Hybrid parallel deterministic solver for DG-MOSFETs

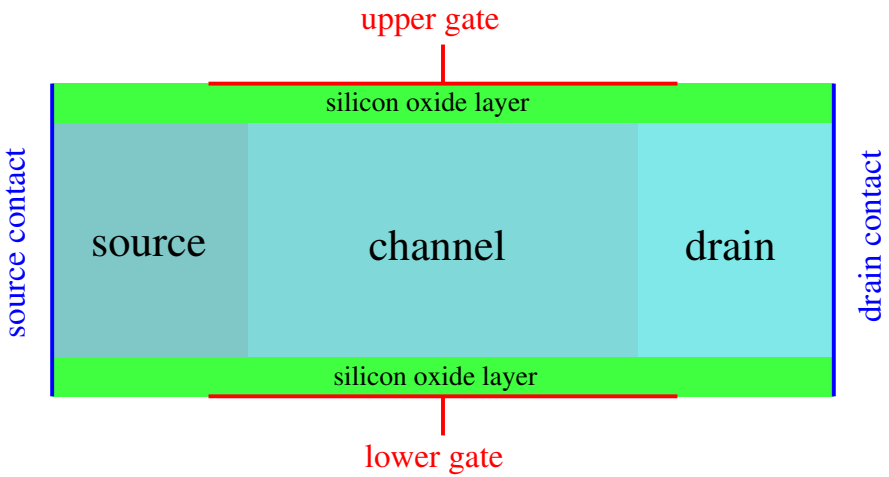
Francesco Vecil, José Miguel Mantas

ECMI 2018, Budapest, 2018/06/19

Outline

Outline

Geometry



About the scaling

In 1971, the Intel 4004 processor had 1000 transistors, whose channel length was 10000 nm. In 2003 the Intel Pentium IV had 50 million. Nowadays, for instance, Intel's i7-4650U has 1.3 billion transistors, whose channel is 22 nm long. The shortest transistor in the market is 14 nm long.

Why is it important?

Smaller MOSFETs allow for the construction of smaller devices with better performances; moreover, they allow silicon and energy saving, due to the lower voltages needed to switch on or off the transistor.

Outline

The modeling

2D model

We assume invariance of the distrubition function along the perpendicular unconfined dimension.

Dimensional coupling

Electrons are considered as **particles** along the unconfined dimension, as **waves** along the confined dimension.

Deterministic model

We use Boltzmann Transport Equations (BTEs) for the unconfined dimension and steady-state Schrödinger equations for the confined dimension. Hence, we provide a **high-dimensional**, **fully-deterministic** solver, whose goal is to provide reference results for Monte-Carlo or macroscopic solvers, which are faster but coarser.

Electron populations

Due to silicon's physical properties, the electron populations is split into three independent valleys, indexed on ν. Moreover, the confinement decomposes ν^{th} valley's electron population into independent energy levels indexed on p.

The modeling

2D model

We assume invariance of the distrubition function along the perpendicular unconfined dimension.

Dimensional coupling

Electrons are considered as **particles** along the unconfined dimension, as **waves** along the confined dimension.

Deterministic model

We use Boltzmann Transport Equations (BTEs) for the unconfined dimension and steady-state Schrödinger equations for the confined dimension. Hence, we provide a **high-dimensional**, **fully-deterministic** solver, whose goal is to provide reference results for Monte-Carlo or macroscopic solvers, which are faster but coarser.

Electron populations

Due to silicon's physical properties, the electron populations is split into three independent valleys, indexed on ν. Moreover, the confinement decomposes ν^{th} valley's electron population into independent energy levels indexed on p.

The modeling

2D model

We assume invariance of the distrubition function along the perpendicular unconfined dimension.

Dimensional coupling

Electrons are considered as **particles** along the unconfined dimension, as **waves** along the confined dimension.

Deterministic model

We use Boltzmann Transport Equations (BTEs) for the unconfined dimension and steady-state Schrödinger equations for the confined dimension. Hence, we provide a **high-dimensional**, **fully-deterministic** solver, whose goal is to provide reference results for Monte-Carlo or macroscopic solvers, which are faster but coarser.

Electron populations

Due to silicon's physical properties, the electron populations is split into three independent valleys, indexed on ν. Moreover, the confinement decomposes ν^{th} valley's electron population into independent energy levels indexed on p.

The modeling

2D model

We assume invariance of the distrubition function along the perpendicular unconfined dimension.

Dimensional coupling

Electrons are considered as **particles** along the unconfined dimension, as **waves** along the confined dimension.

