Solving or Circumventing Eigenvalue Problems in Electronic Structure Theory

August 15-17, 2018
Richmond, Virginia

Day 1 (Wed, August 15)

Opening

08:00-08:55 Registration and Breakfast
08:55-09:00 Welcoming Remarks

Morning Session 1

09:00-09:25 Teresa Head-Gordon (University of California at Berkeley)
Reducing and Eliminating Self-Consistent Field Calculations in Classical and Ab Initio Simulations
09:25-09:50 Bert de Jong (Lawrence Berkeley National Laboratory)
NWChemEx, Current State and Solver Needs
09:50-10:15 Stefano de Gironcoli (SISSA)
Comparing Efficiency of Iterative Eigenvalue Solvers: The Quantum ESPRESSO Experience
10:15-10:50 Coffee Break

Morning Session 2

10:50-11:15 Victor Yu (Duke University)
Recent Developments in the ELSI Infrastructure for Large-Scale Electronic Structure Theory
11:15-11:40 Micael Oliveira (Max Planck Institute for the Structure and Dynamics of Matter)
The CECAM Electronic Structure Library: Past, Present and Future
11:40-12:05 Ben Hourahine (University of Strathclyde)
Experiences with Self-Consistent Tight Binding and ELSI
12:05-01:30 Lunch

Afternoon Session

01:30-01:55 Bruno Lang (University of Wuppertal)
Recent Algorithmic Developments in ELPA
01:55-02:20 Christian Carbogno (Fritz Haber Institute of the Max Planck Society)
Recent Advancements in the ELPA Library: Best Practices in Real Applications
02:20-02:45 Eric Polizzi (University of Massachusetts)
FEAST v4.0 with Applications to Electronic Structure
02:45-03:10 Coffee Break

Hands-On Discussion

03:10-05:00 Hands-On Discussion
05:00-06:00 Break

Evening

06:00- Poster Session

Day 2 (Thu, August 16)

Opening

08:00-09:00 Breakfast

Morning Session 1

09:00-09:25 Edmond Chow (Georgia Institute of Technology)
Exact Density Matrix Purification for Parallel Computations
09:25-09:50 Jack Deslippe (Lawrence Berkeley National Laboratory)
Eigenvalue Problems in Many Body Perturbation Theory
09:50-10:15 Jack Poulson (Google)
Equilibrating Alternating Methods for Low-Rank Approximations with Gaussian Priors via Conic Automorphisms
10:15-10:50 Coffee Break

Morning Session 2

10:50-11:15 Weile Jia (University of California at Berkeley)
Robust Determination of Chemical Potential in the Pole Expansion Method and Fast Time Dependent Density Functional Theory Calculations
11:15-11:40 Carolin Penke (Max Planck Institute for Dynamics of Complex Technical Systems)
Opportunities for ELPA to Accelerate the Solution of the Bethe-Salpeter Eigenvalue Problem
11:40-12:05 Mathias Jacquelin (Lawrence Berkeley National Laboratory)
Task-Based Left-Looking Selected Inversion Algorithm on Shared Memory Systems
12:05-01:30 Lunch

Afternoon Session

01:30-01:55 Jeongnim Kim (Intel)
Empowering Domain- and Application-Specific Libraries on Intel Platforms
01:55-02:20 Nadezhda Reshetnikova (Intel)
Recent Development of Intel MKL Library
02:20-02:45 Piotr Luszczek (University of Tennessee)
Eigenvalue Solvers Considerations on Modern HPC Hardware Platforms
02:45-03:10 Hong Guo (McGill University)
Computational Materials Science by RESCU - a KS-DFT Method for Solving Thousands of Atoms
03:10-03:40 Coffee Break

Hands-On Discussion

03:40-05:30 Hands-On Discussion
05:30-06:00 Break

Evening

06:00- Conference Dinner

Day 3 (Fri, August 17)

Opening

08:00-09:00 Breakfast

Morning Session 1

09:00-09:25 Chris-Kriton Skylaris (University of Southampton)
Electronic Energy Minimization Methods in the ONETEP Program
09:25-09:50 Amartya Banerjee (Lawrence Berkeley National Laboratory)
Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems
09:50-10:15 Jean-Luc Fattebert (Oak Ridge National Laboratory)
Finite Difference DFT Solver: Direct Functional Minimization with Eigensolver in Projected Subspace
10:15-10:50 Coffee Break

Morning Session 2

10:50-11:15 Murat Keceli (Argonne National Laboratory)
SLEPc-SIPs: Massively Parallel Sparse Eigensolver for Electronic Structure Calculations
11:15-11:40 Yunkai Zhou (Southern Methodist University)
Accelerating Eigenvalue Calculations: Shift-Without-Invert and Shift-with-Invert Techniques for Spectrum-Partition
11:40-12:05 Yingzhou Li (Duke University)
Reverse Communication Interface for Iterative Eigensolvers in ELSI
12:05-01:30 Lunch

Afternoon Session

01:30-01:55 Toshiyuki Imamura (RIKEN)
Development of a Dense Eigenvalue Solver for Exa-Scale Systems
01:55-02:20 Nick Papior (Technical University of Denmark)
Bigger and Better: Large Scale NEGF Calculations and Coupling DFT with TB
02:20-02:45 Ayako Nakata (National Institute for Materials Science)
Eigenstate-Analysis using Sakurai-Sugiura Method with O(N)-DFT Code CONQUEST
02:45-03:10 William Huhn (Duke University)
GPU-Accelerated Real Space Electronic Structure Theory on HPC Resources
03:10-04:00 Close

Talks

Reducing and Eliminating Self-Consistent Field Calculations in Classical and Ab Initio Simulations

Teresa Head-Gordon[1-5]

[1]Kenneth S. Pitzer Center for Theoretical Chemistry, [2]Department of Chemistry, [3]Department of Chemical and Biomolecular Engineering, [4]Department of Bioengineering, University of California Berkeley

I will discuss new extended Lagrangian methods that address accuracy and tractability for using atomistic polarizable force fields and DFT in the condensed phase. These models and their implementations are opening up new abilities for allowing larger scales of study for molecular simulation with more complex potential energy surfaces.

[1] Albaugh, A. M. N. Niklasson, T. Head-Gordon (2018). Higher-order Extended Lagrangian Born-Oppenheimer Molecular Dynamics for Classical Polarizable Models. J. Chem. Theory Comput, 14 (2), 499–511.
[2] A. Albaugh, M. Tuckerman, and T. Head-Gordon (2018). Combining an SCF-less solution to many-body polarization with stochastic isokinetic integration. *In preparation.*
[3] A. Albaugh and T. Head-Gordon (2017). A new method for treating Drude polarization in classical molecular simulation. *J. Chem. Theory Comput.* 13(11):5207-5216.
[4] A. Albaugh, A. Niklasson, and T. Head-Gordon (2017). Accurate classical polarization solution with no self-consistent field iterations. *J. Phys. Chem. Lett.* 8, 1714–1723
[5] A. Albaugh, O. N. Demerdash, and T. Head-Gordon (2015). An efficient and stable hybrid extended lagrangian/self-consistent field scheme for solving classical mutual induction. *J. Chem. Phys.* 143, 174104

NWChemEx, Current State and Solver Needs

W.A. de Jong[1]

[1]*Lawrence Berkeley National Laboratory*

NWChem has sustained its ability to exploit powerful parallel supercomputers for over 25 years, primarily through dense high-accuracy many-body numerical simulations. In recent years, high-performance computing platforms have undergone radical changes, think GPUs and alike, with the advent of radically different exascale architectures just around the corner. Computational chemistry has also transitioning, moving towards broadly adopting sparse and reduced scaling methodologies.

To better address the challenges presented by emerging exascale architectures, take advantage of modern software development technologies and properly integrate sparse techniques, NWChem is being redesigned and reimplemented from scratch. The resulting package, called NWChemEx, will be designed to optimally exploit any computing platform with an emphasis on sxascale computing systems. This new C++17 based software platform has been designed to provide a community infrastructure for computational chemistry that is scalable, flexible, and portable and supports a broad range of chemistry research on a broad range of computing systems.

In this talk I will briefly discuss the NWChemEx software development, funded by the DOE ASCR Exascale Computing Project. I will give insights into the science driving the development efforts, the new software infrastructure and the changing needs in solvers.

Comparing Efficiency of Iterative Eigenvalue Solvers: the Quantum ESPRESSO experience

Stefano de Gironcoli[1]

[1]*SISSA, Trieste (Italy)*

The iterative determination of a small fraction of the lowest-lying eigenvalue-eigenvector pairs of large sparse matrices is a fundamental task in modern electronic structure applications that needs to be addressed efficiently in terms of time-to-solution and memory requirements.

