Designing and understanding nanoelectronic devices through petascale simulations

Mathieu Luisier
Integrated Systems Laboratory, ETH Zürich
Overview

- Introduction about Nanoelectronics
  From Moore’s Law to OMEN
- Numerical Simulation of Nanoscale Devices
  Electronic and thermal applications
- Code Implementation
  Physical Models
  Parallelization Scheme
  Numerical Algorithms
- Performance Benchmarks
- Outlook and Conclusion
Motivation: Evolution of Electronic Devices

Number of transistors per chip doubles every 2 years

Intel 45nm

Source
SiGe
High-k
Silicon

Gate
Metal

Drain
SiGe
Motivation: Future of Moore’s Scaling Law

65nm (2005)  
45nm (2007)  
32nm (2009)  
22nm (2011)  
5nm (2020)

1. 3-D Si FinFETs for ever?  
2. What will be the dominant limiting factors when $L_g < 10\text{nm}$?

Gate Length Reduction in planar Si MOSFETs:
=> increase of short-channel effects (SCE)  
=> poor electrostatic control (single-gate)  
=> SOLUTION: 3-D FinFET since 2011

Source: Intel Corporation
Next Generation Devices

Production: around 2020

**Nanowire**
P. Hashemi et al., *EDL* **30**, 401 (2009)

**Graphene**

**III-V UTB**

**BTB Tunneling**

**CNT**
Supratik Guha, IBM Research

**NEEDED:** Fast, cheap, and reliable platform to support the development and accelerate the innovation of novel nanoelectronic devices

Physics-based Numerical Device Simulator **OMEN**
What is OMEN? Computer Aided Design Tool

First Peta-scale Engineering Application

Device Engineering

- **Industrial-Strength** Nano-electronic Device Simulator
- **Multi-Geometry** Capabilities
- Explore, Understand, Explain, Optimize Novel Designs

Physical Models

- **3D Quantum** Transport Solver
- Accurate Representation of the Semiconductor Properties
- Atomistic Description of Devices
- **Multi-Physics** Modeling

OMEN

- Accelerate Simulation Time
- Investigate New Phenomena at the Nanometer Scale
- Move Hero Experiments to a Day-to-Day Basis

Efficient Parallel Computing
\[ H / \psi_E > = E / \psi_E > + \]

Quantum Transport Solver: Basic Principles

Quantum Transport Model

Supercomputer

Nanoelectronic Device Simulator
Overview

- Introduction about Nanoelectronics
  From Moore’s Law to OMEN
- **Numerical Simulation of Nanoscale Devices**
  Electronic and thermal applications
- Code Implementation
- Physical Models
- Parallelization Scheme
- Numerical Algorithms
- Performance Benchmarks
- Outlook and Conclusion
**Simulation Features:**
- Electron transport
- Ballistic limit
- 1,000 to 200,000 cores
(2) CNT FET Simulations

Simulation Features:
- Electron/hole transport
- Ballistic limit
- 100 to 4,000 cores

Expt: A. Franklin @ IBM YH

Publication: IEDM 2011, Nano Letters 2012

$L_g=9\text{nm}$

OMEN Device Structure

Ambipolar Current Flow

$E_f l$  
$E_f r$  
$V_B$

Characteristics

$I_d-V_{gs}$

Experiment  
Simulation

$V_{ds} = -0.01 \text{ V}$

Sonntag, 19. Februar 2012
(3) BTBT Diode Simulations

Simulation Features:
- Electron/hole transport
- Ballistic limit
- 1,000 to 20,000 cores

Discrepancy due to measurement setup

Publication: TECHCON 2010, APL 2012
Simulation Features:
- Phonon transport
- Phonon-phonon scattering
- 1,000 to 40,000 cores
Overview

- Introduction about Nanoelectronics
  From Moore’s Law to OMEN
- Numerical Simulation of Nanoscale Devices
  Electronic and thermal applications

- Code Implementation
  Physical Models
  Parallelization Scheme
  Numerical Algorithms
- Performance Benchmarks
- Outlook and Conclusion
Multi-Dimensional Schrödinger Equation with OBCs

\[ H / \psi_E > = E / \psi_E > \]

Tight-Binding Ansatz for the Wave Function

\[ < r | \psi_E > = \sum_{\sigma,ijk,k_t} C_{ij}^{\sigma}(E,k_t) \Phi_{\sigma}(r - R_{ijk}) e^{ik_t \cdot r_t} \]

\[ (E-H-\Sigma^{RB}+\Sigma^{RS}) \cdot G^R = I \]

\[ G^< = G^R \cdot (\Sigma^{<B}+\Sigma^{<S}) \cdot G^{R\dagger} \]

