Computational Nanoelectronics

Eric Polizzi

Department of Electrical and Computer Engineering,
University of Massachusetts, Amherst, USA

PMAA06-September 8, 2006

Computational Nanoelectronics

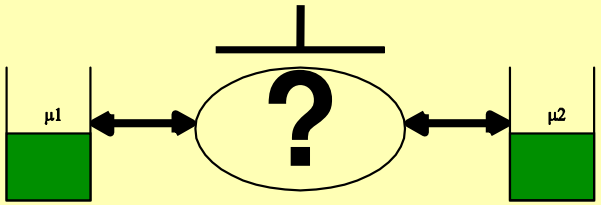
Overview

Ultimate size of traditional MOS device - End of scaling ?

Scaling of processor speed

Scaling of #processors

Nanoelectronics research



Modeling and Simulation

Development of new parallel computing platforms (multicore)

New computational power at hand

algorithm design, library innovation, adaptive library

Computational Nanoelectronics

Overview

needed for

- ***investigating the ultimate size of traditional MOS transistor devices***
- ***leading experiments in the exploration of new class of devices that relies on quantum effects***

the need to account for high degrees of detail and realism (quantitative results)

- *Multi-scale and multi-dimensional aspect of electron transport and electrostatics effects,*
- *High-fidelity treatment of the contact regions,*
- *Information of the material at the atomistic level,*
- *Efficient methodology for the treatment dissipative scattering,*

➔ **Computationally challenging problems**

Computational Nanoelectronics

How to proceed ?

Important features for a Nanoelectronics Simulator:

- ***Addressing high-fidelity simulations of material & transport phenomena***
- ***Flexible to handle arbitrary devices, different physics, materials, physical dimensions and geometries***

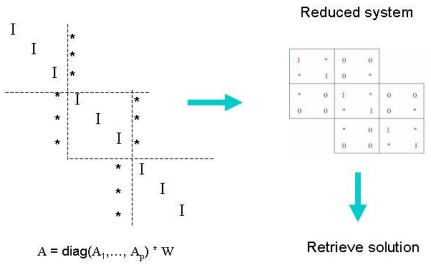
  ➔ ***Prototyping new-class of devices***

NESSIE (1998-2006)

A brief overview

Numerical Parallel Algorithms:

Linear solver (SPIKE), Eigenpairs solver (TraceMin), preconditioning strategies,…



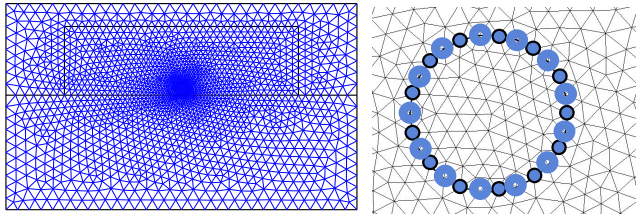
Transport/Electrostatics

Multi-scale, Multi-physics

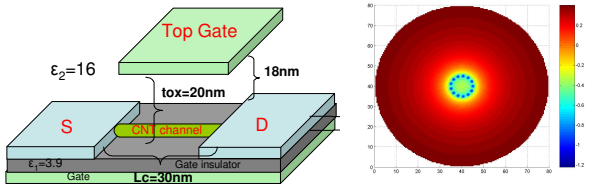
Multi-method

Mathematical Methodologies:

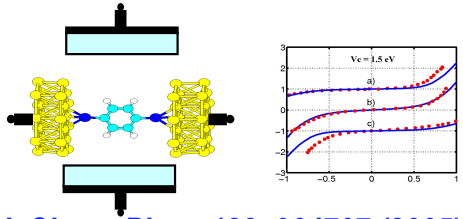
Finite Element Method, mode decomposition, multi-scale, non-linear numerical schemes,…



3D CNTs-3D Molec. Devices

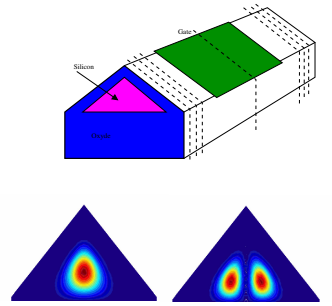


Appl. Phys. Let., 88, 242106 (2006)



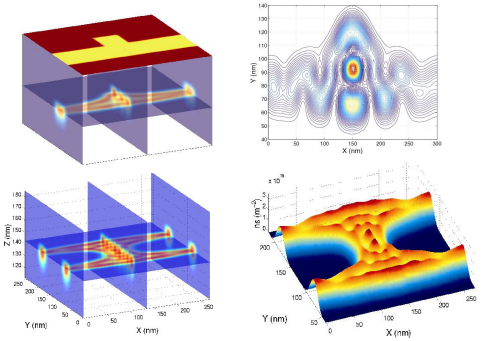
J. Chem. Phys. 123, 064707 (2005)

3D Si Nanowires



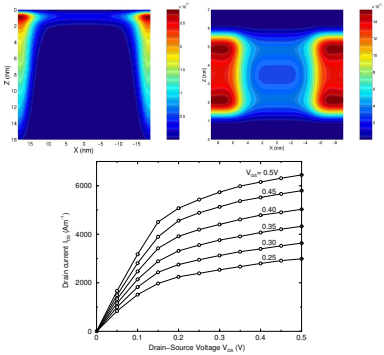
J. Appl. Phys. 96, 2192 (2004)

3D III/V Devices



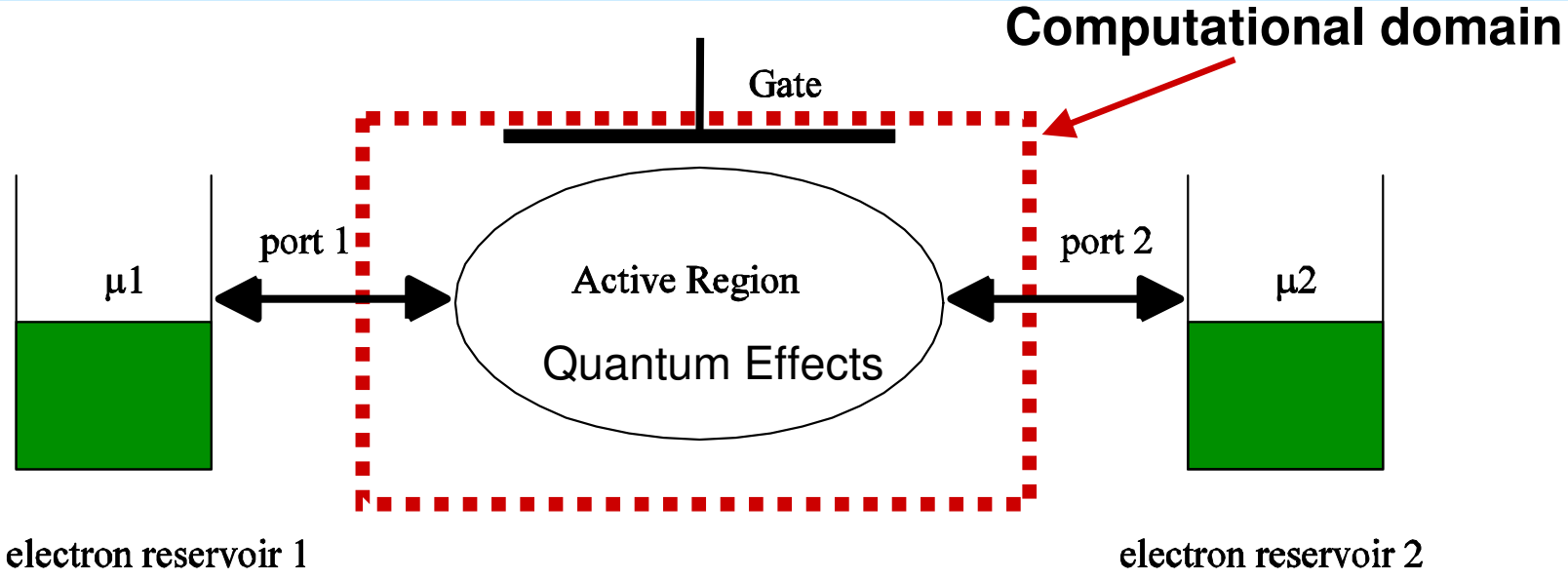
Phys. Rev B,. 66, 245301 (2002)