A number of iterative eigensolvers are implemented in the Quantum ESPRESSO [1] suite of codes, including the block Davidson [2] diagonalization, a simple band-by-band conjugate gradient (CG) solver, a Parallel Orbital-updating (ParO) [3] approach, and the Projected Preconditioned Conjugate-Gradient (PPCG) [4] method.

I will discuss the merits and problems of the different methods in view of their performance in mixed MPI/OpenMP or hybrid CPU/GPU architectures.

[1] www.quantum-espresso.org
[2] ER Davidson, *J of Comput Phys* 17 (1975) 87-94.
[3] Y Pan, XY Dai, S de Gironcoli, XG Gong, GM Rignanese, AH Zhou, *J of Comput Phys* 348 (2017) 482-492.
[4] E Vecharynski, C Yang, JE Pask, *J of Comput Phys* 290 (2015) 73-89.

Recent Developments in the ELSI Infrastructure for Large-Scale Electronic Structure Theory

Victor Yu[1], William Dawson[2], Alberto García[3], Ville Havu[4], Ben Hourahine[5], William Huhn[1], Mathias Jacquelin[6], Weile Jia[6,7], Murat Keçeli[8], Raul Laasner[1], Yingzhou Li[9], Lin Lin[6,7], Jianfeng Lu[9], Jose Roman[10], Álvaro Vázquez-Mayagoita[8], Chao Yang[6], and Volker Blum[1]

[1]Department of Mechanical Engineering and Materials Science, Duke University, United States; [2]RIKEN Center for Computational Science, Japan; [3]Institut de Ciència de Materials de Barcelona, Spain; [4]Department of Applied Physics, Aalto University, Finland; [5]Department of Physics, University of Strathclyde, Scotland; [6]Computational Research Division, Lawrence Berkeley National Laboratory, United States; [7]Department of Mathematics, University of California, Berkeley, United States; [8]Argonne Leadership Computing Facility, Argonne National Laboratory, United States; [9]Department of Mathematics, Duke University, United States; [10]Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Spain

In large-scale molecular and material simulations based on Kohn-Sham density-functional theory (KS-DFT), solving or circumventing eigenvalue problems is often the major computational bottleneck, which limits the achievable system size to roughly several thousand atoms. This is a generic problem that must be addressed by essentially all current electronic structure codes. Conventional KS-DFT implementations solve the eigenvalue problems by direct or iterative diagonalization, using algorithms whose computational complexity scales cubically with respect to the system size. There exist alternative algorithms that circumvent the explicit solution of an eigenvalue problem by, e.g., directly computing the density matrix. With a smaller scaling exponent and a larger prefactor, these methods can potentially outperform diagonalization for systems beyond a thousand atoms. We present an open-source software interface, ELSI (http://elsi-interchange.org) [1], to simplify the access to existing strategies to address the KS eigenvalue problem for different problem classes on different scales. Currently supported algorithms are the massively parallel dense eigensolver ELPA [2], the orbital minimization method implemented in libOMM [3], the pole expansion and selected inversion method PEXSI [4], the shift-and-invert parallel spectral transformation eigensolver implemented in SLEPc [5], and the linear scaling density matrix purification methods implemented in NTPoly [6]. The ELSI interface aims to simplify the use of these methods by providing (a) a unified platform for implementing, testing, and optimizing the solvers, (b) reasonable default settings for a chosen solver, (c) automatic conversion between input and internal working matrix formats, and (d) suggestions on the optimal solver for a given problem. Comparative benchmarks performed on distributed memory supercomputing architectures are presented for system sizes up to ten thousand atoms with KS-DFT, and a hundred thousand atoms with density-functional tight-binding theory. The strengths and limitations of the solvers will be discussed in detail.

[1] V. Yu et al., *Computer Physics Communications* 222, 267-285 (2018).
[2] A. Marek et al., *Journal of Physics: Condensed Matter* 26, 213201 (2014).
[3] F. Corsetti, *Computer Physics Communications* 185, 873-883 (2014).
[4] L. Lin et al., *Journal of Physics: Condensed Matter* 25, 295501 (2013).
[5] M. Keceli et al., *Journal of Computational Chemistry* 37, 448-459 (2016).
[6] W. Dawson and T. Nakajima, *Computer Physics Communications* 225, 154 (2018).

The CECAM Electronic Structure Library: Past, Present and Future

Micael J.T. Oliveira[1] and all the ESL contributors[2]

[1]Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany
[2]esl.cecam.org

With the goal of providing a comprehensive collection of software libraries and tools, the CECAM Electronic Structure Library (ESL) [1,2], a community-driven effort, has allowed developers of electronic structure codes to share the burden of developing and maintaining complex pieces of software. In this talk I will present the different aspects and aims of the ESL and how these can be useful for the electronic structure community. I will review how the library developed since its inception in 2014 and present the current status of the different projects that are part of it. I will also present the ESL Bundle, a key component of the library, and the ESL development road map for the next couple of years.

[1] http://esl.cecam.org
[2] https://gitlab.e-cam2020.eu/esl

Experiences with self-consistent tight binding and ELSI

Ben Hourahine[1]

[1]*SUPA, The Department of Physics, University of Strathclyde, John Anderson Building, 107 Rottenrow Glasgow UK*

DFTB$^+$ [1] is an implementation of the Density Functional Tight Binding (DFTB) method, which approximates full density functional theory (DFT) in the limit of a two center hamiltonian in a non-orthogonal minimal basis. DFTB is derived from an input only Kohn-Sham functional [2] and was originally for non-self consistent solutions [3,4], but other extensions derived from DFT including charge self-consistency [5] and spin polarization [6] have been included into the method. DFTB$^+$ evaluates most of its operations with a sparse compressed block matrix format [1] which is designed to treat molecular and periodic (as well as other) boundary conditions on an equal footing in most of the code. The two major exceptions in conventional ground state calculations are for the linear algebra associated with the generalized eigenvalue problem and, when necessary, the electrostatics of the boundary conditions. In the former case this requires dense linear algebra for conventional solvers, traditionally using either LAPACK or ScaLAPACK [7], with calculations of up to ~20,000 atom systems being demonstrated with sufficient parallel resources.

This presentation describes some of the recent developments in DFTB+ for the forthcomming 18.2 release. Experiences in interfacing with the ELSI [8] solver collection, demonstrating substantial performance improvements, will also be described.

[1] B. Aradi, B. Hourahine and Th. Frauenheim, J. Phys. Chem A *111,* 5678-5684 (2007).
[2] W. M. C. Foulkes and R. Haydock, Phys. Rev. B 39, 12520-12536 (1989).
[3] G. Seifert, H. Eschrig, W. Bieger, Z. Phys. Chem. 267, 529–539 (1986).
[4] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert and R. Kaschner, *Phys. Rev. B* 51, 12947-12957 (1995).
[5] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, Phys. Rev. B 58, 7260-7268 (1998).
[6] C. Kohler, G. Seifert, U. Gerstmann, M. Elstner, H. Overhof, and T. Frauenheim, Phys. Chem. Chem. Phys. 3, 5109-5114 (2001).
[7] DFTB+ 18.1 http://dx.doi.org/10.5281/zenodo.1186900
[8] V. Yu et al., Comput. Phys. Commun. 222, 267 (2018).

Recent algorithmic developments in ELPA

B. Lang

Mathematics and Natural Sciences, University of Wuppertal

The ELPA-AEO project is aimed at adding new features (concerning, e.g., performance and robustness) and functionality to the widely used ELPA library [1] that provides direct methods for moderately-sized standard and generalized eigenvalue problems (SEPs and GEPs, resp.). In this talk we describe efficient new algorithms for the matrix multiplications that occur in the reduction of GEPs $Ax = Bx\lambda$ to equivalent standard SEPs $Cz = z\lambda$. Together with the corresponding back-transformation of the eigenvectors, this reduction is an established component in direct GEP solvers for full matrices, such as available with ELPA. We also report on the development of a reduction routine for banded GEPs. Finally, we briefly point to recent and upcoming ELPA features that enable making the most efficient use of the library.

[1] A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, and H. Lederer., *J. Phys.: Condens. Matter* 26(21), 213201 (2014).

