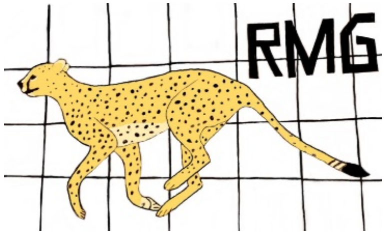
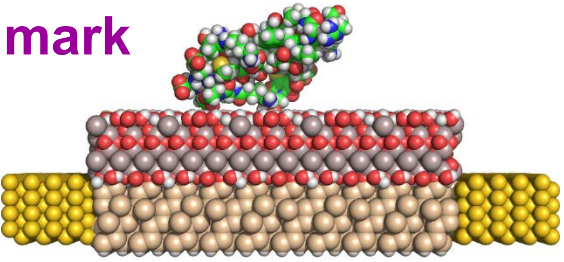
Petascale Quantum Simulations of Nano Systems and Biomolecules

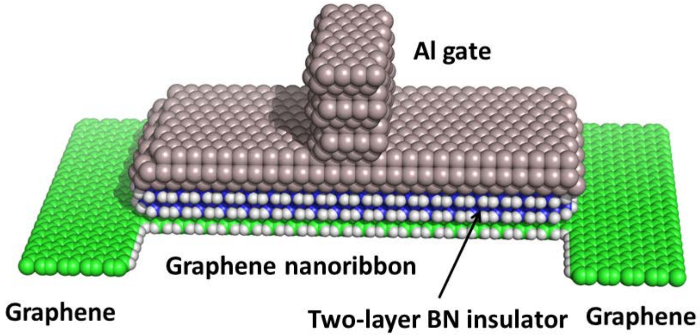
J. Bernholc, E. Briggs, W. Lu,Y. Li and M. Hodak
North Carolina State University, Raleigh



I. RMG – petascale, open-source electronic structure code
   Blue Waters community Portal
   Part of *Sustained Petascale Performance* benchmark



II. Si nanowire-based sensors for potential DNA sequencing
   ~12k atom DFT quantum transport calculations
   Sensitivity versus length & cross section
   Sensitivity enhancement via gate potential



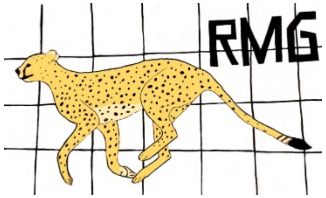
III. Nanoscale transistors for post-Moore's era
   Multi-terminal quantum transport
   Quantum interference and tunneling leakage

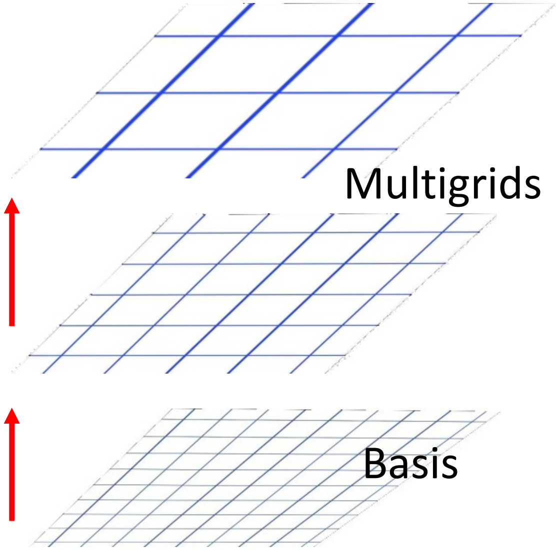
Real-space Multi-Grid method (RMG)

- **Density functional equations solved directly on the grid**
- **Multigrid techniques remove instabilities by working on one length scale at a time**
- **Non-periodic boundary conditions are as easy as periodic**
- **Compact “Mehrstellen” discretization**

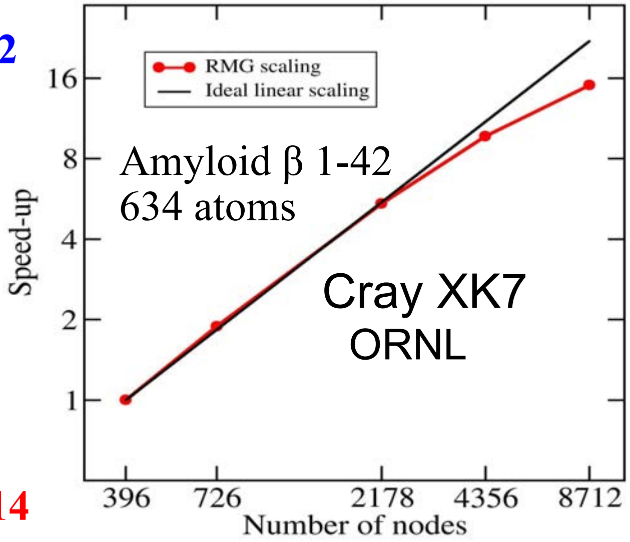
$$A[\phi_i]+B[(V_{eff}+V_{NL})\phi_i]=\varepsilon_i B[S\phi_i]$$