Deterministic model

We use Boltzmann Transport Equations (BTEs) for the unconfined dimension and steady-state Schrödinger equations for the confined dimension. Hence, we provide a **high-dimensional**, **fully-deterministic** solver, whose goal is to provide reference results for Monte-Carlo or macroscopic solvers, which are faster but coarser.

Electron populations

Due to silicon's physical properties, the electron populations is split into three independent valleys, indexed on ν. Moreover, the confinement decomposes ν^{th} valley's electron population into independent energy levels indexed on p.

The modeling

Description of the confinement

A set of 1D Schrödinger eigenvalue problems describe the electrons along z.

$$-\frac{\hbar^2}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{1}{m_{z,\nu}}\frac{\mathrm{d}\psi_{\nu,p}[V]}{\mathrm{d}z}\right]-q\left(V+V_c\right)\psi_{\nu,p}[V]=\epsilon_{\nu,p}[V]\psi_{\nu,p}[V]$$

Description of the unconfined dimension

The BTEs, one for each pair (ν,p), along x reads

$$\frac{\partial f_{\nu,p}}{\partial t}+\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_\nu^{\mathrm{kin}}}{\partial k_x}\frac{\partial f_{\nu,p}}{\partial x}}^{\text{free motion}}-\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_{\nu,p}}{\partial x}\frac{\partial f_{\nu,p}}{\partial k_x}}^{\text{force field}}=\overbrace{\mathcal{Q}_{\nu,p}[f]}^{\text{scatterings}}\ .$$

Electrostatic field

Poisson's equation couples x and z: $-\mathrm{div}_{x,z}\left[\varepsilon_R\nabla_{x,z}V\right]=-\frac{q}{\varepsilon_0}\left(N[V]-N_D\right)$.

Here appears the volume density $N[V]=2\sum_{\nu,p}\int_{\mathbb{R}^2}f_{\nu,p}\,\mathrm{d}\boldsymbol{k}\,|\psi_{\nu,p}[V]|^2$.

The modeling

Description of the confinement

A set of 1D Schrödinger eigenvalue problems describe the electrons along z.

$$-\frac{\hbar^2}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{1}{m_{z,\nu}}\frac{\mathrm{d}\psi_{\nu,p}[V]}{\mathrm{d}z}\right]-q\left(V+V_c\right)\psi_{\nu,p}[V]=\epsilon_{\nu,p}[V]\psi_{\nu,p}[V]$$

Description of the unconfined dimension

The BTEs, one for each pair (ν,p), along x reads

$$\frac{\partial f_{\nu,p}}{\partial t}+\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_\nu^{\mathrm{kin}}}{\partial k_x}\frac{\partial f_{\nu,p}}{\partial x}}^{\text{free motion}}-\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_{\nu,p}}{\partial x}\frac{\partial f_{\nu,p}}{\partial k_x}}^{\text{force field}}=\overbrace{\mathcal{Q}_{\nu,p}[f]}^{\text{scatterings}}\quad.$$

Electrostatic field

Poisson's equation couples x and z: $-\mathrm{div}_{x,z}\left[\varepsilon_R\nabla_{x,z}V\right]=-\frac{q}{\varepsilon_0}\left(N[V]-N_D\right)$.

Here appears the volume density $N[V]=2\sum_{\nu,p}\int_{\mathbb{R}^2} f_{\nu,p}\,\mathrm{d}\boldsymbol{k}\,|\psi_{\nu,p}[V]|^2$.

The modeling

Description of the confinement

A set of 1D Schrödinger eigenvalue problems describe the electrons along z.

$$-\frac{\hbar^2}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{1}{m_{z,\nu}}\frac{\mathrm{d}\psi_{\nu,p}[V]}{\mathrm{d}z}\right]-q\left(V+V_c\right)\psi_{\nu,p}[V]=\epsilon_{\nu,p}[V]\psi_{\nu,p}[V]$$

Description of the unconfined dimension

The BTEs, one for each pair (ν,p), along x reads

$$\frac{\partial f_{\nu,p}}{\partial t}+\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_\nu^{\mathrm{kin}}}{\partial k_x}\frac{\partial f_{\nu,p}}{\partial x}}^{\text{free motion}}-\overbrace{\frac{1}{\hbar}\frac{\partial \epsilon_{\nu,p}}{\partial x}\frac{\partial f_{\nu,p}}{\partial k_x}}^{\text{force field}}=\overbrace{\mathcal{Q}_{\nu,p}[f]}^{\text{scatterings}}\quad .$$

Electrostatic field

Poisson's equation couples x and z: $-\mathrm{div}_{x,z}\left[\varepsilon_R\nabla_{x,z}V\right]=-\dfrac{q}{\varepsilon_0}\left(N[V]-N_D\right)$.