Recent Advancements in the ELPA Library: Best Practices in Real Applications

Danilo Simoes Brambila[1] and Christian Carbogno[1]

[1]*Fritz Haber Institute of the Max Planck Society, Berlin, Germany*

In the ELPA-AEO project [1], a consortium of mathematicians, computer scientist, software designers, as well as first-principles code-developers from chemistry, solid-state physics and material science have teamed up to improve and accelerate the ELPA library [2], which provides highly efficient and highly scalable direct eigensolvers. This includes various new functionalities that facilitate the numerical efficient usage of ELPA, especially for non-expert users. In particular, this required to define and develop a new GET/SET based API that deviates from the traditional LAPACK/SCALAPACK-type interface used in earlier releases.

In this talk, we exemplify these advancements with our recent implementation of this GET/SET API in the ELSI package [3]. For this purpose, we also showcase the impact of the individual ELPA features in real-world applications from chemistry and solid-state physics using large-scale first-principles calculations performed with *FHI-aims* using the aforementioned ELSI/ELPA interface. In particular, this includes the use of *mixed-precision* [4], i.e., switching from single- to double precision routines during one SCF cycle, and the use of *autotuning* to automatically identify the best possible numerical solver settings and code paths during actual production calculations. Eventually we discuss the performance of the new ELPA releases on new architectures, including many-core AVX512 systems and GPU-accelerated installations.

[1] ELPA-AEO: http://elpa-aeo.mpcdf.mpg.de
[2] A. Marek, *et al.*, *J. Phys. Condens. Mat.* 26, 213201 (2014).
[3] V. W.-Z. Yu, *et al.*, *Comp. Phys. Comm.* 222, 267 (2018).
[4] A. Alvermann, *et al.*, *Japan J. Indust. Appl. Math.* (submitted).

FEAST v4.0 with Applications to Electronic Structure

E. Polizzi

Department of Electrical and Computer Engineering,
Department of Mathematics and Statistics,
University of Massachusetts, Amherst

FEAST is a general purpose eigenvalue solver which takes its inspiration from the density-matrix representation and contour integration technique in quantum mechanics [1]. The algorithm gathers key elements from complex analysis, numerical linear algebra and approximation theory, and it can be interpreted as a generalization of shift-and-invert iterations that uses multiple shifts in the complex plane leading to an optimal filter projector [2]. Once a given search domain is selected, FEAST's main computational task consists of a numerical quadrature computation that involves solving independent linear systems along a complex contour. FEAST offers a set of appealing features including remarkable robustness with well-defined convergence rate, and the capability to exploit natural parallelism at three different levels: (i) search domains can be treated separately (no overlap), (ii) linear systems can be solved independently across the quadrature nodes of the complex contour, and (iii) each complex linear system with multiple right-hand-sides can be solved in parallel. Parallel resources can be placed at all three levels simultaneously in order to achieve scalability and optimal use of the computing platform.

During these past few years, the FEAST public software package [3] for solving the eigenvalue problem has reached a significant international visibility. Since the integration of FEAST v2.1 in Intel-MKL in 2013, the solver is now widely accessible to the community. The library offers 'black-box' reverse communication interfaces (RCI) which are both matrix format and linear system solver independent, and can then be fully customized by expert users to allow maximum flexibility for their applications. Consequently, the FEAST library has been very well received by application developers of electronic structure packages.

In our most recent work (which includes the upcoming v4.0 of the software), the solver has been reimplemented to make use of residual inverse iterations. Although, the new filter form is mathematically equivalent to the original FEAST linear projector, it is numerically more efficient and more appealing in a number of new situations. We have demonstrated the effectiveness of the FEAST residual inverse iterations for addressing: (i) the inexact inner-outer iterative approach (IFEAST or FEAST without factorization), (ii) the mixed precision arithmetic iterative procedure, and (iii) the non-linear eigenvalue problem. In addition to all these new features, version v4.0 contains also the PFEAST package including three levels of MPI parallelism which can be integrated with distributed solvers such as cluster pardiso, MUMPS or any custom domain decomposition techniques [4].

[1] E. Polizzi, Phys. Rev. B, 79, p115112, (2009).
[2] P. Tang, E. Polizzi, SIAM Journal on Matrix Analysis and Applications, 35, p354, (2014).
[3] The FEAST eigenvalue solver, http://www.feast-solver.org
[4] J. Kestyn, V. Kalantzis, E. Polizzi, and Y. Saad. In Proceedings of the 2016 ACM/IEEE Supercomputing Conference (SC16).

Exact Density Matrix Purification for Parallel Computations

Edmond Chow[1] and Jie Chen[2]

[1]*School of Computational Science and Engineering, Georgia Institute of Technology*
[2]*IBM T. J. Watson Research*

Density matrix purification is a key strategy of linear scaling methods. This talk advocates using density matrix purification more generally, and specifically without truncation, when the chemical system size is small relative to the size of the parallel resources available.

To accelerate density matrix purification, nonmonotonic purification polynomials have been proposed by Rubensson (2011) and Suryanarayana (2013). The acceleration technique is related to scalings used when computing the matrix sign function or the matrix polar decomposition. Using theoretical results from Nakatsukasa and Higham (2012) on the backward stability of iterations for computing the polar decomposition, we show that nonmonotonic purification is not stable in general. We illustrate how this instability arises, leading to loss of commutativity between the Hamiltonian and the computed density matrix. We also propose a modification to nonmonotonic purification that appears to cure the stability issue without noticeably degrading convergence properties.

We note in passing that there are many types of iterations that converge more rapidly (than those derived from the McWeeny purification polynomial) but these iterations require inverse computations and thus it is not known how to adapt these iterations to the common canonical or varying Fermi level case. Further, an open question is how the above instability, if not addressed, affects the convergence of any outer iterations, like SCF.

Eigenvalue Problems in Many Body Perturbation Theory

Jack Deslippe[1]

[1]*NERSC, Lawrence Berkeley National Lab*

We will discuss the nature and scale of the eigenvalue problems that enter into Many Body Perturbation Theory problems (MBPT - GW based methods in particular) and some of the approaches used to solve them. The scope of these problems include the generation of occupied and empty input orbitals from mean-field codes (typically from Density Functional Theory (DFT)), the construction of the dielectric matrix at multiple frequency or energy responses and the generation of the correlated exciton (electron + hole) excited states. In this presentation, we also present a performance comparison of different exact parallel eigensolver libraries on the Cori (Intel Knights Landing) system at NERSC as used in MBPT applications. We analyze the performance impact of the data layout, MPI + OpenMP scaling and absolute performance.

Equilibrating alternating methods for low-rank approximations with Gaussian priors via conic automorphisms

J. Poulson[1]

[1]*Google Research*

Computing low-rank approximations that generalize the SVD, in the sense of more general loss functions and the addition of regularization/priors, has become standard practice as a means of providing input features of recommendation systems and classifiers (most notably for NLP). In the case of generalized matrix completion with Gaussian priors on the factors, block coordinate descent schemes are effective in practice but can lead to a competition between the loss function and regularization penalty that slows convergence.

It is shown that, for arbitrary loss functions on the low-rank reconstruction combined with equal Frobenius-norm penalties on the factors, the optimal solution will always have equivalent factor Gramians. Indeed, the proof is constructive, as a global gauge transformation on the loss function that minimizes the regularization can be found by computing a particular automorphism of the cone of symmetric positive-definite matrices; in the parlance of Interior Point Methods, this is the automorphism used to compute a Nesterov-Todd scaling point of a primal-dual pair in a semidefinite program.

Given this interpretation, with the exception of rank-deficiency, a linear-time algorithm for equilibrating the low-rank factors at each iteration is immediate. A brief overview of the Jordan algebraic interpretation of automorphisms of symmetric cones and the associated Riemannian geometry will be provided before discussing pragmatic schemes for handling rank-deficiency.

Robust determination of chemical potential in the pole expansion method and fast time dependent density functional theory calculations

Weile Jia[1] , Dong An[1] and Lin Lin[1,2]

[1]Mathematic Department, University of California, Berkeley.
[2]Scalable Solver Group, Computing Research Division, Lawrence Berkeley National Lab

In this talk, I will present our recent work on pole expansion and selected inversion (PEXSI) method and time dependent density functional theory calculation.

PEXSI can be used to approximate the Fermi operator expansion (FOE) by rational matrix functions and reduce the computational complexity to at most quadratic scaling for solving Kohn-Sham density functional theory (KSDFT). Unlike the diagonalization method, the chemical potential often cannot be directly read out from the result of a single step of evaluation of the Fermi operator. Hence multiple sequential evaluations are necessary to compute the chemical potential to ensure the correct number of electrons within a given tolerance. This hinders the performance of FOE methods in practice. In this paper we develop an efficient and robust strategy to determine the chemical potential in the context of the PEXSI method. The main idea of the new method is not to find the exact chemical potential at each self-consistent field (SCF) iteration, but to dynamically and rigorously update the upper and lower bounds for the true chemical potential, so that the chemical potential reaches its convergence along the SCF iteration. Instead of evaluating the Fermi operator for multiple times sequentially, our method uses a two-level strategy that evaluates the Fermi operators in parallel. In the regime of full parallelization, the wall clock time of each SCF iteration is always close to the time for one single evaluation of the Fermi operator, even when the initial guess is far away from the converged solution. We demonstrate the effectiveness of the new method using examples with metallic and insulating characters, as well as results from *ab initio* molecular dynamics.