- **Allows for efficient massively parallel implementation**

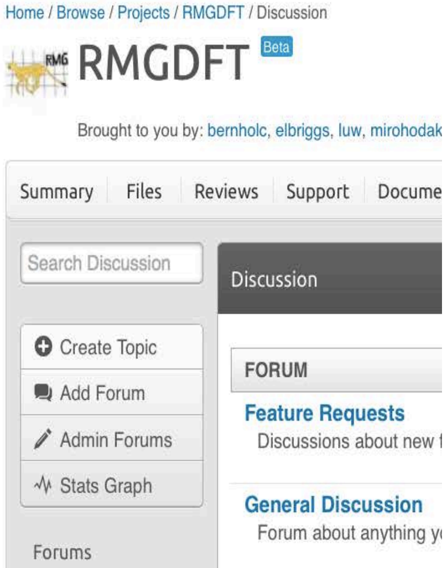




Largest run used **139,392** CPU cores and **8,712** GPU's, > 6.5 PF



1 node = 16 Opteron cores+
1 Nvidia K20x GPU



Home / Browse / Projects / RMGDFT / Discussion

RMGDFT Beta

Brought to you by: bernholc, elbriggs, luw, mirohodak

Summary | Files | Reviews | Support | Docume

Search Discussion

Discussion

Create Topic

Add Forum

Admin Forums

Stats Graph

Forums

FORUM

Feature Requests

Discussions about new

General Discussion

Forum about anything y

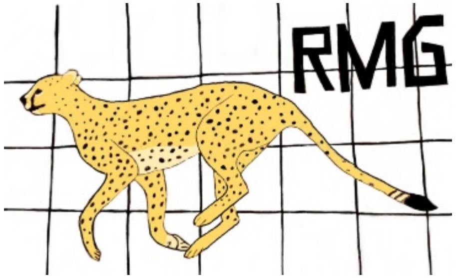
RMG open source

sourceforge.net/projects/rmgdft/

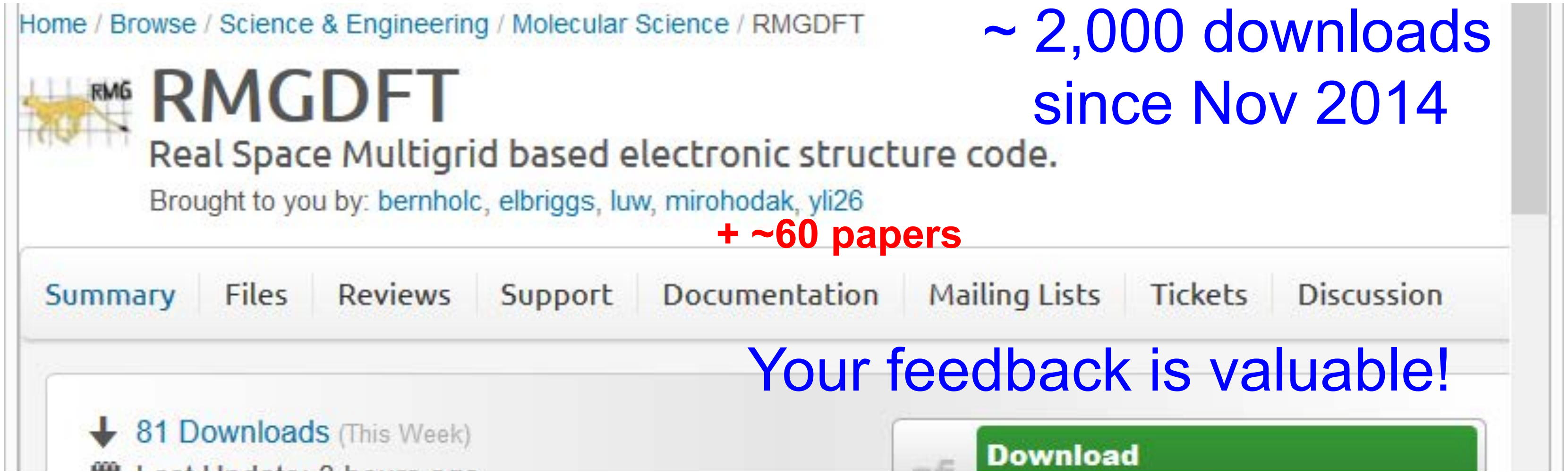
Quantum transport: later in 2017

Performance on 3,872 Cray XK7 (K20x GPU) Blue Water nodes: 1.14 PFLOPS

RMG Code Downloads



- Downloads: https://www.rmgdft.org
- Code & binaries (for Linux distr., Mac, Win)
- Blue Waters portal for Cray XE6/XK7
  https://bluewaters.ncsa.illinois.edu/rmg
- Part of NSF's Sustained Petascale Performance benchmarks
  https://bluewaters.ncsa.illinois.edu/benchmarks
- Graphical User Interface (GUI) for input & output
- Examples, documentation & discussion forum on sourceforge



Home / Browse / Science & Engineering / Molecular Science / RMGDFT

RMGDFT

Real Space Multigrid based electronic structure code.

Brought to you by: bernholc, elbriggs, luw, mirohodak, yli26

81 Downloads (This Week)

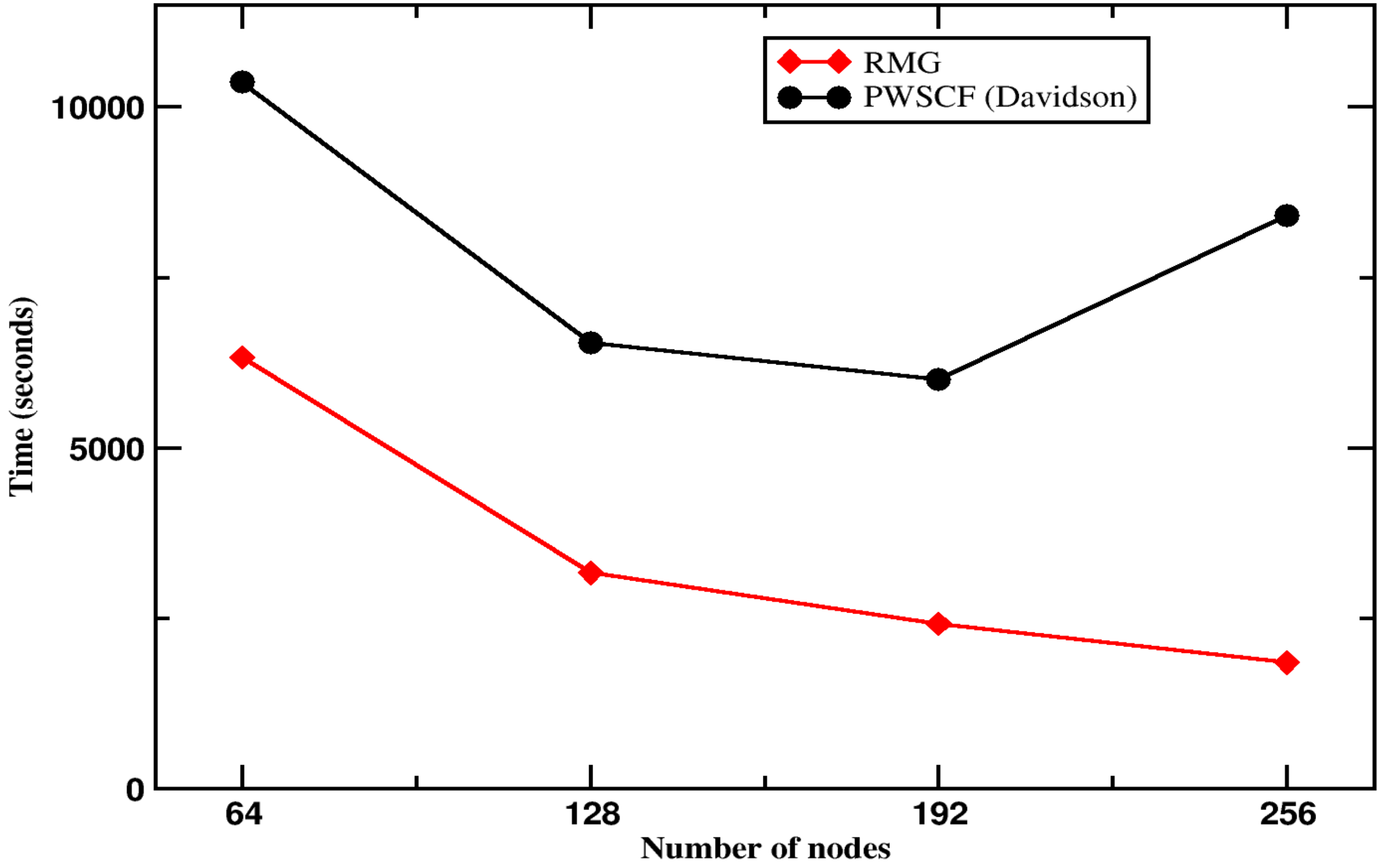
~ 2,000 downloads
since Nov 2014

+ ~60 papers

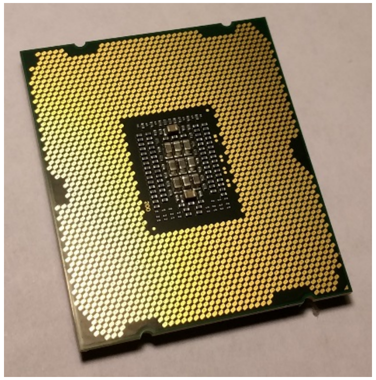
Your feedback is valuable!

