

超大規模量子物質シミュレーター-ELSESES の

FX10 および「京」での超並列計算

星 健夫

鳥取大学大学院工学研究科・JST-CREST(ポストペタ領域)

秋山 洋平・山崎 溪太

鳥取大学大学院工学研究科

超大規模量子物質シミュレーター-ELSESES[1]を、東京大学情報基盤センターFX10 および「京」で最適利用するための技術研究を行った。平成 24 年度成果[2-5]を中心に、背景や関連研究も含めて報告する。本プロジェクトは基盤アルゴリズムから産業応用までの垂直的プロジェクトであるが、個々の論点は数理・情報科学的視点から汎用であり、今後は他分野への水平的波及も目指していきたい。

1. 超大規模量子物質シミュレーター-ELSESES

電子 1 つ 1 つを「波」としてあつかう量子物質計算（電子状態計算）は、量子力学にもとづく盤石な微視的理論基盤をもつ。プログラムも多数配布され、ナノ物質系での応用は、材料・物理・化学・電気・機械など、従来学問分野を超えて広く普及している。

近年筆者(星)らは、計算科学と計算機科学の融合として、1,000 万原子系までの超大規模超並列電子状態計算コード ELSESES[1]の開発と応用に取り組んでいる。1,000 万原子系はシリコン結晶での 70nm 立方領域に相当し、産業で重要なナノ物質研究を対象にできる。また、オーダー N 法であるため¹、さらなる大規模系への展開も期待できる。中核となるのは数種類の革新的大行列数理(クリロフ部分空間)ソルバーであり、金属・半導体など幅広い物質に適用できる。これらは、第一原理にもとづくモデル化(tight-binding 型)理論を用いている。

量子物質計算は次世代産業(ナノ分野)の中核として重要なだけでなく、多様な使われ方を想定する必要があることも特徴である。具体的には、下記 2 点である：

- (i) 幅広いユーザー層（理論や HPC に明るくない実験系・産業系研究者まで）、
- (ii) 幅広いマシン環境（トップクラススパコンから PC まで）。

ELSESES は産学連携型コンソーシアムを主体として、プロジェクト(謝辞欄を参照)なども通じて、開発・応用研究が進められている。筆者(星)らによる最近の研究として、有機 EL 型 π 共役高分子固体であるアモルファス状 poly-(9,9 dioctyl flourene) (aPF) [2][3]²および、

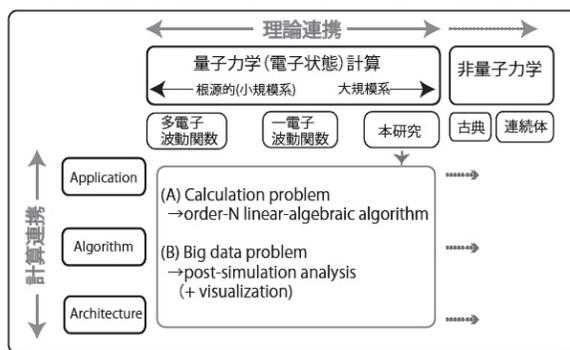
¹ オーダー N 法とは、総演算量が原子数 N に比例する手法の総称である。通常の電子状態計算では計算量が N^3 （またはそれ以上）に比例するため、原子数を 10 倍にすると計算時間が 1000 倍以上に増加し、計算機が高速になっても大規模系への発展が難しい。一方、オーダー N 法では計算量は 10 倍にしか増えない(図 3 (a) 参照)。

² 産学連携研究として行っており、論文[3]の謝辞(データ提供)として石田雅也(住友化学)があげられている。

sp2-sp3 ナノ複合カーボン固体(NCCS) [2]³がある。また、ELSEES を使った最近の応用研究として、西野信也・藤原毅夫(東大)・山崎久嗣(トヨタ自動車)・山元進(東京工科大)との共同研究による電池関連物質研究[4]もある。