2D MOSFETs

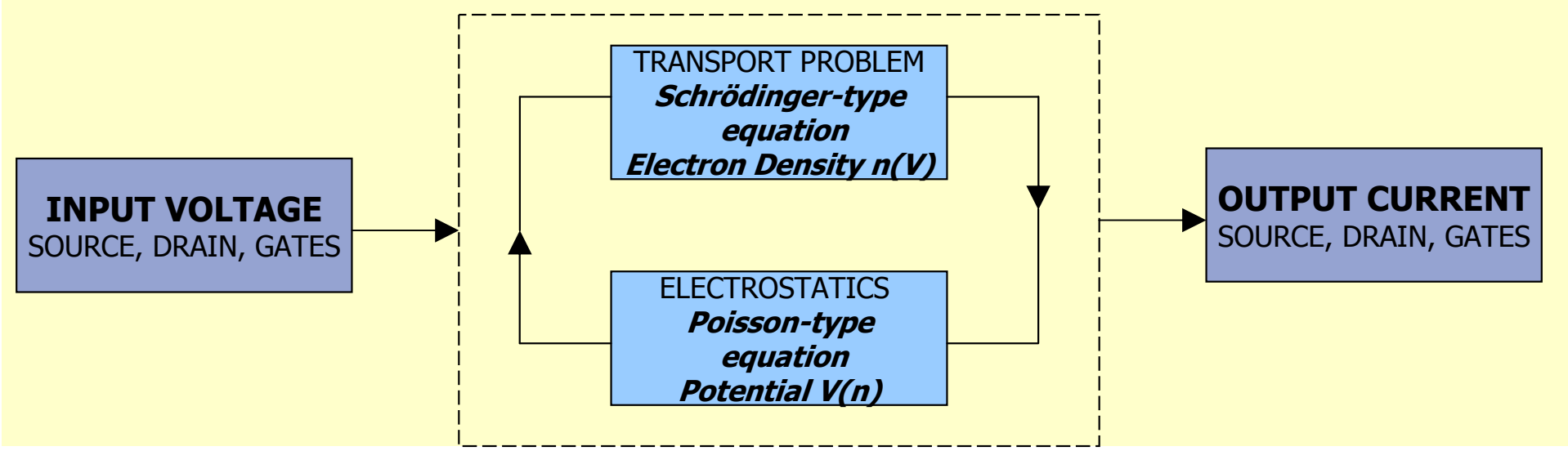


J. Comp. Phys. 202,150 (2004)

Quantum Transport Modeling:
Introduction



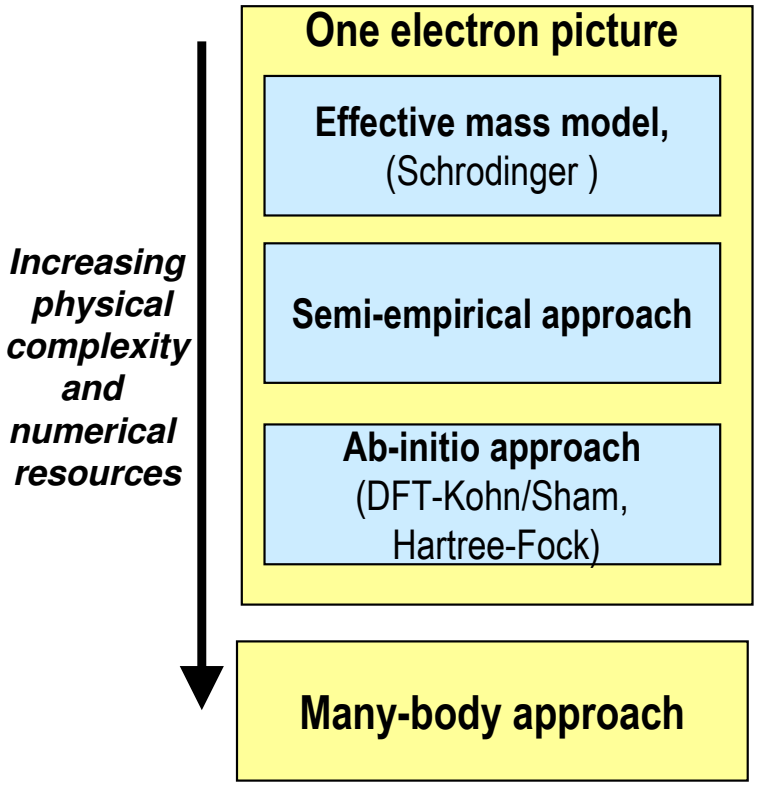
Current-Voltage Characteristics obtained by **self-consistent simulations**: **Transport-Electrostatics**



Quantum Transport Modeling:
Open systems- Ballistic transport

$$(E - H(V))\,\Psi_E = 0; \quad + \quad \text{B.C. (Boundary Conditions)}$$

Electronic structure



Nature of the problem

- **Dirichlet or Neumann B.C.**
  - *Isolated system*
  - *Energy levels are discretized*
- **Periodic B.C.**
  - *Infinite system*
  - *Bandstructure calculations (via Bloch states)*

Eigenvalue problems

- **Mixed B.C. (Energy dependent)**
  - *Open system*
  - *Energy levels form a continuum (Broadening)*

Linear systems

$$\Psi_j = \Psi_j^{in} + \Psi_j^{out} \qquad \left(\frac{\partial}{\partial \eta_j} - K_j(E)\right)\Psi_j^{out} = 0$$

Quantum Transport Modeling:
Open systems – FEM discretization

- **Variational form of the problem** (existence and uniqueness: Ben Abdallah et al.):

$$\text{Find } \Psi \in H_g^1 = \{\Psi : \Omega \to C;\ \Psi \in H^1(\Omega);\ \Psi = 0 \text{ on } \gamma_0\},\ \text{such that } \forall \varphi \in H_g^1 :$$

$$\frac{\hbar^2}{2}\int_\Omega \frac{1}{m^*}\nabla\Psi\nabla\varphi d\Omega + \int_\Omega(-qV-E)\Psi\varphi d\Omega - \sum_j \frac{\hbar^2}{2}\int_{\gamma_j}\frac{1}{m^*}\left(K_j\Psi\right)\varphi d\gamma_j = L_{j0}^{ex}$$

- **Finite Element Discretization**: $\Psi(\mathbf{x}) \simeq \sum_{i=1}^{N} \psi_i \omega_i(\mathbf{x})$

$$\left(E[\mathrm{S}] - [\mathrm{H}(\mathrm{V})] - \sum_{\mathrm{j}} [\Sigma_{\mathrm{j}}(\mathrm{E})]\right)\Psi_{j0}^{ex} = \mathrm{L}_{j0}^{ex}$$

For the calculation of electron and current densities, it is necessary to solve a large number of those linear systems- Integration over the energy

Numerical Techniques:
Computational Challenges in Nanoelectronics Simulations

Realistic Device Simulations

- *Multidimensional geometries*
- *Detailed description of the electronic structure*
- *Dissipative Scattering*
- *Etc…*

➔ **High numerical costs**

➔ **Huge computer resources**

For only one point in the I-V curve	Full 2D	Full 3D
Matrix size	$O(10^4)$	$O(10^6)$
linear systems to solve by iteration	$O(10^3)$	$O(10^3)$
# self-consistent iterations	O(10)	O(10)
Simulation time	O(**hours**)	O(**days**)

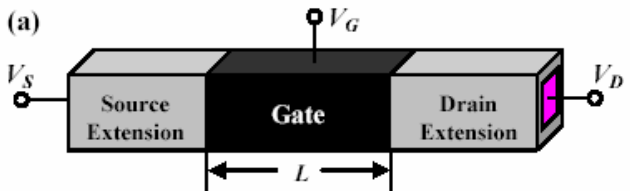
Standard strategy

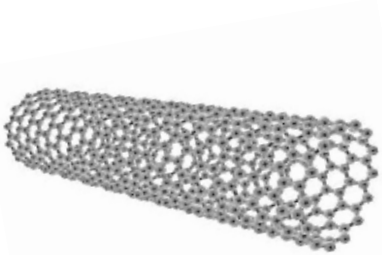
- *Each processor handles many linear systems, one at a time*
- ***Sequential algorithms are inefficient for obtaining I-V curves on large systems***

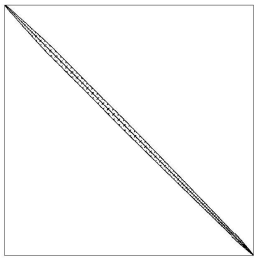
Proposed strategy

- ***Development of high performance numerical parallel algorithms enabling large-scale nanoelectronics simulations such as the linear solver SPIKE***

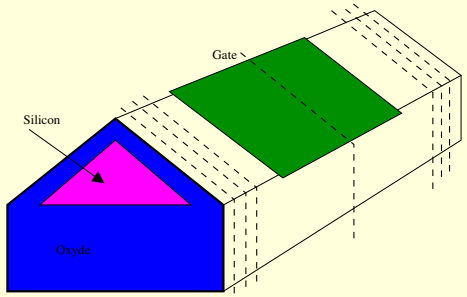
Numerical Techniques:
Matrix representation

