MAX DRIVING THE EXASCALE TRANSITION

siesta

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

CSIC

ICMAB

EXCELENCIA SEVERO OCHOA

Electronic-Structure Solvers in SIESTA: Features and Performance

Alberto García (ICMAB-CSIC, 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

The basic core of SIESTA

$$\psi_i(\boldsymbol{r}) = \sum_\mu \phi_\mu(\boldsymbol{r}) c_{\mu i},$$

$$\sum_{\nu\beta} (H^{\alpha\beta}_{\mu\nu} - E_i S_{\mu\nu}\delta^{\alpha\beta}) c^{\beta}_{\nu i} = 0$$

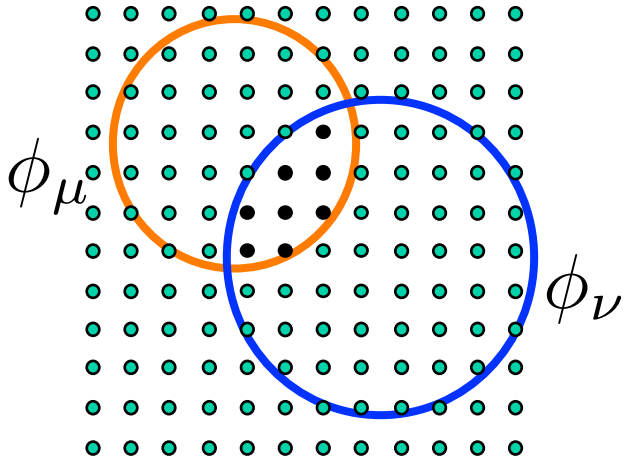
Generalized eigenvalue problem

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

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

Density matrix

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$$



The SOLVER step takes most of the CPU time

Diagonalization-based solvers

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

$$\sum_{\nu\beta}(H_{\mu\nu}^{\alpha\beta} - E_i S_{\mu\nu}\delta^{\alpha\beta})c_{\nu i}^{\beta} = 0$$

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

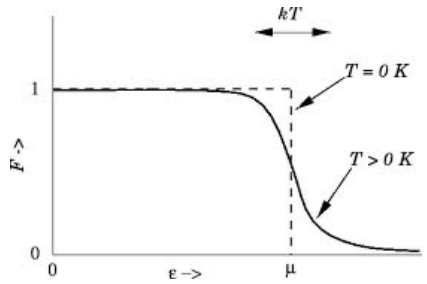
Still competitive for low-cardinality basis sets

Direct solution for the density matrix

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

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

Fermi-Dirac function



Fermi Operator Expansion (FOE)

$$p(H) = \frac{c_0}{2} I + \sum_{j=1}^{n_{pl}} c_j T_j(H)$$

Calculation of the DM involves only (sparse) matrix-vector multiplications

CheSS library
(originally in BigDFT)

Linear-scaling

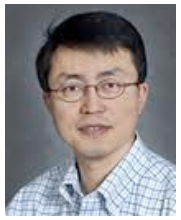
Stephan Mohr (BSC)

- Number of terms in the expansion can be large
- Efficiency increases for contracted basis sets.
- Exploring on-the-fly contraction

Direct solution for the density matrix

PEXSI: Pole Expansion plus Selected Inversion

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



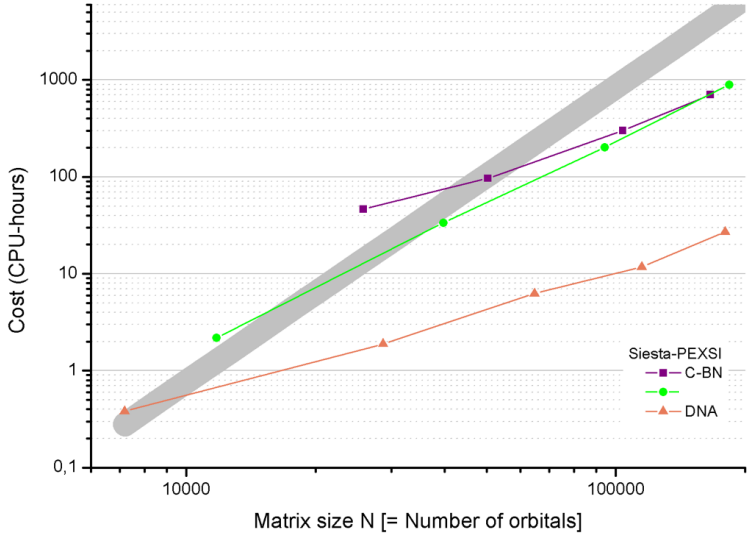
$$\hat{\rho} = Im\left(\sum_{l=1}^{P} \frac{\omega_l}{H-(z_l+\mu)S}\right)$$