計算科学の観点での大局的な位置づけを、図1(a)に示す。数原子系からマクロ系までを対象とする計算物質科学では、単一の理論手法で全てをカバーできず、種々の連携が重要となる。「理論連携」(図の横軸)の他、計算することを軸とした物理・数理・アーキテクチャ(HPC)研究間の「計算連携」(図の縦軸)も重要となる。後者を「Application-Algorithm-Architecture co-design」と言うこともある。また、計算の概要(物理学的用語を使わない説明)を図1(b)に示す。具体的には、「行列生成+大行列(線形代数的)ソルバー+物理量計算」(の繰り返し)を一体と考え、最適なアルゴリズム・ワークフローをテーラーメイドすることである。

(a) 計算物質科学における本研究の位置づけ



(b) 計算の概要(物理学用語を使わない説明)

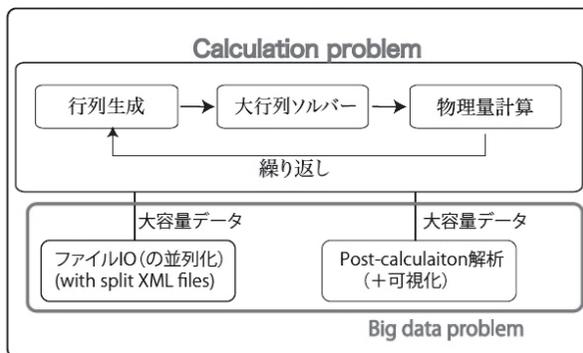


図1: (a) 計算物質科学における本研究の位置づけ。(b) 計算の概要(物理学的用語を使わない説明)。

数理・情報科学的視点から課題を大別すると、

- (i) Calculation 問題・・・計算量が多い問題(入力データ量 << 出力データ量)。
- (ii) Big data 問題・・・データ量が多い問題(入力データ量 >> 出力データ量)。

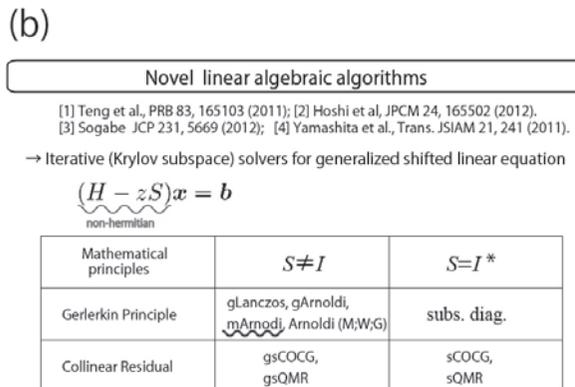
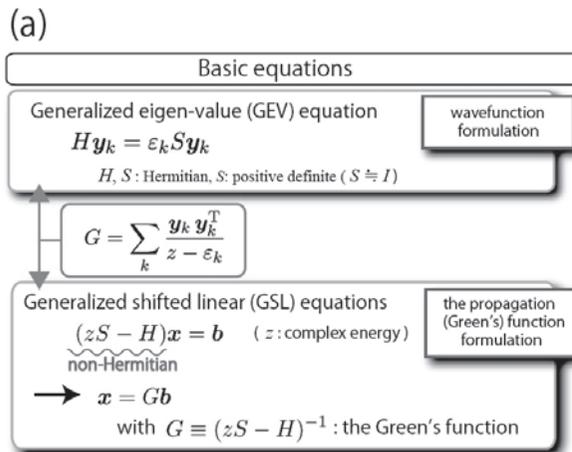
にわけられる。電子状態計算は通常(i)に相当する。入力データは原子座標(原子数 N に対して

³ 日本初の超強度材料として産業利用が期待される、ナノ多結晶ダイヤモンド(NPD, Irifune, et al., Nature 421, 599 (2003))の形成過程研究である。

3N 個の実数データ)のみであり、出力データはそれより遥かに膨大な電子状態データになる。しかし、「京」などでの超大規模計算から膨大なデータが生成され、post-simulation 解析としての(ii)も重要になる(本記事の 3 を参照)。