Ballistic (Wave Function)

\[ Ax=b \]

Scattering (NEGF)

AB=C
Physical Models: Quantum Transport (ii)

Carriers Localized around Atom Positions

\[ \rho(r) = F \rho \sum_{i,k} \int dE |C_i(E,k)|^2 \delta(r - r_i) \]

Current along Bonds Connecting two Atoms

\[ J(r) = F \sum_{ij,k} \int dE \text{Im}\{C_i(E,k) \cdot H_{ij} \cdot C_j(E,k)\} (r_j - r_i) \delta(r - r_i) \]

Solve Poisson Equation on FEM Grid

\[ \Delta V(r) = -\rho(r)/\varepsilon(r) \]

Repeat till \( \rho(r) \) and \( V(r) \) Convergence
Objective:
• Nanoelectronic Device Simulations with Quantum Transport and Atomistic Basis

Approach:
• Multi-Level parallelism
  • Voltage
  • Momentum
  • Energy
  • Space
• Parameter sweep over voltages
• Dynamic load balancing in double integral
• Leverage of existing linear solvers (Pardiso, MUMPs, SuperLU, Umfpack, …)
• Novel:
  • Development of new solvers (Block Cyclic Reduction) with Computational Interleaving between BC and sparse LSE

Parallelization Scheme

Quad-Level Parallelisation Scheme
Tested on multiple platforms
Overview

• Introduction about Nanoelectronics
  From Moore’s Law to OMEN
• Numerical Simulation of Nanoscale Devices
  Electronic and thermal applications
• Code Implementation
  Physical Models
  Parallelization Scheme
  Numerical Algorithms
• Performance Benchmarks
• Outlook and Conclusion
Benchmarks: End-to-end Device Simulations

Double-Gate InAs BTBT FET

Single-Gate MQW III-V HEMT (MIT)

Same code executable for both applications: no specific tuning

Specifications:

- unsymmetric single-material structure
- electron and hole current flow
- $sp^3s^*_{tight}$ binding with SO coupling
- $N_A=54,272$ atoms in active region
- $\text{sizeof}(A)=542,720$ in $Ax=b$ (|| on 9 CPU)

Specifications:

- symmetric multi-quantum-well structure
- electron flow only, mainly in $s$-InAs
- $sp^3d^5s^*_{tight}$ binding without SO
- $N_A=55,226$ atoms in active region
- $\text{sizeof}(A)=552,260$ in $Ax=b$ (|| on 9 CPU)
Band-to-band Tunneling Transistor

Double Precision **Strong Scaling** up to 221,400 Cores

- 4 parallel levels
- maximum of 11,070 cores per bias
- ~20 years on a single core
- <1 hour on 221,400 cores
- almost ideal speed-up till 221,400 cores
- **1.28 PFlop/s**
- **55.4% of peak**

![Graph showing performance and efficiency](image)

- **96% || efficiency**
- **1.28 PFlop/s**
- **4 parallel levels**
- **maximum of 11,070 cores per bias**
- **~20 years on a single core**
- **<1 hour on 221,400 cores**
- **almost ideal speed-up till 221,400 cores**
- **1.28 PFlop/s**
- **55.4% of peak**
High Electron Mobility Transistor

Double and Mixed Precision Scheme

Strong Scaling from 2,700 up to 221,400 Cores

- 4 parallel levels
- maximum of 11,070 cores per bias
- 5 Poisson iterations
- mixed: last Poisson iteration in double precision
- 1.27 PFlop/s double
- 54% of peak
- 1.44 PFlop/s mixed
Overview

• Introduction about Nanoelectronics
  From Moore’s Law to OMEN
• Numerical Simulation of Nanoscale Devices
  Electronic and thermal applications
• Code Implementation
  Physical Models
  Parallelization Scheme
  Numerical Algorithms
• Performance Benchmarks

• Outlook and Conclusion
Conclusion

- **Nanoelectronic Device Simulations**
  Required: quantum transport simulator

- **OMEN Simulation Approach**
  Good agreement with experimental data
  Dedicated to large variety of nanoscale devices
  Multi-geometry and multi-physics approach
  Sustained performance up to 1.44 PFlop/s

- **Future Work and Challenges**
  Development of new physical models
  Code modifications to benefit from GPUs
Evolution of Nanoelectronic Device Simulation

Time to compute 1 Poisson Iteration for 1 Bias Point on 11,070 cores

- NEGF: most popular technique, but not most efficient
- WF: computationally more efficient
- BCR: 20% faster than MUMPS and allows comp. interleaving
- as compared to standard techniques, OMEN 10.7x faster (double precision)