Implementation of Parallel Matrix Diagonalization for Ab-Initio Molecular Dynamics Program using ScaLAPACK

Indra Gunawan[a,b], Masao Obata[a], Makoto Nakamura[a], Muhamad A. Martoprawiro[b], Tatsuki Oda[a,c]

[a]Grad. School of Natural Sci. and Tech., Kanazawa University, Kakuma, Kanazawa, 920-1192 Japan, Email: ehmind@yahoo.com

[b]Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132 Indonesia

[c]Institute of Science and Engineering, Kanazawa University, Kakuma, Kanazawa 920-1192 Japan, E-mail: oda@cphys.s.kanazawa-u.ac.jp

Abstract. *A matrix diagonalization is one of the most time-consuming parts of the ab-initio molecular dynamics programs. The bigger system to be simulated, the bigger matrix needs to be diagonalized to obtain the eigenvalues and eigenvectors. To raise this matter, we have applied matrix diagonalization provided by Scalable Linear Algebra PACKage (ScaLAPACK) to the ab-initio molecular dynamics program. We analyzed the efficiency of RS routine provided by EISPACK and PDSYEVX and PZHEEVX by ScaLAPACK.*

Keywords: matrix diagonalization, ScaLAPACK, parallel calculation, ab-initio molecular dynamics

1 Introduction

Approach based on quantum mechanics is an essentially important tool in the analysis for materials. The calculations using ab-initio methods involve inter-atomic interaction that able to get more accurate result [1]. However, the problem is that the ab-initio calculations require considerable computer resources. This is the one of reasons why the high-speed computer is needed. At present, the only high-speed computer is still not enough to meet these computing needs and then these computers can run in parallel to be able to work faster [2].

Car-Parrinello (CP) method [3] is one of ab-initio molcular dynamics (MD) methods. This has contributed for long time the study on structural and electronic property in crystal, clusters, biological matter, etc.. The CP method uses a classical Lagrangian, and envolves the electronic wavefunction and the ionic coordinates. This Lagrangian has to be solved with the Lagrange multiplier which is introduced to realize the constraint of orthonormalixation among wave functions.

In the original method, such Lagrange multiplier can be solved in relatively simpler way [3], while in the formulation with the ultrasoft pseudo potential or similar scheme, the multiplier forms the matrix equation given by

$$S + \lambda R + R^{\dagger}\lambda^{\dagger} + \lambda T \lambda^{\dagger} = 1, \tag{1}$$

where λ_{ij} is propotional to the Lagrange multiplier Λ_{ij}[4, 5], The square matrices of R, S, T have the size of dimention to the number of electronic states included. Definition of R, S, T can be refered to the original work [4]. Solving the equation (1), eigenvalues and eigenvectors for the Hermitian part of R are needed as numerical values. This matrix diagonalization is time-consuming in application of large system, then it is necessary to reduce the time consumption. There are so many methods

Table 1: Properties of computer in investigation

Computer Name	Num. of Proc. Used	Proc. Specifications	Memory
Kaga100	8	Intel (R) Xeon(R) CPU X5680 3.33GHz	24 GB

Table 2: Target routines for investigation

Name of subroutine	Package	Matrix type	Data type	Computation
RS	EISPACK	Symmetric	Real	Single
PDSYEVX	ScaLAPACK	Symmetric	Double	Parallel
PZHEEVX	ScaLAPACK	Hermitian	Double complex	Parallel

that can be used to solve the problem of matrix diagonalization, which can run on a single computer or in a parallel computer.

Irrespective of method, the matrix diagonalization which contains the matrix multiplication needs a number of operations propotional to the cubic of system size. The goal of this work is to solve the problem of time consuming that occurs in the matrix diagonalization in our architecture. This paper describes the work when we have applied matrix diagonalization provided by Scalable Linear Algebra PACKage (ScaLAPACK)[6] to our ab-initio molecular dynamics program [7]. We reported efficiency of ScaLAPACK compared with EISPACK [8] and the growth pattern in consumption time for diagonalizing matrix.

2 Investigation of running time on package routine

The properties of computer machine that used for our investigation are summarized in the Table 1 and the routines investigated are listed in Table 2. EISPACK contains nine subroutines from nine classes of matrices for calculating the eigenvalues and eigenvectors. From them we investigated one subroutine, RS, for real symmetric matrix. ScaLAPACK contain routines for solving systems of linear equations, least squares problems, and eigenvalue problems. We use two subroutines from ScaLAPACK; PDSYEVX and PZHEEX. The former is for solving eigenvalue problems for symmetric double matrix and the latter for Hermitian matrix. ScaLAPACK will run on any machine where both the Basic Linear Algebra Subprograms (BLAS) and Basic Linear Algebra Communications Subprograms (BLACS) are available.

The time consumption is shown in Figure 1. In RS routine, time consumption is growing as the bigger dimension of matrix. The PDSYEVX and PZHEEVX have similar time consumption with RS in small size dimension where less equal than 2560. But starting from 2816 matrix dimension, the time consumption is decreased then grows again. It is shown that the parallelization of ScaLAPACK have main advantage, compared with EISPACK. During the test we encountered memory problems in large matrix dimension. This problem appears while the dimension matrix is 6144 in PDSYEVX and 7168 in PZHEEVX, and does not occur while the dimension until 7168.

