MAX DRIVING THE EXASCALE TRANSITION

siesta

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

ICN2 Institut Català de Nanociència i Nanotecnologia

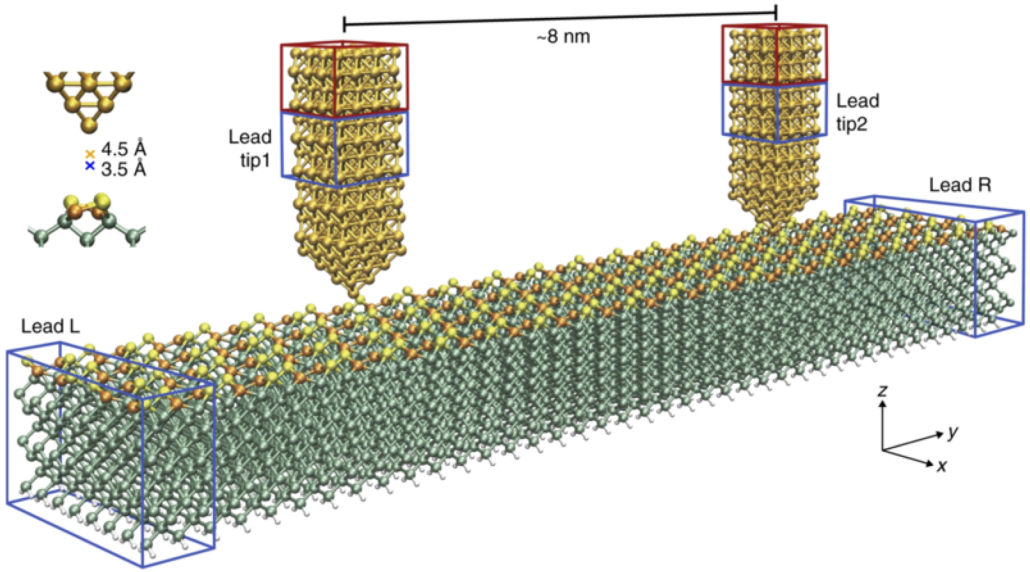
HPC-enabled very large scale quantum simulations in materials with SIESTA

Pablo Ordejón (ICN2, Barcelona)

MaX "*Materials Design at the Exascale*", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEDI-2018-1, grant agreement 824143

Large Scale Quantum Simulations

Electronic transport in nanoscale devices



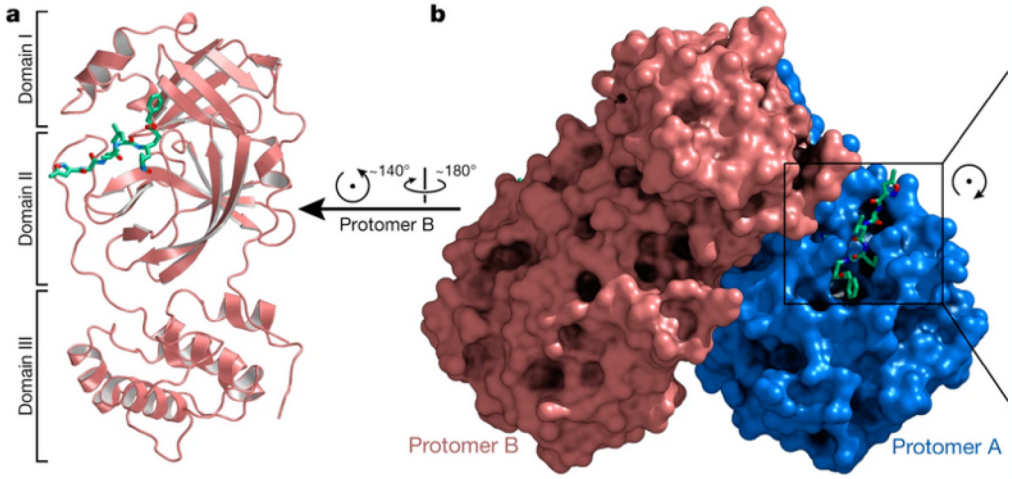
Four-terminal electron transport on Ge surface and gold tips