Performance and scalability: RMG vs. plane waves

3000 atoms (1000 water molecules, Vanderbilt USPP, initial electronic quench)



RMG and GPU accelerators



Programming models CPU vs GPU

CPU - high clock speed, smaller number of powerful execution units.
Memory latency hidden by caches and out of order execution.
Good single threaded performance.

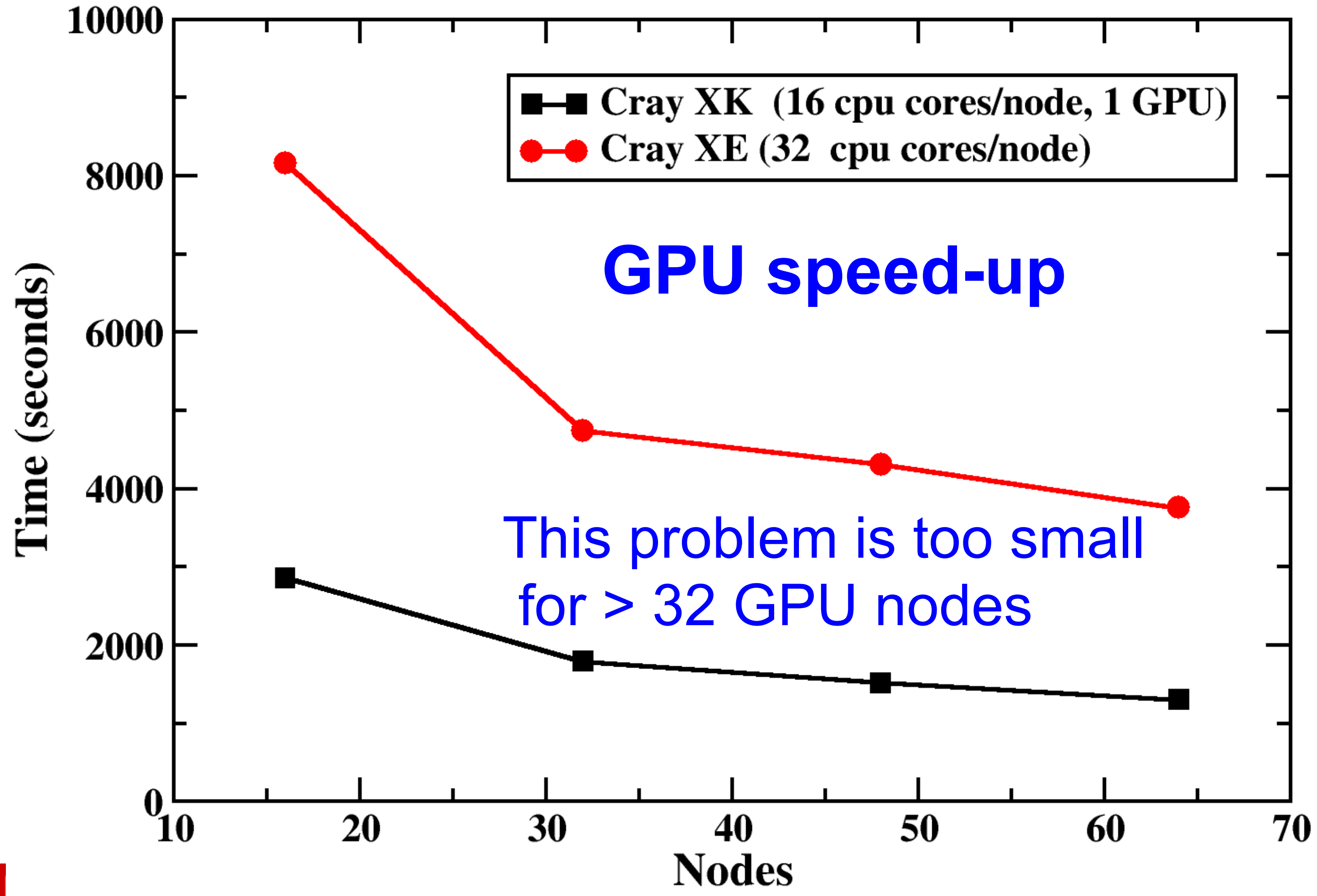
GPU – lower clock speed, large number of weaker execution units.
Memory latency hidden by high thread counts.
Poor single-threaded performance.

Large fraction of RMG codebase is well adapted for the GPU model.

Large data sets required though. CPUs work better for very small problems.

GPU Speed-up

Bulk copper 256 atom cell initial electronic quench



GPU speed-up

This problem is too small for > 32 GPU nodes

GPU performance workstation level

Workstation calculation

Dual Xeon E5-2630v2 workstation. Total of 12 CPU cores and 32GBytes RAM.
New Nvidia GP100 Pascal, 16GBytes, HBM memory and 5.3 TFLOPS double precision.

Test problem 256 atom copper cell

Vanderbilt ultrasoft pseudopotentials with 18 beta functions/atom.
Total of 1536 electronic orbitals. PBE XC functional.

Execution time dominated by eigensolver and large matrix operations

CPU only run required 94.4 seconds/SCF step.
CPU/GPU run required 19.7 seconds/SCF step.

A single GPU produced a speedup by a factor of 4.8!