In Figure 2, the log-log plot for the growing time is also presented, showing interesting properties on time growth clearly. The time growth follows the formular,

$$time\ growth\ \cong\ \alpha N^A, \tag{2}$$

where the N and the scaling factor A repredent matrix size and gradient of the line in the figure,

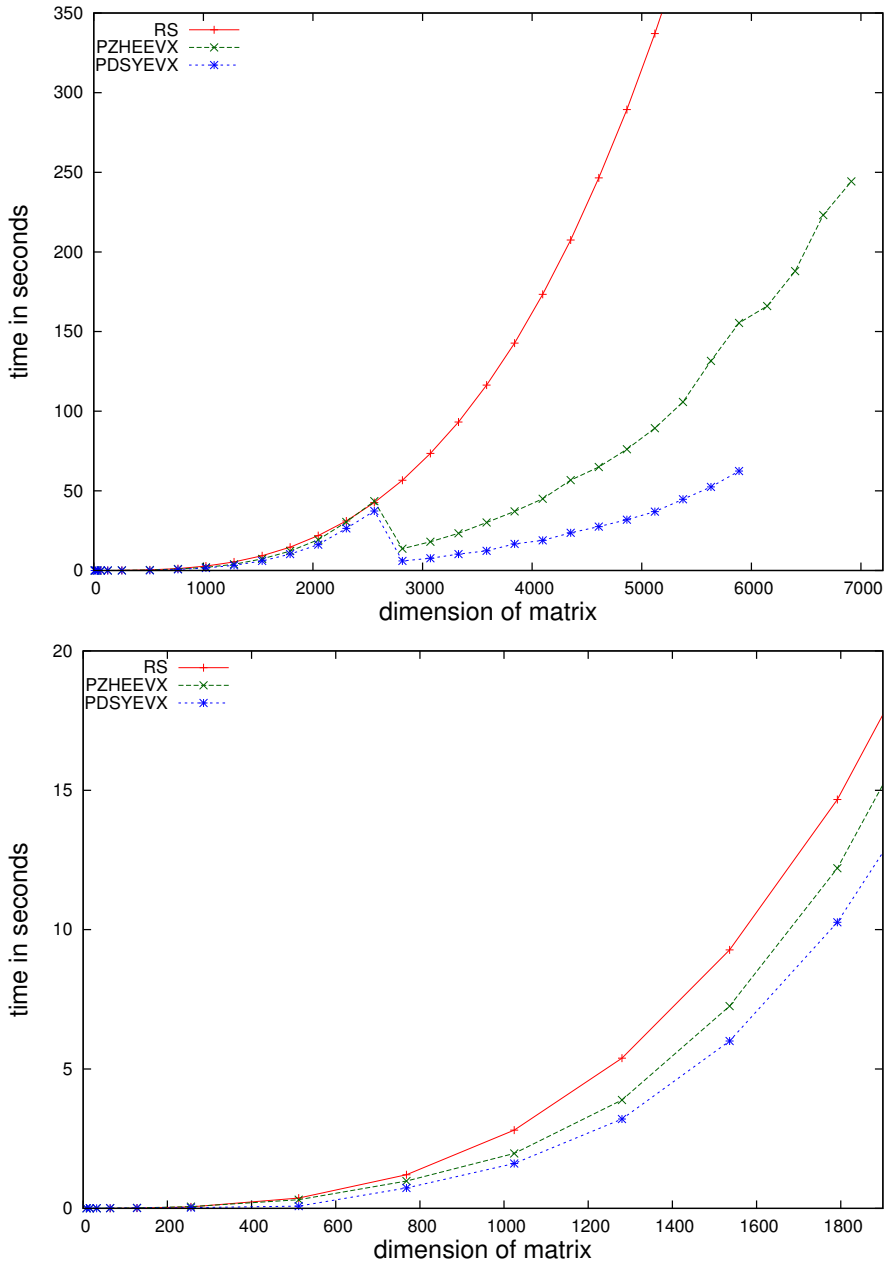

Figure 1: Running time of symmetric eigensystems in EISPACK (RS), symmetric eigensystems and Hermitian eigensystems (PDSYEVX and PZHEEVX) in ScaLAPACK. Enlargement (lower panel) at small dimensions.