Real-time time-dependent density functional theory (RT-TDDFT) is known to be hindered by the very small time step (attosecond or smaller) needed in the numerical simulation due to the fast oscillation of electron wavefunctions, which significantly limits its range of applicability for the study of ultrafast dynamics. In this paper, we demonstrate that such oscillation can be considerably reduced by optimizing the gauge choice using the parallel transport formalism. RT-TDDFT calculations can thus be significantly accelerated using a combination of the parallel transport gauge and implicit integrators, and the resulting scheme can be used to accelerate any electronic structure software that uses a Schrodinger representation. Using absorption spectrum, ultrashort laser pulse, and Ehrenfest dynamics calculations for example, we show that the new method can utilize a time step that is on the order of 10-100 attoseconds in a plane wave basis set, and is no less than 5-10 times faster when compared to the standard explicit 4th order Runge-Kutta time integrator.

Opportunities for ELPA to Accelerate the Solution of the Bethe-Salpeter Eigenvalue Problem

Peter Benner[1], Andreas Marek[2], Carolin Penke[1]

[1]*Max-Planck-Institute for Dynamics of Complex Technical Systems (Computational Methods in Systems and Control Theory)*
[2]*Max Planck Computing and Data Facility*

The Bethe-Salpeter eigenvalue problem arises in quantum chemistry and is characterized by a peculiar block structure that involves complex conjugated transposes as well as regular non-conjugated complex transposes. The matrices are dense and easily become very large. This calls for high performance parallel algorithms that scale well on many nodes of modern supercomputers. The immanent symmetry of the spectrum should be preserved. A promising direct method is found in the BSEPACK library [1] which is based on ScaLAPACK routines for skew-symmetric eigenvalue problems. The ELPA library [2] is a modern alternative to ScaLAPACK for solving symmetric eigenvalue problems. We investigate how the performance of BSEPACK can benefit from using ELPA. The adaptation of the ELPA algorithm to skew-symmetric matrices yields further runtime improvements.

[1] M. Shao, F. H. da Jornada, C. Yang, J. Deslippe, and S. G. Louie, *Linear Algebra Appl.* 488, 148-167 (2016).
[2] T. Auckenthaler, V. Blum, H.-J. Bungartz, T. Huckle, R. Johanni, L. Kraemer, B. Lang, H. Lederer and P. R. Willems, *Parallel Computing* 37, 783-794 (2011).

Task-based Left-Looking Selected Inversion algorithm on Shared Memory Systems

Mathias Jacquelin

Lawrence Berkeley National Laboratory

In this work, we introduce the left-looking variant of the selected inversion algorithm, and present an efficient implementation of the algorithm for shared memory machines using a task parallel method. We demonstrate that with the task scheduling features provided by OpenMP 4.0, the left-looking selected inversion algorithm can scale well both on the Intel Haswell multicore architecture and on the Intel Knights Landing (KNL) manycore architecture up to 16 and 64 cores, respectively. On the KNL architecture, we observe that the maximum parallel efficiency achieved by the left-looking selected inversion algorithm can be as high as 62% even when all 64 cores are used, despite the inherent asynchronous nature of the computation and communication patterns in sparse matrix operations.

Compared to the right-looking selected inversion algorithm, the left-looking formulation facilitates efficient pipelining of operations along different branches of the elimination tree, and can be a promising candidate for future development of massively parallel selected inversion algorithms on heterogeneous architectures.

Eigenvalue Solvers Considerations on Modern HPC Hardware Platforms

Piotr Luszczek[†], Mark Gates[†] , Azzam Haidar[‡], Jakub Kurzak[†], Hatem Ltaief[*], Stanimire Tomov[†] and Jack Dongarra[†,¶,§]

[†]University of Tennessee, USA, [‡]NVIDIA Corporation, [§]Manchester University, UK, [*]AUST, Saudi Arabia, [¶]Oak Ridge National Lab., USA

We show optimizations of the two-stage algorithm for reducing a dense symmetric matrix to tri-diagonal form through similarity transformations on multicore processors using runtime task scheduling based on dataflow principles[7]. By restructuring the task-based code, the data reuse inside the cache hierarchy increases[2]. Haidar et al. the two-stage approach may be extended to reduction from dense to bi-diagonal form for square matrices[3]. For rectangular matrices, communication-avoiding methodology is used for efficient bi-diagonal reduction[6].

The two-stage algorithm may be used for computing for singular vectors[4]. They achieved speedup over existing codes despite requiring more than twice as many floating-point operations when compared with the classical one-stage approach. Similar trend applies to computing eigenvectors of a symmetric matrix[5] . Cache-cognizant techniques may be used for accelerating eigenvector procedure for nonsymmetric matrices[1].

[1] M. Gates, A. Haidar, and J. Dongarra. Accelerating computation of eigenvectors in the nonsymmetric eigenvalue problem. In *Proceedings of 11th International Meeting High Performance Computing for Computational Science*, VECPAR 2014, University of Oregon and Hilton Conference Center, Eugene, OR, USA, June 30 – July 3 2014. Springer.

[2] A. Haidar, H. Ltaief, and J. Dongarra. Parallel reduction to condensed forms for symmetric eigenvalue problems using aggregated fine-grained and memory-aware kernels. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, pages 8:1–8:11, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0771-0. Doi: 10.1145/2063384.2063394. URL http://doi.acm.org/10.1145/2063384.2063394.

[3] A. Haidar, H. Ltaief, P. Luszczek, and J. Dongarra. A comprehensive study of task coalescing for selecting parallelism granularity in a two-stage bidiagonal reduction. In *Proceedings of the IEEE International Parallel and Distributed Processing Symposium*, Shanghai, China, May 21-25 2012. ISBN 978-1-4673-0975-2.

[4] A. Haidar, J. Kurzak, and P. Luszczek. An improved parallel singular value algorithm and its implementation for multicore hardware. In *SC13, The International Conference for High Performance Computing*, Networking, Storage and Analysis, Denver, Colorado, USA, November 17-22 2013.

[5] A. Haidar, P. Luszczek, and J. Dongarra. New algorithm for computing eigenvectors of the symmetric eigenvalue problem. In *The 15th IEEE International Workshop on Parallel and Distributed Scientific and Engineering Computing* (PDSEC 2014), Phoenix, AZ, USA, May 23 2014.

[6] H. Ltaief, P. Luszczek, A. Haidar, and J. Dongarra. Enhancing parallelism of tile bidiagonal transformation on multicore architectures using tree reduction. In R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Waśniewski, editors, *Proceedings of 9th International Conference*, PPAM 2011, volume 7203, pages 661–670, Toruń, Poland, 2012.

[7] P. Luszczek, H. Ltaief, and J. Dongarra. Two-stage tridiagonal reduction for dense symmetric matrices using tile algorithms on multicore architectures. In *IPDPS 2011: IEEE International Parallel and Distributed Processing Symposium*, Anchorage, Alaska, USA, May 16-20 2011.

Computational Materials Science by RESCU - a KS-DFT method for solving thousands of atoms

Vincent Michaud-Rioux, Ying-Chin Chen, Saeed Bohloul and Hong Guo

Centre for the Physics of Materials & Department of Physics, McGill University, Montreal, Canada, H3A 2T8

A major effort in materials physics is to develop a first principles method that can accurately, efficiently and comfortably calculate condensed phase materials comprising thousands of atoms. Solving large systems is necessary when dealing with materials involving interfaces, surfaces, dilute impurities, grain boundaries, dislocations, domains, solvents etc. So far, the most well-known software packages of Kohn–Sham density functional theory (KS-DFT) can nicely solve problems at several hundred atoms level using modest computers.

Here I shall present our effort in developing a general-purpose KS-DFT solver called RESCU. We demonstrate that RESCU can compute electronic structure for systems comprising many thousands of atoms on modest computer resources, for metals, semiconductors, insulators, DNA-in-water, Moire patterns in 2D heterjunction, dilute doped III-nitrides, etc. For these problems and for supercells with up to 14,000 atoms, RESCU converges KS-DFT at GGA level in a few to ten wall-clock hours; at hybrid-functional level, supercells with 1200 atoms were solved. RESCU achieves good efficiency without compromising KS-DFT accuracy. I shall discuss the computational issues behind the efficiency gain and present several materials physics problems solved by RESCU.