2. 超並列オーダーN数値アルゴリズムと「京」コンピュータ

「Calculation problem」(図 1 (b))の基盤の数値アルゴリズムとして、超並列オーダーN計算理論構築に取り組んだ。ここでは、従来型の問題設定である一般化固有値問題($H\mathbf{y} = eS\mathbf{y}$)ではなく、一般化シフト型線形方程式($(zS-H)\mathbf{x} = \mathbf{b}$)を数理的基礎としている(図 2(a))。これはグリーン関数(伝搬関数)表示の量子力学に対応している。



*Takayama et al., JPSJ73, 1519 (2004); Takayama et al. PRB 73, 165108(2006); Sogabe et al. ETNA31, 126 (2008)

第 2 図: (a)基礎方程式。従来型(一般化固有値問題)と本研究(一般化シフト型線形方程式)の比較。(b)新しいクリロフ部分空間解法の一覧。

図 2(b)にここ数年来の研究を通じて構築した、一般化シフト型線形方程式に対する革新的解法をまとめた。具体的には、クリロフ部分空間理論に基づく新しいソルバーを 6 種、構築した。これらは大別すると、Gerlerkin 原理にもとづく 4 種の解法(gLanczos, gArnoldi, mArnoldi, Arnoldi(M, W, G))と共線残差定理(Collinear Residual Theorem)に基づく 2 種の解法

(gsCOCG, gsQMR)に分類される[3, 5, 6, 7]。これらは、曾我部知広（愛知県立大）、張紹良（名古屋大）、宮田考史（名古屋大）らとの共同研究による。さらに理論的發展として、Gerlerkin原理にもとづく4種の解法が統一的理論体系に組み込めることを示した[8]。

応用計算として、上記アルゴリズム(主として mArnoldi 法)を、分散メモリ型超並列計算(MPI/OpenMP 複合並列計算)として実装し、「京」コンピュータの 98, 304 CPU コアまでで実行した[2, 3]。図 3(a)にオーダー N スケーリングを示し、(b) では 1000 万原子系での並列性能(‘strong scaling’)を示した[2][3]。(b)での並列効率は、NCCS 系が 95%、aPF 系が 79%であった。通信コスト・メモリコストなどのアーキテクチャ上の工夫も重要であり[2]、Application-Algorithm-Architecture co-design の例と言える。

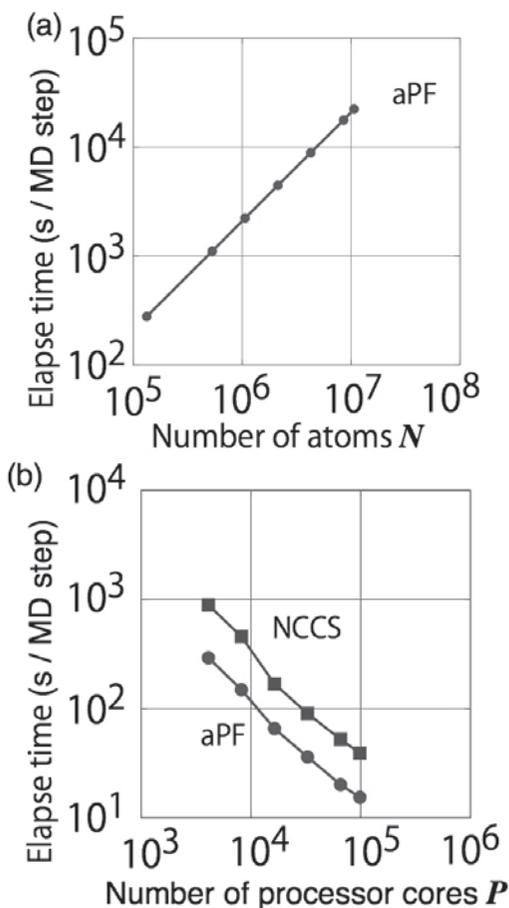


図 3 : ELSES によるベンチマーク。(a) オーダー N スケーリング[3]。(b) 「京」における、約 1,000 万原子系での並列効率 (ストロングスケーリング) [2]。物質系としては、有機 EL 型 π 共役高分子系であるアモルファス状 poly-(9,9 diocetil flourene) (aPF) および、sp²-sp³ ナノ複合カーボン固体(NCCS)をとりあげた。