(GP100 received in March 2017, code not optimized yet, further improvement expected)

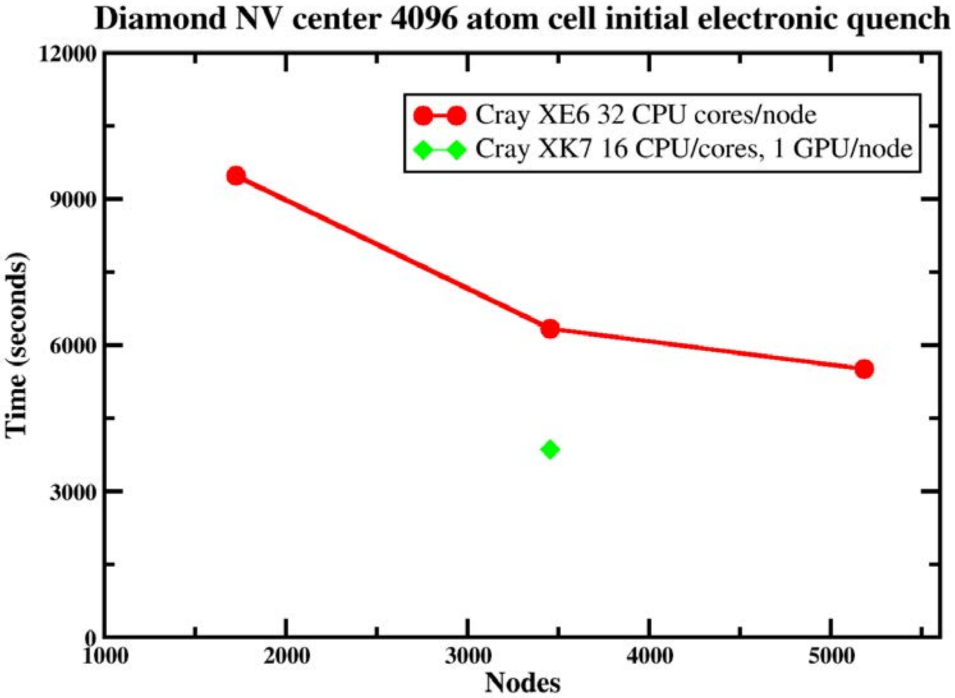
GPU performance: supercomputer level

Quantum computing candidate
Nitrogen-vacancy complex in diamond: spin-polarized DFT.
4095 atoms, 18432 orbitals. Norm-conserving pseudopotentials (FHI).

Cray XE6 results using 5184 nodes
165888 CPU cores (32 cores/node)
Initial electronic quench
5571.6 seconds, 52.8 GFlops/node

Cray XK7 results using 3456 nodes
55296 CPU cores (16 cores/node),
3456 Nvidia GPU accelerators (1/node)
Initial electronic quench
3860.64 seconds, 113.0 GFlops/node

Cray XE6 1728 nodes
99.8 Gflops/node 30% of peak.

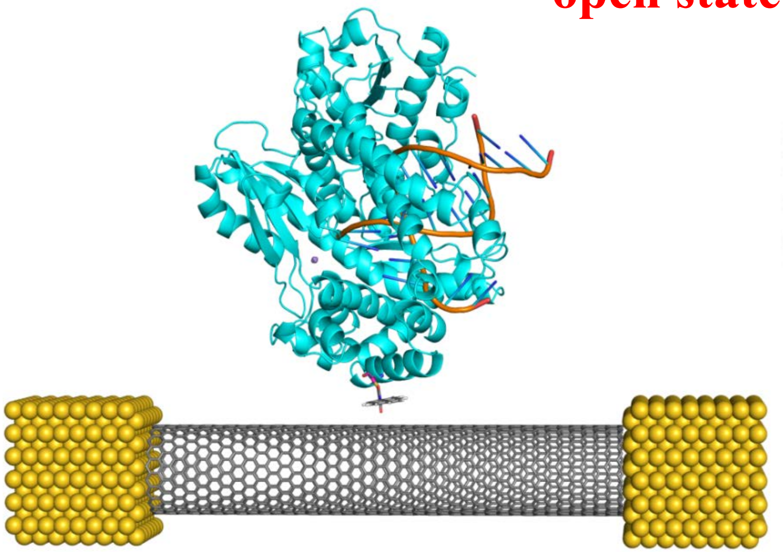


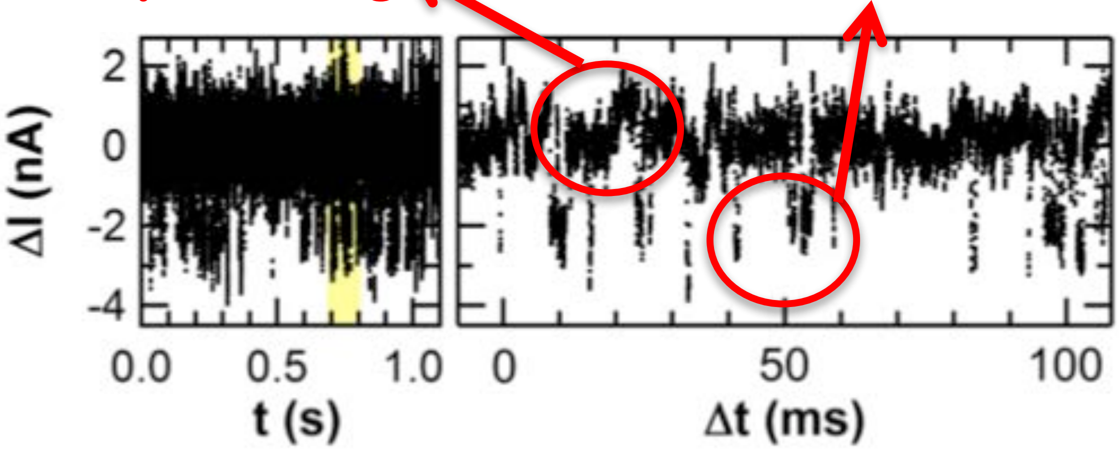
- Part of NSF's Sustained Petaflops Performance benchmark
  https://bluewaters.ncsa.illinois.edu/spp-methodology
  https://bluewaters.ncsa.illinois.edu/spp-benchmarks

Introduction to DNA Replication

- Klenow Fragment (KF) from DNA Polymerase I is an enzyme that synthesizes double-stranded DNA from single-stranded templates
  - High replication rate, 100s of base pairs per second
  - Low error frequencies, 1 error in 10^6 base pairs

- Experiments observe a different current when the enzyme has just added a DNA base
  - Potential for DNA sequencing, if different bases lead to different currents.





