libNEGF: a portable library for quantum transport

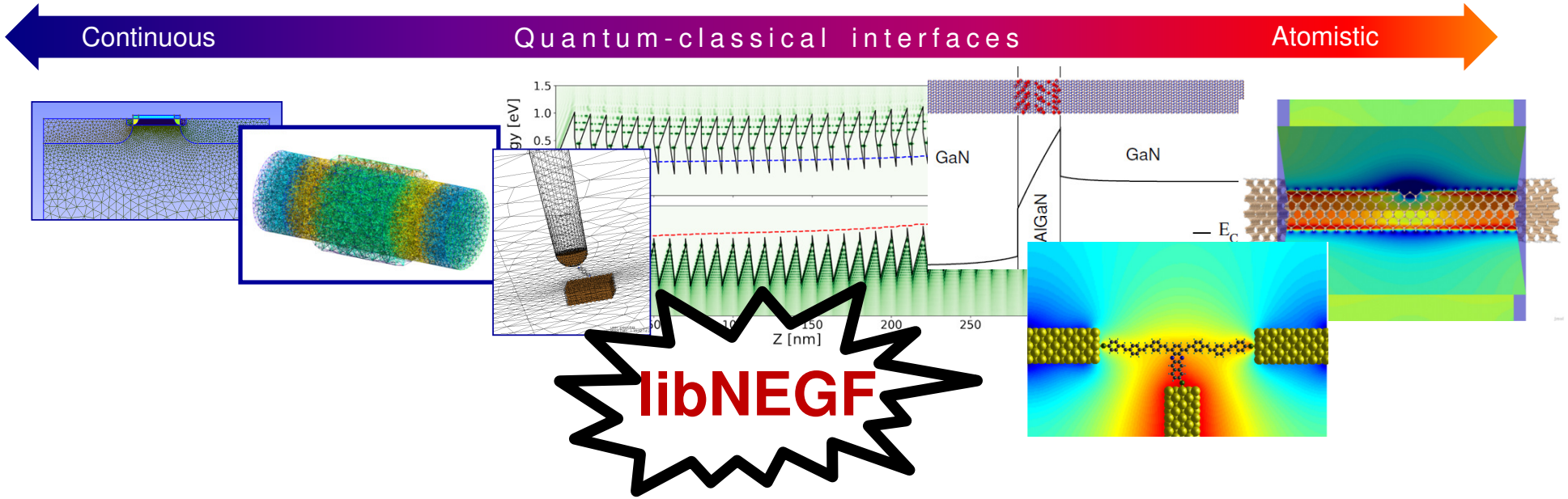
Alessandro Pecchia, CNR-ISMN

Daniele Soccodato, Matthias Auf der Maur, University of 'Tor Vergata'
Christoph Conrads, Edoardo Di Napoli, Forschungszentrum Julich

EERA-CASTIEL2 Webinar

EuroHPC Joint Undertaking
Co-funded by the European Union

*lib*NEGF as a general purpose library



- Interface any **local** Hamiltonian (FEM, ETB, DFTB, Wannier, Wavelets, LCAO-DFT)
- Basis-set agnostic
- Beyond coherent transport (**el-phonon**, **el-photon**, …)
- Ported to Nvidia GPU (CUDA)