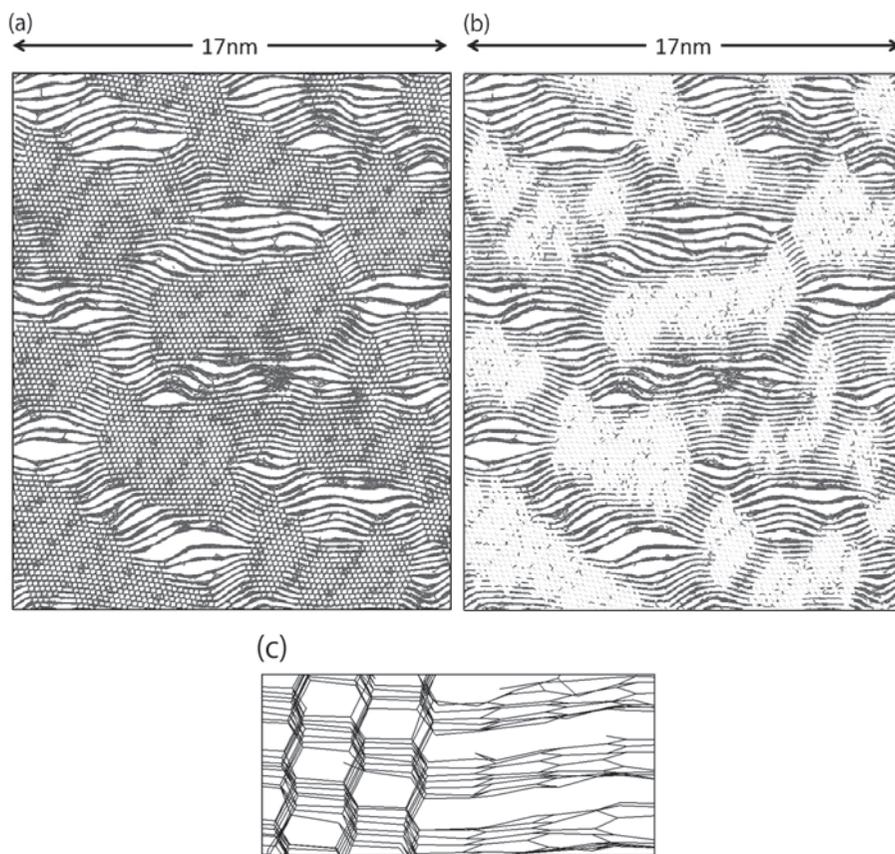
これらソルバー群は数理的に汎用な解法であり、他物理分野にも適用可能である。例えば、図 2(b)にある基本数理原理である共線残差定理[9](レビューとして[10])はドイツの数学者 Frommer によるものであるが、QCD の計算例[9][11]とともに論じされている。我々の研究とし

ては大規模系(最初の論文は[12])の他、多体問題 (full-CI 型計算) [13][14]がある。さらに、shell model 計算[15]、GW 計算[16]、TDDFT 計算[17]など、多彩な分野での応用研究がある。

3. 超並列量子解析および可視化

複雑なナノ物質計算から物理的知見をえるための結果解析は、膨大な電子状態データを対象とする「big data」問題である。出力データである電子状態データは電子を（空間分布をもった）「波」として記述し、1000 万原子系では、TB から原理的には PB 以上となりうる。

