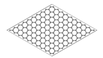
ELSI: A unified software interface for Kohn-Sham electronic structure solvers

Victor Wen-Zhe Yu[1], Alberto García[2], William Paul Huhn[1], Mathias Jacquelin[3], Weile Jia[3], Murat Keçeli[4], Raul Laasner[1], Yingzhou Li[5], Lin Lin[3,6], Jianfeng Lu[5], Álvaro Vázquez-Mayagoita[7], Chao Yang[3], Haizhao Yang[8], and Volker Blum[1]

[1] Department of Mechanical Engineering and Materials Science, Duke University; [2] Laboratorio de Estructura Electrónica de Materiales, Institut de Ciència de Materials de Barcelona; [3] Computational Research Division, Lawrence Berkeley National Laboratory; [4] Chemical Sciences and Engineering Division, Argonne National Laboratory; [5] Department of Mathematics, Duke University; [6] Department of Mathematics, University of California, Berkeley; [7] Argonne Leadership Computing Facility, Argonne National Laboratory; [8] Department of Mathematics, National University of Singapore

See also: Computer Physics Communications, 222, 267-285 (2018)

The O(N³) Kohn-Sham Eigenvalue Problem



Kohn-Sham Density-Functional Theory (DFT) [1]:

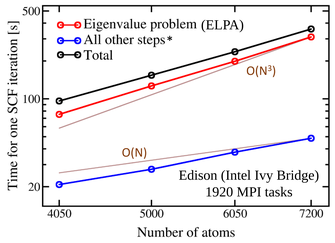
- Predictive theory for materials and molecules
- More than 30,000 scientific publications per year report results of DFT-based calculations and DFT method developments
- Frequently among the most used applications on the world's leading HPC/supercomputing resources

"Cubic Wall":
K.-S. Eigenvalue Problem - O(N³) with System Size N

Example system: Graphene supercell models

Electronic structure method: DFT-PBE [2], FHI-aims [3] all-electron simulation, "light" settings, ELPA [4,5] eigensolver for massively-parallel eigenvalue solutions.

Scalability limit: O(N³) solution of eigenvalue problem dominates over remaining operations (common bottleneck for semi-local K.-S. DFT).



*Key steps: evaluation of the Hartree potential, numerical integration of the Hamiltonian matrix, update of the electron density and its gradient.

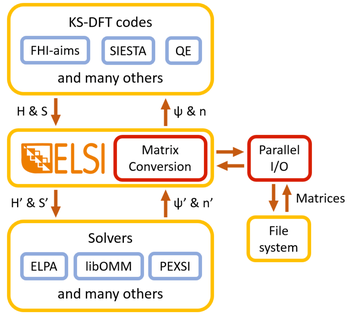
The ELSI Project: ELectronic Structure Infrastructure

ELSI provides an open platform to the electronic structure community for collaboration on solutions to the algebraic problems that are often the bottleneck of large K.-S. DFT simulations. Current features include:

- Interface for eigensolvers and density matrix solvers
- Solvers: ELPA, libOMM, PEXSI, SIPs
- Matrix formats: BLACS, PEXSI, SIESTA
- Programming interfaces: Fortran and C/C++
- Parallel matrix I/O

Features that will be available in future versions are:

- Reverse communication interface (RCI) for iterative eigensolvers
- Decision layer to automate solver selection
- Software bundle distribution with CECAM Electronic Structure Library (esl.cecam.org)



Solvers Supported in ELSI

Direct diagonalization: ELPA (Eigenvalue soLvers for Petaflop-Applications) [4,5] solves the K.-S. eigenvalue problem with dense linear algebra, including Cholesky transformation, reduction to banded and tridiagonal form, tridiagonal solver, and efficient back-transformations.

Iterative diagonalization: SIPs (Shift-and-Invert Parallel Spectral transformation, available as part of the SLEPc library) [6,7] partitions the eigenspectrum into a number of slices that can be solved independently by the iterative Krylov-Schur method. Slices are processed simultaneously, leading to a decent parallel efficiency. Additional iterative solvers, e.g., Davison and conjugate-gradient, are being implemented into ELSI via a reverse communication interface fashion.