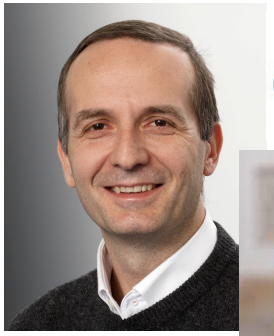
Fortran2003/2008/CUDA

http://www.github.com/libnegf

Tight Development Collaboration

CNR ISMN
ISTITUTO PER LO STUDIO DEI MATERIALI NANOSTRUTTURATI

E·C·E



E. Di Napoli

JÜLICH
Forschungszentrum

A. Buttari

IRIT

B. Aradi

DF+TB

BCCMS
Bremen Center for Computational Materials Science

L. Genovese, W. Dawson

HANAMI

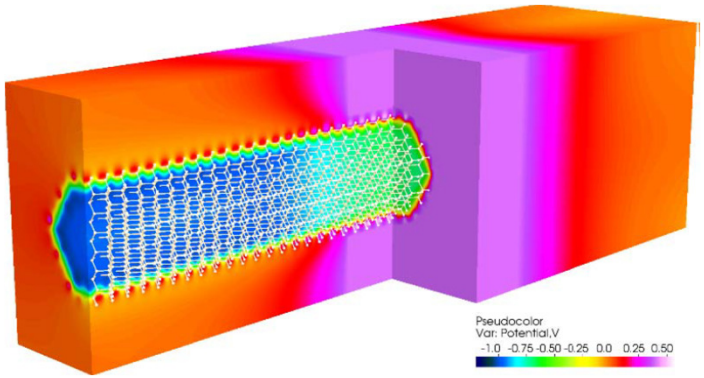
BigDFT

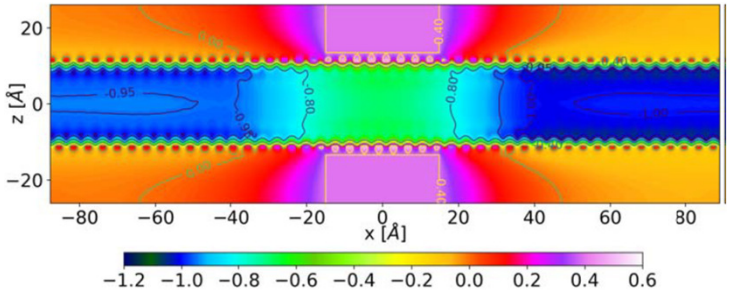
tC
tiberCAD
multiscale simulation

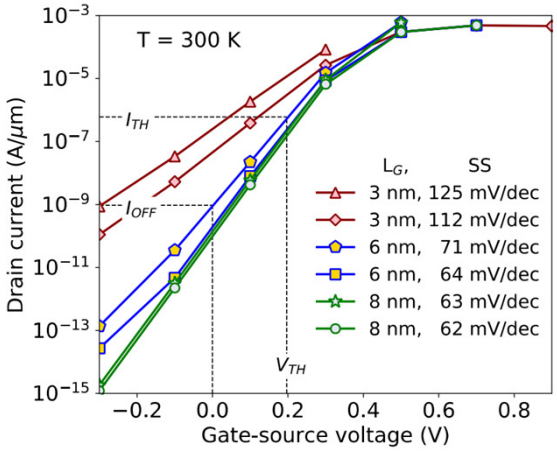
Università di Roma
Tor Vergata

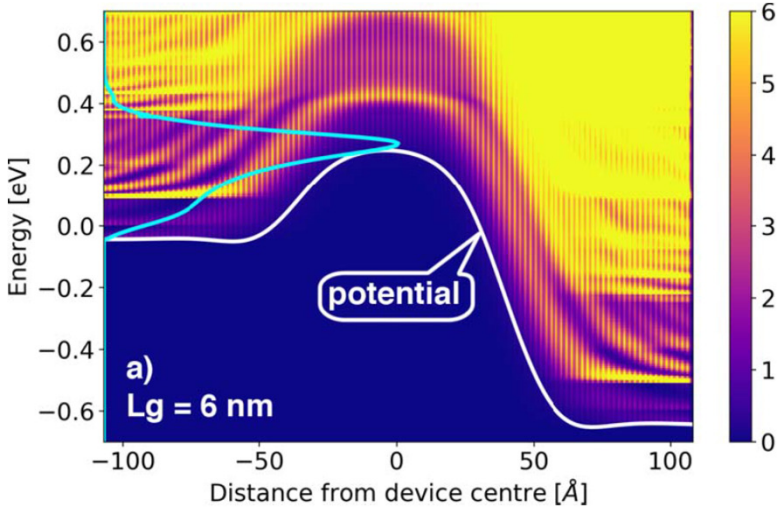
What's for ?