Here appears the volume density $N[V]=2\displaystyle\sum_{\nu,p}\int_{\mathbb{R}^2} f_{\nu,p}\,\mathrm{d}\boldsymbol{k}\,|\psi_{\nu,p}[V]|^2$.

The modeling

The scattering operator

The scattering operator reads,

$$\mathcal{Q}_{\nu,p}[f] = \sum_{s} \sum_{\nu',p'} \int_{\mathbb{R}^2} \left[S^{s}_{(\nu',p',\boldsymbol{k}')\to(\nu,p,\boldsymbol{k})} f_{\nu',p'}(\boldsymbol{k}') - S^{s}_{(\nu,p,\boldsymbol{k})\to(\nu',p',\boldsymbol{k}')} f_{\nu,p}(\boldsymbol{k}) \right] \mathrm{d}\boldsymbol{k}'.$$

Electron-phonon interactions

For the seven electron-phonon interactions, scattering rates read (up to constants)

$$S^{s,\mathrm{ph}} = \int_0^{L_z} |\psi_{\nu,p}|^2 |\psi_{\nu',p'}|^2 \,\mathrm{d}z \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu',p'}(\boldsymbol{k}') - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) \pm \text{some energy}\right).$$

Surface roughness

For the SR phenomenon, scattering rates have form (up to constants)

$$S^{s,\mathrm{SR}} = \left| \int_0^{L_z} |\psi_{\nu,p}(x,z)|^2 \,\Delta V(x,z)\,\mathrm{d}z \right|^2 \cdot \frac{1}{\left(1 + \dfrac{|\boldsymbol{k}-\boldsymbol{k}'|^2}{2}\right)^{3/2}} \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}')\right)$$

The modeling

The scattering operator

The scattering operator reads,

$$\mathcal{Q}_{\nu,p}[f] = \sum_{s} \sum_{\nu',p'} \int_{\mathbb{R}^2} \left[S^{s}_{(\nu',p',\boldsymbol{k}')\to(\nu,p,\boldsymbol{k})} f_{\nu',p'}(\boldsymbol{k}') - S^{s}_{(\nu,p,\boldsymbol{k})\to(\nu',p',\boldsymbol{k}')} f_{\nu,p}(\boldsymbol{k}) \right] \mathrm{d}\boldsymbol{k}'.$$

Electron-phonon interactions

For the seven electron-phonon interactions, scattering rates read (up to constants)

$$S^{s,\mathrm{ph}} = \int_0^{L_z} |\psi_{\nu,p}|^2 |\psi_{\nu',p'}|^2 \,\mathrm{d}z \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu',p'}(\boldsymbol{k}') - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) \pm \text{some energy}\right).$$

Surface roughness

For the SR phenomenon, scattering rates have form (up to constants)

$$S^{s,\mathrm{SR}} = \left| \int_0^{L_z} |\psi_{\nu,p}(x,z)|^2 \,\Delta V(x,z)\,\mathrm{d}z \right|^2 \cdot \frac{1}{\left(1 + \frac{|\boldsymbol{k}-\boldsymbol{k}'|^2}{2}\right)^{3/2}} \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}')\right)$$

The modeling

The scattering operator

The scattering operator reads,

$$\mathcal{Q}_{\nu,p}[f] = \sum_{s} \sum_{\nu',p'} \int_{\mathbb{R}^2} \left[S^{s}_{(\nu',p',\boldsymbol{k}')\to(\nu,p,\boldsymbol{k})} f_{\nu',p'}(\boldsymbol{k}') - S^{s}_{(\nu,p,\boldsymbol{k})\to(\nu',p',\boldsymbol{k}')} f_{\nu,p}(\boldsymbol{k}) \right] \mathrm{d}\boldsymbol{k}'.$$

Electron-phonon interactions

For the seven electron-phonon interactions, scattering rates read (up to constants)

$$S^{s,\mathrm{ph}} = \int_0^{L_z} |\psi_{\nu,p}|^2 |\psi_{\nu',p'}|^2 \,\mathrm{d}z \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu',p'}(\boldsymbol{k}') - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) \pm \text{some energy}\right).$$

Surface roughness

For the SR phenomenon, scattering rates have form (up to constants)

$$S^{s,\mathrm{SR}} = \left| \int_0^{L_z} |\psi_{\nu,p}(x,z)|^2 \, \Delta V(x,z) \,\mathrm{d}z \right|^2 \cdot \frac{1}{\left(1 + \frac{|\boldsymbol{k}-\boldsymbol{k}'|^2}{2}\right)^{3/2}} \cdot \delta\left(\epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}) - \epsilon^{\mathrm{tot}}_{\nu,p}(\boldsymbol{k}')\right)$$

Outline

Meshes

Magnitudes are adimensionalized. Wave-vector space uses ellipsoidal variables $(\tilde{k}_x, \tilde{k}_y) = \frac{\sqrt{m_e \kappa_B T_L}}{\hbar} \sqrt{2w(1+\alpha_\nu w)} \left(\sqrt{m_{x,\nu}} \cos(\phi), \sqrt{m_{y,\nu}} \sin(\phi)\right)$. Globally, the problem spans on a 7-dimensional space:

(i). The **valley** $\nu \in \{0, 1, 2\}$.