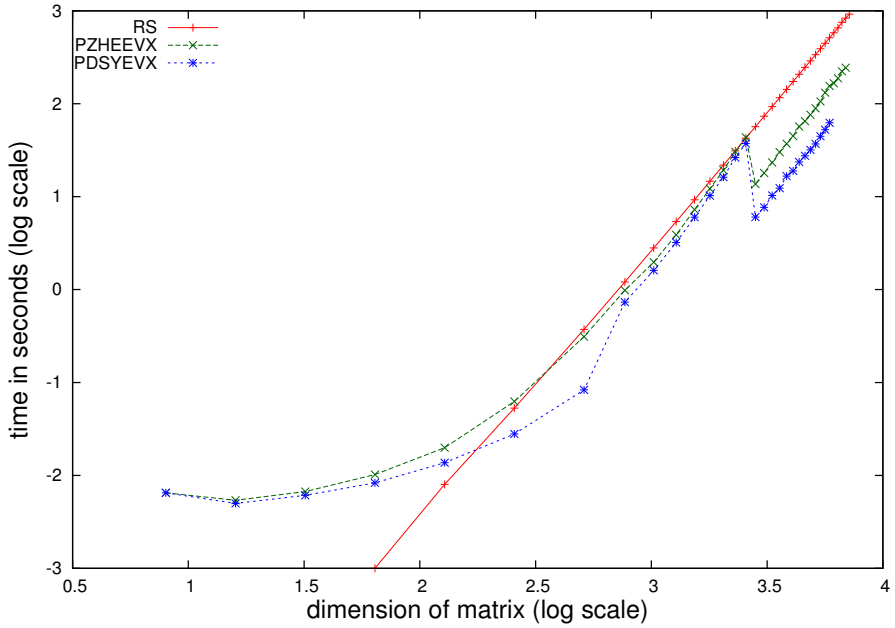

Figure 2: Running time in log-log plot scale for Symmetric Eigensystems in EISPACK (RS), Symmetric Eigensystems and Hermitian Eigensystems in ScaLAPACK (PDSYEVX and PZHEEVX).

Table 3: Scaling factor,A, estimated from the gradient of line found in Figure 2.

Matrix Dimension	A(RS)	A(PDSYEVX)	A(PZHEEVX)
8-64	2.92	0.13	0.23
128-768	2.92	2.06	2.17
1024-2560	2.92	3.46	3.40
2816-7168	2.92	3.12	3.22

respectively. In Table 3 the gradient is listed for the range of matrix dimension. In RS, the scaling factor (A=2.92) does not change in the range investigatted. It means that the time growth is proportional to the scale near the cubic one. The latter scale is a consequence of standard one in the algorithm of matrix diagonalization. Different with RS, PDSYEVX and PZHEEVX show that, the value of A increases from a small number for the small matrix to higher for the bigger matrix then stable around 3.

3 Discussion

As shown by the present investigation on scaLAPACK the routine gets efficiency at larger matrix dimensions, when the matrix dimension of 3000 is allowed, for example, the 1500-atoms Si system in which every atom needs the couple of orbitals at least, can be a target on study. Following a similar consideration the efficency on diagonalization will be offered also in systems which contains 750 water molcules (four valence orbitals in a water), the Fe(5ML)/MgO(5ML)/Fe(5ML) film [9] with 4×4 in-plane superlattice configuration, three small peptides which hold the transition metal

clusters [10], 6 units of MOR type zeolite network [11], liquid oxygen with 250 molecules, etc.. It is noted that the limitation on application is not only at the size of matrix dimention, but also at the computational time for preparing matrices (R, S, T), in equation (1).

4 Summary

We have investigated efficiency of the routine for matrix diagonalization; RS from EISPACK and PDSYEVX and PZHEEVX from scaLAPACK. The time growth with respect to the size of matrix clearly shows the cubic scaling feature which is built in the procesure of matrix diagonalization. In the present work, the ScaLAPACK showed the advantage after the dimension of matrix reaches to 2816 or more, when taking 8 degrees in parallel computation. The ScaLAPACK have a good opportunity to improve CP-MD performance for the large systems with more than about 3000 matrix dimension.

References

[1] D. Marx and J. Hutter (2009). *Ab Initio Molecular Dynamics, Basic Theory and Advanced Methods*. Cambridge University Press, New York.

[2] Peter S. Pacheco (1997). *Parallel Programming with MPI*. Morgan Kaufmann Publishers, Inc., California.

[3] R. Car and M. Parrinello (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory *Physical Review Letters*, **55**, 2471–2474.

[4] K. Lasasonen, A. Pasquarello, R. Car, C. Lee, D. Vanderbilt (1993). Car-Parinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials *Physical Review B*, **47**, 10242–10153.

[5] P. E. Blöch (1994). Projector augmented-wave method *Physical Review B*, **50**, 17953–17979.

[6] J. Choi, et al, (1996). ScaLAPACK: a portable linear algebra library for distributed memory computers - design issues and performance. *Computer Physics Communications*, **97**, 1–15.

[7] Tatsuki Oda (2002). Metallic Car-Parinello Dynamics. *J. Phys. Soc. Jpn.*, **71**, 519–524.

[8] Burton S. Garbow (1974). Eispack - A Package of Matrix Eigensystem Routines* *Computer Physics Communications*, **7**, 179–184.

[9] W. H. Butler (2001). Spin-dependent tunneling conductance of Fe/MgO/Fe sandwiches *Physical Review B*, **63**, 054416(1–12).

[10] T.Oda, H.Nagao, A.Sugiyama, and K.Wada (2005). Electonic structure and magnetism at the active site in Ferredoxin: Ab initio approach to $(Fe_2S_2)^{2+}$ complex with 1st peptide shell, *Polyhedron*, **24**, 2550–2556.

[11] The structure of MOR type zeolite can be found at http://www.iza-structure.org/.