Density matrix methods (circumventing explicit eigensolution): libOMM (Orbital Minimization Method) [8-10] computes the density matrix from a set of Wannier functions, which are obtained by directly minimizing an appropriately-constructed energy functional in the occupied subspace. PEXSI (Pole EXpansion and Selected Inversion) [11-13] efficiently evaluates certain selected elements of the Fermi-Dirac function of the Hamiltonian, yielding the density matrix. PEXSI exploits the sparsity of matrices to reduce the scaling to O(N)-O(N²), depending on the dimensionality of the system.

ELPA (eigensolution)
Dense, direct eigensolver
O(N³)

SIPs (eigensolution)
Sparse, iterative eigensolver
≤O(N²)

libOMM (density matrix)
Orbital minimization
O(N³), with reduced prefactor

PEXSI (density matrix)
Pole expansion, selected inversion
≤O(N²)

ELSI Application Programming Interface (API) Overview

Example: ELSI used in typical SCF calculations

elsi_init

Choose solver, input matrix format, and parallel mode.

elsi_set_mpi

Pass MPI information to ELSI.

elsi_set_{blacs|csc}

Set up dense or sparse matrix storage in ELSI.

SCF loop
  user code prepares Hamiltonian matrix
  elsi_{dm|ev}_{real|complex}_{dense|sparse}
  Compute the density matrix or eigenvalues/eigenvectors.
  user code checks convergence criterion
  exit loop if converged
end SCF loop

elsi_finalize

Finalize ELSI interface.

- Designed for rapid integration into a variety of electronic structure codes
- Supports:
  - Eigensolvers and density matrix solvers
  - Real and complex matrices
  - Dense and sparse matrix formats
  - 1D and 2D data distributions
  - Periodic/non-periodic and spin-polarized/non-spin-polarized calculations
- Offers full flexibility to tune solver settings and get additional outputs
- New solvers may be easily added

Using ELSI as the interface layer, successful connection has been established between the solvers supported in ELSI and various electronic structure codes (DFTB+ [14], DGDFT [15], FHI-aims, and SIESTA [16] as first steps).

Benchmark on High-Performance Computing (HPC) Facilities

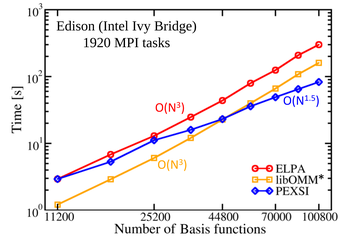
Benchmark system: Graphene supercell models ranging from 800 to 7200 atoms.

Electronic structure method: DFT-PBE, FHI-aims all-electron simulation, "light" settings.

- O(N³) methods are in general more suitable for small-and-medium-sized calculations.
- PEXSI outperforms ELPA and libOMM in (extra-)large-scale calculations due to its lower scaling exponent, i.e., O(N) for 1D, O(N^1.5) for 2D, O(N²) for 3D systems.

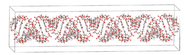
Portability

Compilation and execution of ELSI and the solvers have been demonstrated with Cray, GNU, IBM, Intel, and PGI compilers on a broad list of architectures, including Cori and Theta (Intel Knights Landing), Edison (Intel Ivy Bridge), Mira (IBM BlueGeneQ), Summitdev (IBM Power 8 + NVIDIA P100), Titan (AMD Bulldozer + NVIDIA K20X), and many computer clusters and laptops around the world.



*libOMM timings correspond to the optimal case where the orbital minimization can converge in one iteration, e.g. near the end of an SCF cycle.

Connection to Electronic Structure Codes

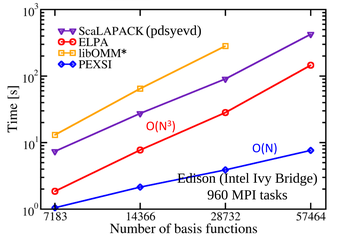


The ELSI interface is specifically designed so that it can be easily inserted into the workflow of SCF calculations in electronic structure codes. One recent success is the integration of ELSI into the SIESTA code (with only about 100 lines of code added), providing SIESTA users a simple and unified access to the solvers supported in ELSI.