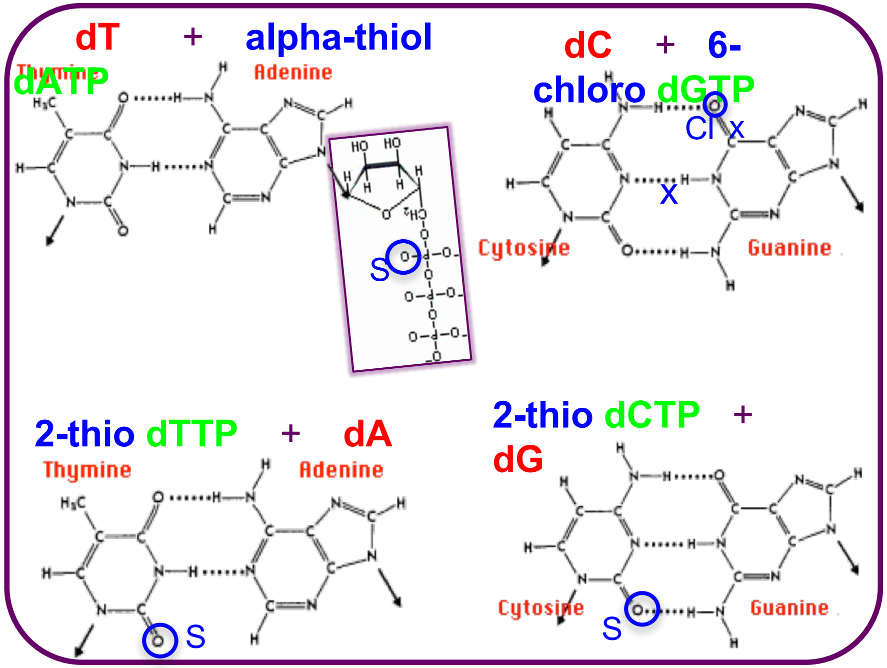
open state: not synthesizing

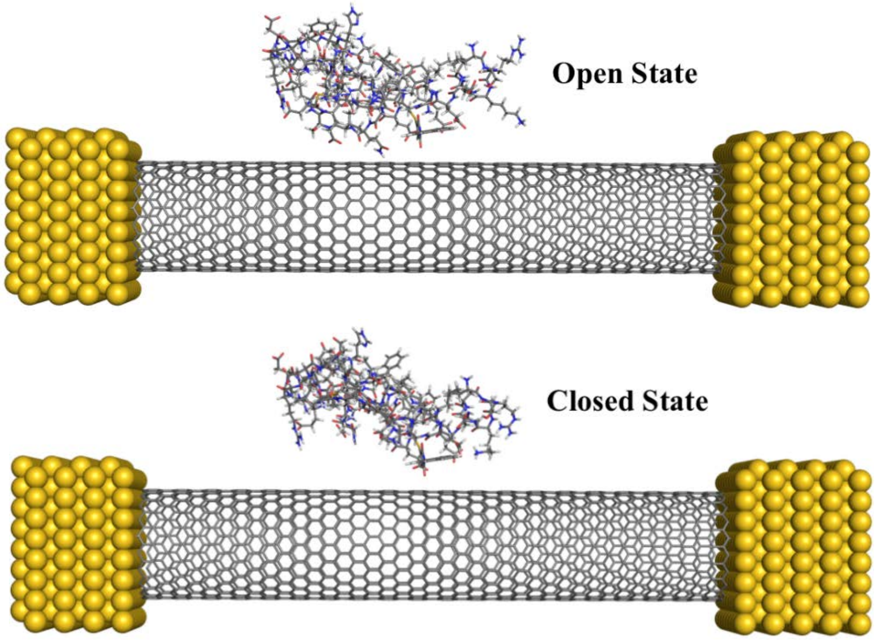
closed state: synthesizing

T. J. Olsen, Y. Choi, P. C. Sims, O. T. Gul, B. L. Corso, C. Dong, W. A. Brown, P. G. Collins, and G. A. Weiss, JACS 135, 7855 (2013)

Our previous results on nanotube sensing

- Gate potential scanning helps to distinguish between DNA bases
- Nucleotide substitutions can be used to enhance the current difference,
  - Calculated signals can identify T and G, but cannot distinguish between C and A





K. M. Pugliese, O. T. Gul, Y. Choi, T. J. Olsen, P. C. Sims, P. G. Collins, and G. A. Weiss, JACS 137, 9587 (2015)

Self-Consistent Electron Transport for >10k Atoms

Coupling between lead and conductor

Coupling between lead and conductor

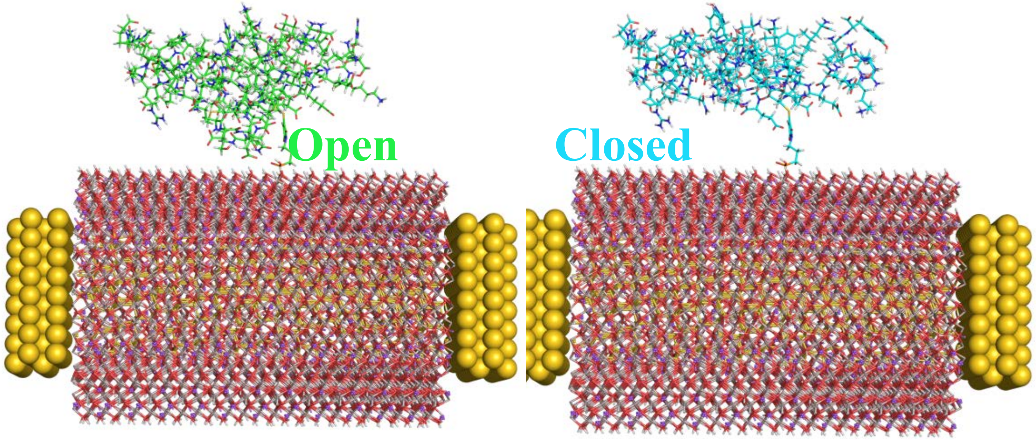
- **Non-equilibrium Green's function (NEGF) method in a basis of variationally optimized orbitals** *Lu, Meunier, Bernholc, Phys. Rev. Lett. 2005*
- **Optimal basis set:**
  - **Variationally-optimized localized orbitals: a smaller and more complete basis set than in other localized-orbital methods.** *Fattebert & Bernholc, Phys. Rev. B 2000*
- **Green's functions are calculated iteratively with linear scaling.**
- **Multi-level parallelization**
  - **MPI between nodes**
  - **Full multi-core and multi-GPU support**
  - **Scales to thousands of nodes**

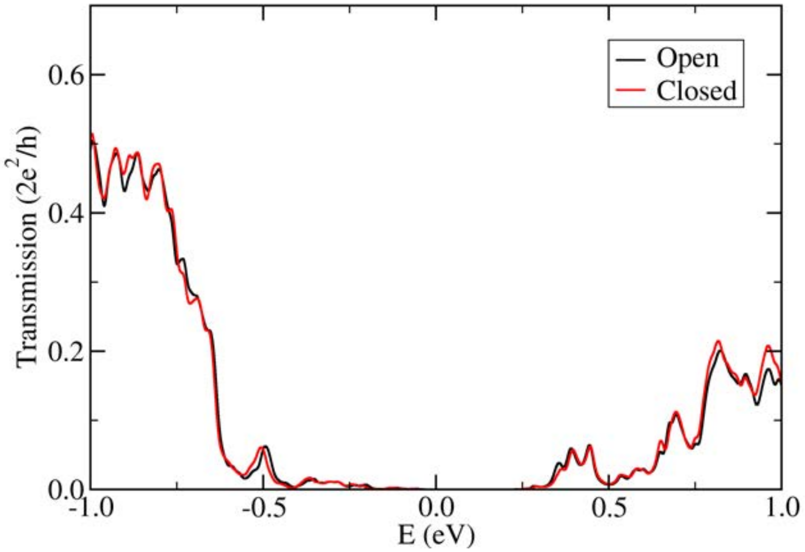
We can calculate conductances and I-V curves for multi-terminal devices from first principles (DFT) for large systems (>10,000 atoms)