4924 atoms

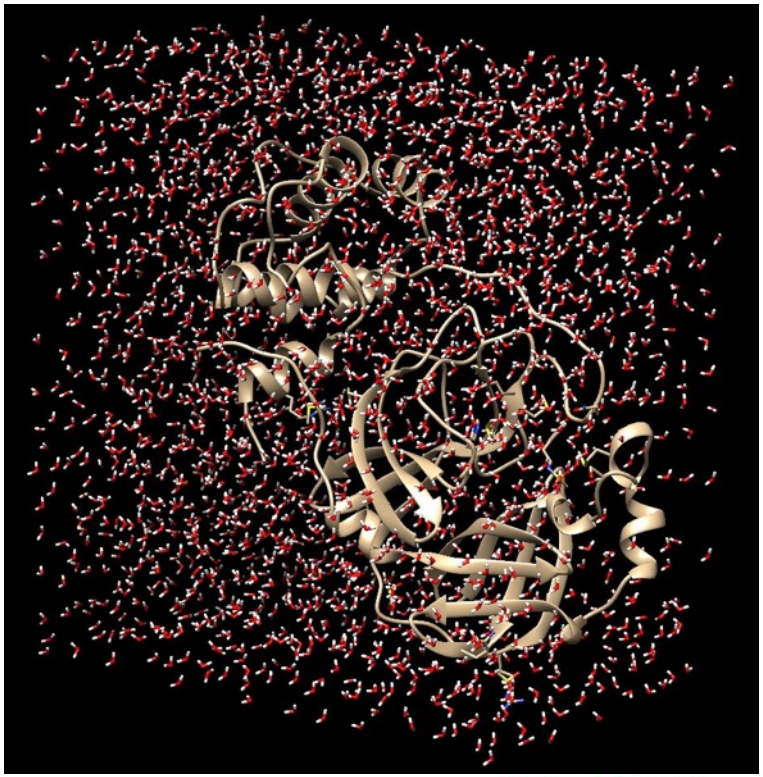
M. Kolmer et al Nature Commun. 2019

Large Scale Quantum Simulations

SARS CoV-2 M^{pro} - interaction with inhibitor drugs



Jin et al, Nature 582, 289 (2020)