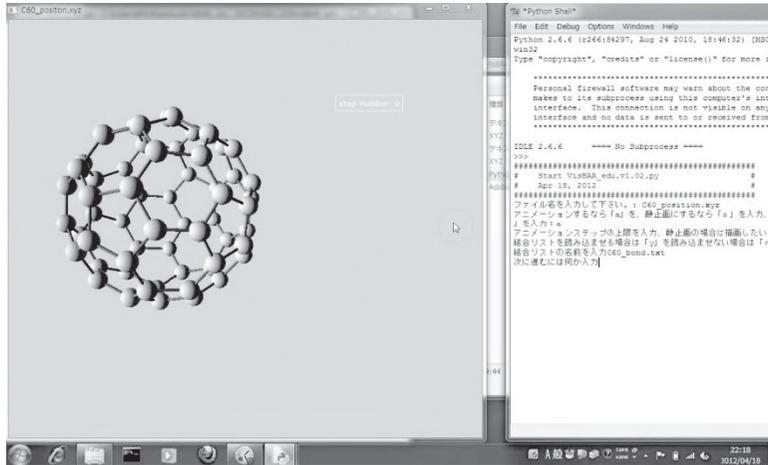
このような「big data」解析には、データ分散性を意識した数値解析手法が必要である。例として、10nm スケール系である sp²-sp³ ナノ複合カーボン固体(NCCS)に取り組んだ[2]。結果を図4に示す。



第4図: COHP, π COHP (量子解析) 理論を用いた、sp²-sp³ ナノ複合カーボン固体系(NCCS)の可視化解析[1]。(a) sp²/sp³ ドメイン両方の可視化。(b) sp² ドメインのみの可視化。(c) sp²/sp³ ドメイン境界の拡大図。

図4では、グリーン関数をもちいた量子解析手法である π COHP 理論[2]を提案し、sp²(グラファイト的)領域と sp³(ダイヤモンド的)領域を、定量的・自動的に判別した。我々の大規模計算は、一般化シフト型線形方程式を通じてグリーン関数を計算する手法であるため、データ分散性を完全に保持したまま（ノード間通信を一切使わずに）上記解析が実行された。

さらに、数値解析と可視化の一体処理も必須である。これを実現するために、Python による独自可視化ツール「VisBAR」⁴を開発した。関連して、機能を極力縮小した約 100 行のツール「VisBAR_edu」⁵(図 5)も開発し、学部教育で活用している。このようなツールは、他研究者が独自に可視化コードを開発する際の有用なソースとなるであろう。



第 5 図: Python による自作可視化コード「VisBAR_edu」のスクリーンショット。
可視化しているのはフラーレン(C60)。

4. 分割 XML ファイルによる超並列ファイル I/O 機構の利用

大規模系における技術的論点として、ファイル I/O の並列化があげられる。1,000 万原子系の入力データは XML 形式で約 1.8GB である⁶。この場合、ファイル I/O コストが無視できない。分割 XML ファイルを基礎として、FX10 および「京」での超並列ファイル I/O 機構（「ランクディレクトリ」機構）を用いることで回避できた。具体的には、初期処理にかかる時間が 20 分以上から 1 分へと劇的に加速された。

XML ファイルは標準的な情報技術であるが、分割ファイル I/O を用いる立場でもメリットがある。XML 形式では全ての情報をタグで論理的に定義するため、分割しても well-defined な XML ファイルであるからである⁷。

一般に、ファイル分割数=ノード数 (MPI 並列度数)、と設定する。例えば 4 並列むけ入力ファイルは、

```
sample_000000.xml
sample_000001.xml
sample_000002.xml
```

⁴ VisBAR(=Visualization tool with Ball, Arrow and Rod) [2]

⁵ http://www.damp.tottori-u.ac.jp/~hoshi/visbar_edu/

⁶ 1 原子あたりの原子座標データは、(x, y, z) の 3 成分の倍精度実数が必要となる。概算すると 1000 万原子系では、(8 バイト) × (3 成分) × (1000 万) = 約 0.24GB、となる。XML ファイルはタグなどの情報も含まれるため、実際のファイルは 1.8GB となった。