For sufficiently big problems

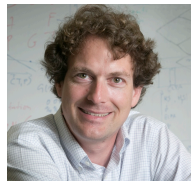
(quasi-)1D:	$\mathcal{O}(N)$
(quasi-)2D:	$\mathcal{O}(N^{3/2})$
3D:	$\mathcal{O}(N^2)$

(Due to sparsity of the target density matrix)

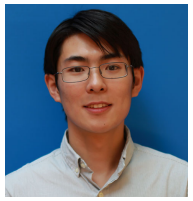
Relatively small number of poles (20-30)
Trivially parallelizable over them



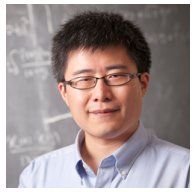
ELSI initiative to integrate solver libraries



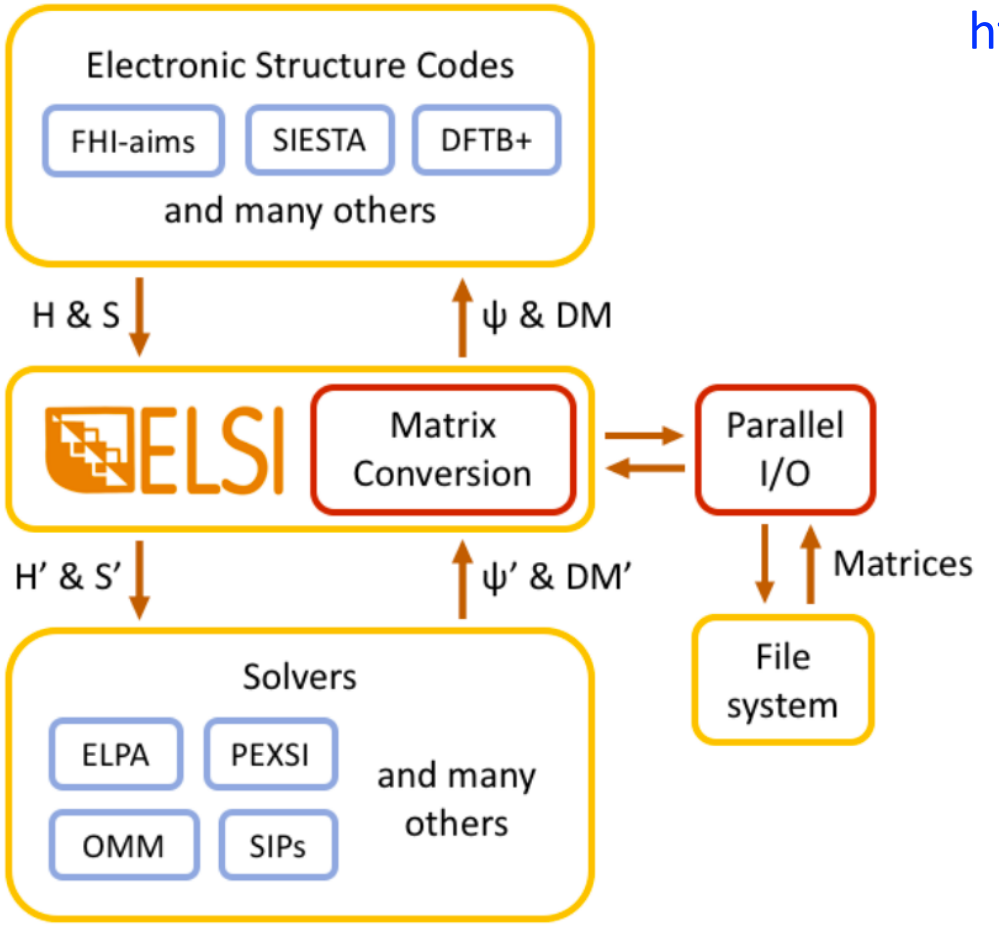
Volker Blum, Duke



Lin Lin, Berkeley



Jiangfen Lu, Duke

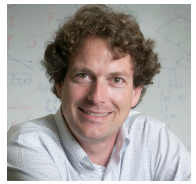


https://elsi-interchange.org

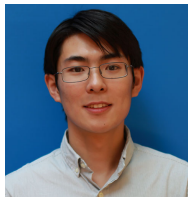
Interface in Siesta:

Collaboration with Victor Yu (Duke)

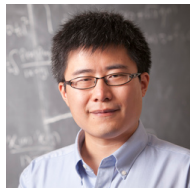
ELSI initiative to integrate solver libraries



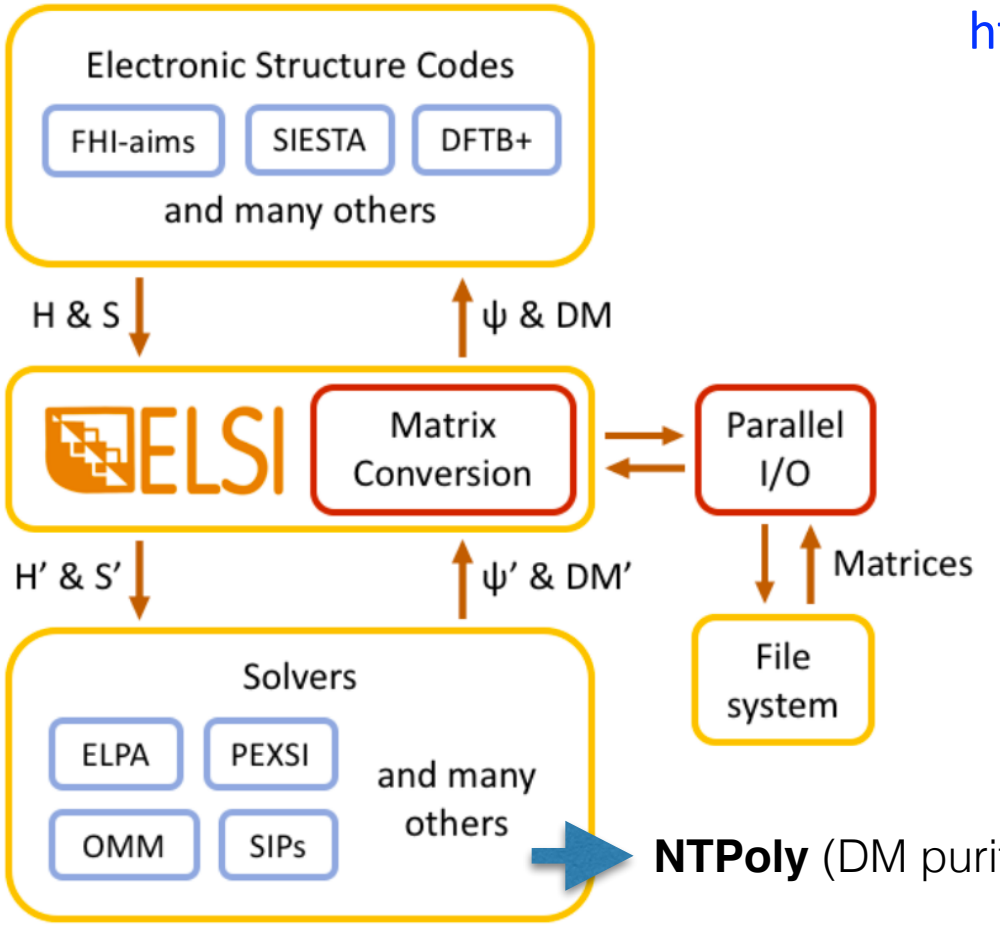
Volker Blum, Duke



Lin Lin, Berkeley



Jiangfen Lu, Duke



NTPoly (DM purification, O(N))