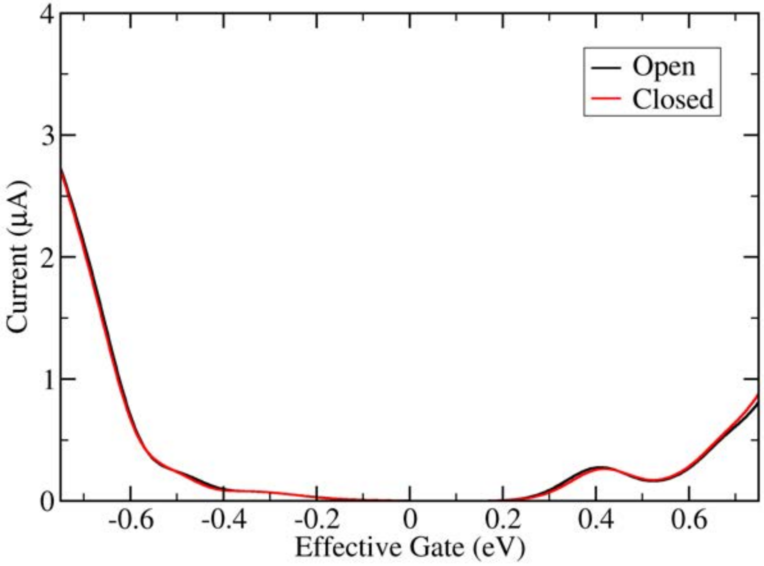
$$I(V) = \frac{2e^2}{h} \int_{-\infty}^{\infty} T(E,V)[f(E-\mu_L) - F(E-\mu_R)]\,dE$$

Long Si nanowire with Al_2O_3 coating

- **Nanowire (NW) size:**
  - **Si diameter: 5 nm, Length 5.64 nm**
  - **Al_2O_3 thickness: 1nm**
- **A transport gap of 0.5 eV appears for this nanowire.**
- **Transmission is nearly zero at the Fermi level, no tunneling current.**
- **Currents are calculated with source-drain bias of 100 meV.**