⁷ 伝統的な科学分野データフォーマット (例えば原子構造を記述する XYZ 形式) では、各行に記すべき量が決まっている、など、1 ファイルであることが前提として定義されている。

```
sample_000003.xml
sample_basic.xml
```

の5種のファイルとなる。通し番号付きのファイルには、原子構造データ（主に原子の座標データ）が記載されている。F分割した場合は、（全原子数）/F程度の原子構造データ、それぞれのファイルに格納される。並最後の「basic」ファイルは、全ノードで読み込むべき共有データ（数十行のみ）が記述されている。「京」でのランクディレクトリ機構をもちいた、ジョブスクリプトファイルの例をあげる。

```
#!/bin/bash -x
#
#PJM --rsc-list "node=8x8x8"
#PJM --rsc-list "elapsed=03:00:00"
#PJM --rsc-list "rscgrp=small"
#PJM --stg-transfiles all
#PJM --mpi "use-rankdir"
#PJM --stgin "rank=* ./elses %r:./"
#PJM --stgin "rank=* ./config.xml %r:./"
#PJM --stgin "rank=* ./sample_basic.xml %r:./"
#PJM --stgin "rank=* ./sample_%06r.xml %r:./"
#PJM --stgout "rank=0 ./Output.txt ./out/"
#PJM --stgout "rank=0-7 ./log-node%06r.txt ./out/"
#PJM --stgout "rank=* ./std-file.%r ./out/"
#PJM -s
#
. /work/system/Env_base
export PARALLEL=8
export OMP_NUM_THREADS=$PARALLEL
mpixec -n 512 -of-proc std-file --mca mpi_print_stats 1 lpgparm
-s 4MB -d 4MB -h 4MB -t 4MB -p 4MB ./elses ./config.xml
（注：最後の2行は、実際のファイルでは1行で書く。）
```

5. 展望：行列データ公開・擬似アプリ・複合数値ソルバー・クラウドサービス

本研究の展望を述べる。物理・数値・HPC分野の連携（Application-Algorithm-Architecture co-design）は、ポストペタ（ポスト「京」）時代にもならみ、さらに重要となろう。物質科学の展望としては、さらなる加速だけでなく、電子物性への適用など、質的な発展も重要となる。

数値と物理の接点として、疎行列データの公開を開始している[18]。疎行列問題では、行列と解法の両者の性質が重要であり、実問題のデータ公開は研究を促進させると期待している。

物理分野のニーズに合わせた展望としては、種々の数値手法を複合する、複合数値ソルバー研究があげられる。基盤となる新しい解法（シーズ）としては、クリロフ部分空間以外の研究（例えばシルベスタ慣性定理に基づく研究[19]）も開始している。

本研究以外に大行列固有値問題の研究としては、例えば櫻井鉄也(筑波大) [20, 21, 22]・今村俊幸(理研) [23, 24, 25]・山本有作(神戸大) [26, 27]らの研究がある。海外の関連ライブラリとしては、ELPA[28]、FEAST[29]などがある。こうした新しい数理的発展を必要に応じて複合した「数値計算エンジン」を構築し、各種スパコン上の物理シミュレーションへと生かしていくことが重要となろう。

数値計算エンジンと実アプリとの中間に位置する「擬似アプリ」(小型アプリ)のひな形も試験公開している。ここでいう擬似アプリとは、単体で動作し、

(a) 実アプリ開発者が数値計算エンジンの性能評価ができること、
(b) 数値計算エンジンを実アプリへの実装をするさいの「手本」となること、
(c) 数値計算エンジン開発者が実アプリからのニーズを理解することを助けること、
の3点を意図している。実アプリは複雑で長過ぎるため、どうしても個別的に開発することになる。上記のような比較的短い「擬似アプリ」を通じて知見を共有し、Application, Algorithm, Architectureの連携研究が発展することが期待できる。