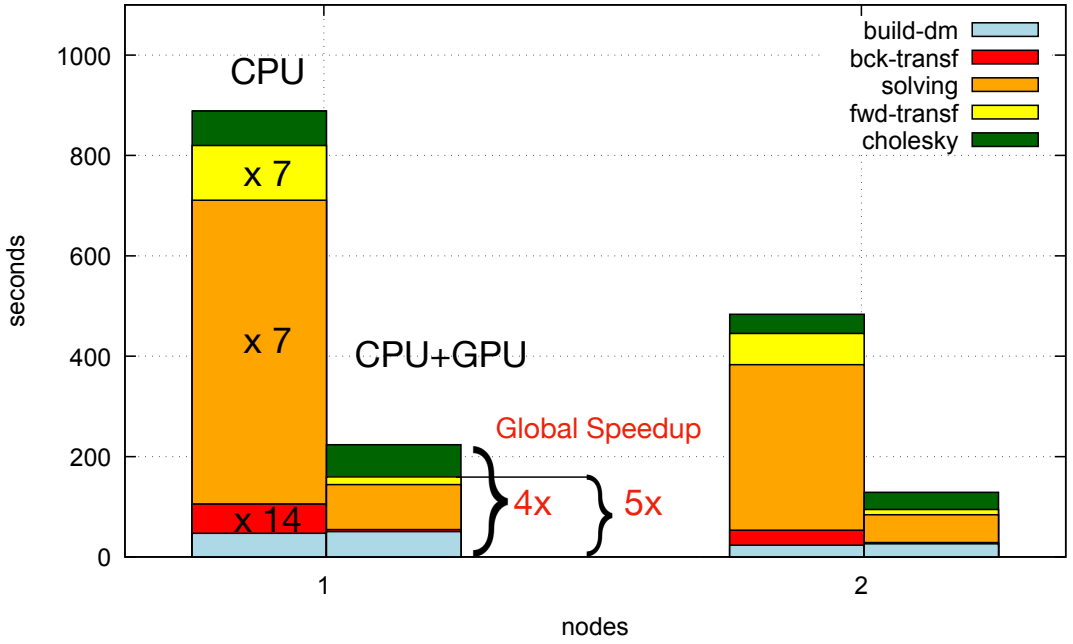
https://elsi-interchange.org

Interface in Siesta:

Collaboration with Victor Yu (Duke)

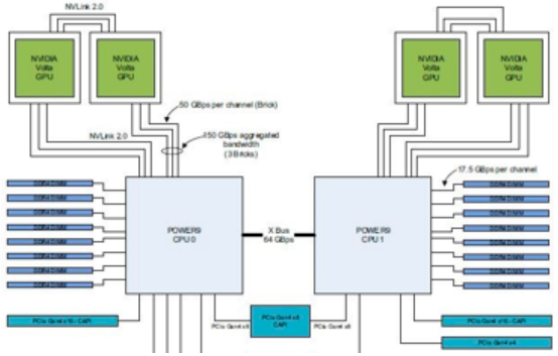
GPU acceleration for diagonalization

ELSI-ELPA GPU acceleration

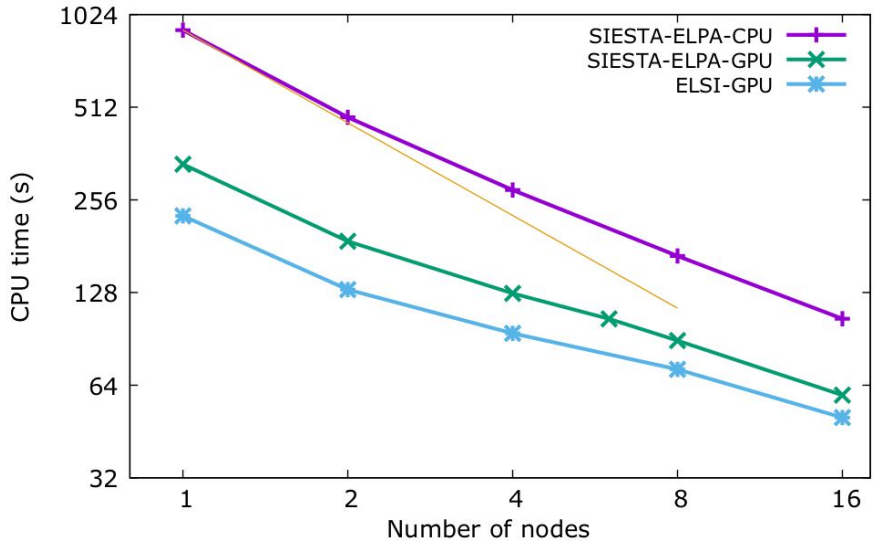


Future enhancements in ELPA (better kernels) and in ELSI (e.g. build-DM stage) are integrated in SIESTA automatically