Gate All-around SiNW *nano* MOSFETs





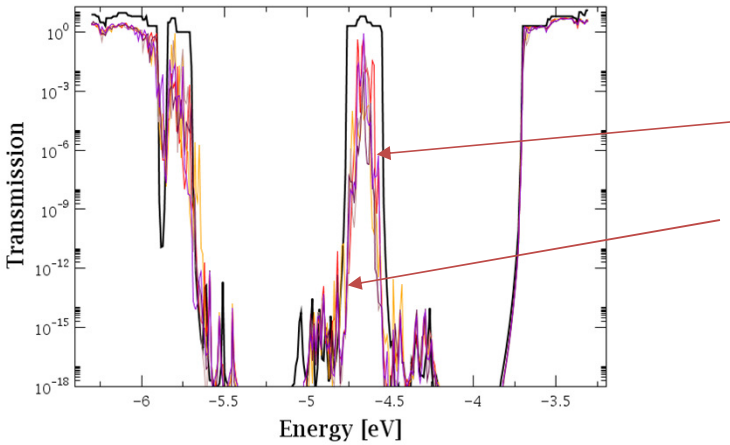


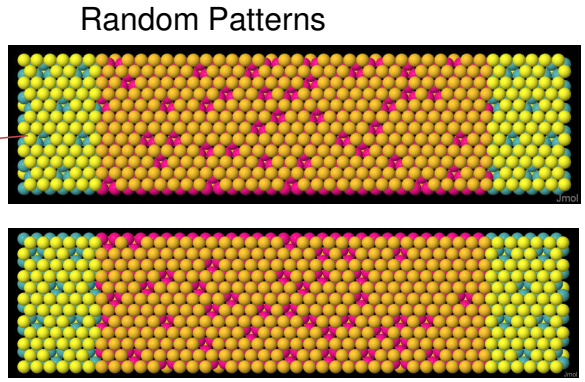


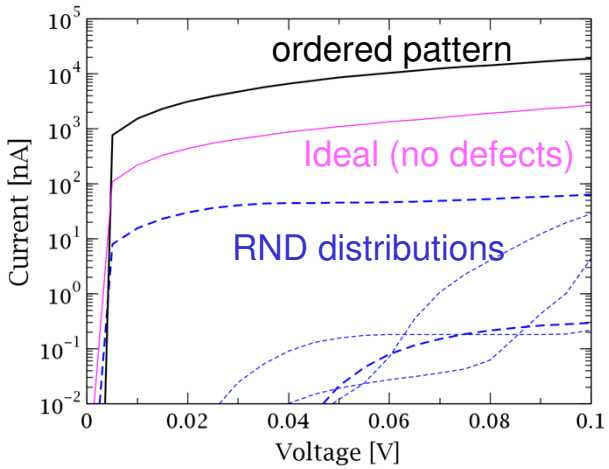
S. Markov et al. IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 66, NO. 3 (2019)

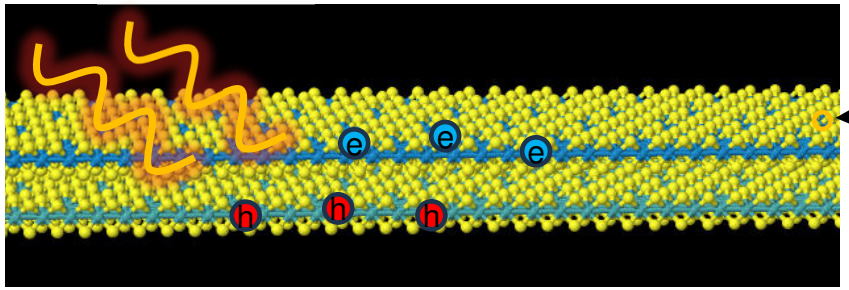
Address nanoscale transport / PV devices