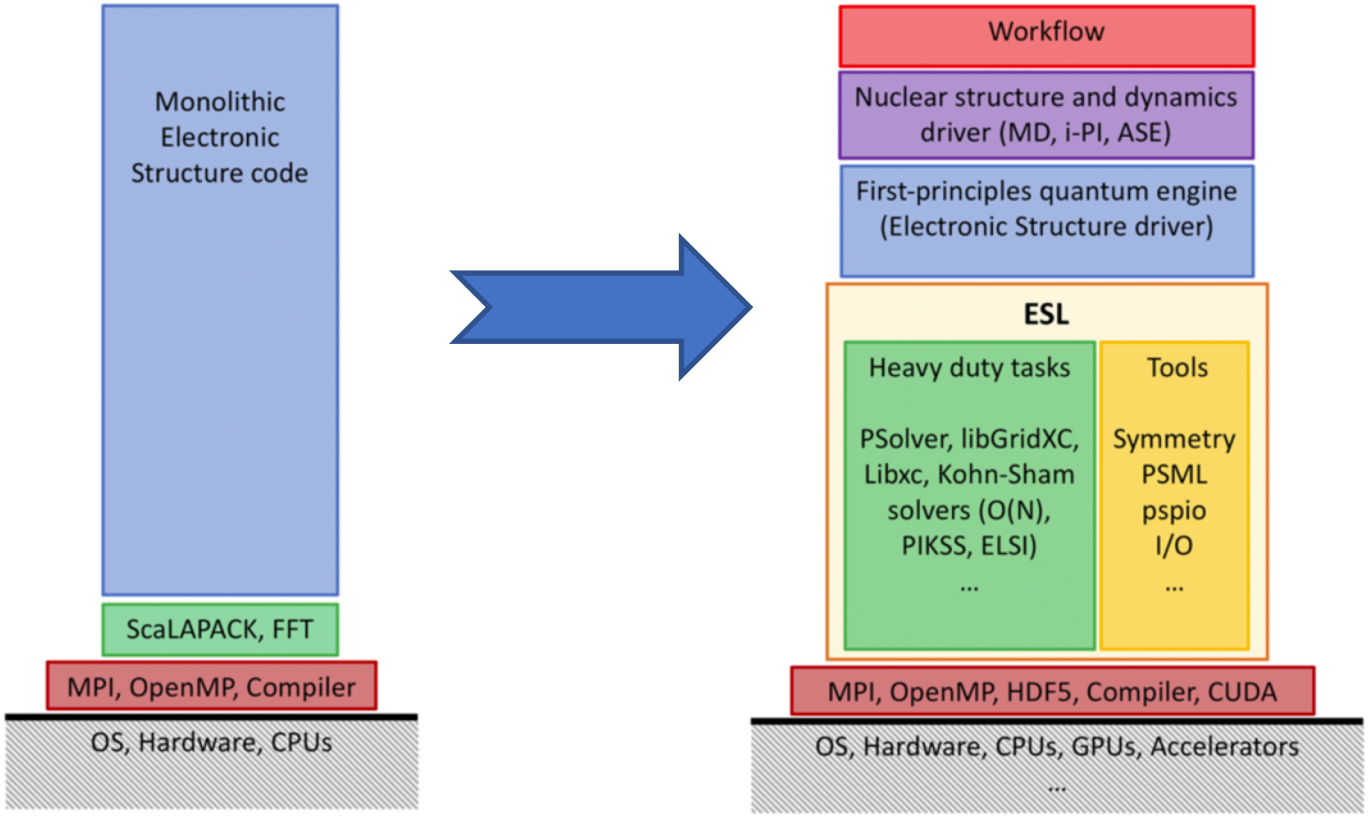


Protomer B in solvation

~8800 atoms

SIESTA development for HPC and scalability

Changing paradigm



The SIESTA code

SIESTA is a DFT code, density-functional theory

(like many others in many ways)

Aim from inception: EFFICIENCY

$$\psi_n(r) = \sum_\mu c_{n\mu}\phi_\mu(r)$$

$$h_{\nu\mu} = \int d\vec{r}\ \phi_\nu^*(\vec{r})\ \hat{h}\ \phi_\mu(\vec{r})$$

$$\hat{h}_{\mu\nu} c_{n\mu} = \varepsilon_n\, \hat{s}_{\mu\nu}\, c_{n\mu}$$

$$\rho_{\mu\nu} = \sum_i c_{\mu i} n_i c_{i\nu}$$

Pioneer LINEAR-SCALING DFT code (or Order-N, O(N))

meaning: computational cost (CPU & memory) scaling linearly with number of atoms

The Solvers

We need to obtain the density from the Hamiltonian

Direct eigenvalue/eigenvector solvers

$$\hat{h}_{\mu\nu} c_{n\mu} = \varepsilon_n \hat{s}_{\mu\nu} c_{n\mu}$$

$$\rho_{\mu\nu} = \sum_i c_{\mu i} n_i c_{i\nu}$$

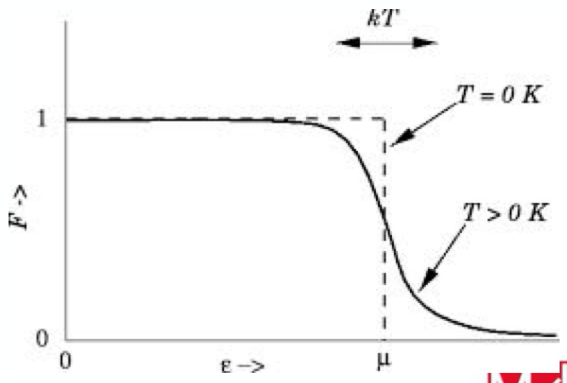
$$\rho(\boldsymbol{r}) = \sum_{\mu\nu} \rho_{\mu\nu} \phi^*_\nu(\boldsymbol{r}) \phi_\mu(\boldsymbol{r})$$

Obtaining the density directly

Fermi-Dirac function

$$\hat{\rho} = f_\beta(\hat{H} - \mu)$$

$$f_\beta(\epsilon_i - \mu) = \frac{2}{1 + \mathrm{e}^{\beta(\epsilon_i - \mu)}}$$



Diagonalization-based solvers

Need to use DIRECT solvers, as the number of desired eigenvectors is a substantial fraction of the matrix size

SIESTA uses pre-packaged libraries for this pure math problem:

- ScaLaPACK
  - pdsyev, pzheev and related drivers
  - MRRR
- ELPA: Alternative transformation sequence + optimizations
  https://elpa.mpcdf.mpg.de/

- Conversion of H and S to dense form
- Cholesky decomposition to reduce to standard eigenproblem
- Transformation to tri-diagonal form
- Solution of tri-diagonal problem
- Back-transformation

Cubic scaling with matrix size — Quadratic scaling for memory

Method of choice for "small" problems (some hundreds of atoms)