(ii). The **energy level** $p \in \{0, \ldots, N_{\text{sbn}} - 1\}$.

(iii). The **longitudinal dimension** (unconfined) $x_i = i \times \underbrace{\frac{1}{N_x - 1}}_{\Delta x}$.

(iv). The **transversal dimension** (confined) $z_j = j \times \underbrace{\frac{1}{N_z - 1}}_{\Delta z}$.

(v). The **energy** $w_\ell = (\ell + 0.5) \times \underbrace{\frac{w_{\max}}{N_w - 1}}_{\Delta w}$

(vi). The **angle** $\phi_m = m \times \underbrace{\frac{2\pi}{N_\phi}}_{\Delta \phi}$.

(vii). The **time** step, adapted following a Courant-Friedrichs-Lewy condition.

Outline

The Newton scheme

Applying a Newton-Raphson scheme to the adimensionalized, discretized Schrödinger-Poisson block leads to iteratively solving a linear system followed by an eigenvalue/eigenvector problem.

The linear system

At iteration k, we refine the potential V by $L^{(k)}\, V^{(k+1)} = R^{(k)}$, where

$$L^{(k)}\, V^{(k+1)} = -\operatorname{div}\left[\varepsilon_{\mathrm{R}}\, \nabla V^{(k+1)}\right] + \int \mathcal{A}^{(k)}(x,z,\zeta)\, V^{(k+1)}(x,\zeta)\, \mathrm{d}\zeta$$

$$R^{(k)} = -N^{(k)}(x,z) + \int \mathcal{A}^{(k)}(x,z,\zeta)\, V^{(k)}(x,\zeta)\, \mathrm{d}\zeta.$$

The Schrödinger equation

We compute eigenvalues and eigenvectors of a tridiagonal symmetric matrix:

$$d_j = \left(\frac{1/4}{m_{\mathrm{z},\nu,i,j-1}} + \frac{1/2}{m_{\mathrm{z},\nu,i,j}} + \frac{1/4}{m_{\mathrm{z},\nu,i,j+1}} \right) \frac{1}{\Delta z^2} - V_{i,j}, \qquad e_j = -\left(\frac{1/4}{m_{\mathrm{z},\nu,i,j}} + \frac{1/4}{m_{\mathrm{z},\nu,i,j+1}} \right) \frac{1}{\Delta z^2}.$$

The Newton scheme

Applying a Newton-Raphson scheme to the adimensionalized, discretized Schrödinger-Poisson block leads to iteratively solving a linear system followed by an eigenvalue/eigenvector problem.

The linear system

At iteration k, we refine the potential V by $L^{(k)}\, V^{(k+1)} = R^{(k)}$, where

$$L^{(k)}\, V^{(k+1)} = -\operatorname{div}\left[\varepsilon_{\mathrm{R}}\, \nabla V^{(k+1)}\right] + \int \mathcal{A}^{(k)}(x,z,\zeta)\, V^{(k+1)}(x,\zeta)\, \mathrm{d}\zeta$$

$$R^{(k)} = -N^{(k)}(x,z) + \int \mathcal{A}^{(k)}(x,z,\zeta)\, V^{(k)}(x,\zeta)\, \mathrm{d}\zeta.$$

The Schrödinger equation

We compute eigenvalues and eigenvectors of a tridiagonal symmetric matrix:

$$d_j = \left(\frac{1/4}{m_{\mathrm{z},\nu,i,j-1}} + \frac{1/2}{m_{\mathrm{z},\nu,i,j}} + \frac{1/4}{m_{\mathrm{z},\nu,i,j+1}} - V_{i,j} \right), \qquad e_j = \left(-\frac{1/4}{m_{\mathrm{z},\nu,i,j}} + \frac{1/4}{m_{\mathrm{z},\nu,i,j+1}} \right).$$

Outline

Schemes

Time discretization

We use the Total-Variation-Diminishing Runge-Kutta scheme of order 3. It is robust, but its explicitness constraints the time stepping.