Transport in MoS_2 with S-vacancies

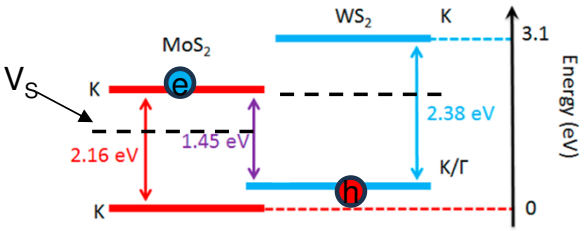








Type II band alignment



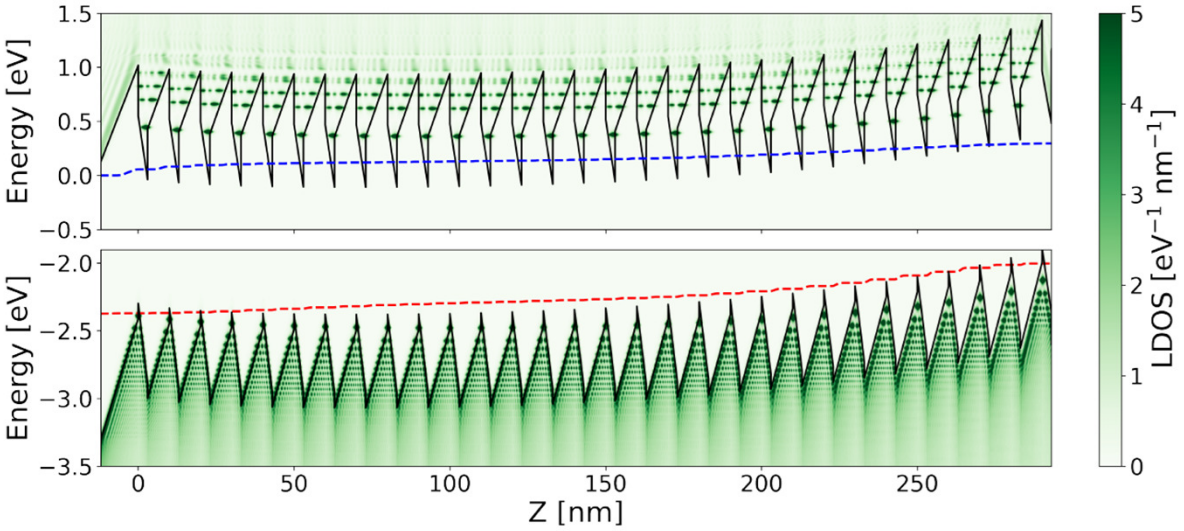
X. Hong et al. *Nat. Nan.* **9**, pages682–686 (2014)

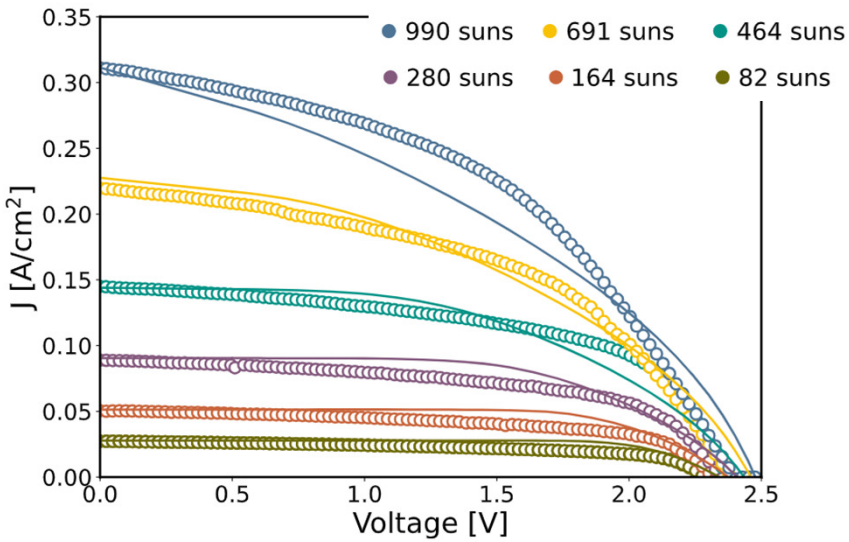
A. Pecchia et al., to be submitted (2025)

What's for ?