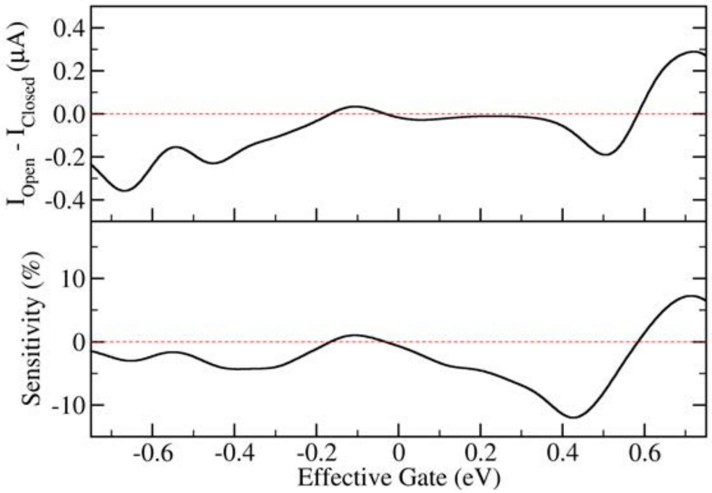


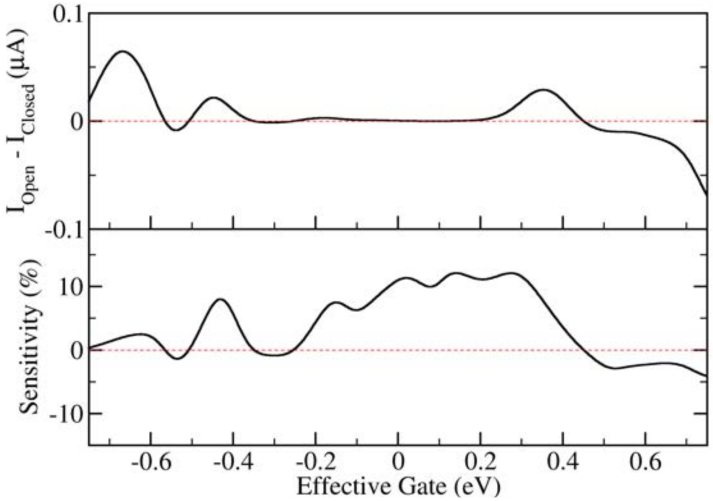




Si nanowire with Al_2O_3 coating: sensitivity

$$\text{Sensitivity} = \frac{I_{open} - I_{closed}}{I_{open} + I_{closed}}$$

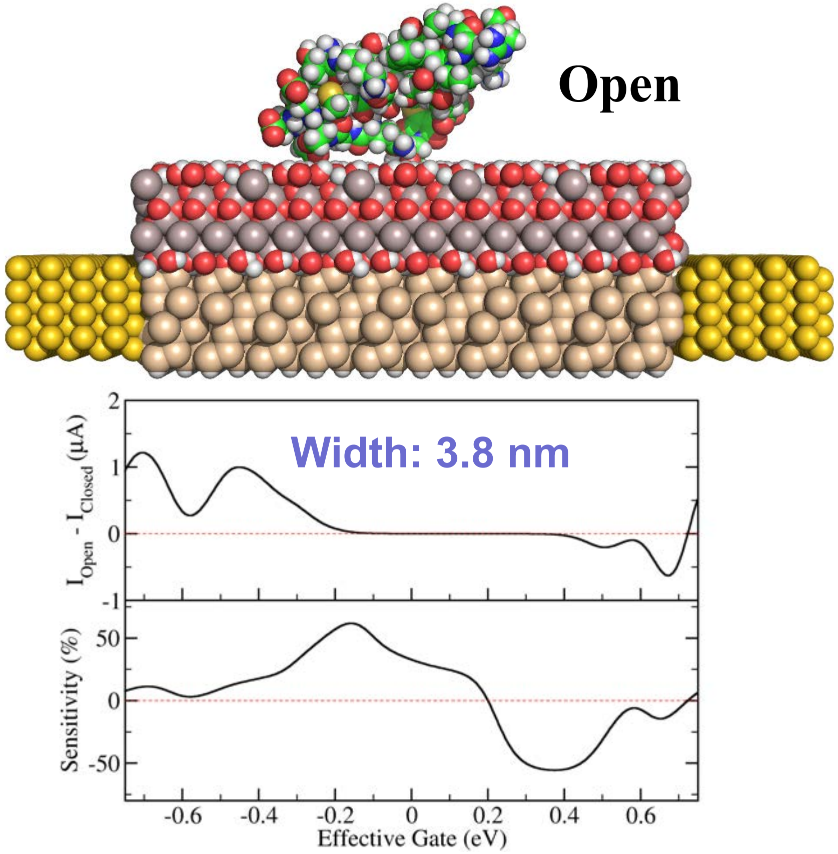

Si nanowire, diameter = 5 nm, length = 2.8 nm

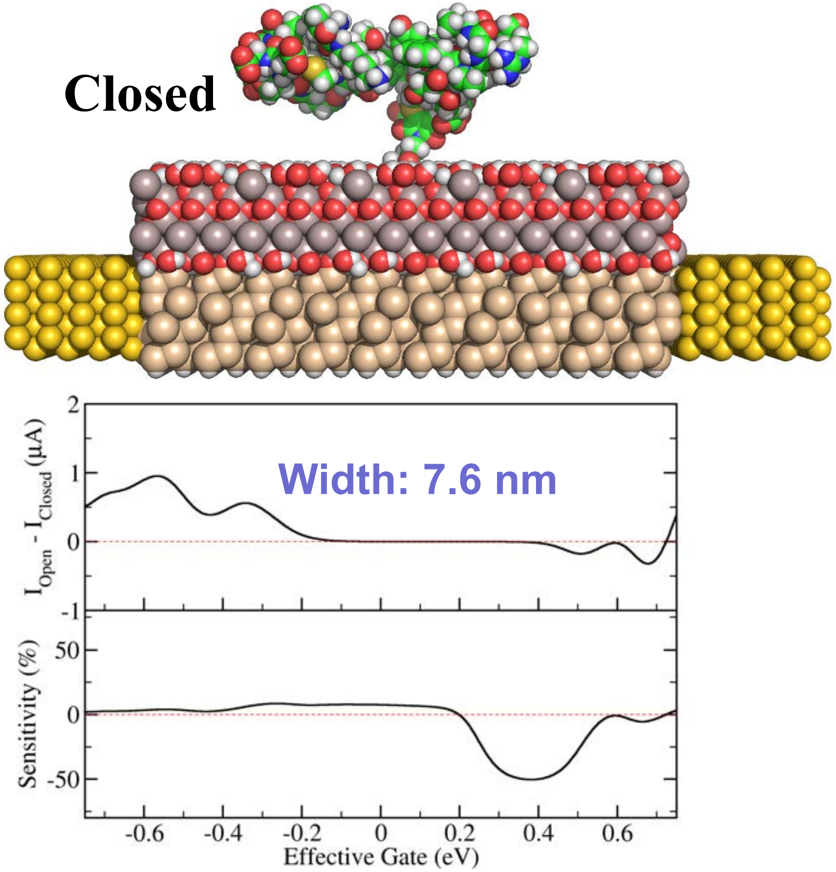

Si nanowire, diameter = 5 nm, length = 5.6 nm

- **Short Si nanowire (2.8 nm):**
  - **At effective gate of 0.4 V, the sensitivity is about 10%, the current difference between open and closed states is 0.18 μA.**
  - **At negative gate, the surface states dominate: not suitable for sensing.**
- **Long Si nanowire (5.6 nm):**
  - **At effective gate of -0.43 V or 0.35 V, the sensitivity is 8% and the current difference is 0.02 μA**
  - **In subthreshold region, the sensitivity is 10%, but the current is very small (0.3 nA)**

Si Slab with Al_2O_3 Passivation

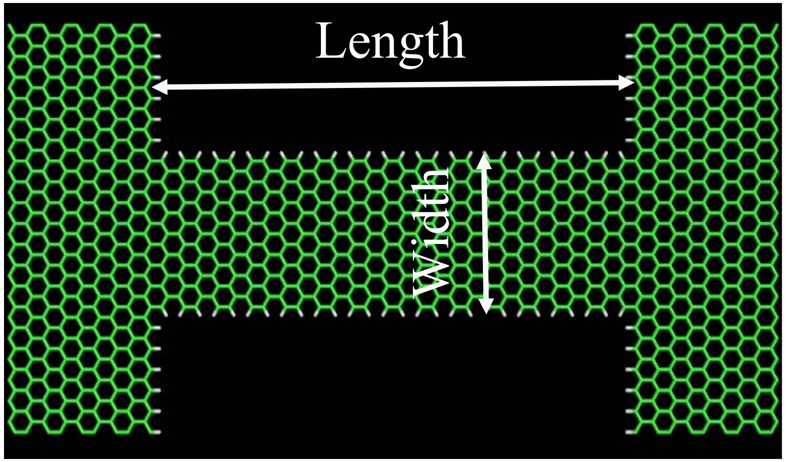
Si slab thickness: 1 nm, length: 6.7 nm, Al_2O_3 thickness: 1 nm





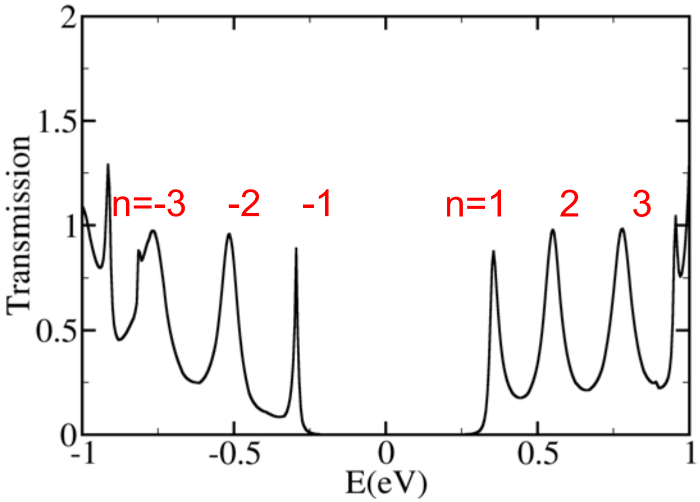
- **Transport gap of 0.6 eV for both widths**
- **Subthreshold (-0.2 to 0.4 V): sensitivities are large but currents are small.**
- **Turn-on state (gate at -0.3 and 0.5 V):**
  - **Sensitivity is reduced when slab width increases, because the base current becomes larger.**
  - **The absolute current difference does not change dramatically.**

Nanoribbon Confinement and Transmission



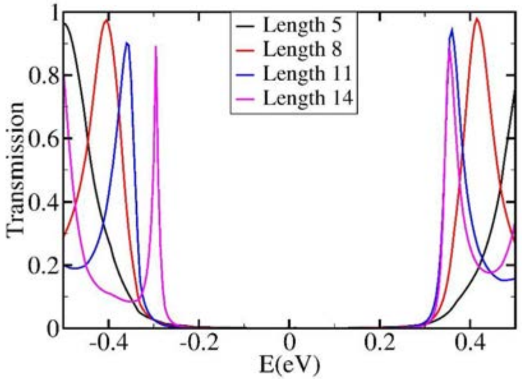
- Leads: bulk graphene
- Conductor part: armchair nanoribbon
  width from 3 to 15 atomic layers
  length from 5 to 14 atomic layers
- Dangling bonds are saturated with hydrogens
- Number of atoms in NEGF calculations
  800 to 1300 atoms depending on width/length

L = 14, W = 12 atomic layers

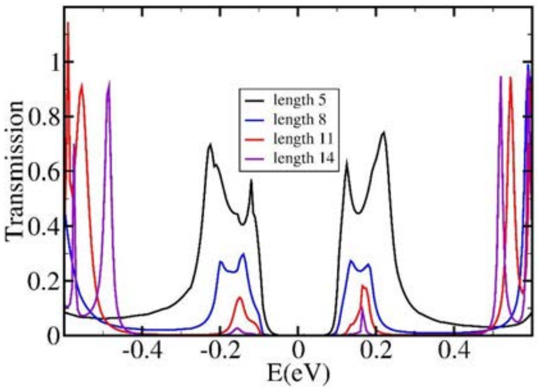


- Finite-size GNRs confined between contacts lead to Fabry-Perot interference patterns
  Liang, Park et al, Nature 2001; Rickhaus, Maurand et al, Nat. Com. 2013; Yannouleas, Romanovsky, Landman, JPC 2015, Sci. Rep. 2015
- The interference patterns depend on length, width, GNR index and passivation.
- Interfaces can also induce localized states.