Partial derivatives

We use fifth-order WENO (non-oscillatory) schemes to approximate them.

Scattering operator

Explicit, but numerically costly, formulae are obtained.
For example, for the electron-phonon elastic phenomena, we have

$$Q^{\mu,\text{gain}}_{\nu,p,i,\ell} = \mathcal{C}^{\mu} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \cdot \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \times s_{\nu}(w_{\ell}) \cdot \text{LI}\left[\tilde{\Phi}^{s}_{\nu,p',i,\cdot}\right]\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right)$$

and

$$Q^{\mu,\text{loss}}_{\nu,p,i,\ell,m} = -\mathcal{C}^{\mu}\, 2\pi \cdot \Phi^{s}_{\nu,p,i,\ell,m} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \times \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \cdot s_{\nu}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right) \cdot$$

Schemes

Time discretization

We use the Total-Variation-Diminishing Runge-Kutta scheme of order 3. It is robust, but its explicitness constraints the time stepping.

Partial derivatives

We use fifth-order WENO (non-oscillatory) schemes to approximate them.

Scattering operator

Explicit, but numerically costly, formulae are obtained.
For example, for the electron-phonon elastic phenomena, we have

$$\mathcal{Q}^{\mu,\text{gain}}_{\nu,p,i,\ell} = \mathcal{C}^{\mu} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \cdot \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \times s_{\nu}(w_{\ell}) \cdot \text{LI}\left[\tilde{\Phi}^{s}_{\nu,p',i,\cdot}\right]\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right)$$

and

$$\mathcal{Q}^{\mu,\text{loss}}_{\nu,p,i,\ell,m} = -\mathcal{C}^{\mu}\, 2\pi \cdot \Phi^{s}_{\nu,p,i,\ell,m} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \times \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \cdot s_{\nu}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right) \cdot$$

Schemes

Time discretization

We use the Total-Variation-Diminishing Runge-Kutta scheme of order 3. It is robust, but its explicitness constraints the time stepping.

Partial derivatives

We use fifth-order WENO (non-oscillatory) schemes to approximate them.

Scattering operator

Explicit, but numerically costly, formulae are obtained.
For example, for the electron-phonon elastic phenomena, we have

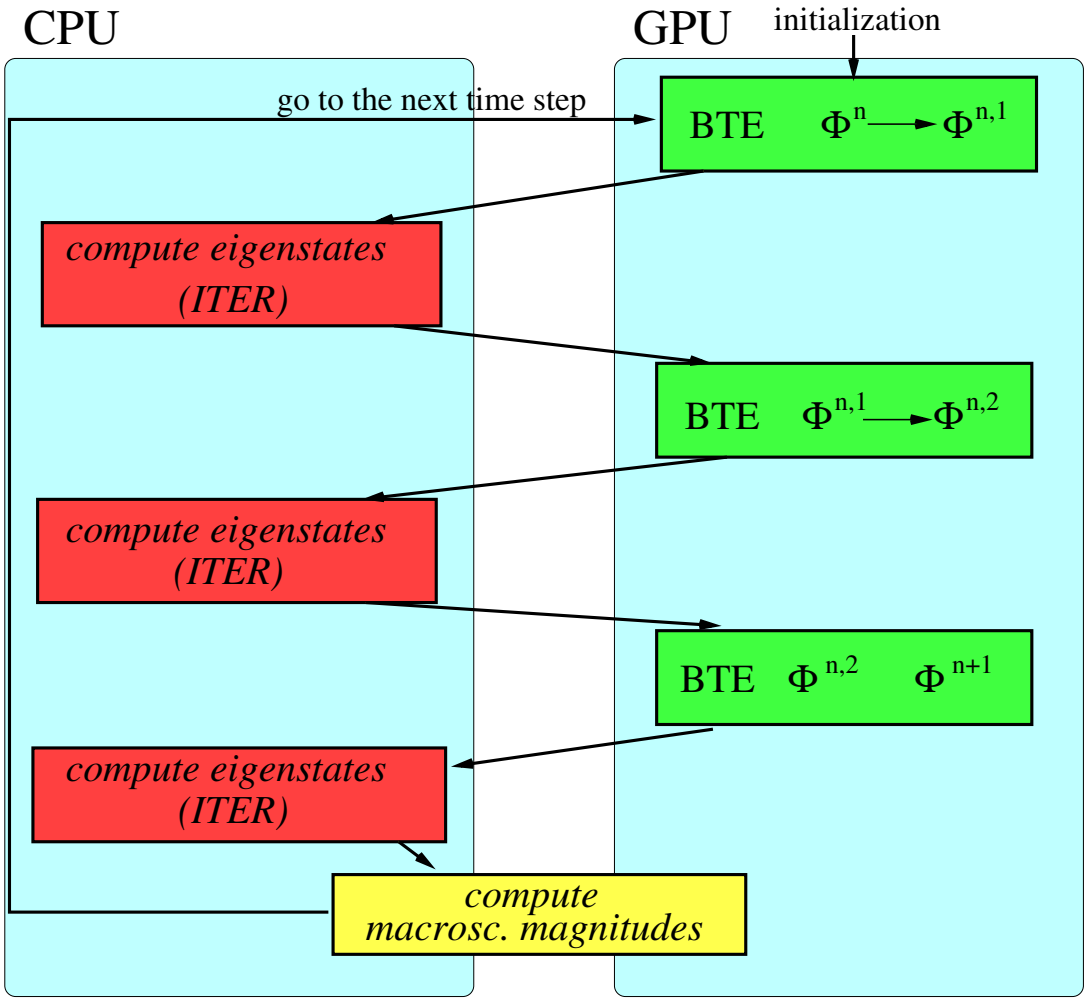
$$\mathcal{Q}^{\mu,\text{gain}}_{\nu,p,i,\ell} = \mathcal{C}^{\mu} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \cdot \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \times s_{\nu}(w_{\ell}) \cdot \text{LI}\left[\tilde{\Phi}^{s}_{\nu,p',i,\cdot}\right]\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right)$$