大局的な展望としては、産業利用のためのスパコン上の「研究型クラウドサービス」とし、物質(材料)科学における産学連携研究拠点に昇華させたい。そこでは、(i)ハードウェア・(ii)ソフトウェア・(iii)オープンデータ・(iv)コミュニティ(ソフト開発者およびソフトウェアユーザー(企業系研究者など))、の全てが密に連携していることを想定する。そこでソフトウェアは、大規模系計算と高精度(小規模系)計算を組み合わせた「マルチスケール型量子物質計算」が、特に有用となろう。オープンデータは、(許される範囲で)計算結果を戦略的に蓄積し再利用可能とすることで、産業利用を促進させることを念頭におく。ただし、量子計算データ量は膨大となるため、これらを扱う情報(big data)技術が必要となる。スパコンも物質(材料)科学も伝統的に日本が強い分野であり、両者が融合すれば、世界標準の「研究型クラウドサービス」へと押し上げることができる。

謝辞

本研究は、東京大学情報基盤センター「スーパーコンピューター若手利用者推薦」平成24年度後期利用に基づく。「京」コンピュータ利用は、課題番号hp120170およびhp120280を通じて行った。予算補助として、JST-CREST「ポストペタスケールに対応した階層モデルによる超並列固有値解析エンジンの開発」・JST-ASTEP「ナノスケール材料向け超大規模電子構造計算プログラムの実用化研究開発」、科研費(No. 23104509, 23540370)を利用した。

参考文献

- [1] ELSSES(=Extra Large-Scale Electronic Structure calculation): <http://www.elses.jp/>
- [2] T. Hoshi, Y. Akiyama, T. Tanaka and T. Ohno, ‘Ten-million-atom electronic structure calculations on the K computer with a massively parallel order- N theory’, J. Phys. Soc. Jpn. 82, 023710, 4pp. (2013).
- [3] T. Hoshi, S. Yamamoto, T. Fujiwara, T. Sogabe, S.-L. Zhang, ‘An order- N electronic structure theory with generalized eigenvalue equations and its application to a

- ten-million-atom system', J. Phys.: Condens. Matter 24, 165502, 5pp. (2012)
- [4] S. Nishino, T. Fujiwara, H. Yamasaki, S. Yamamoto, T. Hoshi, 'Electronic structure calculations and quantum molecular dynamics simulations of the ionic liquid PP13-TFSl, Solid State Ionics 225, 22-25 (2012).
- [5] T. Sogabe, T. Hoshi, S.-L. Zhang, T. Fujiwara, 'Solution of generalized shifted linear systems with complex symmetric matrices', J. Comp. Phys. 231, 5669-5684 (2012).
- [6] H. Teng, T. Fujiwara, T. Hoshi, T. Sogabe, S.-L. Zhang, S. Yamamoto, 'Efficient and accurate linear algebraic methods for large-scale electronic structure calculations with nonorthogonal atomic orbitals', Phys. Rev. B 83, 165103, (2011).
- [7] 山下達也, 宮田考史, 曾我部知広, 星健夫, 藤原毅夫, 張紹良, '一般化固有値問題に対する Arnoldi(M, W, G) 法', 日本応用数理学会論文誌 21, 241 (2011).
- [8] T. Hoshi, S. Yamamoto, T. Fujiwara, T. Sogabe, S-L. Zhang, Y. Akiyama, T. Ohno, 'Ten-million-atom electronic structure calculations with novel linear-algebraic algorithm and the K computer' (invited talk), International Symposium on Computing: Quantum Simulation and Design (ISC-QSD), Osaka University Hall, 11-13 Oct. (2012).
- [9] A. Frommer, 'BiCGStab(1) for Families of Shifted Linear Systems', Computing 70, 87 (2003).
- [10] 曾我部知広, 張紹良, '大規模シフト線形方程式の数値解法: クリロフ部分空間の性質に着目して', 応用数理 19, 163-178, (2009).
- [11] J. C.R. Bloch, T. Brey, A. Frommer, S. Heybrock, K. Schafer, T. Wettig, 'Short-recurrence Krylov subspace methods for the overlap Dirac operator at nonzero chemical potential', Comp. Phys. Comm. 181, 1378 (2010)
- [12] R. Takayama, T. Hoshi, T. Sogabe, S.-L. Zhang, and T. Fujiwara, 'Linear algebraic calculation of the Green's function for large-scale electronic structure theory', Phys. Rev. B 73, 165108 (2006)
- [13] S. Yamamoto, T. Fujiwara, and Y. Hatsugai, 'Electronic structure of charge and spin stripe order in $\text{La}_{2-x}\text{Sr}_x\text{NiO}_4$ ($x=1/3, 1/2$)', Phys. Rev. B 76, 165114 (2007).
- [14] S. Yamamoto, T. Sogabe, T. Hoshi, S.-L. Zhang and T. Fujiwara, 'Shifted conjugate-orthogonal conjugate-gradient method and its application to double orbital extended hubbard model', J. Phys. Soc. Jpn. 77, 114713 (2008).
- [15] T. Mizusaki, K Kaneko, M. Honma, T. Sakurai, 'Filter diagonalization of shell-model calculations', Phys.Rev. C 82, 024310 (2010)
- [16] F. Giustino, M. L. Cohen, and S. G. Louie, 'GW method with the self-consistent Sternheimer equation', Phys. Rev. B 81, 115105 (2010)
- [17] 篠原康, 二村保徳, 矢花一浩, 櫻井鉄也, 'Krylov 部分空間法に基づくシフト線形方程式による TDDFT の線形応答計算', 日本物理学会, 2012 年 3 月.
- [18] ELSEs matrix library; <http://www.elses.jp/matrix/>
- [19] D. Lee, T. Miyata, T. Sogabe, T. Hoshi, S.-L. Zhang, 'Eigenvalue computation for a particular need arising from electronic structure calculation', The 8th East Asia SIAM Conference (EASIAM 2012), Taiwan, 25-28. Jun. (2012).