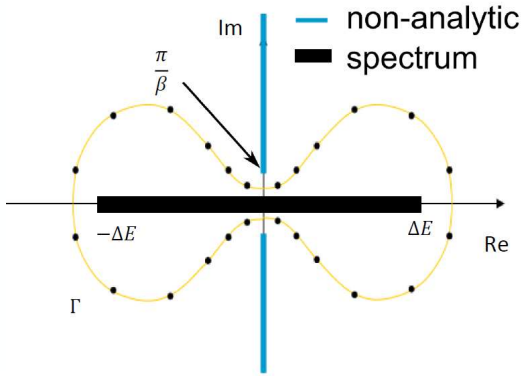
Direct solution – Fermi Operator

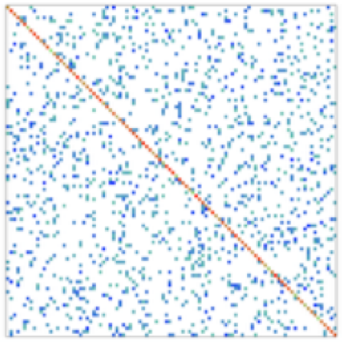
PEXSI: Pole Expansion plus Selected Inversion

(Lin Lin, Chao Yang, et al., Berkeley)

Lin, García, Huhs, Yang, JPCP 2014

$$f_\beta(\epsilon_i - \mu) = \frac{2}{1 + e^{\beta(\epsilon_i - \mu)}} \approx Im \sum_{l=1}^{P} \frac{\omega_l}{\epsilon_i - (z_l + \mu)}$$





$$\hat{\rho}(x) \approx \sum_{ij} \phi_i(x) Im \left(\sum_{l=1}^{P} \frac{\omega_l}{H - (z_l + \mu)S} \right) \phi_j(x)$$

$$\begin{aligned} \text{(quasi-)1D:} \quad & \mathcal{O}(N) \\ \text{(quasi-)2D:} \quad & \mathcal{O}(N^{3/2}) \\ \text{3D:} \quad & \mathcal{O}(N^2) \end{aligned}$$

- Relatively small number of poles (20-30)
- Trivially parallelizable over poles
- Only selected elements of inverse are needed

(Due to sparsity of the target density matrix)

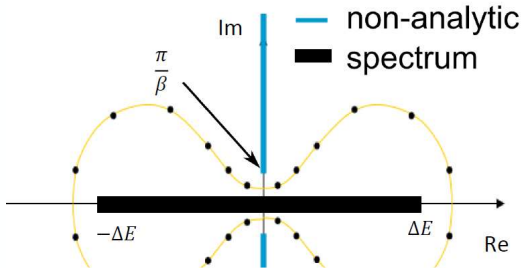
Direct solution – Fermi Operator

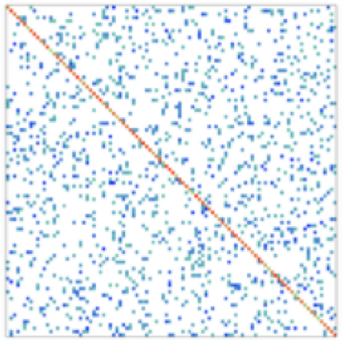
PEXSI: Pole Expansion plus Selected Inversion

(Lin Lin, Chao Yang, et al., Berkeley)

Lin, García, Huhs, Yang, JPCP 2014

$$f_\beta(\epsilon_i - \mu) = \frac{2}{1 + e^{\beta(\epsilon_i - \mu)}} \approx Im \sum_{l=1}^{P} \frac{\omega_l}{\epsilon_i - (z_l + \mu)}$$