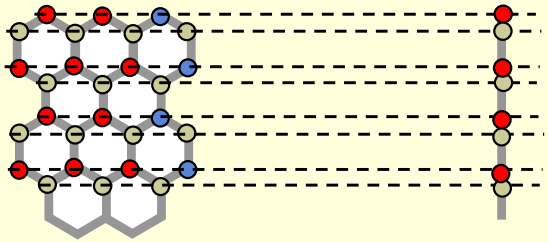


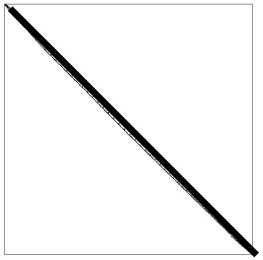




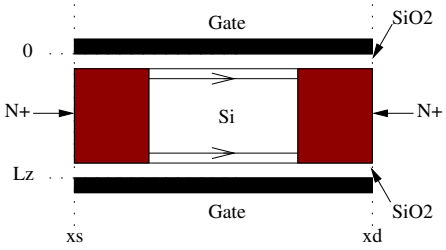
Sparse banded linear systems

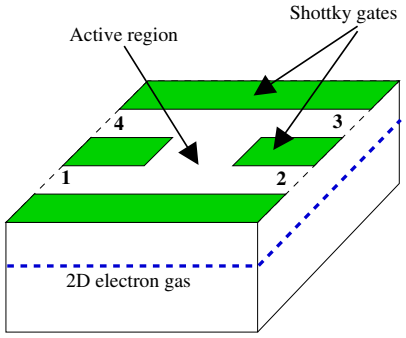


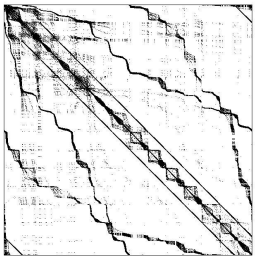




Dense banded linear systems







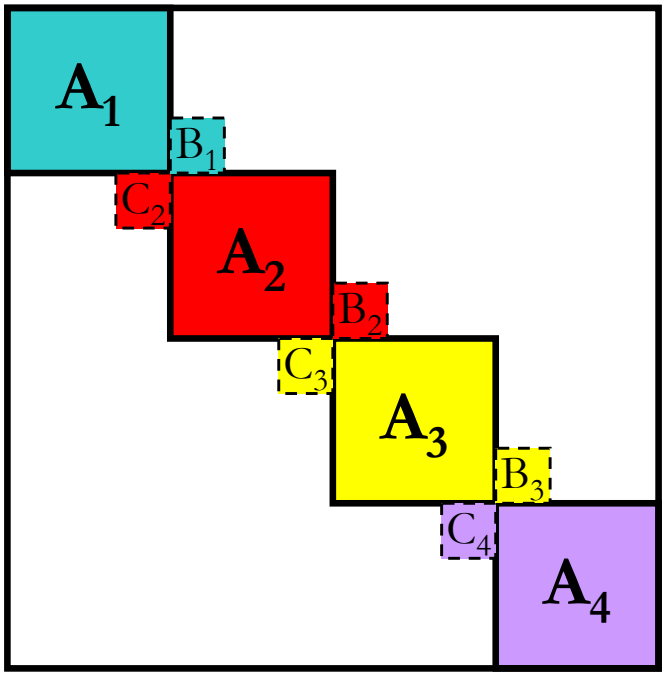
General sparse linear systems

➔ ***Computational Nanoelectronics is in need of efficient parallel banded solvers***

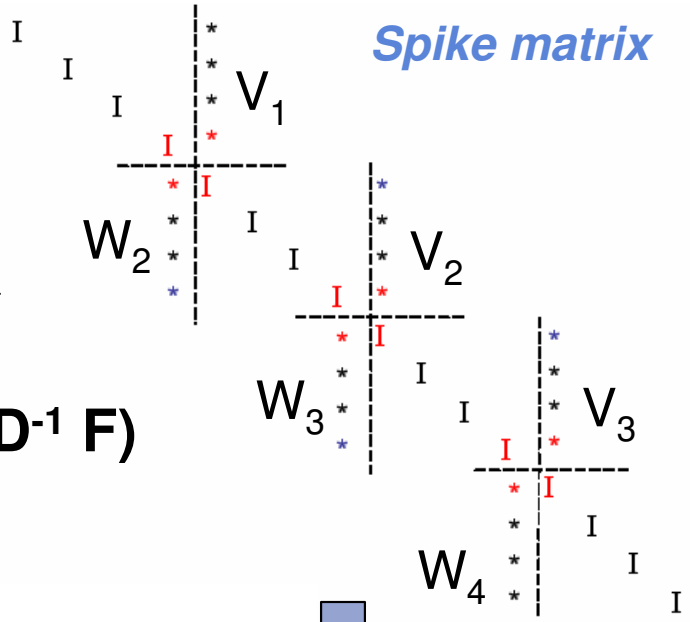
SPIKE

Overview of the general algorithm

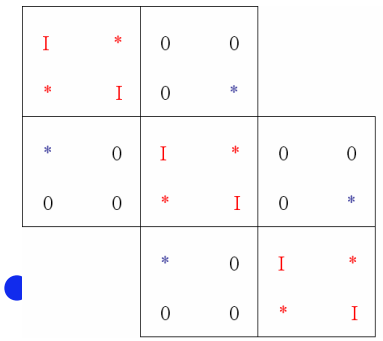
- **SPIKE is a parallel banded system solver proposed by A. Sameh (77-78), recently revisited.**



AX=F ➔ SX=(D^{-1} F)



Reduced system

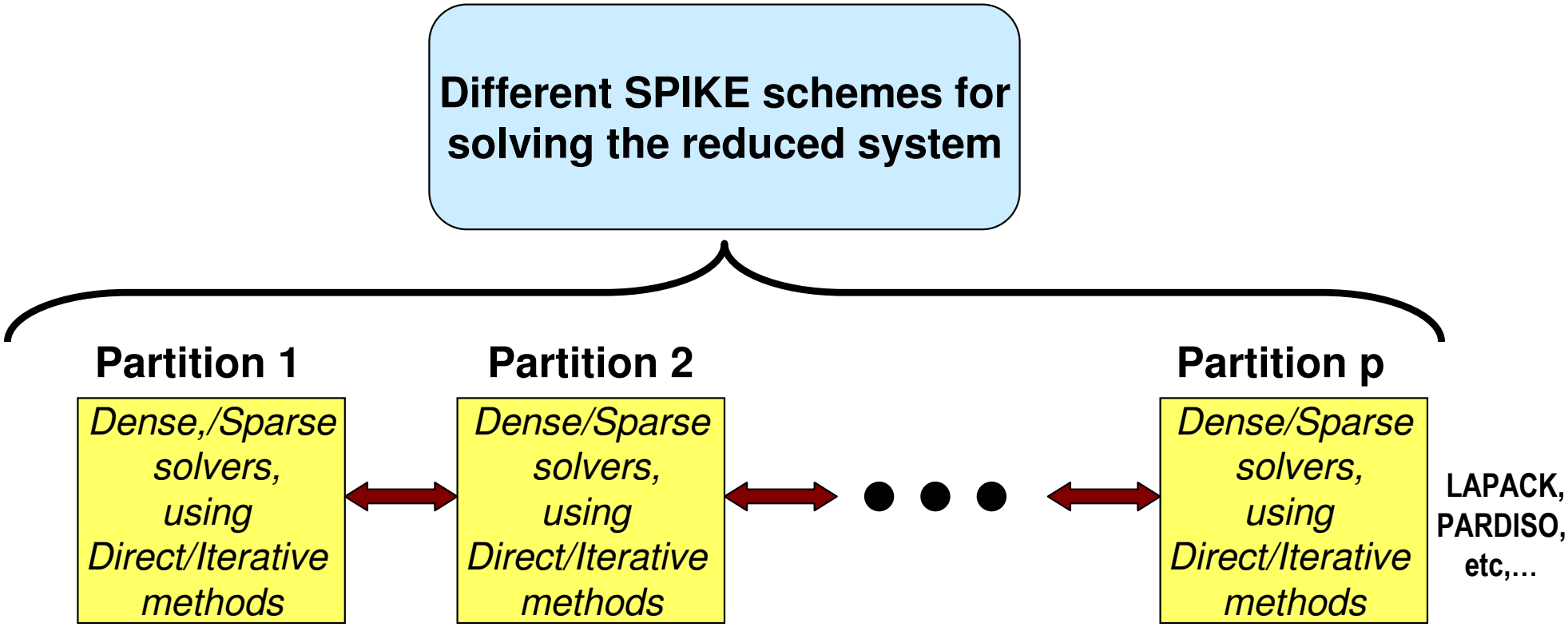


Retrieve solution

$$A_j\ X_j = F_j - \begin{bmatrix} 0 \\ I_m \end{bmatrix} B_j X_{j+1}^{(t)} - \begin{bmatrix} I_m \\ 0 \end{bmatrix} C_j X_{j-1}^{(b)}$$

SPIKE:

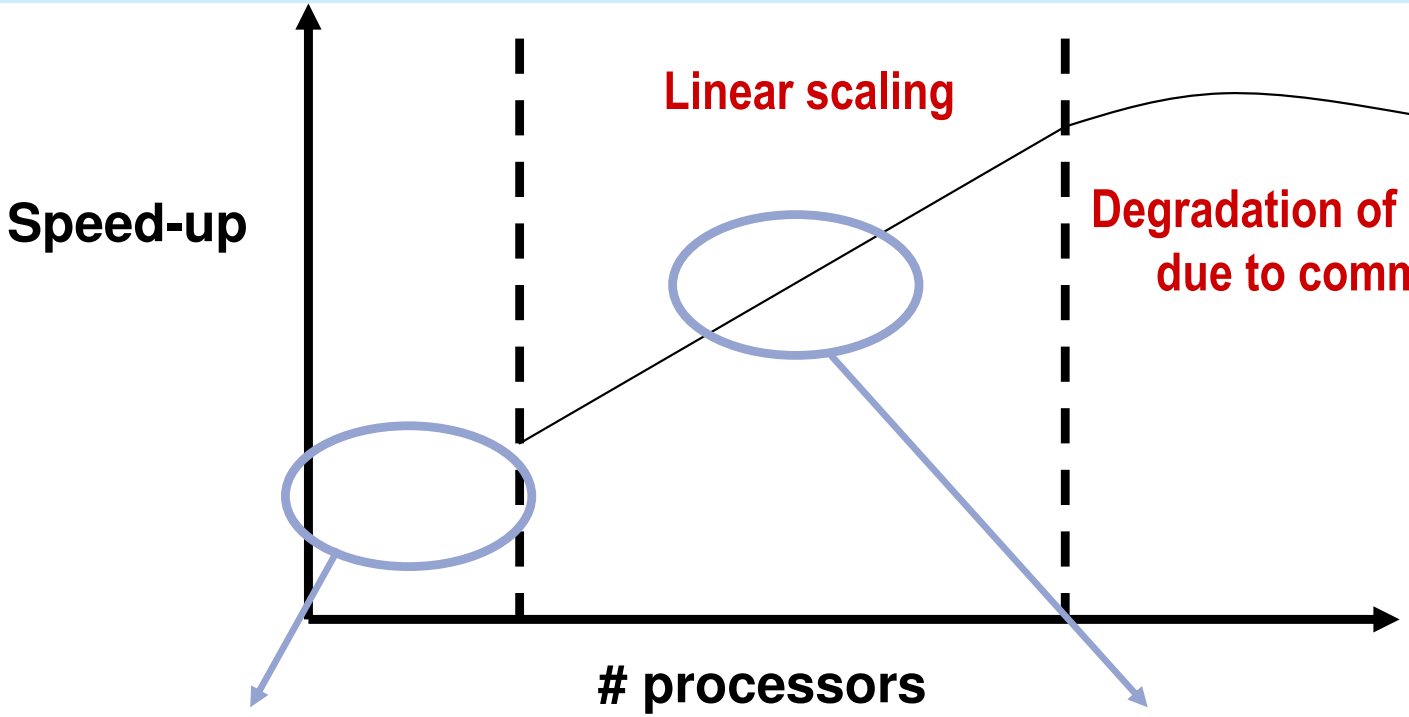
New Implementation - A robust parallel environment

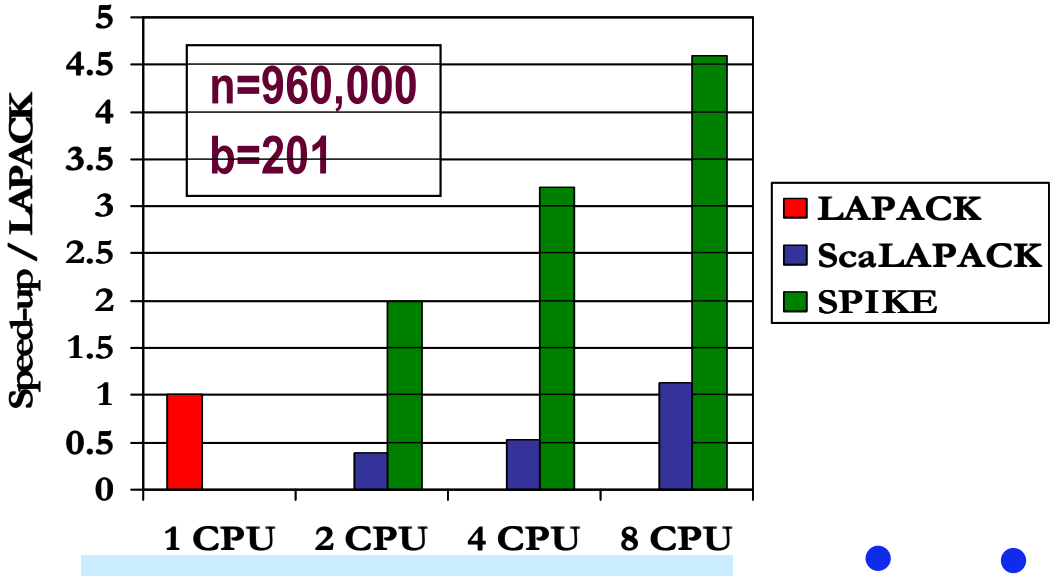


- ***SPIKE is essentially a family of domain decomposition techniques. The diagonal blocks A_j correspond to independent subdomains, and solving the resulting reduced system corresponds to solving the interface problem,***
- ***SPIKE is a hybrid and polyalgorithm (more than 40 different algorithms)***

SPIKE:

From small to large number of processors





SPIKE exhibits significant speed improvement as compared to banded ScaLAPACK

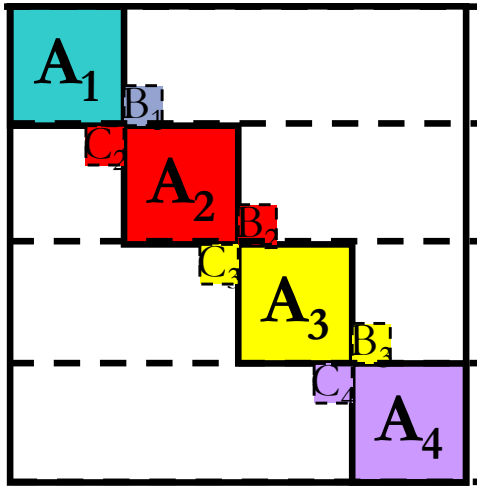
Polizzi, Sameh- Parallel Computing,32, p117 (2006)

SPIKE is ideally suited for both small and large number of processors (multicore technology)

SPIKE:

From small to large number of processors

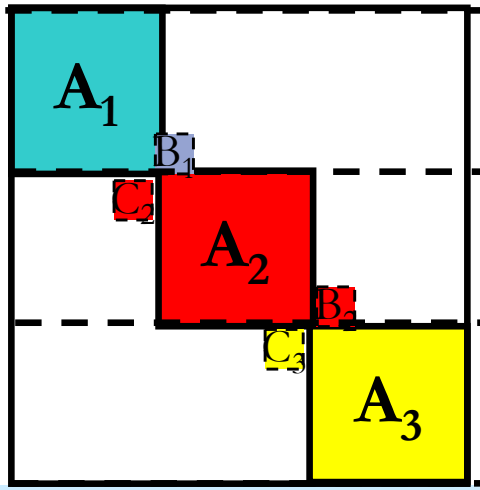
Traditional partitioning



Processors	Factorizations (w/o pivoting)
1	LU
2	LU and UL
3	LU and UL
4	UL

New partitioning

P processors → (P-2)/2 Partitions

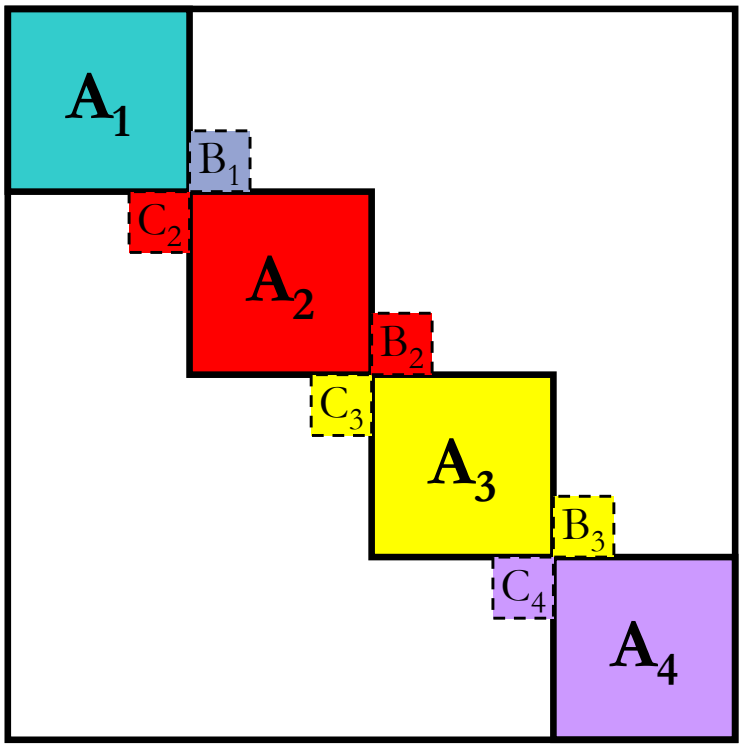


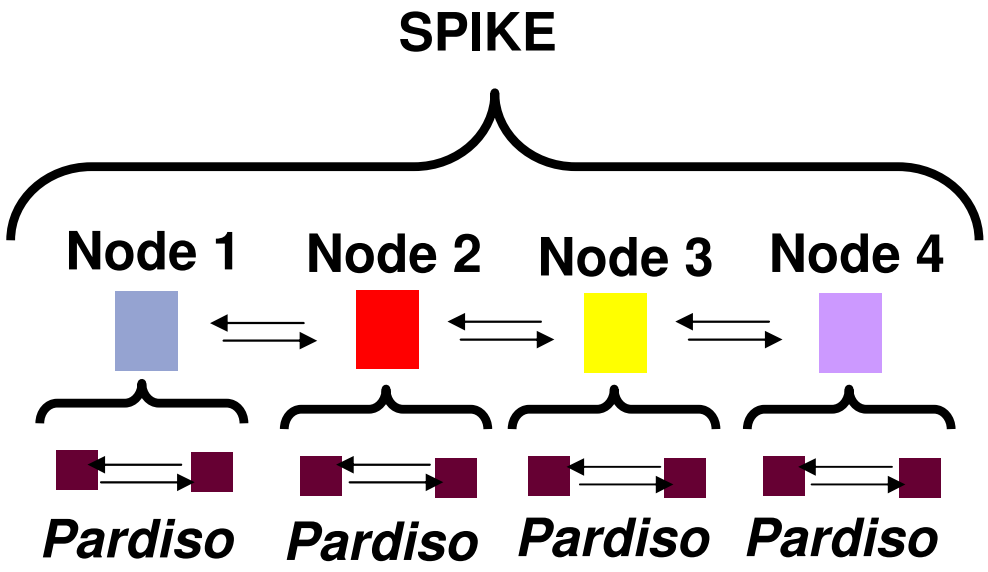
Processors	Factorizations (w/o pivoting)
1	LU
2,3	LU (p=2); UL (p=3)
4	UL