Multi QW solar cells based on InGaN/GaN

tiberCAD
multiscale simulation



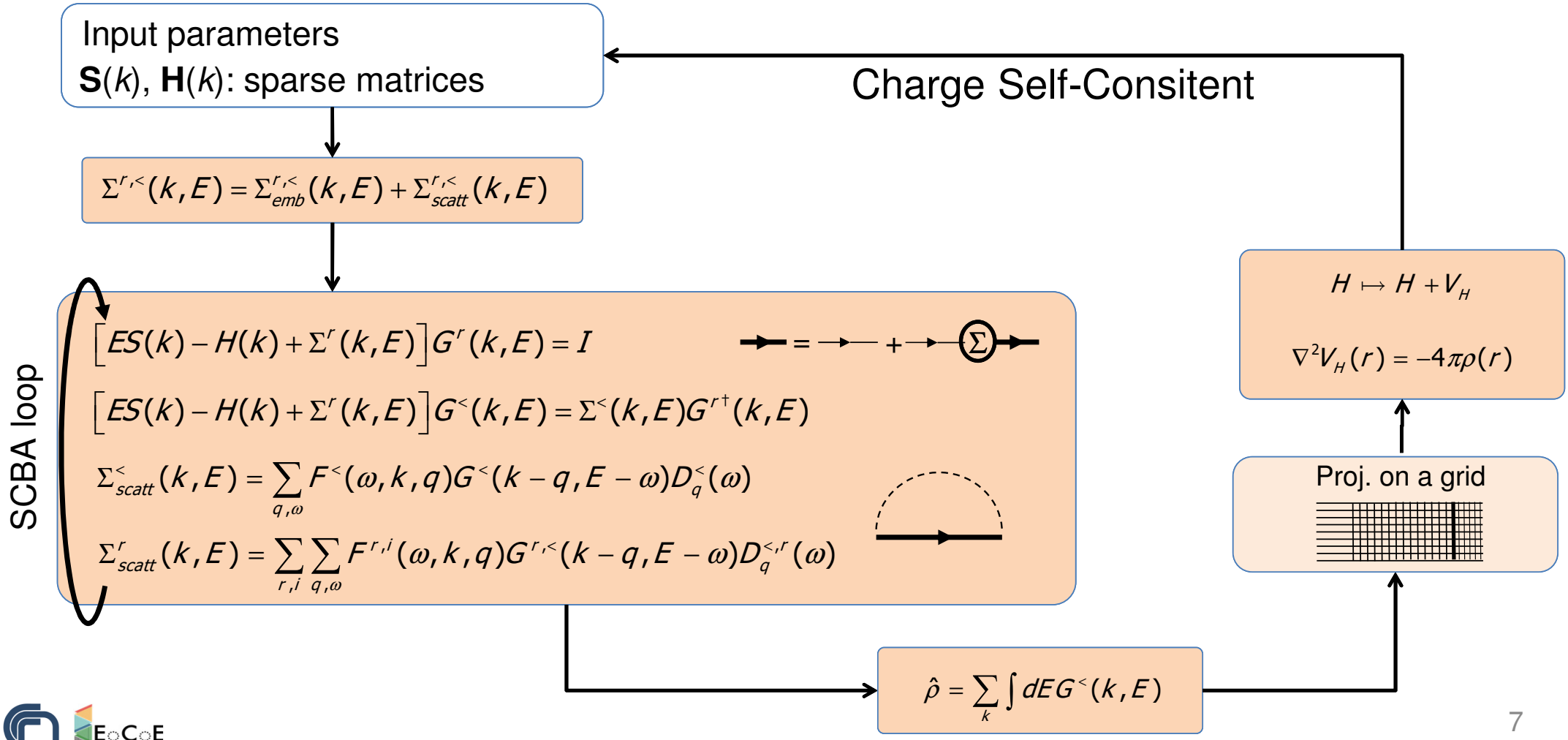


Density of states and J-V characteristics of the 30 QW active region

https://github.com/tiberlab

D. Soccodato et al., Phys. Status Solidi A 2025, 2500035

NEGF equations + External Self-consistent loop



Implemented Self-energies

- Elastic deformation potential scattering

$$\Sigma^{<}_{\mu\mu}(k,E)=\frac{1}{A}\sum_{\alpha,\lambda,q}\frac{\hbar^2\Omega_{cell}}{2\hbar\omega_{q\lambda}m_\alpha}\Xi^2_\lambda q^2(2N_q+1)G^{<}_{\mu\mu}(k-q,E)\approx\frac{k_BT\Xi^2_{LA}}{\rho c^2_{LA}}\left\langle G^{<}_{\mu\mu}(k,E)\right\rangle_{BZ}\approx\frac{k_BT\Xi^2_{LA}}{\rho c^2_{LA}}G^{<}_{\mu\mu}(k,E)$$