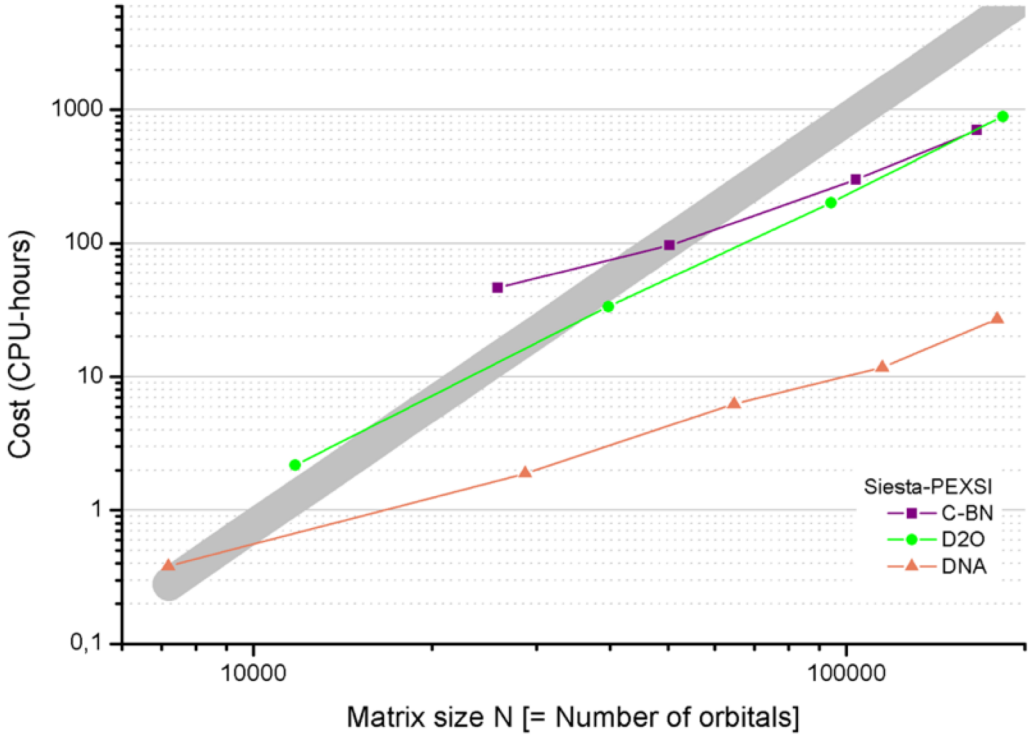




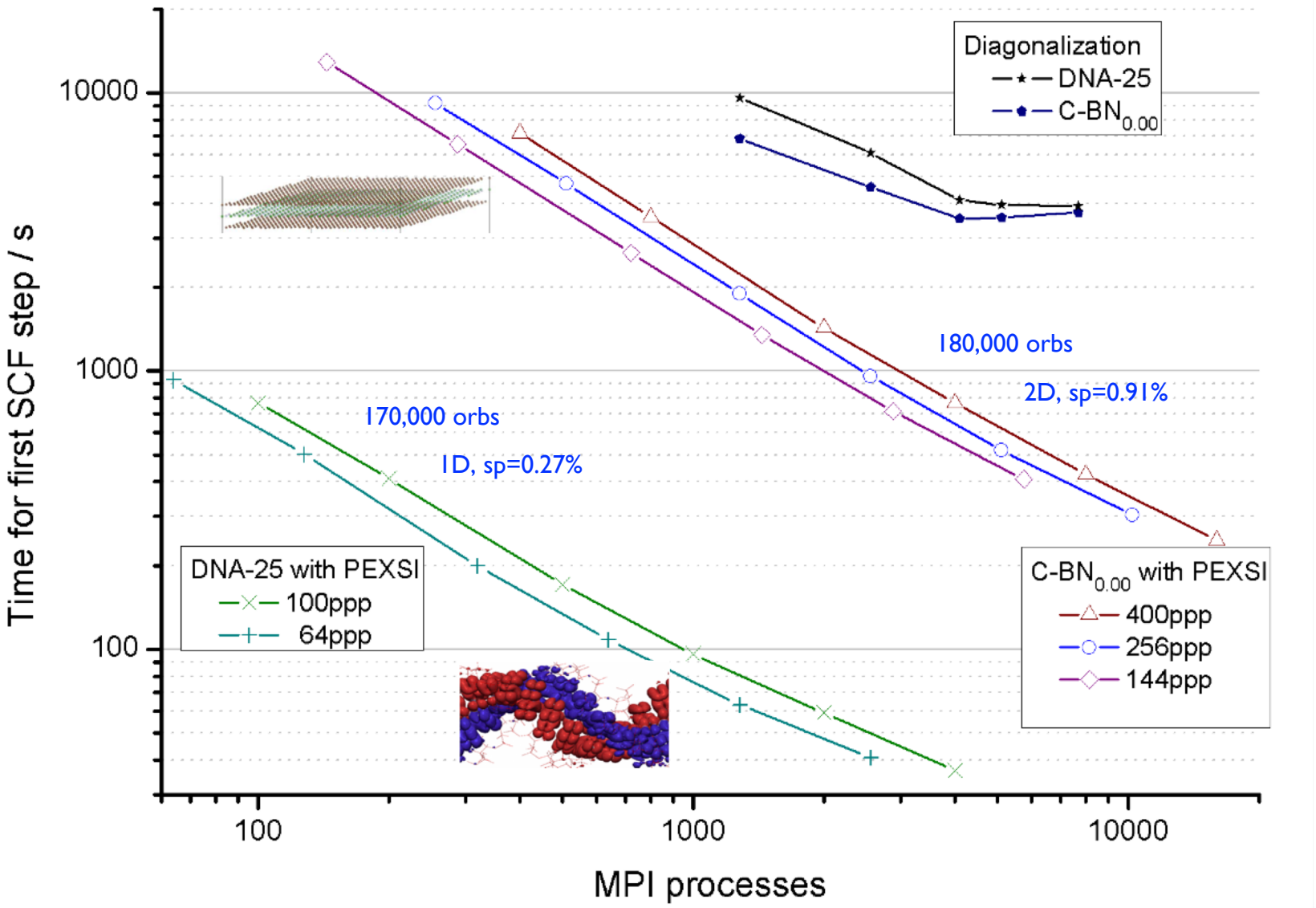
$$\hat{\rho}(x) \approx \sum_{ij} \phi_i(x)$$