SPIKE:

Banded systems that are sparse within the band

- **We make use of SPIKE "on-the-fly" scheme**
- ***A SPIKE hybrid scheme exhibits better performance than other parallel direct sparse solvers used alone such as Pardiso, SuperLU, MUMPS***
- **Multilevel of Parallelism can be used to reduce the complexity of the reduced system**





Polizzi, Sameh- Computers & Fluids (2006)

SPIKE:

Banded systems that are sparse within the band

N=432,000, b= 177, nnz= 7, 955, 116, sparsity of the band: 10.4%

Test (a)	Reord.	Fact.	Solve	Total	Residual
SPIKE 2-nodes	9.48	1.55	4.54	15.57	10^{-7}
SPIKE 4-nodes	5.68	0.77	2.3	8.76	10^{-7}
Trunc. SPIKE (8)	0	0.84	0.14	0.98	10^{-14}

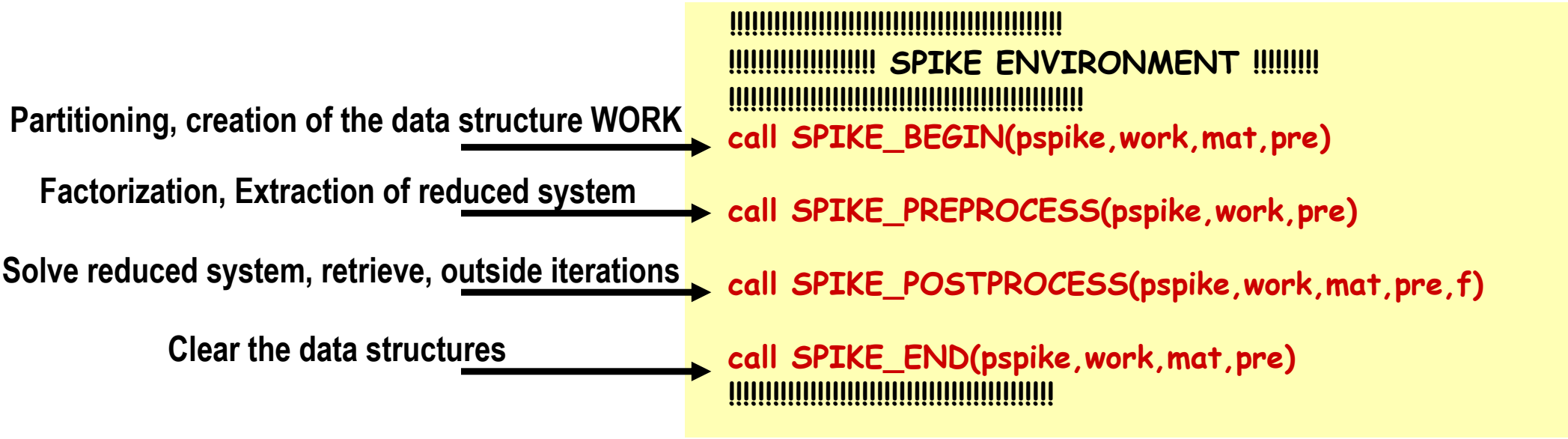
- ***MUMPS: time (2-nodes) = 21.35 s; time (4-nodes) = 39.6 s (memory swap)***
- ***For narrow banded systems, SPIKE will consider the matrix dense within the band. Reordering schemes for minimizing the bandwidth can be used if necessary.***

N=471,800, b= 1455, nnz= 9, 499, 744, sparsity of the band: 1.4%

Test (b)	Reord.	Fact.	Solve	Total	Residual
SPIKE 2-nodes	18.69	11.06	23.06	52.82	10^{-7}
SPIKE 4-nodes	8.41	6.54	10.85	25.81	10^{-7}

➔ ***Good scalability using "on-the-fly" SPIKE scheme***

SPIKE:
Current status



- *SPIKE v.2 contains ~10K lines of MPI-Fortran90,*
- *Documentation (user guide) is ready,*
- *More than 40 different schemes for SPIKE available (best choices depend on the properties of the linear systems and/or architectures of the machines).*
- *New options: customized partitioning, and customized preconditioner.*
- *A SPIKE_TUNE module is currently developed by Intel for automatic selection of the best SPIKE algorithm*
- *SPIKE will be soon available on-line*

Computational Nanoelectronics

Preconditioning strategy

$$(E[S] - [H(V)] - \sum_j [\Sigma_j(E)])\mathbf{X} = \mathbf{F}, \quad \forall E$$

- Traditional strategy:
  - Iterative solvers such as QMR, GMRES, BicgStab, with heuristic preconditioners such as ILUT, SSOR require **~80-100 iterations**.

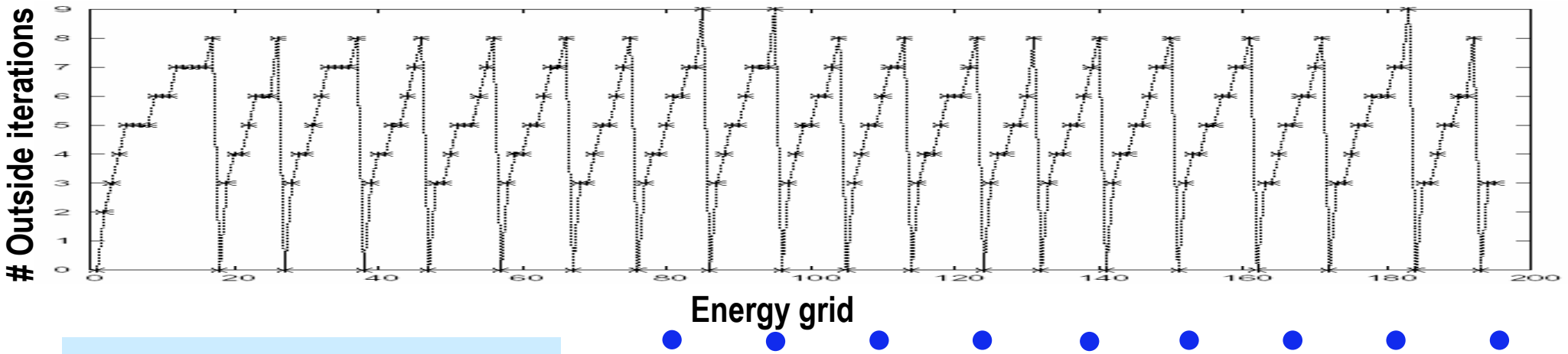
- We propose to use **$(E_1[S]-[H]-[\Sigma_{E1}])$ as preconditioner** for solving **$(E_2[S]-[H]-[\Sigma_{E2}])$**, if ***(E2 -E1) < δE***,
  - *the preconditioner is **solved in parallel using SPIKE** and updated if # of iterations > Nmax*
  - ***Fast algorithm allowing refinement of the uniform energy grid***

SPIKE as Preconditioner

SPIKE "Preprocessing" on
$[M]=(E_1[S]-[H]-[\Sigma_{E1}])$

ITERATIVE METHOD

- SPIKE solver M z = r
- Matrix-vector mult. y=A*x



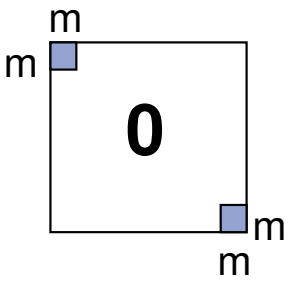
Computational Nanoelectronics:

Towards large-scale nanoelectronics simulations

Large sparse (banded) real symmetric (or Hermitian)- linear scaling

$$(E[S] - [H(V)] - \sum_j [\Sigma_j(E)])\mathbf{X} = \mathbf{F}, \quad \forall E$$

complex symmetric- quadratic scaling



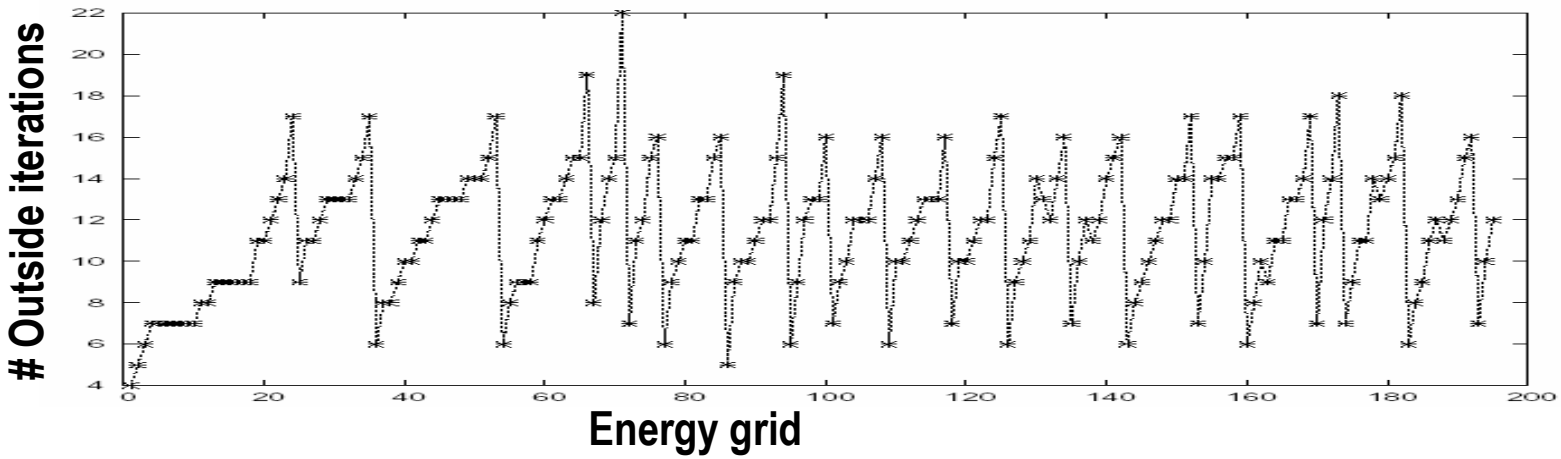
- **Large-scale simulations:** *the size m of the dense blocks in [Σ] increases significantly*

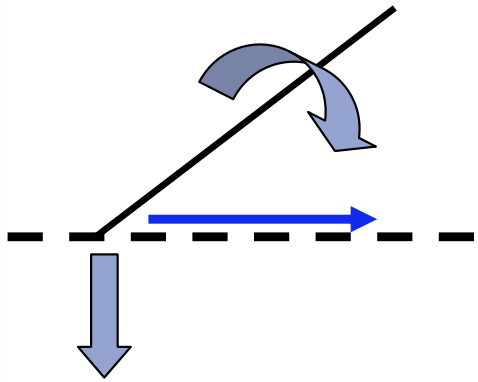
- **SPIKE as a solver and preconditioner:**

(E_1[S]-[H]) is a good preconditioner for **(E_1[S]-[H]-[Σ_{E1}])**

  - ***~6 outer-iterations of BiCGstab***
  - *[Σ_{E1}] is required only in mat-vecs, which can be done on the fly for large systems*

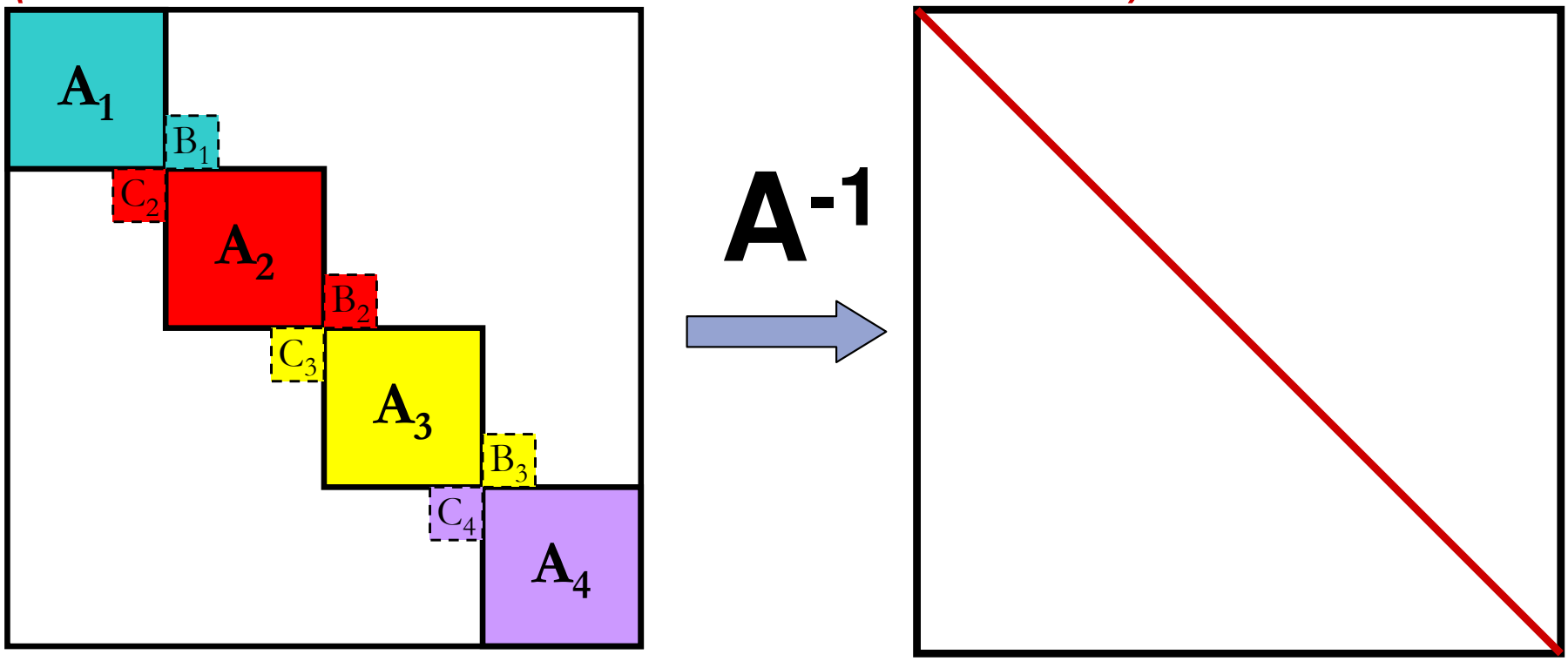
Can be improved !



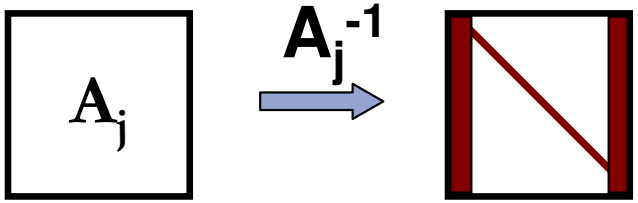


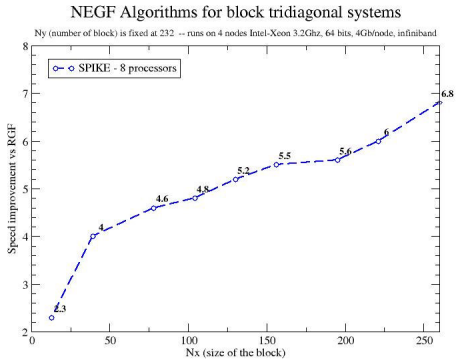
SPIKE:
Additional scheme

SPIKE can be used for computing the diagonal elements of the inverse
(alternative to standard electronic structure calculations).



- **Basic computational module for each A_j**





Computational Nanoelectronics
Directions

- **Modeling Challenges**
  - *Dissipative scattering processes*
  - *Atomistic modeling*
  - *Time dependent*
- **Computational Challenges**
  - *Improvement of the numerical techniques (multiscale, high order discretization schemes, preconditioning strategies…)*
  - ***In Need of fast parallel algorithms (SPIKE, etc…)***
- **Applications: Emerging Electronics Devices**
  - *Arbitrary devices, materials and geometries*
  - *CNT and carbon graphene using fully atomistic real-space mesh techniques*
  - *Full 3D simulations of Nanowire (Si and III-V) and dissipative scattering*

NESSIE as a platform for exploring innovative modeling strategies, the essential numerical methods and high performance algorithms

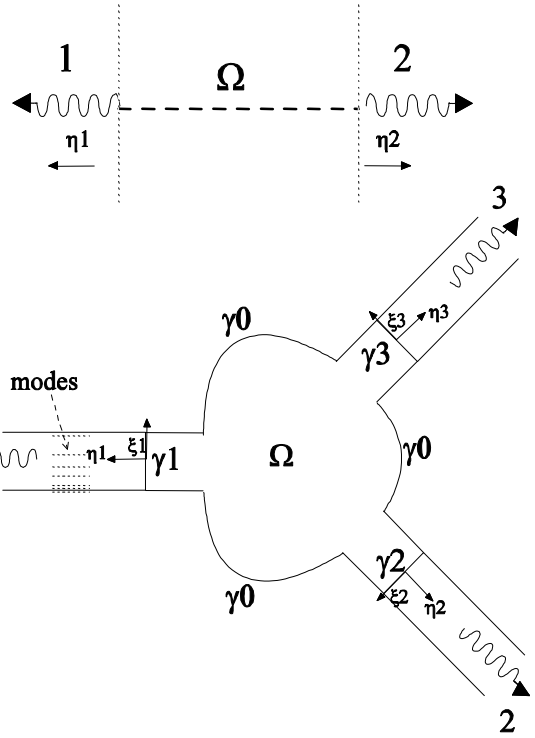
Acknowledgment: Intel support

Backup

Quantum Transport Modeling:
Open systems - derivation of the open B.C.