- Inelastic non-polar Optical within deformation potential

$$\Sigma^{<}_{\mu\mu}(k,E)=\frac{\hbar^2\Xi^2_o}{2\hbar\omega_o\rho}\left[(N_q+1)G^{<}_{\mu\mu}(k,E-\hbar\omega_o)+N_qG^{<}_{\mu\mu}(k,E+\hbar\omega_o)\right]$$

- Inelastic polar Optical Frölich coupling

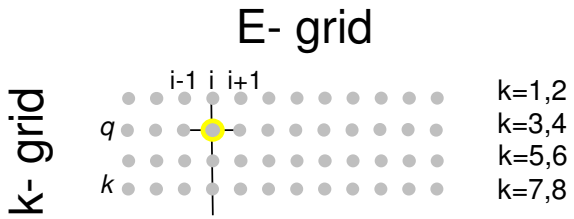
$$\Sigma^{<}_{\mu\nu}(k,E)=\sum_q F(|z_\mu-z_\nu|,k,q)\left[(N_o+1)G^{<}_{\mu\nu}(q,E-\hbar\omega_o)+N_oG^{<}_{\mu\nu}(q,E+\hbar\omega_o)\right]$$

- Electron-photon self-energy

$$\Sigma^{<}_{\mu\nu}(k,E)=M_{\mu\sigma}\left[(N_\gamma+1)G^{<}_{\sigma\tau}(k,E-\hbar\omega_\gamma)+N_\gamma G^{<}_{\sigma\tau}(k,E+\hbar\omega_\gamma)\right]M_{\tau\nu}$$