Acknowledgements: I thank Prof. Lei Zhang for participating some earlier work in RESCU and Prof. Xiaobin Chen for contributing to the phonon package. I wish to thank many other researchers who helped us improving the RESCU.

Electronic energy minimization methods in the ONETEP program

Chris-Kriton Skylaris

Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

The ONETEP [1] code can perform density functional theory (DFT) calculations with linear-scaling computational effort with respect to the number of atoms while, uniquely, retains near-complete basis set accuracy. Strictly localised orbitals (Non-Orthogonal Generalised Wannier Functions - NGWFs) are optimised *in situ*, in a basis set of periodic sinc functions. For insulators, linear-scaling is achieved by taking advantage of the exponential decay of the Wannier functions in electronic systems with a band gap. In order to have linear-scaling behaviour in the SCF procedure we need to avoid diagonalization. Instead, we use a direct energy minimisation technique based on conjugate gradients where our objective function maintains the essential physical conditions that the one-particle density matrix needs to obey [2]. The method takes advantage of the inherent sparsity of our matrices using sparse matrix storage and sparse matrix operations. Metallic systems pose a challenge for Kohn-Sham DFT calculations as their lack of band gap does not allow a numerically stable description of "occupied" and "virtual" molecular orbitals and neither do they follow exponential decay at zero temperature. Mermin's extension of DFT to finite electronic temperature [3] overcomes these difficulties by using fractional occupancies for the molecular orbitals. In this approach, the Helmholtz free energy rather than the energy of the electronic system is computed. The electronic entropy is a function of the fractional occupancies. We have developed a direct energy minimization method for metallic systems for the localised NWGFs of the ONETEP code. In its original formulation [4] our method contains a cubic scaling computational bottleneck due to a required diagonalization of the Hamiltonian matrix but with greatly reduced impact on computational cost due to the minimal size of the Hamiltonian matrix and the use of parallel eigensolvers. Our more recent developments based on Fermi Operator Expansion (FOE) methods completely avoid the diagonalization. Thus, our Annealing and QUenching Algorithm FOE or AQuA-FOE [5] is a linear-scaling DFT method for metallic systems as exponential decay is re-established at finite electronic temperature, including also the necessary calculation of the entropy and the chemical potential.

[1] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne. *J. Chem. Phys*. **122** (2005) 084119.
[2] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne. *J. Phys. Condens. Matter* **20** (2008) 294207.
[3] N. D. Mermin, *Phys. Rev.* **137** (1965) A1441.
[4] A. Ruiz-Serrano, and C.-K. Skylaris. *J. Chem. Phys.* **139** (2013) 054107.
[5] J. Aarons, and C.-K. Skylaris. *J. Chem. Phys.* **148** (2018) 074107.

Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems

Amartya S. Banerjee[1,2*], Lin Lin[1,3] , Phanish Suryanarayana[5] , Chao Yang[1] , and John E. Pask[4]

[1]*Computational Research Division, Lawrence Berkeley National Lab, Berkeley, CA 94720, USA.*
[2]*Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee, Knoxville ,TN 37996, USA.*
[3]*Department of Mathematics, University of California, Berkeley, CA 94720, USA.*
[4]*Physics Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA.*
[5]*School of Civil and Environmental Engineering, Georgia Institute of Technology, GA 30332, USA.*

I will describe a novel iterative strategy for Kohn Sham Density Functional Theory calculations aimed at large systems, that applies to metals and insulators alike. This technique avoids diagonalization of the Kohn-Sham Hamiltonian on every self-consistent field (SCF) iteration, and instead, employs a two-level Chebyshev polynomial filter based complementary subspace strategy to: 1) compute a set of vectors that span the occupied subspace of the Hamiltonian; 2) reduce subspace diagonalization to just partially occupied states; and 3) obtain those states in an efficient manner via an inner Chebyshev-filter iteration. These steps reduce the cost of large metallic calculations significantly, while eliminating the subspace diagonalization cost for insulating systems altogether. Unlike linear-scaling methods however, this approach does not invoke additional approximations related to nearsightedness and therefore, is more widely applicable to a larger variety of systems at different electronic temperatures.

I will describe the implementation of this strategy within the framework of the Discontinuous Galerkin (DG) electronic structure method and demonstrate that this results in a computational scheme that can effectively tackle systems containing tens of thousands ofelectrons, with chemical accuracy, within a few minutes of wall clock time per SCF iteration on large-scale computing platforms. I will discuss possible future applications of this computational strategy, particularly its use in the study of materials defects using abinitio molecular dynamics. I will end with a discussion of ongoing and future work, particularly in relation to alternate computational strategies (based on the so called Spectrum Slicing technique) that are expected to further push the limits of ab initio calculations of large non-insulating systems.

[1] A.S. Banerjee, L. Lin, P. Suryanarayana, C. Yang and J.E. Pask (2018): Two-level Chebyshev filter based complementary subspace method: pushing the envelope of large-scale electronic structure calculations; Journal of Chemical Theory and Computation; Volume 14 (6), Pages 2930–2946.
[2] A.S. Banerjee, L. Lin, W. Hu, C. Yang and J.E. Pask (2016): Chebyshev polynomial filtered subspace iteration in the Discontinuous Galerkin method for large-scale electronic structure calculations; The Journal of Chemical Physics; Volume 145 (15), 154101.

Finite difference DFT solver: Direct functional minimization with eigensolver in projected subspace

J.-L. Fattebert[1] and D. Osei-Kuffuor[2]

[1]*Computational Sciences & Engineering Division, Oak Ridge National Laboratory*
[2]*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*

When discretizing the Kohn-Sham equations on a mesh by finite differences or using a "fine" numerical basis set (Plane Waves, Finite Elements,…), "matrix-free" algorithms are typically used to compute the eigenvectors associated with the lowest eigenvalues of the Hamiltonian, avoiding building very large matrices. This set of eigenvectors can be viewed as an auxiliary basis set for which a much smaller matrix eigenvalue problem can then be setup and solved to determine the occupation numbers for each eigenvector.

In a direct minimization approach, the steepest descent directions associated with the nonlinear Kohn-Sham energy functional can be readily computed, preconditioned, and used to update the subspace associated with the lowest eigenvalues of the Hamiltonian, while the occupations numbers can be updated after each wave functions update. Also, since the solution of the Kohn-Sham equations can be represented by any linearly independent set of vectors spanning the invariant subspace associated with the lowest eigenvalues of the Hamiltonian, a search for an approximate solution made of localized orbitals can be carried out by a minimization with localization constraints. A fully O(N) scheme can be obtained if in addition, a sparse solver can be used for computing the single particle density matrix representing the occupations of these localized orbitals. For that, we propose a matrix divide-and-conquer parallel approach, where local principal submatrices corresponding to the closest orbitals in space are used to compute the density matrix elements in a distributed way [1,2].

[1] Osei-Kuffuor, D. and J.-L. Fattebert, "Accurate and Scalable O(N) Algorithm for First-Principles Molecular-Dynamics Computations on Large Parallel Computers", Phys. Rev. Lett. 112, 046401 (2014)
[2] J.-L. Fattebert, D. Osei-Kuffuor, E.W. Draeger, T. Ogitsu, W.D. Krauss, "Modeling Dilute Solutions Using First-Principles Molecular Dynamics: Computing more than a Million Atoms with over a Million Cores", SC16: International Conference for High Performance Computing, Networking, Storage and Analysis, Salt Lake City, Utah, November 2016, p. 12-22

Acknowledgement: This material is based upon work supported by the U.S. Department of Energy, Office of Science under contract DE-AC05-00OR22725, and supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

SLEPc-SIPs:
Massively Parallel Sparse Eigensolver for Electronic Structure Calculations

Murat Keçeli[1], Hong Zhang[2], Fabiano Corsetti[3], Carmen Campos[4], Jose E. Roman[5], Álvaro Vázquez-Mayagoitia[1], Peter Zapol[6], Albert F. Wagner[7]

[1]*Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[2]*Mathematics and Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3] *Departments of Materials and Physics and the Thomas Young Centre for Theory and Simulation of Materials, Imperial College London, London, SW7 2AZ, United Kingdom*
[4]*D. Sistemes Informàtics i Computació, Universitat Politècnica de València, Camí de Vera s/n, València 46022, Spain*
[5]*Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[6] *Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[7]*Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*

Shift-and-Invert Parallel spectral transformation (SIPs) method as implemented in the SLEPc library offers two layers of parallelization for the real and symmetric generalized eigenvalue problem. The first layer is based on the computation of eigenpairs independently as long as the eigenvalues are not degenerate. The second layer is based on the concurrent linear algebra operations as implemented in the PETSc framework and other libraries required for sparse matrix factorization. The non-uniformity of the eigenvalue spectrum and particularly the large gaps between the eigenvalues in quantum chemistry applications weakens the strong-scaling efficiency in the first layer of parallelization. In this talk, I will describe the techniques we developed to improve the load-balance once the approximate knowledge of the eigenvalues is provided, the performance comparison of this method with dense eigensolvers, and the implementation details of the integration of SLEPc-SIPs in SIESTA ab initio molecular dynamics package.