- **CASE 1 (1D device)** Frensley (1990)

$$\left.\frac{\partial}{\partial \eta_j}\right|_0 \Psi_j^{out}(\eta_j) = ik_j \Psi_j^{out}(0)$$



- **CASE 2 (2D/3D device)** Lent, Kirkner (1990)

$$\left.\frac{\partial}{\partial \eta_j}\right|_0 \Psi_j^{out}(\eta_j, \xi_j) = i \sum_m k_j^m \chi_j^m(\xi_j) \, \langle \chi_j^m(\xi_j) | \Psi_j^{out}(0, \xi_j) \rangle_{L^2(\gamma_j)}$$

- **GENERAL CASE** (G_j: Green's function of the reservoirs j)
  - Inglesfield (1981)–with Neumann B.C. for G_j on γ_j

$$\left.\frac{\partial}{\partial \eta_j}\right|_0 \Psi_j^{out}(\eta_j, \xi_j) = -2 \int_{\gamma_j} G_j^{-1}(\xi_j; \xi_j') \Psi_j^{out}(0, \xi_j') \, d\xi_j'$$

  - Fisher (1990)– with Dirichlet B.C. for G_j on γ_j

$$\left.\frac{\partial}{\partial \eta_j}\right|_0 \Psi_j^{out}(\eta_j, \xi_j) = \frac{1}{2} \int_{\gamma_j} \left(\frac{\partial^2}{\partial \eta_j \partial \eta_j'} G_j(\eta_j, \xi_j; \eta_j' \xi_j') \right) \Psi_j^{out}(0, \xi_j') \, d\xi_j'$$

Quantum Transport Modeling:

From Jellium to Atomistic models, why using real-space mesh techniques?

In contrast to traditional methods such as: empirical tight-binding, LCAO, plane wave scheme

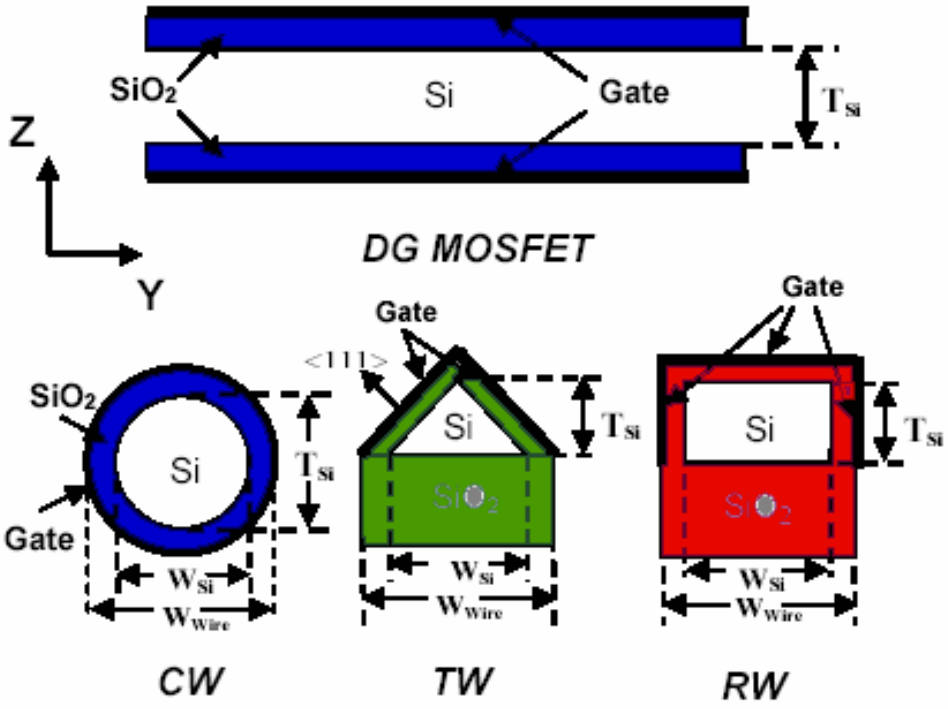
Computational Materials Science

- *They do not suffer from numerical truncation errors of finite expansions when scaled to larger systems.*
- *Consistency while solving the Poisson equation for electrostatics*
- *Produce very sparse matrices (banded) and are cast as linear scaling methods.*

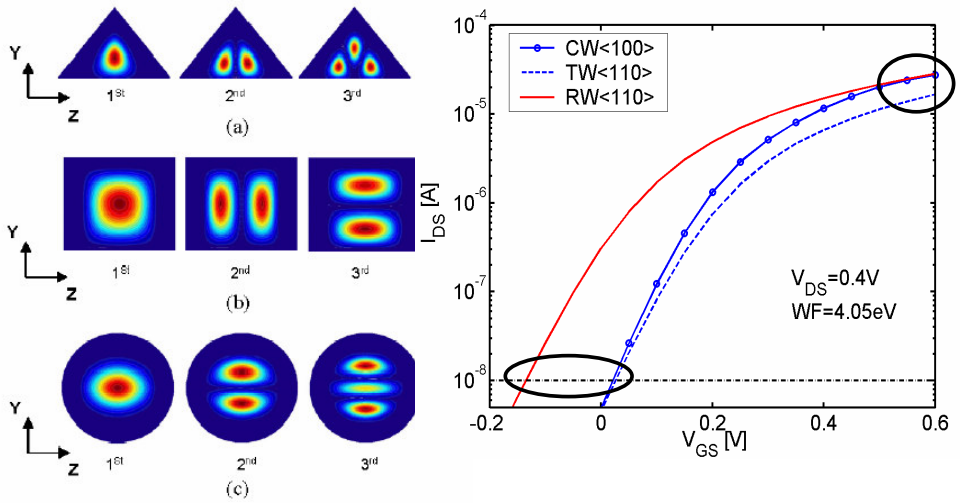
Computational Electronics

- *Ease to derive boundary conditions using a mathematical framework*
- *Ability to consider discontinuities in materials, and the coupling between different physical domains: jellium/atomistic*
- *FEM is a local basis-oriented approach, with ability to accommodate irregular device geometries, and to consider non-uniform meshes.*

Simulation of Emerging Electronic Devices:
Silicon Nanowire Transistors

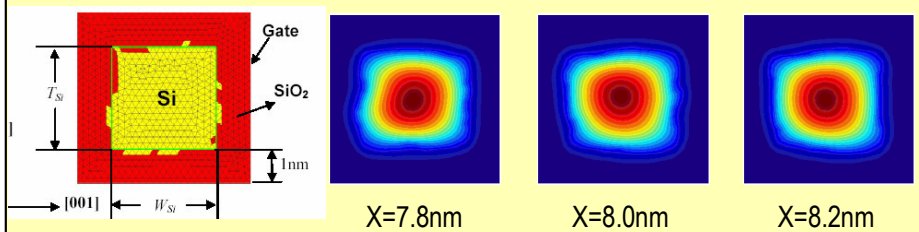


J. Appl. Phys. 96, 2192 (2004)

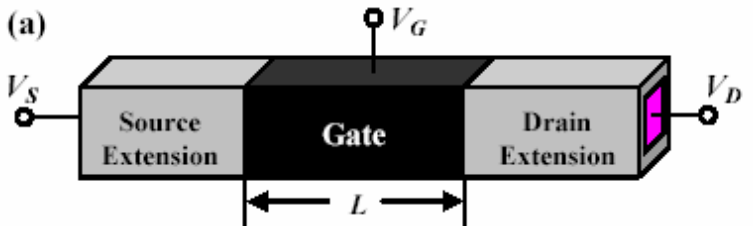


- **3D Mode space approach**
- **Asymptotic model (fast uncoupled mode approach or quasi-full dimensional model)**
- **Full 3D simulation to account for gate leakage in progress**