$$\begin{aligned} \text{(quasi-)1D:} &\quad \mathcal{O}(N) \\ \text{(quasi-)2D:} &\quad \mathcal{O}(N^{3/2}) \\ \text{3D:} &\quad \mathcal{O}(N^2) \end{aligned}$$

(Due to sparsity of the target density

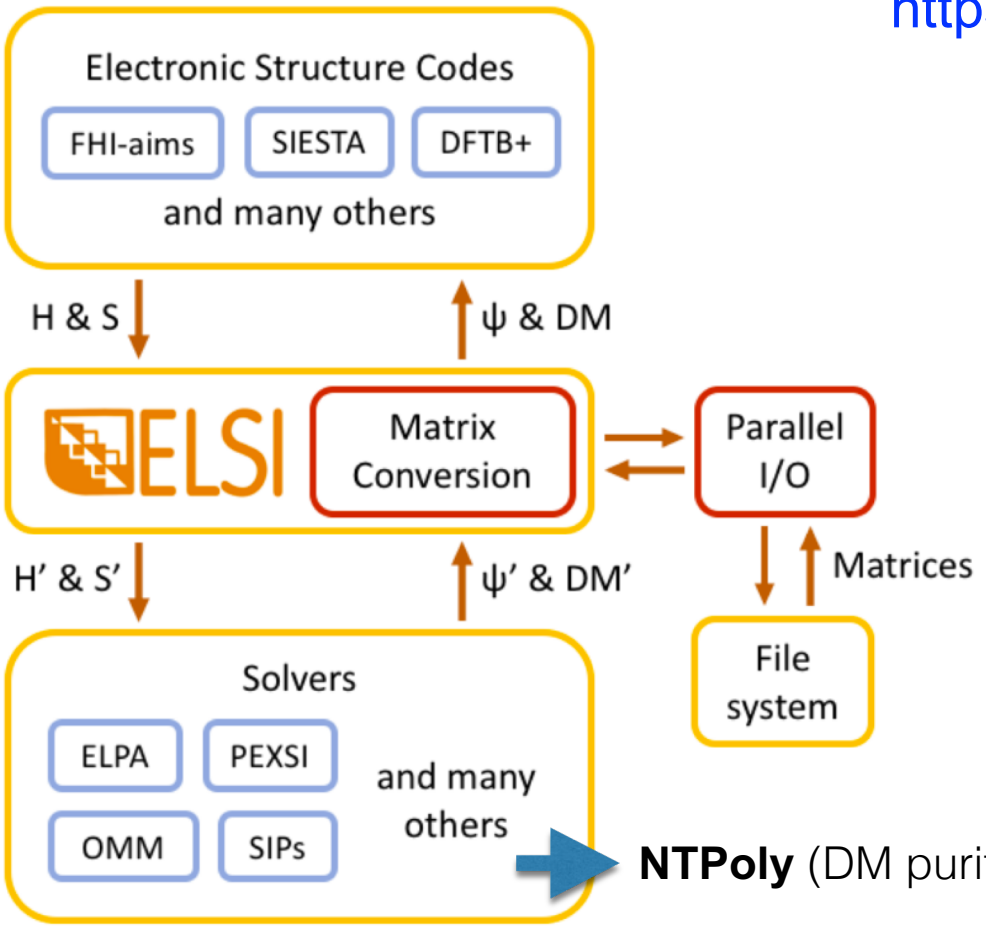


Massive scalability: PEXSI solver



ELSI initiative to integrate solver libraries

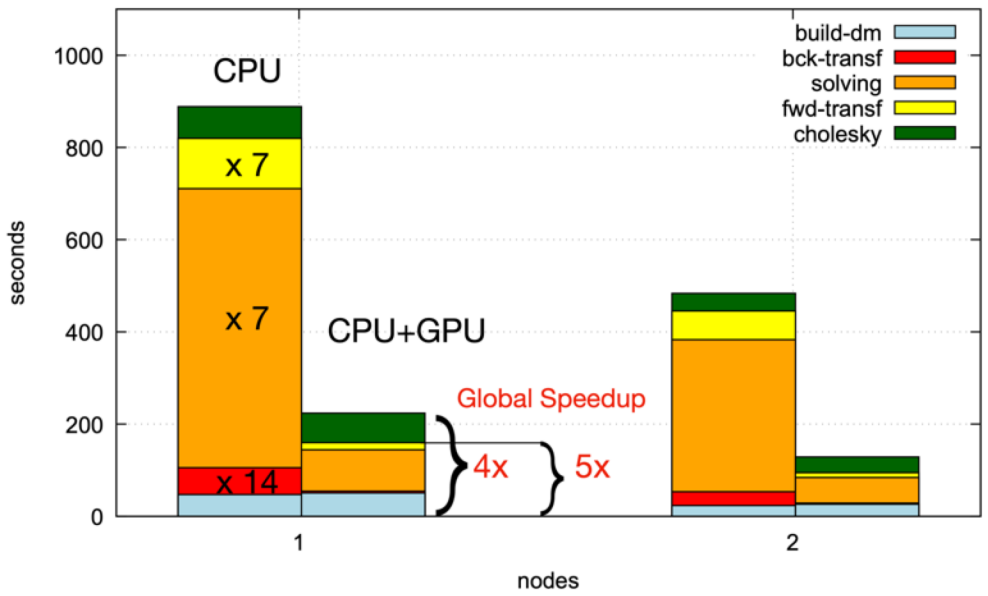
https://elsi-interchange.org