and

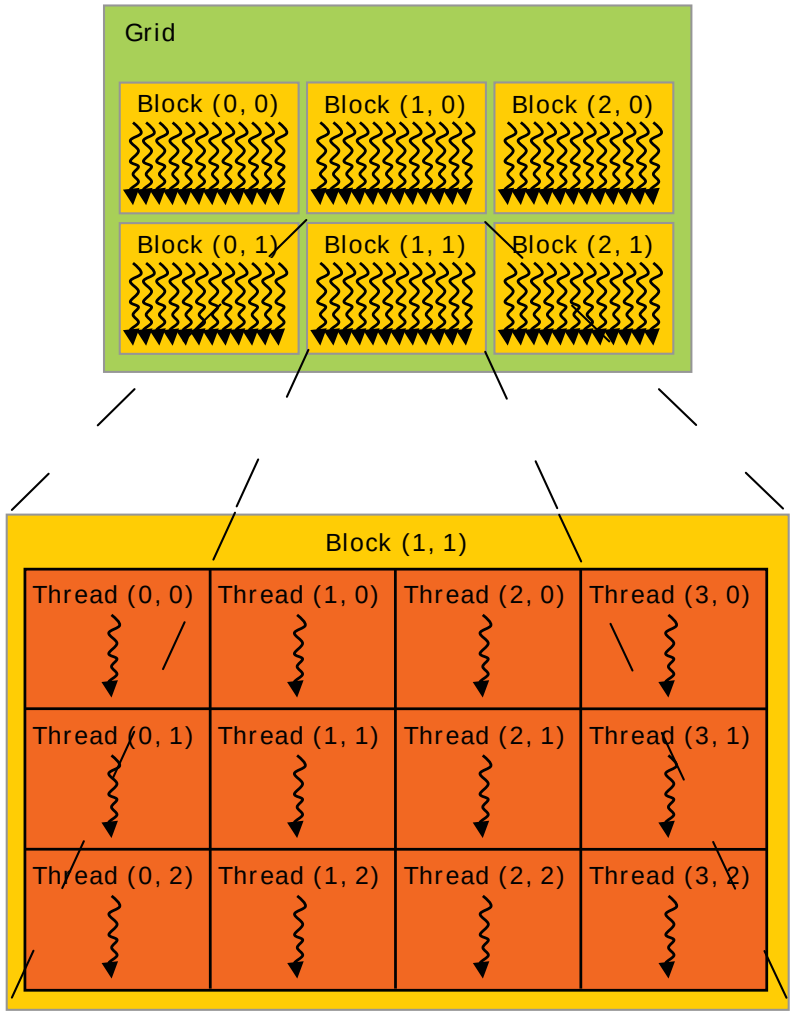
$$\mathcal{Q}^{\mu,\text{loss}}_{\nu,p,i,\ell,m} = -\mathcal{C}^{\mu}\, 2\pi \cdot \Phi^{s}_{\nu,p,i,\ell,m} \sum_{p'=0}^{N_{\text{sbn}}-1} \mathcal{W}_{\nu,p,\nu,p',i} \times \mathbf{I}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell} \geq 0\right) \cdot s_{\nu}\left(\Gamma^{0}_{\nu,p,\nu,p',i,\ell}\right).$$

Outline

Overall design of the solver



Cuda programming model on GPU



(from book CUDA C Programming Guide)

Some remarks on the Cuda implementation

Fine grain

Fine-grain paradigm exploited as much as possible: many threads, each of them with a light weight.