Accelerating eigenvalue calculations: Shift-without-invert and shift-with-invert techniques for spectrum-partition

Yunkai Zhou[1]

[1]*Math Department, Southern Methodist University*

We present two spectrum partition methods based on ARPACK for large symmetric eigenvalue problems. One of them utilizes shift-without-invert, which is less traditional but easier to implement; the other utilizes the more traditional shift-with-invert techniques. We discuss ways for automatic partition of unknown spectrum, adaptive choices of shifts, and ways to keep the cost for orthogonalization of eigenvectors minimal.

Reverse Communication Interface for Iterative Eigensolvers in ELSI

Yingzhou Li[1] , Jianfeng Lu[1,2] and ELSI Team

[1]*Department of Mathematics, Duke University*
[2]*Department of Chemistry and Department of Physics, Duke University*

ELectronic Structure Infrastructure (ELSI) provides and enhances open-source software packages which solve or circumvent the Kohn-Sham eigenvalue problems in the self-consistent field cycle of density-functional theory. In this talk, we introduce the recent development of a reverse communication interface (RCI) for several major iterative eigensolvers in ELSI. The RCI enables the usage of iterative eigensolver for Kohn-Sham eigenvalue problem with or without an explicit form of the Hamiltonian matrix. When basic linear algebra operations are available, RCI can be adopted efficiently, and different iterative methods can be applied to solve different problems without any modification of the code. In addition to the easy use, numerical results demonstrate the efficiency of the reverse communication interface. This is joint work with ELSI team.

Development of a dense eigenvalue solver for Exa-scale systems

T. Imamura[1]

[1]*RIKEN (Center for Computational Science)*

Since 2010, a new project 'H4ES' has been initiated in the Japanese numerical linear algebra community to develop eigenvalue software led by six PI's, Prof. Sakurai, Prof. Zhang, Prof. Yamamoto, Prof. Hoshi, Prof. Kuramashi and the author. The project has been supported by one of the national grant CREST JST, 'Development of System software Technologies for post-Peta Scale High Performance Computing.' We organized two mini research groups devoting to the development of a dense eigenvalue solver 'EigenExa' and a sparse eigenvalue solver 'z-Pares' at the University of Tsukuba and the University of Electro-Communication (later at RIKEN AICS), respectively.

The dense solver project inherited the results from the Earth Simulator system [1], which was a world-largest vector supercomputer in the early 2000's. The numerical algorithms adopted were quite traditional ones; i) parallel Householder tridiagonalization, ii) Cuppen's divide and conquer algorithm, and iii) back-transformation corresponding to the Householder tridiagonalization. The reference codes were optimized by accordance with a combination of old-fashioned vectorization technique and a non-modern cache tuning technique. Although the codes were written in an MPI+OpenMP hybrid fashion, we needed to build scratch codes at the first target system of the T2K cluster [2]; then the kernel code was ported to the K computer with the help of Fujitsu. We have published the EigenExa library version 2.4p1, which is the latest release of our eigensolver [3]. Through co-design with application users, we decided to exploit several algorithmic and implementation techniques to remove bottleneck which comes from data communication (memory and node interconnection).

Since 2014, RIKEN AICS (currently, R-CCS) has started a national project to develop a flagship system; we say post-K system, which will be empowered by a brand-new HPC processor, 'armv8-A+SVE.' For the seamless transition from the K computer to the emerging system, we are improving the EigenExa library towards other supercomputer systems in Japan, such as HOKUSAI-GW at RIKEN and Oakforest-PACS at the JCAHPC handled by U-Tokyo and U. Tsukuba. Since another aspect on the extreme computing implies not only capability computing but capacity computing, we recognized the necessity of the high-performance eigensolver with a broad variety of parallelism and parallel scaling. It is a quite challenging work for applied mathematics, computer science, and engineering.

[1] S. Yamada, et al. High-Performance Computing for Exact Numerical Approaches to Quantum Many-Body Problems on the Earth Simulator, SC06, 2006
[2] T. Imamura, et al. Development of a High Performance Eigensolver on the Peta-Scale Next Generation Supercomputer System, PNST, AESJ, Vol. 2, pp.643-650, 2011
[3] EigenExa Homepage. http://www.aics.riken.jp/labs/lpnctrt/en/projects/eigenexa/

Bigger ~~and~~ is better: large scale NEGF calculations and coupling DFT with TB

Nick R. Papior, Gaetano Calogero, Mads Brandbyge, August, MolSSI

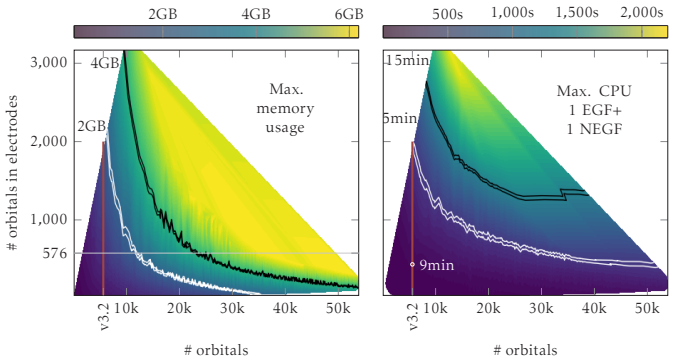

Figure 1: Left, TranSiesta memory for varying electrode (y-axis) and device size (x-axis). Right, TranSiesta time usage for equivalent systems. A linear scaling for constant electrode widths are found.

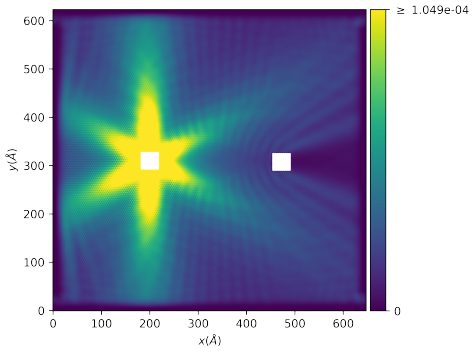

Figure 2: DFT+TB calculation with an accurate DFT STM-tip which is connected to a TB parameterized infinite graphene system to investigate the far field currents.

We will present the recent developments of $N_\mathfrak{e} \geq 1$ electrode DFT+NEGF simulations enabling simulations of multiprobe physics. In particular the implementation in [1, 2] and the efficiency and scalability of the code (TranSiesta) will be emphasized. In Fig. 1 the memory and time requirements are shown for a system of increasing size along the transverse and transport directions, y, x respectively. A red line indicates the prior TranSiesta memory/time requirements.

Our findings suggests a reduction of the number of point on the contour integration with increased performance. Additionally we find that proper selection of k-points for certain materials may improve non-equilibrium convergence.

Finally we will present results based on large scale DFT+NEGF calculations coupled to extreme scale DFT-parameterized tight-binding calculations simulating infinite graphene systems. Such calculations can be used to simulate realistic device configurations with near DFT precision. An example is seen in Fig. 2 which depicts the bond-currents in an infinite graphene sample by injection from an Au-tip and collected with another Au-tip located $\sim$ 28 nm away. DFT calculations are used to calculate the local potential perturbations around both tip-graphene interface, empty boxes in figure.

[1] Nick R. Papior and Mads Brandbyge. “Computational Tools and Studies of Graphene Nanostructures”. PhD thesis. 2016, p. 176. DOI: 10.11581/DTU:00000025.

[2] Nick Papior et al. “Improvements on non-equilibrium and transport Green function techniques: The next-generation TranSiesta”. In: *Computer Physics Communications* 212 (2017). DOI: 10.1016/j.cpc.2016.09.022.