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



Proper binding of GPUs to MPI ranks



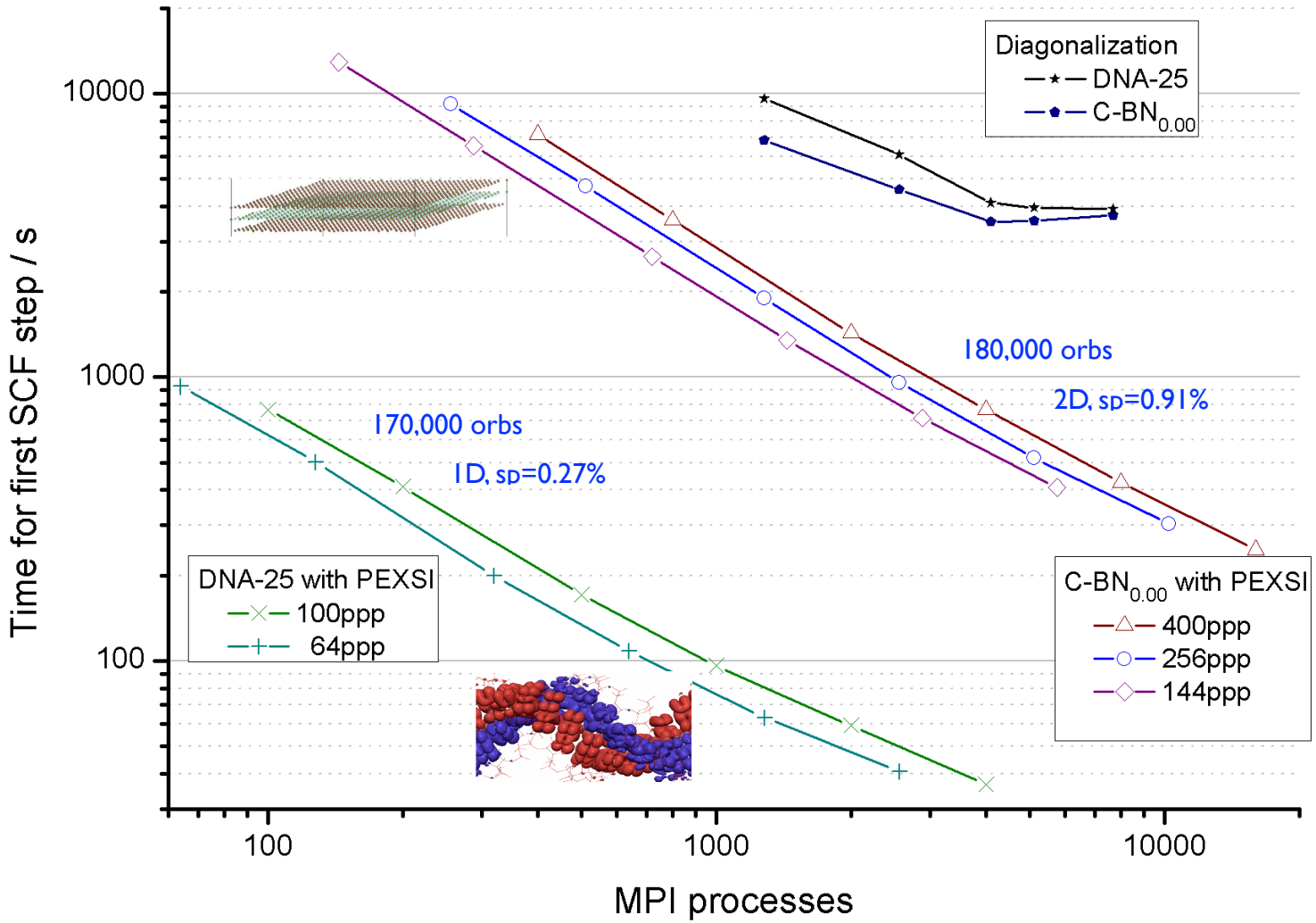
System: Si quantum dot, with approx 35000 orbs

Massive scalability: PEXSI solver

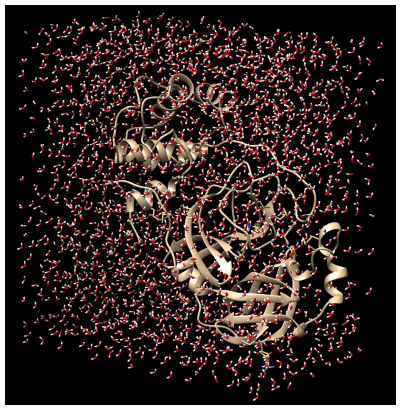
$$\hat{\rho} = \; Im\left(\sum_{l=1}^{P} \frac{\omega_l}{H-(z_l+\mu)S}\right)$$

PEXSI offers:

- Three levels of parallelization (over orbitals, poles, and chemical potential values)
- A reduced memory footprint (only sparse matrices are stored)
- Reduced complexity (maximum O(N^2) size scaling)