Shared memory

Use of block's shared memory to minimize reads from DRAM or to load data from DRAM in a coalescent manner.

Bank conflicts

Attention on avoiding bank conflicts when accessing shared memory.

Overlap

Overlapping computations between GPU and CPU whenever data are independent.

Some remarks on the Cuda implementation

Fine grain

Fine-grain paradigm exploited as much as possible: many threads, each of them with a light weight.

Shared memory

Use of block's shared memory to minimize reads from DRAM or to load data from DRAM in a coalescent manner.

Bank conflicts

Attention on avoiding bank conflicts when accessing shared memory.

Overlap

Overlapping computations between GPU and CPU whenever data are independent.

Some remarks on the Cuda implementation

Fine grain

Fine-grain paradigm exploited as much as possible: many threads, each of them with a light weight.

Shared memory

Use of block's shared memory to minimize reads from DRAM or to load data from DRAM in a coalescent manner.

Bank conflicts

Attention on avoiding bank conflicts when accessing shared memory.

Overlap

Overlapping computations between GPU and CPU whenever data are independent.

Some remarks on the Cuda implementation

Fine grain

Fine-grain paradigm exploited as much as possible: many threads, each of them with a light weight.

Shared memory

Use of block's shared memory to minimize reads from DRAM or to load data from DRAM in a coalescent manner.

Bank conflicts

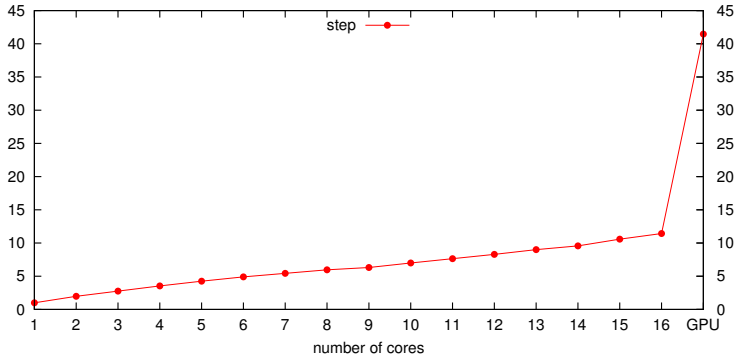
Attention on avoiding bank conflicts when accessing shared memory.

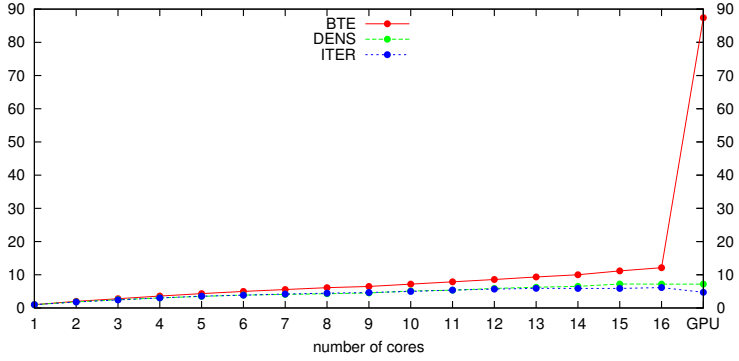
Overlap

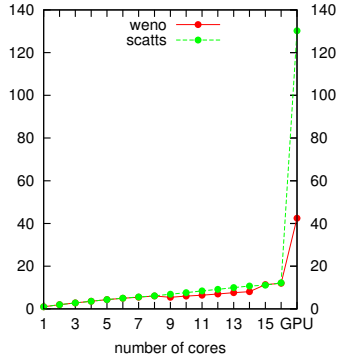
Overlapping computations between GPU and CPU whenever data are independent.