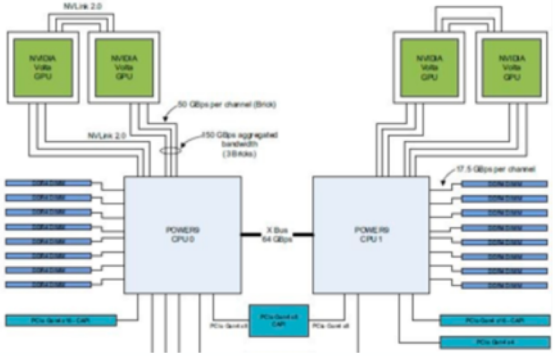


GPU acceleration for diagonalization

Marconi-100 (CINECA): 32 CPUs+ 4 GPUs /node

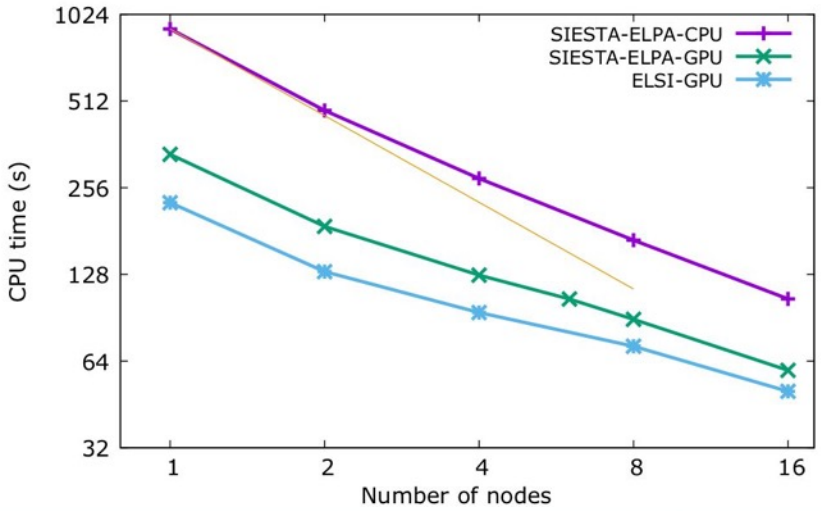
ELSI-ELPA GPU acceleration





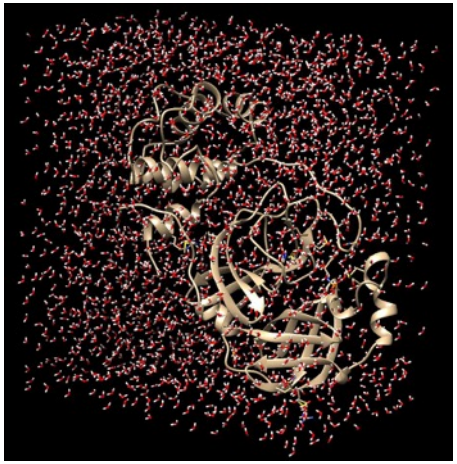
Proper binding of GPUs to MPI ranks

Future enhancements in ELPA (better kernels) and in ELSI (e.g. build-DM stage) are integrated in SIESTA automatically