Eigenstate-analysis using Sakurai-Sugiura method with O(*N*)-DFT code CONQUEST

A. Nakata[1,2], Y. Futamura[2], T. Sakurai[2], D. R. Bowler[1,3] and T. Miyazaki[1,3]
[1]*WPI-MANA, National Institute for Materials Science, Japan*
[2]*Department of Computer Science, University of Tsukuba, Japan*
[3]*Department of Physics & Astronomy, University College London, UK*

We have proposed an efficient way to calculate the electronic structure of large systems by combining a large-scale first-principles density functional theory code, CONQUEST [1], and an efficient interior eigenproblem solver, the Sakurai-Sugiura method [2]. The electronic Hamiltonian and self-consistent charge density of large systems are obtained by CONQUEST and the eigenstates of the Hamiltonians are then obtained by the Sakurai-Sugiura method. The applicability of this combination on systems with 10,000+ atoms will be demonstrated [3].

In CONQUEST, electronic density can be optimized by two methods. One is the conventional diagonalization method obtaining eigenstates of the electronic Hamiltonian, which scales cubically to the number of atoms N in a target system. Another method is the density matrix minimization (DMM) method, in which total electronic energy is minimized with respect to the density matrix. By introducing a cutoff for the density matrix, the DMM calculation can scale linear to N, i.e., order-N (O(N)). We don't calculate eigenstates in the O(N) calculations so that it is not possible to analyze the electronic structures based on one-electron wave functions (the eigenstates of the Hamiltonian).

To analyze the electronic structures of large-systems, we have introduced Sakurai-Sugiura method, which can provide the eigenstates in a specified energy region with high parallel efficiency. We first optimize the electronic Hamiltonian by CONQUEST and calculate the eigenstates of the Hamiltonian afterword only in the energy region of interest by the SS method. This procedure enables us to investigate the electronic structures of large systems quite efficiently.

This procedure is also applicable to improve the unoccupied states in the recently-developed multi-site calculation method [4,5]. The multi-site method uses the local molecular-orbital like functions to express the density matrix. The method reduces the computational cost dramatically without losing the accuracy of the occupied states, but the description of the unoccupied states becomes worse. By applying the SS method after optimizing the occupied states by the multi-site method, we succeeded in reproducing the unoccupied states with high accuracy.

[1] http://www.order-n.org
[2] T. Sakurai and H. Sugiura, *J. Comput. Appl. Math.* 159, 119-128 (2003).
[3] A. Nakata, Y. Futamura, T. Sakurai, D. R. Bowler, T. Miyazaki, *J. Chem. Theory Comput.* 13, 4146-4153 (2017).
[4] A. Nakata, D. R. Bowler and T. Miyazaki, *J. Chem. Theory Comput.* 10, 4813-4822 (2014).
[5] A. Nakata, D. R. Bowler and T. Miyazaki, *Phys. Chem. Chem. Phys.* 17, 31427 (2015).

GPU-Accelerated Real Space Electronic Structure Theory on HPC Resources

William P. Huhn[1], Björn Lange[1], Victor Yu[1], Mina Yoon[2], and Volker Blum[1]

[1]*Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708*
[2]*Center for Nanophase Materials Science, Oak Ridge National Laboratory, Oak Ridge, TN, 37831*

We present our experiences implementing GPU acceleration in the massively parallel, real space FHI-aims electronic structure code for computational materials science. We highlight how FHI-aims achieves efficient scaling on HPC resources using real-space operations and demonstrate how the vectorized implementation originally designed for CPUs was ported to GPUs with minimal code rewrite. This is made possible by a domain decomposition scheme which divides the integration grid into groups ("batches") of points, which can be distributed in an embarrassingly parallel fashion amongst tasks, reducing the problem to dense serial linear algebra on each task. The dense serial nature of the workload on each task allowed us to employ cuBLAS to accelerate various real-space operation (Hamiltonian integration, charge density update, force and stress tensor calculations) with minimal code divergence between the CPU and GPU version.

To assess the performance of our GPU implementation, we performed benchmarks on three different architectures using a 103-material test set. We find that operations which are particularly reliant on dense serial linear algebra show dramatic speedups from GPU acceleration: in particular, SCF iterations including force and stress calculations are observed to exhibit speedups ranging from 4.5x to 6.6x. This translates to an expected overall speedup between 3x-4x for the entire calculation, as is shown in the example of diamond silicon calculations. Additional calculations were performed on OLCF's Titan for a 375- atom Bi2Se3 bilayer to verify that the GPU code scales for large-scale calculations.

Posters

P1: Reverse Communication Interface for Iterative Eigensolvers in ELSI
Yingzhou Li, Jianfeng Lu and ELSI Team

P2: Highly efficient Poisson equation solver for linear scaling Hartree-Fock exchange energy evaluation
Álvaro Vázquez-Mayagoitia, Junteng Jia and Robert Distasio

P3: NOPAIN: Efficient Evaluation of Quantum Nonlocal Operators
Ying-Chih Chen and Hong Guo

P4: Dirac electrons in large-scale moiré supperlattice
Chen Hu, Vincent Michaud-Rioux, Xianghua Kong, and Hong Guo

P5: Systematic search for two-dimensional ferromagnetic materials
Yu Zhu, Xiang-Hua Kong, and Hong Guo

P6: Assessment of localized and randomized algorithms for electronic structure
Jonathan E. Moussa and Andrew D. Baczewski

P7: Susceptibility-based density mixing for in the CASTEP plane-wave DFT code
Vincent E. Sacksteder IV

P8: Cubic scaling algorithms for RPA correlation using interpolative separable density fitting
Kyle Thicke and Jianfeng Lu

P9: The Self-Consistent Field in Kohn-Sham Density Functional Theory: A Review of Methods and Algorithms
Nick Woods, Mike Payne, and Phil Hasnip

P10: Fast TDDFT Calculations with the Parallel Transport Gauge
D. An, W. Jia, L. Lin, and L.W. Wang

P11: GPU-Accelerated Real Space Electronic Structure Theory on HPC Resources
William P. Huhn, Björn Lange, Victor Yu, Mina Yoon, and Volker Blum

Highly efficient Poisson equation solver for linear scaling Hartree-Fock exchange energy evaluation

Álvaro Vázquez-Mayagoitia (alvaro@anl.gov), Junteng Jia and Robert Distasio

We present an efficient Poisson equation (PE) solver using a conjugate gradient algorithm with an efficient Incomplete Cholesky Factorization based preconditioner. Also we show that this method in conjunction with asynchronous communication (to overlap communication and computation) is possible to reach linear scaling in the evaluation of Hartree-Fock energy, which is an important contribution in hybrid exchange correlation functionals in Density Functional Theory. Overall, this approach can accelerate accurate *ab intio* molecular dynamic simulations in the Car-Parinello code (CP.x) in Quantum Espresso software. In many core processors, the time to solution per PE has an speed up of 3.5x for IBM BG/Q, and 2.5x for Intel Xeon Phi KNL, with respect the original version.

NOPAIN: Efficient Evaluation of Quantum Nonlocal Operators

Ying-Chih Chen[1] and Hong Guo[1]
[1]*Center for the Physics of Materials and Department of Physics, McGill University, Montreal, QC, H3A 2T8, Canada*

The computation efficiency of numerical atomic orbitals (NAOs) is high due to its localized property, i.e., the orbital spatial value is strictly zero beyond a given radius cutoff. If two NAOs do not spatially overlap, the matrix element of any local operator is zero. This makes NAOs attractive for the research of order-N methods, for which the computational scaling is linear with the basis size N. However, the matrix element of a nonlocal operator is not always zero if two NAOs do not spatially overlap. All matrix elements should be calculated to ensure accuracy. An artificial distance truncation is usually applied for periodic systems. The artificial distance truncation assumes that if the distance between two atoms is larger than a manually defined distance, the corresponding matrix element is zero. This assumption is in general not true since the Coulomb operator, as well as many other nonlocal operators, are long-range. Its spatial interaction strength decays slowly and such approximation degrades accuracy.

In this work, we solve the problem by developing a novel method which combines plane waves (PWs) and NAOs to efficiently evaluate nonlocal operators under periodic boundary conditions. Our method is denoted as NOPAIN (**N**onlocal **O**perators with **P**lane w**A**ves **I**ntegrated into **N**umerical atomic orbitals)[1]. The problem of distance-truncation is avoided, and the computation complexity is largely reduced with the NOPAIN method. This method is applicable to large systems. The general formalism has been implemented in the RESCU package[2], a MATLAB-based density-functional theory (DFT) software, within the hybrid functional approach where the nonlocal operator is the (short-range) Fock exchange. Comparison of electronic structures of a wide range of semiconductors to a pure PW scheme validates the accuracy of NOPAIN. The computational complexity is asymptotically quadratic to the number of electrons.