Parallelization strategy

2D MPI cartesian grid: Energies and k-points



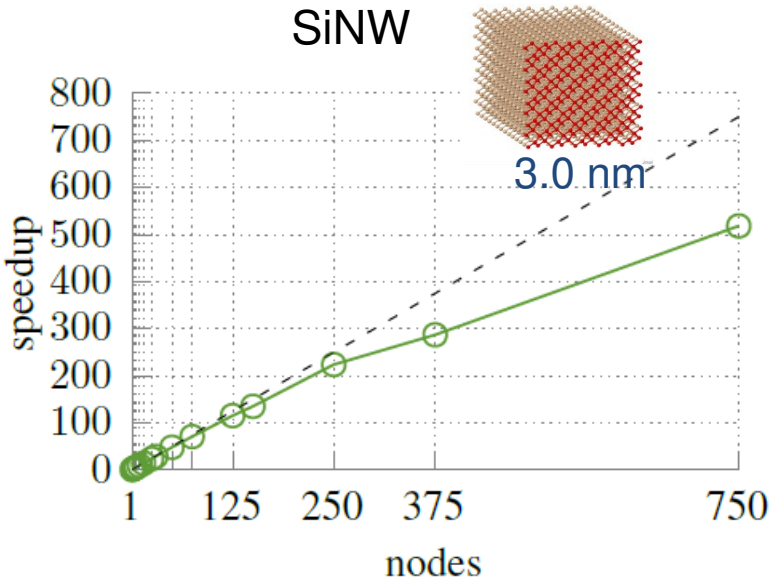
Typical size of the grid

k-mesh: 8x8 grid = 64 points
E-grid: ~100 points

⇒ 6400 points

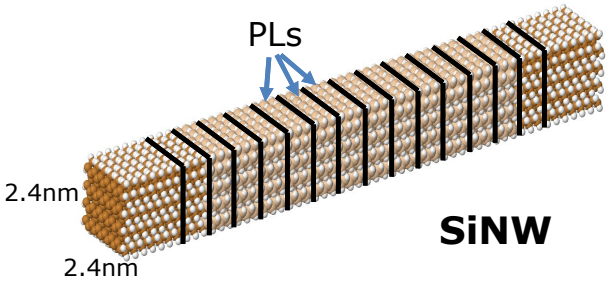
Spatial domain distribution under development

Ballistic runs (no scattering)

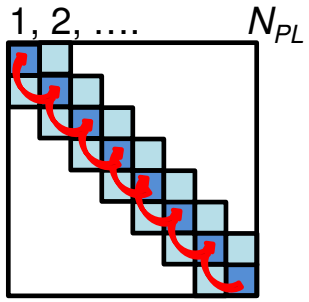


Strong scaling benchmarks

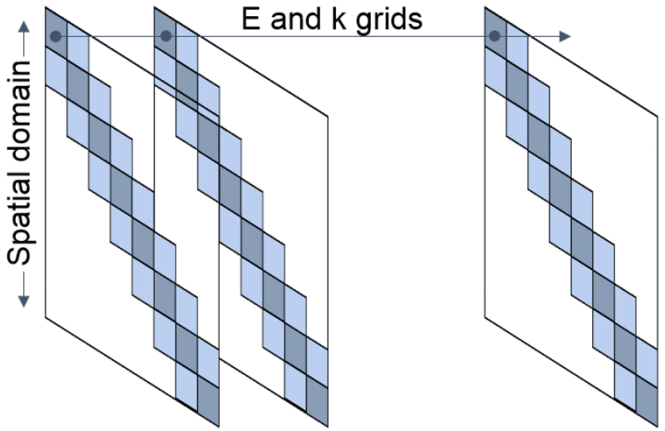
Block-Recursive algorithm for G^r and $G^<$



