Collaborative Research: SI2-SSI: ELSI-Infrastructure for Scalable Electronic Structure Theory

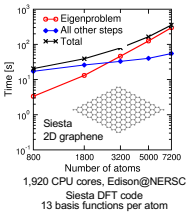
Award #: 1450280
1450372

PIs: Volker Blum[1], Lin Lin[2,3] Co-PI: Jianfeng Lu[1] Senior Personnel: Chao Yang[2], Álvaro Vázquez-Mayagoitia[4], Fabiano Corsetti[5]
Coauthors: Carmen Campos[6], William Dawson[7], Alberto García[8], Ville Havu[9], Ben Hourahine[10], William P. Huhn[4], Mathias Jacquelin[2], Weile Jia[2], Murat Keçeli[4], Raul Laasner[1], Yingzhou Li[1], Jonathan Moussa[11], Jose E. Roman[6], Victor Wen-zhe Yu[1]

[1]Duke University, USA, [2]Lawrence Berkeley National Laboratory, USA, [3]University of California, Berkeley, USA, [4]Argonne National Laboratory, USA, [5]Synopsys QuantumWise, Denmark, [6]Universitat Politècnica de València, Spain, [7]RIKEN Center for Computational Science, Japan, [8]Institut de Ciència de Materials de Barcelona (ICMAB-CSIC), Spain, [9]Aalto University, Finland, [10]University of Strathclyde, UK, [11]Molecular Sciences Software Institute, USA

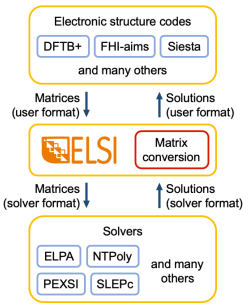
The O(N³) Kohn-Sham Eigenvalue Problem

- Kohn-Sham density-functional theory (KS-DFT) [1]: Workhorse for molecular and materials simulations
- Large number of software implementations
- Running on world's leading HPC resources
- O(N³) complexity of a dense eigensolver
- O(N) in almost all other operations with localized basis sets
- Common bottleneck for semi-local KS-DFT and electronic structure theory in general



ELSI: ELectronic Structure Infrastructure

ELSI [2,3] (https://elsi-interchange.org) provides an integrated software interface to state-of-the-art eigensolvers and density matrix solvers, facilitating large-scale electronic structure simulations.



- Eigensolvers (ELPA [4,5], EigenExa [6], SLEPc-SIPs [7,8], LAPACK [9], MAGMA [10]) and density matrix solvers (libOMM [11], NTPoly [12], PEXSI [13-15]) via a common API
- Reverse communication interface (RCI) for iterative eigensolvers
- Matrix formats: Dense and sparse matrix in 1D/2D block-cyclic or arbitrary distribution, with parallel interconversions in between
- Parallel solution for spin-polarized and/or periodic systems
- Adopted by DFTB+ [16], DGDFT [17], FHI-aims [18], and Siesta [19] codes

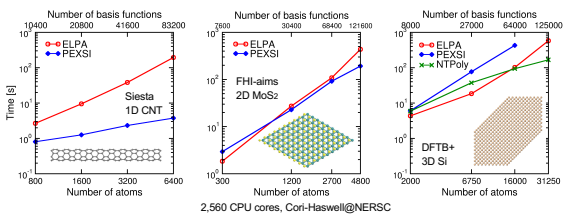
ELSI is included in the software bundle distribution of the CECAM Electronic Structure Library project (https://esl.cecam.org)

Application Programming Interface (API)

```
elsi_init
elsi_set_parameters
do geometry loop
   do SCF loop
      elsi_{ev|dm}
   end SCF loop
   elsi_reinit
end geometry loop
elsi_finalize
```

- Designed for rapid integration into a variety of electronic structure codes
- Compatible with common workflows
  - Self-consistent field (SCF)
  - Multiple SCF cycles (geometry relaxation or molecular dynamics)
- Supports density matrix solvers and eigensolvers on equal footing
- All technical settings adjustable for experienced users

Performance Benchmarks



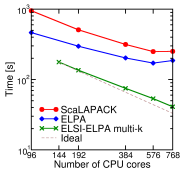
- ELPA eigensolver O(N³): Small-to-medium-sized systems
- PEXSI density matrix solver ≤ O(N²): Large, low-dimensional systems
- NTPoly density matrix solver O(N): Large systems with an energy gap