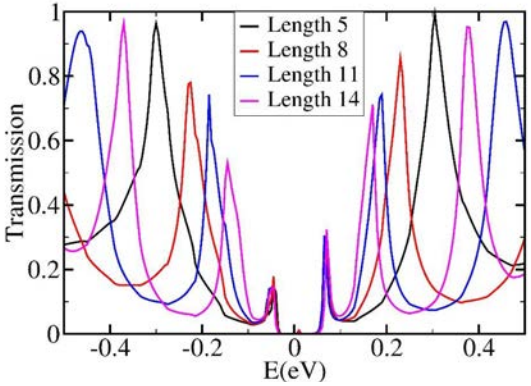
Nanoribbon Families and Suitability for Transistors



- Width = 12, mod(n,3) = 0
- Band gap is large with no localized states inside band gap for a finite length nanoribbon.
- OFF state current can be very small because of the big band gap
- It is a good candidate for a transistor, which requires high ON/OFF ratio.



- Width = 13, mod(n,3) = 1
- Pure nanoribbon has a large band gap, while there are localized states inside the band gap for a finite length nanoribbon.
- Transmission peaks inside the gap decrease dramatically with length
- OFF-state current can be very small because of the big band gap

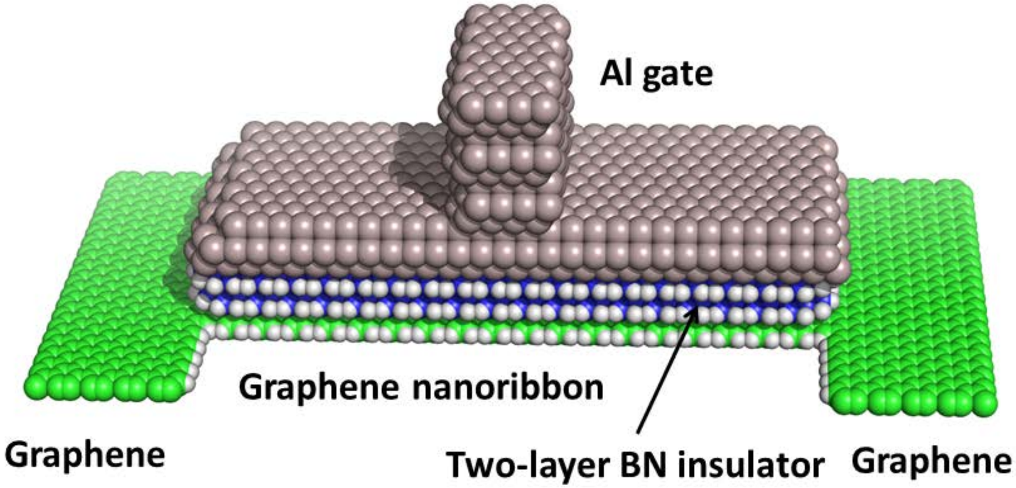


- Width = 14, mod(n,3) = 2
- Small bandgap family
- It is hard to have a small current for OFF state
- It is not a good candidate for a transistor, which requires high ON/OFF ratio.

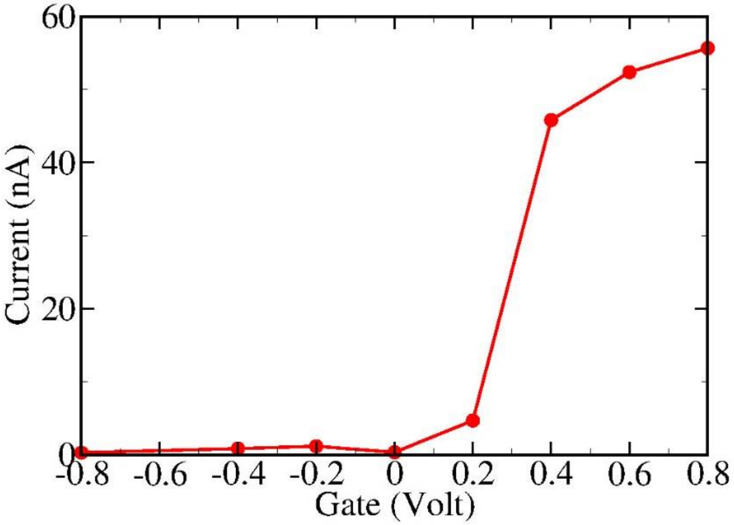
- For a moderate-size device, nanoribbons with mod(n,3) = 0 are the best choice.
- The first transmission peak appears at a small energy for a long ribbon.
- The nanoribbon needs to be long enough to avoid direct tunneling current.

Nanoribbon Transistor Simulation

Best nanoribbon for transistor: mod(n,3) = 0 family of armchair nanoribbons



- Nanoribbon: 18 atomic-layers wide
- 40 atomic-layers long
- Graphene sheet as source and drain
- 2 layers of BN as the insulator
- A few layers of Al as the gate
- 4536 atoms in the NEGF calculation



- Source-drain bias: 50 mV
- At a positive gate of 0.4 V, the device is turned on with ON/OFF ratio = 500
- The small ON/OFF ratio in part due to too small band gap in calculations (DFT artefact).

Summary

- **RMG -- a petaflops-capable open source electronic structure code**
  - **Scales to 20k nodes and 200k CPU cores**
  - **Effective use of GPUs and multi-core CPUs**
  - **Pseudopotential libraries: ultrasoft & norm-conserving**
  - **Cray installation files, Linux, Windows and Mac binaries**
  - **Graphical user interface (GUI)**
  - **Released under GPL: www.rmgdft.org**
  - **Blue Waters community Portal: https://bluewaters.ncsa.illinois.edu/rmg**
  - **Part of NSF's *Sustained Petascale Performance* benchmarks**

- **Applications**
  - **Monitoring of DNA replication using nanotubes and Si nanowires**
    - **Potential for DNA sequencing**
  - **Nanoscale devices for Post Moore's Law era**
    - **Nanoribbon-based transistors are feasible**
      - **Quantum interference patterns in small structures**
      - **Specific nanoribbon indices are required**