3-diagonal block-dense representation

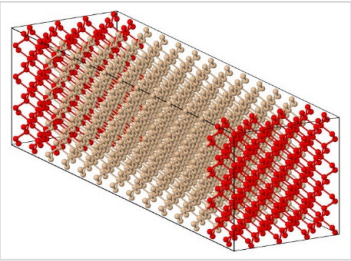


Communications needed for $\Sigma^{r,<}_{scatt}(k,E)$

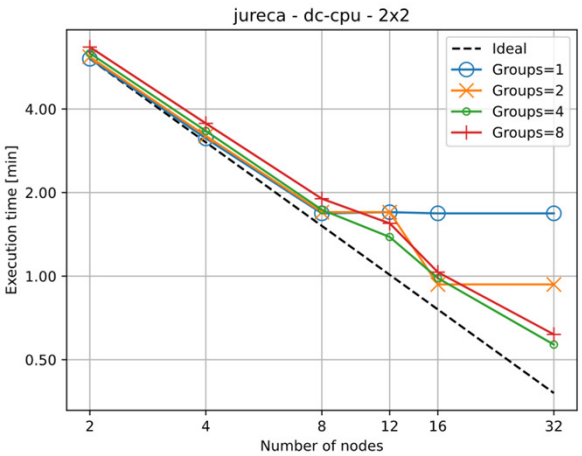


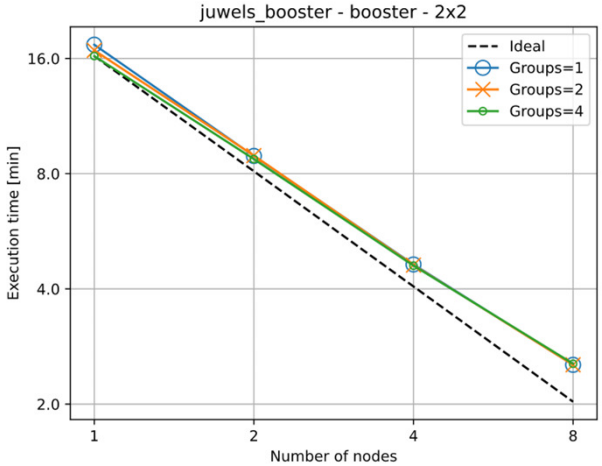
- *E, k* grids are independent of spatial domain
- Parallelize on spatial domain => 3D cartesian grid

Benchmarks – Strong Scaling



Layer size: 288x288
Number of Layers: 10





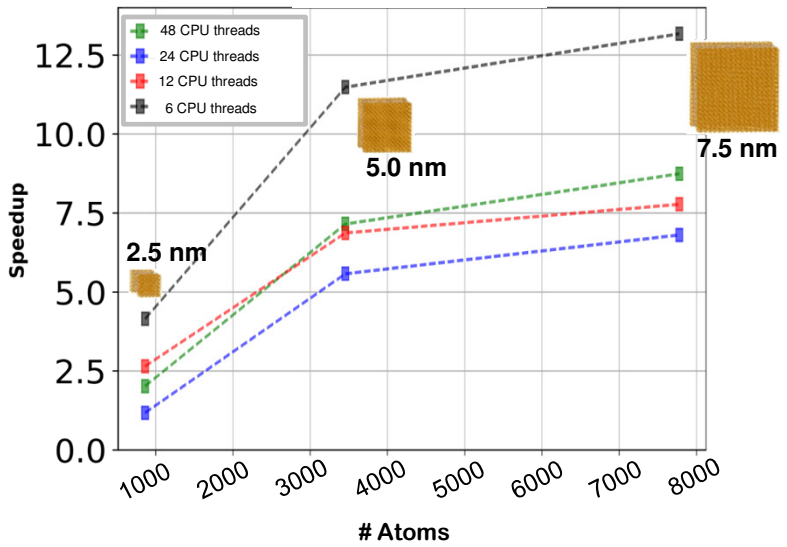
67%

- Benchmarks run on a small Si system with a 2x2 cell with 10 Layers
  - Full inelastic calculation including 1 PO phonon
  - Solution of the SCBA loop
  - Calculation of the layer-to-layer current

Block-recursive routines on GPUs

All recursive numerical algorithms implemented in libNEGF have been ported to GPU

We use a more flexible CUDA implementation interfacing cuBLAS/cuSolver kernels



Speedup of GPU vs CPU for a single MPI task
GPU vs multithreaded CPU: Nvidia A100 vs Intel Xeon 8168

Profiling of selected intensive parts in **single-precision**

Size	Mat Size	CPU (threaded)	GPU	Tens.-C	Speedup
4.8 nm	5832	119 s	1.4 s	0.42 s	x30 - x120
6.5 nm	10368	450 s	4.0 s	1.19 s*	x100 - x400
9.2 nm	20736	3600 s	30.0 s	8.2 s*	x120 - x440

Peak performance on tens.C exceeds 90 Tflops/s

Numerical issues to run the whole NEGF algorithms in sp.

⇒ Mixed precision solutions are needed

Summary

- libNEGF is an LGPL library for NEGF available on github
- Active development involving CNR, JSC, CNRS-IRIT
- API that can interface F2003, C++, Python
- Runs on Intel/NVIDIA arch. (Juwels Booster, LEONARDO, JEDI, etc.)
  Work in progress: porting to LUMI (AMD).
- Codes currently interfaced:
  - dftb+ (semi-empirical DFT method, F2003)
  - tiberCAD (FEM, ETB methods, C++)
  - bigDFT (ab-intio wavelet method, work in progress)

THANK YOU