References

[1] Kohn and Sham, Phys. Rev., 140: A1133-1138 (1965).
[2] Yu et al., Comput. Phys. Commun., 222: 267-285 (2018).
[3] Yu et al., arXiv: 1912.13403 (2019).
[4] Auckenthaler et al., Parallel Computing, 37: 783-794 (2011).
[5] Marek et al., J. Phys. Condens. Matter, 26: 213201 (2014).
[6] Imamura et al., Prog. Nucl. Sci. Technol., 2: 643-650 (2011).
[7] Hernandez et al., ACM Trans. Math. Softw., 31: 351-362 (2005).
[8] Kecelli et al., J. Comput. Chem., 37: 448-459 (2016).
[9] Anderson et al., LAPACK users' guide (1999).
[10] Tomov et al., Parallel Computing, 36: 232-240 (2010).
[11] Corsetti, Comput. Phys. Commun., 185: 873-883 (2014).
[12] Dawson and Nakajima, Comput. Phys. Commun., 225: 154-165 (2018).
[13] Lin et al., Commun. Math. Sci., 7: 755 (2009).
[14] Lin et al., J. Phys. Condens. Matter, 25: 295501 (2013).
[15] Jia and Lin, J. Chem. Phys., 147: 144107 (2017).
[16] Aradi et al., J. Phys. Chem. A, 111: 5678 (2007).
[17] Hu et al., J. Chem. Phys., 143: 124110 (2015).
[18] Blum et al., Comput. Phys. Commun., 180: 2175-2196 (2009).
[19] Soler et al., J. Phys. Condens. Matter, 14: 2745-2779 (2002).
[20] Davidson, J. Comput. Phys., 17: 87-94 (1975).
[21] Vecharynski et al., J. Comput. Phys., 290: 73-89 (2015).
[22] Zhou et al., J. Comput. Phys., 219: 172-184 (2006).

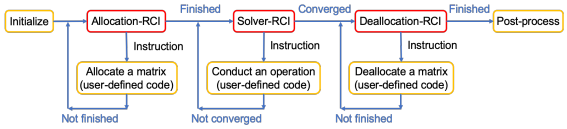
Connection to Electronic Structure Code

- Performance improvement from the use of the extra level of parallelization over *k*-points in Siesta using the ELSI interface, compared to the existing diagonalization schemes in Siesta (standard ScaLAPACK and ELPA)
- Bulk Si with H impurities (1,040 atoms, 13,328 basis functions, 8 *k*-points)



RCI Framework for Iterative Eigensolvers

Iterative eigensolvers are supported in ELSI now, through a reverse communication interface (RCI) framework that separates linear algebra operations from the core iterative algorithms. ELSI-RCI will have use cases in, e.g., DFT codes with planewave basis sets.



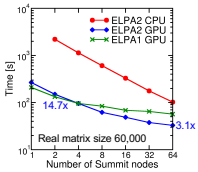
Methods implemented:

- Davidson [20]
- Orbital minimization method (OMM) [10]
- Projected preconditioned conjugate gradient (PPCG) [21]
- Chebyshev filtering [22]

ELPA Developments

A proof-of-concept version of the GPU ELPA2 eigensolver (two-stage tridiagonalization) has been completed, with enhanced robustness, efficiency, and scalability on distributed-memory (multi-nodes) hybrid CPU-GPU architectures.

- Added a CUDA kernel for Householder transformations in the computation of eigenvectors
- Eliminated a restriction on the block size of the distributed matrices
- Optimized GPU memory usage and CPU-GPU communication



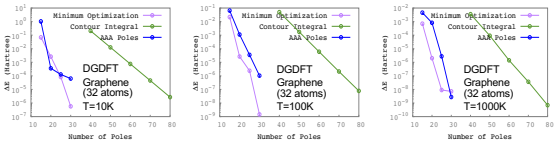
1 node = 42 CPU cores + 6 GPUs
Summit@Oak Ridge

PEXSI Developments

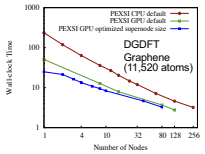
The pole expansion and selected inversion (PEXSI) method expands the density matrix as a sum of rational matrix functions ("poles"). A new set of poles based on the "adaptive Antoulas-Anderson" (AAA) method is now available in PEXSI. Key advantages are:

- Compared to poles based on minimax optimization, density matrix, energy density matrix, and free energy density matrix can be calculated altogether
- Compared to poles based on contour integral, the number of poles is greatly reduced without sacrificing accuracy

Shown below are total energy differences compared to diagonalizaton.



Work is in progress to port the parallel selected inversion algorithm in PEXSI to GPUs. The strong scaling of the GPU-accelerated selected inversion (single pole) on the Summit supercomputer is shown on the right. A speedup up to 10x is observed for a 11,520-atom test system.



1 node = 42 CPU cores + 6 GPUs
Summit@Oak Ridge

Acknowledgments

NSF, MOLSSI, OAK RIDGE National Laboratory, NERSC, Argonne National Laboratory, MPCDF, Electronic Structure Library, CECAM, NVIDIA