Benchmark system: DNA supercell models ranging from 715 to 5720 atoms [13].

Electronic structure method: DFT-PBE, SIESTA, DZP basis, Troullier-Martins norm-conserving pseudopotentials.

- ELPA offers faster eigensolutions compared with the pdsyevd solver in ScaLAPACK.
- Matrices generated from 1D systems are highly sparse, favoring PEXSI over eigensolvers.
- ELPA, libOMM, and PEXSI timings also include the time spent on matrix redistributions, which can be done within around 1% of the computation time.



*libOMM timings correspond to the total time to converge the minimization.

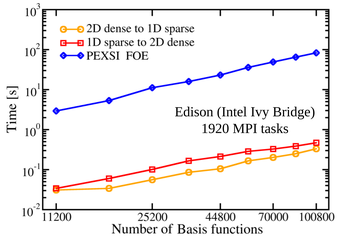
Matrix Format Conversion and Redistribution

Supported Matrix Format

- 2D block-cyclic distributed dense (BLACS)
- 1D block-cyclic distributed sparse CSC (SIESTA)
- 1D block distributed sparse CSC (PEXSI)
- CSR implicitly available due to symmetry
- Additional formats per user's request

Parallel matrix format conversion and redistribution are performed by ELSI automatically, powering seamless interfacing between electronic structure codes and solvers.

- Only non-zero matrix elements transferred by MPI communication
- Information regarding the sparsity pattern reused throughout the whole SCF cycle
- Generic matrix redistribution routines can be quickly extended for new formats

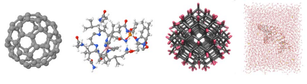


Matrix redistribution performance in comparison with the Fermi operator expansion (FOE) in PEXSI. Time spent on redistribution is about 1% of time spent on PEXSI computation.

The ELSI "Decision Layer": Automating Solver Selection

Selection of the "best" solver should consider:

- **Basis set:** Planewave vs. localized orbitals (size, sparsity, and condition number of the matrices)
- **Eigenspectrum:** All-electron vs. pseudopotential (width and distribution of the spectrum)
- **Physical property:** Insulator vs. semiconductor vs. metal (band gap)
- **Desired output:** Eigenvalues, eigenvectors, density matrix, energy-weighted density matrix, etc
- **Computer:** Compilers and external libraries, number of parallel tasks, available accelerators, etc



Proposal for the ELSI Decision Layer

Benchmark → An ELSI "meta-benchmark set" consisting of structures used in performance benchmarks for 20 electronic structure codes and solvers.

Profile → Benchmark with different codes on different architectures for a time-to-solution profile.

Decision → Extract information that can be used to estimate the "best" solver for new problems.

References and Acknowledgments

[1] Kohn and Sham, Phys Rev, 140: A1133-1138 (1965).
[2] Perdew et al, Phys Rev Lett, 77: 3865-3868 (1996).
[3] Blum et al, Comput Phys Commun, 180: 2175-2196 (2009).
[4] Auckenthaler et al, Parallel Computing, 37: 783-794 (2011).
[5] Marek et al, J Phys Condens Matter, 26: 213201 (2014).
[6] Hernandez et al, ACM Trans Math Softw, 31: 351-362 (2005).
[7] Keceli et al, J Comput Chem, 37: 448-459 (2016).
[8] Mauri et al, Phys Rev B, 47: 9973 (1993).
[9] Ordejon et al, Phys Rev B, 51: 1456 (1995).
[10] Corsetti, Comput Phys Commun, 185: 873-883 (2014).
[11] Lin et al, Commun Math Sci, 7: 755 (2009).
[12] Lin et al, J Phys Condens Matter, 25: 295501 (2013).
[13] Lin et al, J Phys Condens Matter, 26: 305503 (2014).
[14] Aradi et al, J Phys Chem A, 111: 5678 (2007).
[15] Hu et al, J Chem Phys, 143: 124110 (2015).
[16] Soler et al, J Phys Condens Matter, 14: 2745-2779 (2002).

This work is supported by the National Science Foundation under grant number 1450280. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors gratefully acknowledge the following organizations:

Argonne National Lab
Oak Ridge National Lab