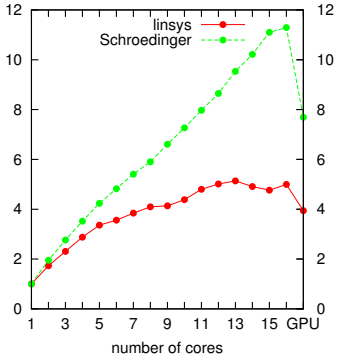
Outline

Performances of the hybrid code

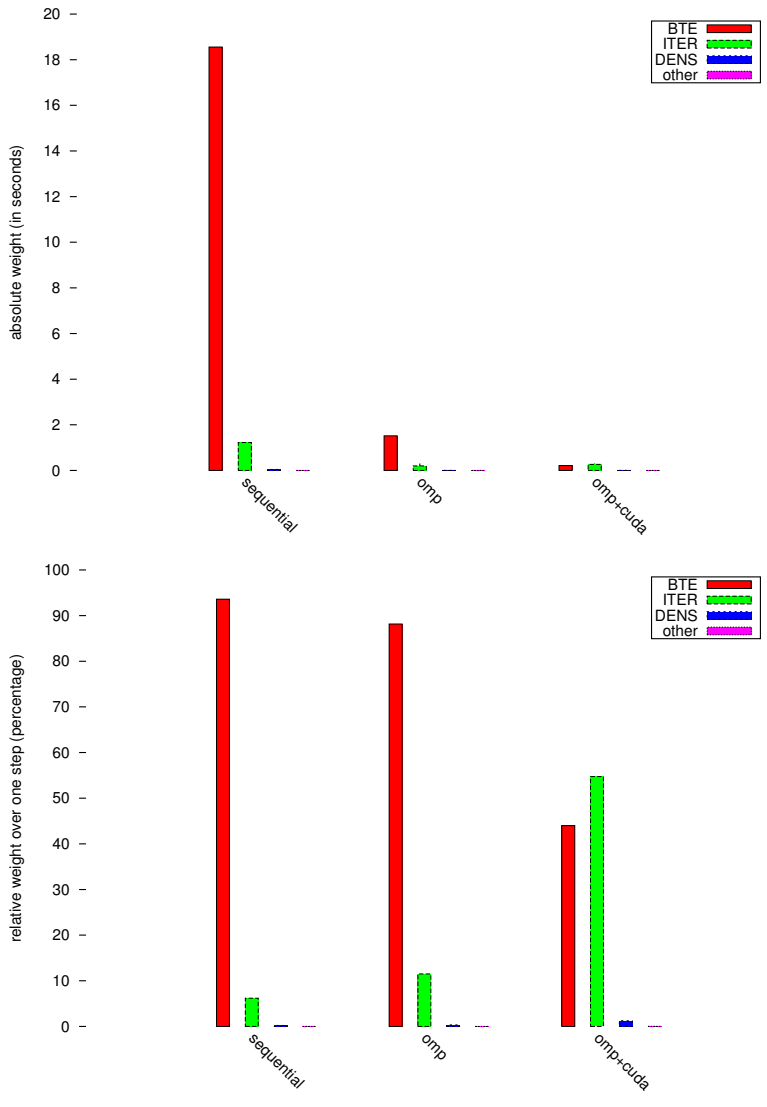








Relative weights



Cuda kernels

rank	kernel/function name	phase	avg exec time	Gflops/s
1	GPU_integrate_PHONONS_loss	BTE	32.8 ms	539
2	GPU_approx_partf_PHI	BTE	10.7 ms	598
3	GPU_approx_partf_W	BTE	10.6 ms	284
4	GPU_approx_partf_X	BTE	6.66 ms	389
5	GPU_set_fluxes_a3	BTE	2.92 ms	226
6	GPU_compute_integrated_pdf_energy	DENS	1.81 ms	9
7	GPU_integrate_PHONONS_gain	BTE	2.47 ms	275
8	GPU_perform_RK_2_3	BTE	3.59 ms	28
9	GPU_perform_RK_3_3	BTE	3.59 ms	28
10	GPU_perform_RK_1_3	BTE	2.90 ms	11
11	GPU_compute_Wm1	BTE	.297 ms	16
12	GPU_integrated_phitilde	DENS	.160 ms	2

Top performance on Tesla 40(c) GPU is 1430 Gflops/s.

Middle-term to-do list for the code

Surface roughness

Complete the analysis of the results with the surface roughness. (In progress, in collaboration with José Miguel Mantas and María José Cáceres.)

GPU implementation

Fully implement the solvers on the GPU to avoid memory transfer between host and graphic card. (In progress, in collaboration with José Miguel Mantas, Pedro Alonso and Antonio Vidal.)

Middle-term to-do list for the code

Surface roughness

Complete the analysis of the results with the surface roughness. (In progress, in collaboration with José Miguel Mantas and María José Cáceres.)

GPU implementation

Fully implement the solvers on the GPU to avoid memory transfer between host and graphic card. (In progress, in collaboration with José Miguel Mantas, Pedro Alonso and Antonio Vidal.)

GRAZIE!

The authors acknowledge Spanish projects **MTM2011-27739-C04-02** and **MTM2014-52056-P** and the European Fund for Development.