We apply our method to two realistic systems which require the hybrid-functional to accurately calculate the band gap and the band alignment. Both systems contain hundreds to a thousand atoms in the supercell which could hardly be investigated at the hybrid-functional level before even using supercomputers. The first is concerning band engineering of GaSbN alloy for solar fuel applications[3]. We show that the band gap of GaN can be efficiently reduced by a dilute Sb-doping. The second is the interface between a single layer black phosphorous (sBP) and c-HfO2. We predict the band offset is 1.29/2.18 eV for the valence/conduction band edge, thus suitable for 2D field-effect transistors applications[1].

[1] Y-C. Chen, J-Z. Chen, V. Michaud-Rioux and H. Guo, *Phys. Rev. B* **97**, 075139 (2018).
[2] V. Michaud-Rioux, L. Zhang, and H. Guo, *J. Comp. Phys.* **307**, 593-614 (2016).
[3] Q. Shi and Y-C. Chen, F. A. Chowdhury, Z. Mi, V. Michaud-Rioux, and H. Guo, *Phys. Rev. Mat.* **1**, 034602 (2017).

Dirac electrons in large-scale moiré supperlattice

Chen Hu, Vincent Michaud-Rioux, Xianghua Kong, and Hong Guo

Department of Physics, McGill University, Montreal, Canada

Moiré patterns in van der Waals (vdW) heterostructures bring novel physical effects to the materials. We report theoretical investigations of the moiré pattern formed by graphene (Gr) on hexagonal boron nitride (*h*BN). For both the two-dimensional (2D) flat-sheet and the freestanding three-dimensional (3D) wavelike film geometries, the behaviors of Dirac electrons are strongly modulated by the local high-symmetry stacking configurations of the moiré pattern. In the 2D flat sheet, the secondary Dirac cone (SDC) dispersion emerges due to the stacking-selected localization of SDC wave functions, while the original Dirac cone (ODC) gap is suppressed due to an overall effect of ODC wave functions. In the freestanding 3D wavelike moiré structure, we predict that a specific local stacking in the moiré superlattice is promoted at the expense of other local stackings, leading to an electronic structure more similar to that of the perfectly matching flat Gr/*h*BN than that of the flat-sheet 2D moiré pattern. To capture the overall picture of the moiré superlattice, supercells containing 12 322 atoms are simulated by first principles.

Systematic search for two-dimensional ferromagnetic materials

Yu Zhu,[1] Xiang-Hua Kong,[2] and Hong Guo[1,2]

[1]*Nanoacademic Technologies Inc., Brossard, QC, Canada J4Z 1A7*
[2] *Department of Physics, McGill University, Montreal, Canada H3A 2T8*

A systematic materials informatics search for two-dimensional (2D) ferromagnetic materials (2DFM) is performed in the Inorganic Crystal Structure Database (ICSD). The key indicator in the search is that the candidates have a 2D network composed of magnetic atoms and heavy atoms. From the 187,093 entries in the ICSD, the search identified 15 candidates whose magnetic order are then determined by first-principles calculations and Curie temperatures obtained by either analytical solution or Monte Carlo simulation of the 2D Ising model. For the 15 candidates, eight materials have ferromagnetic order including three known structural prototypes and the two 2DFM materials discovered in recent experiments. More importantly, the search produced two new structural prototypes and a very promising material, Cr_3Te_4, which has a calculated Curie temperature one order of magnitude higher than the others.

Assessment of localized and randomized algorithms for electronic structure

Jonathan E. Moussa[1,2] and Andrew D. Baczewski[2]

[1]*Molecular Sciences Software Institute, Virginia Tech, Blacksburg, VA 24060, USA*
[2]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185, USA*

We compare the theoretical and practical costs of localized and randomized linear-scaling electronic structure algorithms with established quadratic-scaling and cubic-scaling algorithms for tight-binding calculations of large copper clusters. Linear-scaling algorithms are competitive at high temperatures relevant for warm dense matter, but their cost prefactors are prohibitive at ambient temperatures. We study several hybridized algorithms that combine localized and randomized algorithms to reduce their costs. The simplest hybridized algorithm is unable to improve performance, but a more sophisticated algorithm using recent concepts from structured linear algebra shows promising initial results on a square-lattice orthogonal tight-binding model.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Susceptibility-based density mixing for in the CASTEP plane-wave DFT code

Vincent E. Sacksteder IV[1]

[1]*Royal Holloway University of London, Department of Physics*

CASTEP is an ab initio DFT code which scales to the largest supercomputers and has a large feature set. I have implemented within CASTEP a susceptibility object which calculates the system's linear response to any arbitrary charge or spin perturbation. This object has been used as an alternative to Pulay and Broyden density mixing schemes which accelerate convergence of the Kohn-Sham DFT equations. This code's memory scaling, CPU scaling, and parallelization are the same as when using Pulay and Broyden algorithms. The advantage of this approach to density mixing is that it automatically adapts to the system, without external tuning parameters. The CASTEP susceptibility object has also been used for many body perturbation theory (MBPT) alternatives to the DFT.

Cubic scaling algorithms for RPA correlation using interpolative separable density fitting

Kyle Thicke[1] and Jianfeng Lu[2]

[1]*Department of Mathematics, Duke University*
[2]*Departments of Mathematics, Physics, and Chemistry, Duke University*

We present a new cubic-scaling algorithm for the calculation of the RPA correlation energy. Our scheme splits up the dependence between the occupied and virtual orbitals in χ^0 by use of Cauchy's integral formula. This introduces an additional integral to be carried out, for which we provide a geometrically convergent quadrature rule. Our scheme also uses the newly developed Interpolative Separable Density Fitting algorithm to further reduce the computational cost in a way analogous to that of the Resolution of Identity method.

[1] J. Lu and K. Thicke, *J. Comput. Phys.* 351, 187-202 (2017).

The Self-Consistent Field in Kohn-Sham Density Functional Theory: A Review of Methods and Algorithms

Nick Woods[1], Mike Payne[1], and Phil Hasnip[2]

[1]Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge
[2]Condensed Matter Theory, Department of Physics, University of York

An infimum of the Kohn-Sham energy functional can be found by solving the associated Euler-Lagrange equations – the Kohn-Sham equations. One of two popular methods for finding this infimum, along side direct minimisation of the Kohn-Sham functional, is to solve the Kohn-Sham equations iteratively using a non-linear solver. Doing so yields a ground state particle density that is so-called *self-consistent*.

Determining the optimal algorithms to find a self-consistent particle density (or, more generally, to find an infimum of the Kohn-Shan functional) is the focus of this work. In order to investigate this issue, a test suite of Kohn-Sham density functional theory input systems is designed specifically to target weaknesses in contemporary algorithms, and to exploit difficulties relating to the underlying mathematical framework (i.e. sources of ill-conditioning). A sample of methods from literature is then implemented in the plane-wave, pseudopotential software CASTEP [1]. These methods are benchmarked on the test suite, and a set of analysis tools allows us to fairly assess the strengths and weaknesses of each approach, and determine which methods can be considered state-of-the-art.

[1] S. J. Clark et al. *Z. Kristallogr*. 220, 567–570, (2005).

Fast TDDFT Calculations with the Parallel Transport Gauge

D. An[1], W. Jia[1], L. Lin[1,2], and L.W. Wang[3]

[1]*Department of Mathematics, University of California, Berkeley, Berkeley, CA*
[2]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA*
[3]*Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA*

Real-time time-dependent density functional theory (RT-TDDFT) is known to be hindered by the very small time step (attosecond or smaller) needed in the numerical simulation due to the fast oscillation of electron wavefunctions, which significantly limits its range of applicability for the study of ultrafast dynamics. In this work, we propose that such oscillation is mainly caused by the non-optimal gauge choice (degrees of freedom irrelevant to physical observables) of the Schrodinger equation, and can be considerably reduced by optimizing the gauge choice using the parallel transport formalism. This parallel transport dynamics can be simply interpreted as the dynamics driven by the residual vectors, analogous to those defined in eigenvalue problems in the time-independent setup.

We demonstrate the effectiveness of parallel transport using numerical results for 1D Schrodinger equations, as well as three TDDFT examples: absorption spectrum, ultrashort laser pulse, and Ehrenfest dynamics calculations. We show that the new method can utilize a time step on the order of 10~100 attoseconds, and is no less than 5~10 times faster when compared to the standard explicit 4th order Runge-Kutta time integrator. We analyze the parallel transport dynamics in the context of the singularly perturbed linear Schrodinger equation, and demonstrate its superior performance in the near adiabatic regime.