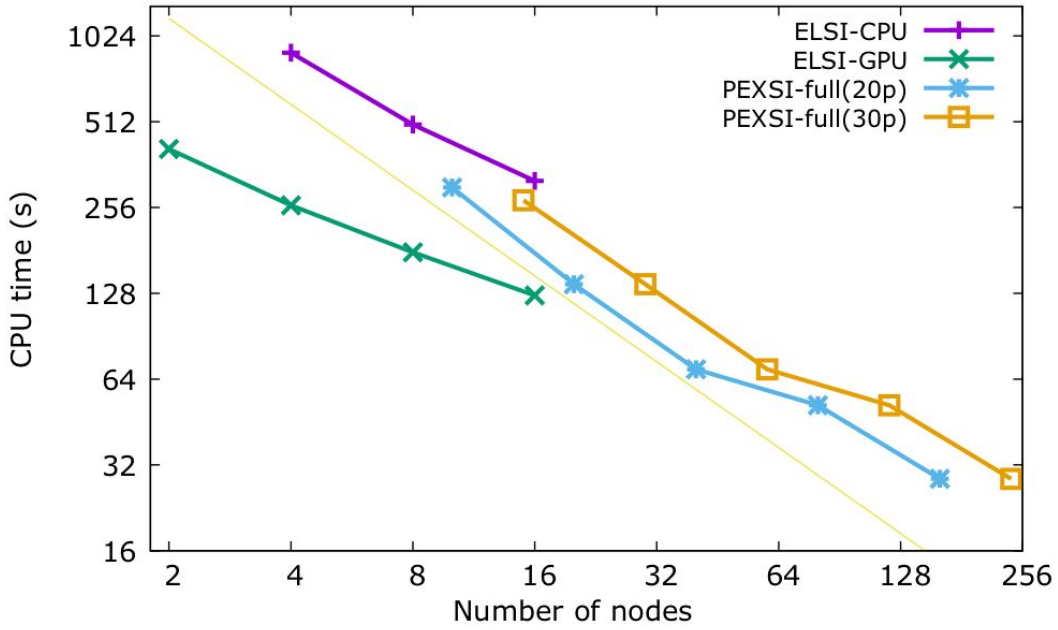


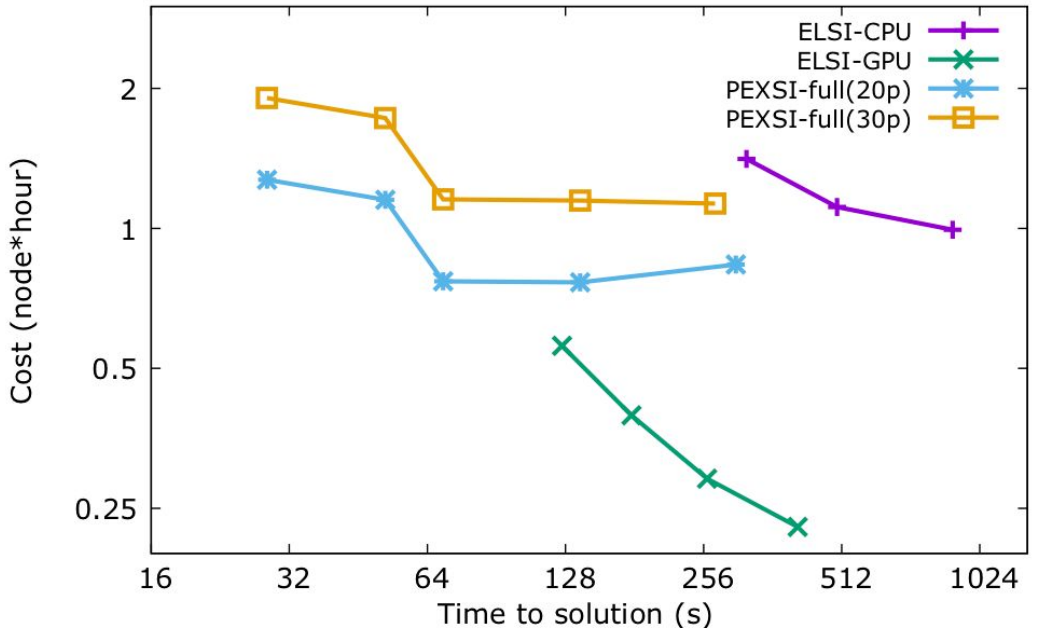
Comparison of global efficiency of solvers for a very large problem



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/