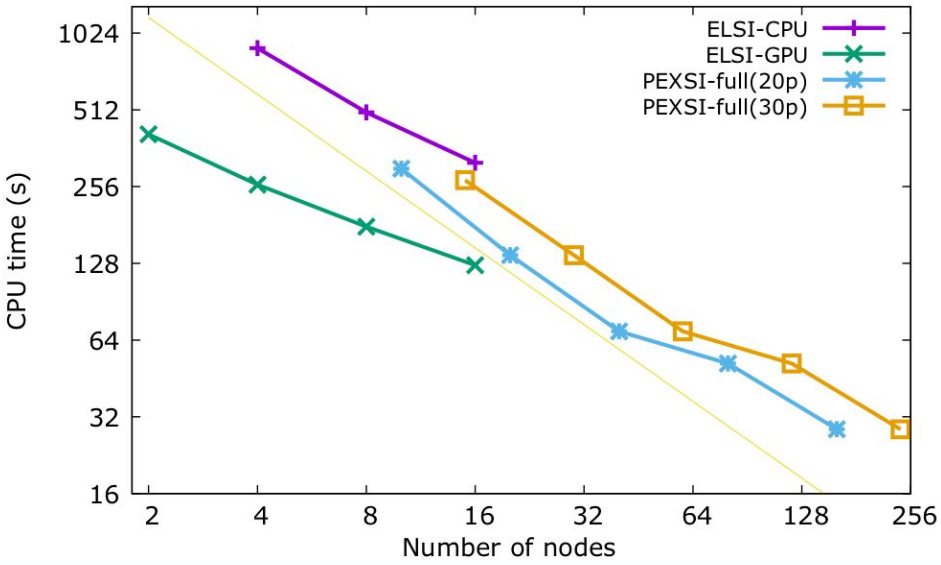
System: Si quantum dot, with approx 35000 orbs

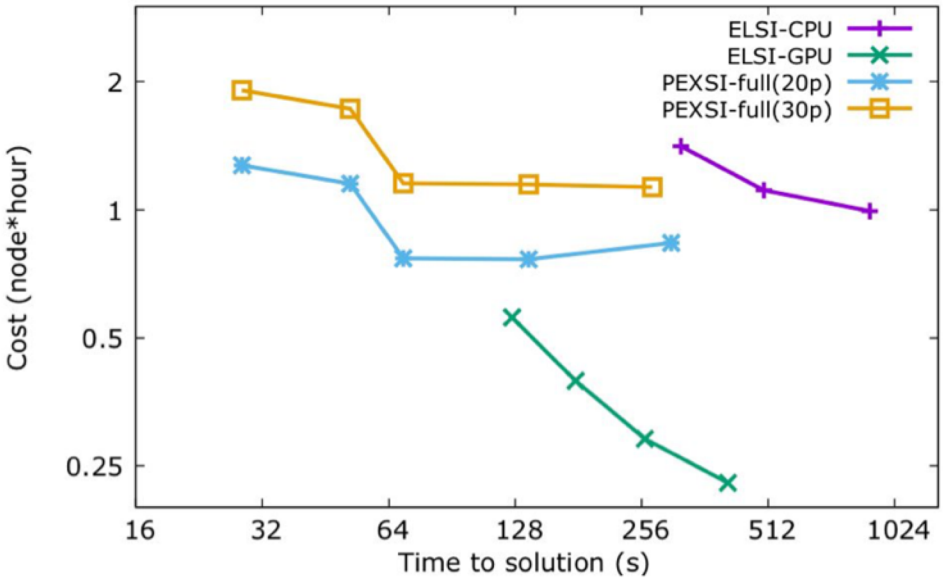
Comparison of global efficiency of solvers for a very large problem



SARS CoV-2 M^{pro} with solvation water molecules

Approx. 8800 atoms; 58000 orbitals





Work on GPU acceleration of PEXSI library is under way

DRIVING THE EXASCALE TRANSITION

THANKS

Follow us on:

company/max-centre/

@max_center2

http://www.max-centre.eu/

siesta

siesta-project.org
gitlab.com/siesta-project