- [20] Y. Futamura, T. Sakurai, S. Furuya and J.-I. Iwata, ‘Efficient Algorithm for Linear Systems Arising in Solutions of Eigenproblems and its Application to Electronic-Structure Calculations’, Proc. VECPAR2012, LNCS 7851, pp.226-235 (2012)
- [21] 櫻井鉄也, フィルタ対角化による大規模固有値問題の解法, システム制御情報学会誌 55 (8), pp. 333-338 (2011)
- [22] <http://h4es.cs.tsukuba.ac.jp/>
- [23] T. Imamura, S. Yamada, M. Machida, “Development of a High Performance Eigensolver on the Peta-Scale Next Generation Supercomputer System”, Progress in Nuclear Science and Technology, the Atomic Energy Society of Japan, Vol. 2, pp.643-650 (2011) .
- [24] T. Imamura, S. Yamada, M. Machida, “Eigen-K: high performance eigenvalue solver for symmetric matrices developed for K computer”, 7th International Workshop on. Parallel Matrix Algorithms and Applications (PMAA2012), June 2012, London UK.
- [25] 今村俊幸, 山田進, 町田昌彦, 「Eigen-Exa: ポストペタ スケール環境での 密行列固有値ソルバー開発」, 第17回計算工学会講演会, 京都教育文化センター, 2012年5月30日
- [26] Y. Takahashi, Y. Hirota and Y. Yamamoto, ‘Performance of the Block Jacobi Method for the Symmetric Eigenvalue Problem on a Modern Massively Parallel Computer’, Proceedings of ALGORITMY 2012, pp. 151-160 (2012).
- [27] 廣田悠輔, 橋本拓也, 山本有作, ‘倍精度正方行列特異値分解アルゴリズムのGPGPU上での性能・精度評価’, 情報処理学会論文誌コンピューティングシステム (ACS), Vol. 5, No. 5, pp. 163-176 (2012).
- [28] <http://elpa.rzg.mpg.de/>
- [29] <http://www.ecs.umass.edu/~polizzi/feast/>
- [30] Eigen Test: http://www.damp.tottori-u.ac.jp/~hoshi/eigen_test/