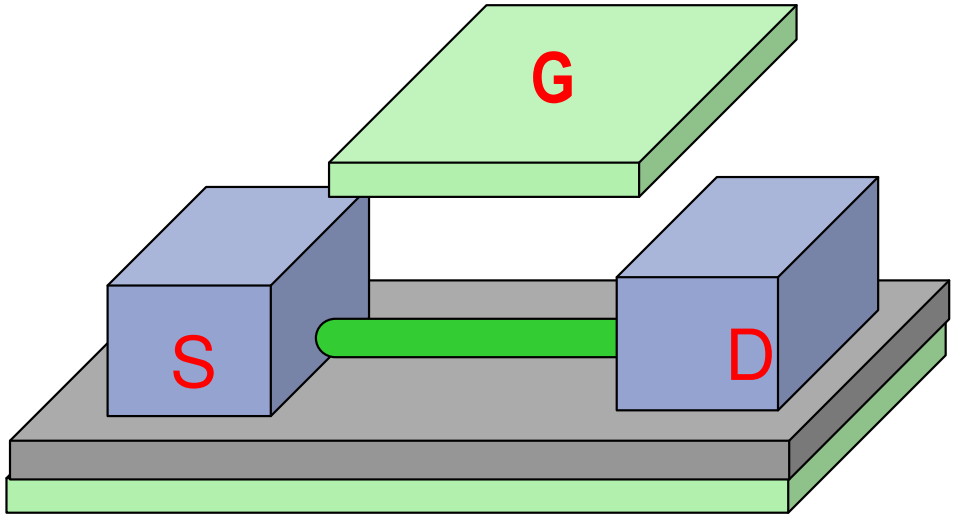
In addition: geometrical model for the treatment of surface roughness scattering



Appl. Phys. Let., 87, 043101 (2005)



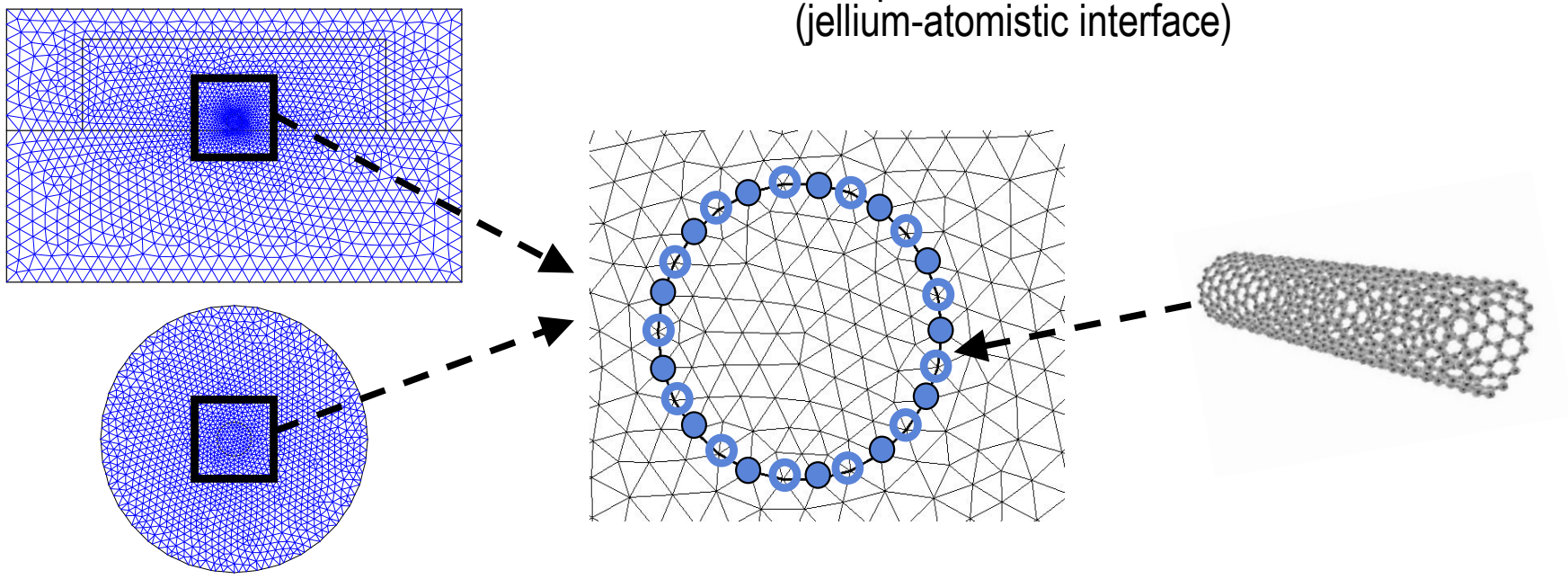
Simulation of Emerging Electronic Devices:
Carbon Nanotube Transistors



- **Huge influence of 3D electrostatics on 1D CNT**
- **Multiscale-Electrostatics approach for CNTFET:**
  - At large scale (transistors), the Carbon atom sites appear as point charges.
- **Atomistic transport model:**
  - **Currently uses tight-binding p_z orbitals**, Mode space or real-space approach; Empirical approach for handling metal-nanotube contacts.
  - **PDE-based approach under investigation**: DFT Kohn/Sham with pseudopotential ➔ hybrid transport model for metal-nanotube contact (jellium-atomistic interface)

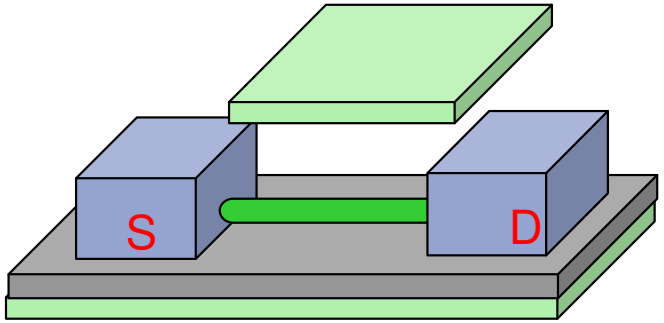
Planar device cross section

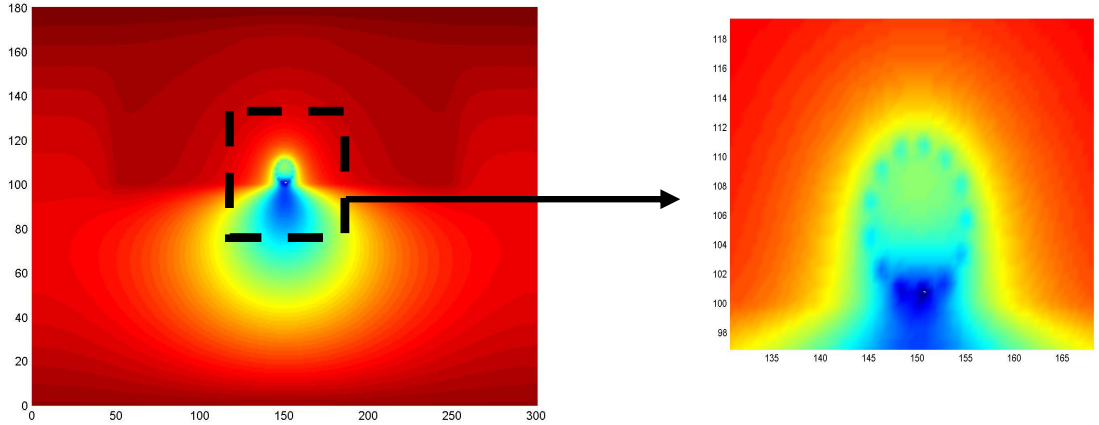
Coaxial device cross section



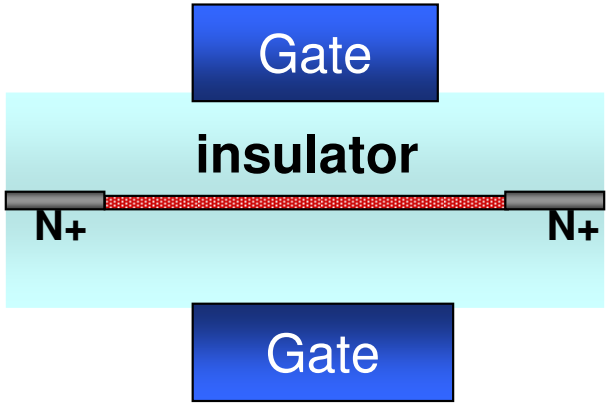
Simulation of Emerging Electronic Devices:
Carbon Nanotube Transistors

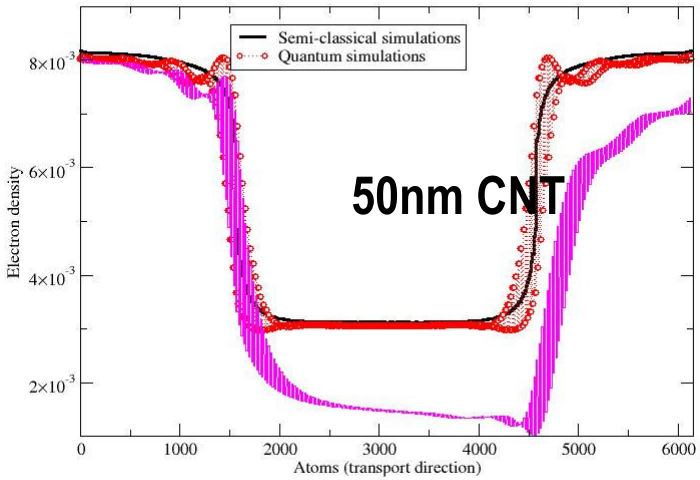
Planar device





Coaxial device





Appl, Phys. Let –to appear- (2006)