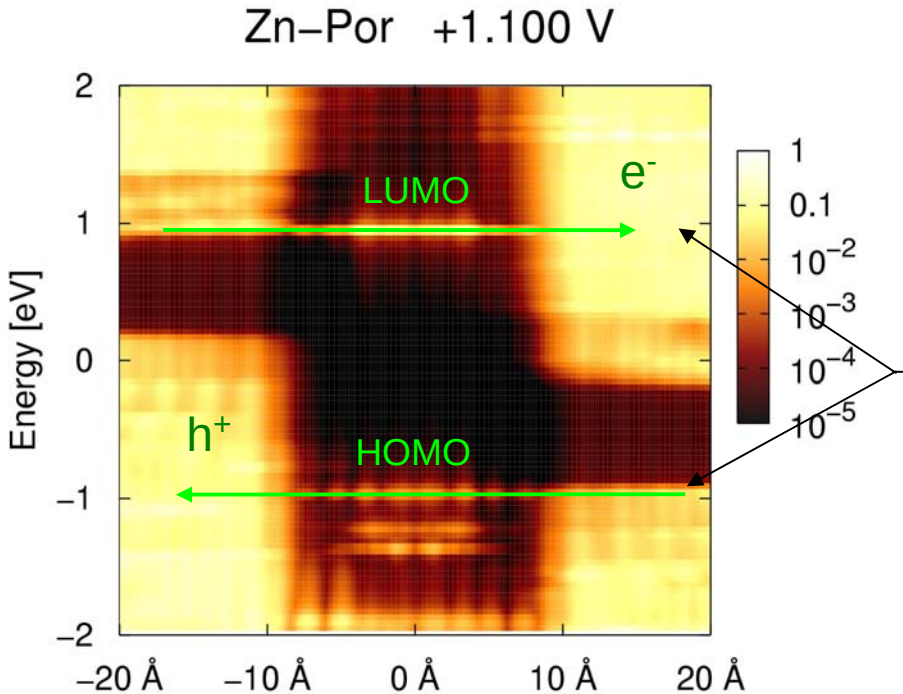


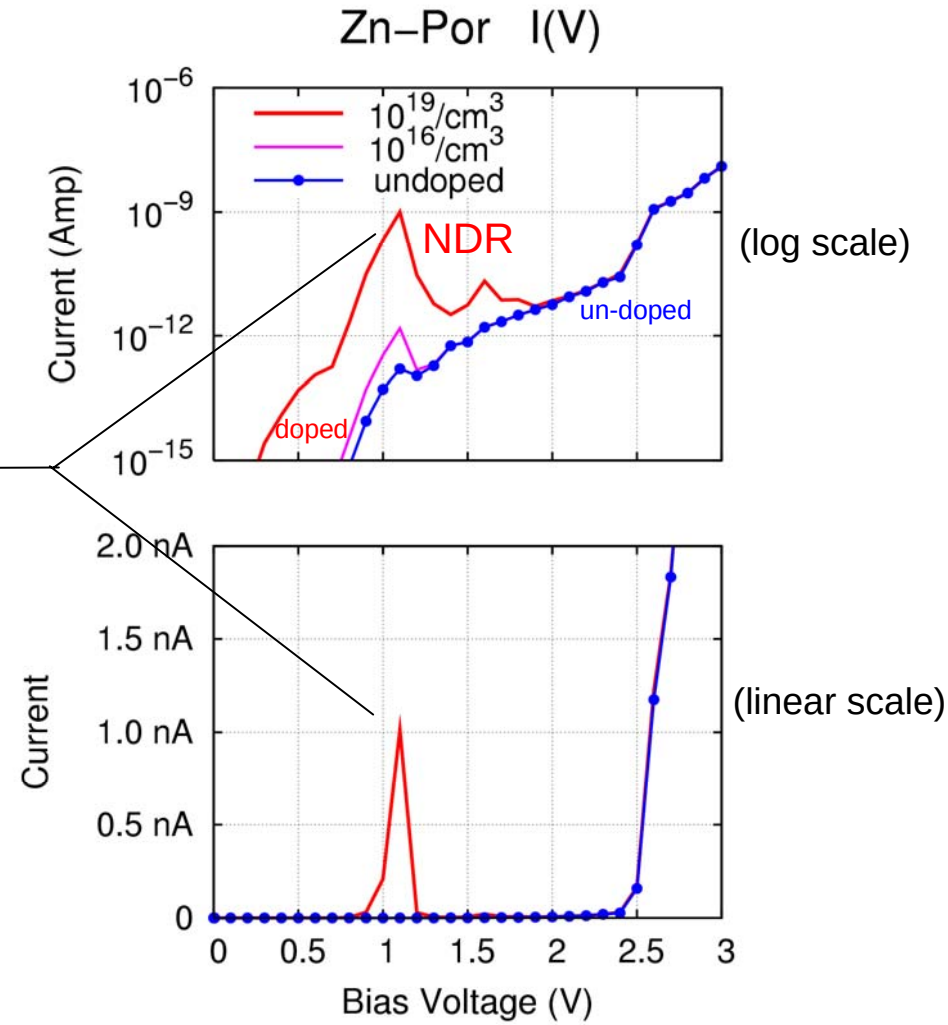


- for bias > 2.1 V the HOMO establishes an open channel
- 2.1V is the turn-on voltage

Zn-Porphyrin NDR







- in the doped p-n case excess holes on the right lead AND excess electrons on the left lead are subject to resonant tunneling
- for bias > 1.1 V both the HOMO and LUMO go into the Si gaps and the current decreases

Summary

- Real space multigrid method (RMG):
  scales up to ~10k processors.
- Hybrid simulation method for biology: DFT + TF
- Linear scaling in memory, comunication, and calculations is required for large systems:
  - nearly O(N) with optimized localized orbitals.
- Transport simulations for nanoscale devices
  - Multiscale parallelization
- Applications to
  - Surface structure determination
  - Simulations of PrP^C
  - Quantum transport in molecular devices
    mechanism of negative differential resistance (NDR)

Mehrstellen in 1D 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

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