First Principles Molecular Dynamics Simulations of Solution

Group Representative
Masaru Hirata Japan Atomic Energy Research Institute, Senior Scientist

Author
Takashi Ikeda Japan Atomic Energy Research Institute, Research Scientist

The detailed process of the optimization of the first principles molecular dynamics (MD) code CPMD used in the project is given. The performance of our tuned code is examined for ambient water that is one of key materials of the solution sciences. We conclude that the first principles molecular dynamics simulations for the system of up to 1000 atoms are feasible using 32 nodes.

Keywords: Solution, First Principles Molecular Dynamics, BLAS, OpenMP, Ambient Water

1. Code optimization

Our project named "First principles Molecular Dynamics Simulations of Solution" actually started from February 2003. Thus, in this report we mainly give the detailed process of the code optimization to which we have been devoted for almost two months since the project was initiated.

The Car-Parrinello Molecular Dynamics (CPMD) code uses plane-waves as basis functions for describing electronic orbitals and the code itself is constructed by employing LAPACK and BLAS libraries as the ingredients. Thus, the high computational cost is paid for some BLAS routines and parallel FFT. It is known that a primary routine is normally parallel FFT for the simulations of about 200 atoms which can be now performed routinely on usual supercomputers. On the contrary, it turns out that for large scale simulations of about 1000 atoms about 70 % of the computational cost is spent for one of the BLAS routine DGEMM used frequently, e.g., for orthogonalizing the orbitals. This is partly due to the change of the inter-node parallelization from MPI to OpenMP. A typical subroutine call of DGEMM is written as

```
CALL DGEMM('N','N', 2*NGW,NSTATE,NSTATE ,
-1.0D0,A,2*NGW,B,NSTATE,1.0D0, C, 2*NGW).      (1)
```

The corresponding lines in the original DGEMM is shown in Fig. 1. Here, K=N=NSTATE and M=2*NGW. The number of NSTATE and NGW changes depending on both the number of atoms contained in your system and the number of nodes you use. However, the number of NSTATE typically ranges from ~10 to ~100 while that of NGW ranges from ~10000 to ~100000. Considering these typical numbers, we modified the DGEMM as sketched in Fig. 2; The most outer-loop is parallelized with OpenMP and the unrolling of the

```
      DO 90, J = 1, N
         IF( BETA.EQ.ZERO )THEN
            DO 50, I = 1, M
               C( I, J ) = ZERO
50          CONTINUE
         ELSE IF( BETA.NE.ONE )THEN
            DO 60, I = 1, M
               C( I, J ) = BETA*C( I, J )
60          CONTINUE
         END IF
         DO 80, L = 1, K
            IF( B( L, J ).NE.ZERO )THEN
               TEMP = ALPHA*B( L, J )
               DO 70, I = 1, M
                  C( I, J ) = C( I, J ) + TEMP*A( I, L )
70             CONTINUE
            END IF
80       CONTINUE
90    CONTINUE
```

Fig. 1 The original DGEMM routine used, e.g., for orthogonalizing electronic orbitals.

```
      MODK=MOD(K,8)
!$OMP parallel do private(J,I,L,T0,...,T7)
!$OMP+ shared(MODK)
      DO J=1,N
        IF(BETA.NE.ONE) THEN
          DO I=1,M
            C(I,J)=BETA*C(I,J)
          ENDDO
        ENDIF
        DO L=1,MODK
          T0=ALPHA*B(L,J)
          DO I=1,M
            C(I,J)=C(I,J)+T0*A(I,L)
          ENDDO
        ENDDO
        DO L=MODK+1,K,8
          T0=ALPHA*B(L  ,J)
          ...
          T7=ALPHA*B(L+7,J)
          DO I=1,M
            C(I,J)=C(I,J)+T0*A(I,L  )+...+T7*A(I,L+7)
          ENDDO
        ENDDO
      ENDDO
```

Fig. 2 A part of modified DGEMM corresponding to the lines shown in Fig. 1 (BETA ≠ ZERO).

intermediate loop is done explicitly. After the modification 77.62 % of the peak performance is obtained for the routines concerned. As the overall performance we achieve the vector operation ratio 99.31 % and the parallelization ratio 99.8835 % satisfying the requirement of using up to 32 nodes.

2. Performance

Ambient water is, needless to say, one of the key materials in the science of solution. To date we have used a model consisting of 32 or 64 water molecules in periodically replicated cubic cell of the side of about 10 Å for first principles MD simulations of ambient water. Here, we examine to what extent we can actually increase the system size and how much computation time is required in reality to carry out the simulations of such a large system. Figure 3 shows a summary of elapsed time per MD step for 128, 256, and 512 water molecules on the Earth Simulator. Although we are able to perform the simulations of 512 water molecules when more than 16 nodes are available, necessary computation time is still well out of reach even using 32 nodes to observe fast relaxation processes of the rate of ~10 ps occurring in solution. We, therefore, conclude that at present the first principles molecular dynamics simulations of the system of up to about 1000 atoms are feasible routinely using 32 nodes. Even though such a system is not large enough to model multi-component solvents, we are capable of studying fundamental processes of chemical reactions occurring in, for instance, aqueous solutions. Thus, our project is now going along this line.

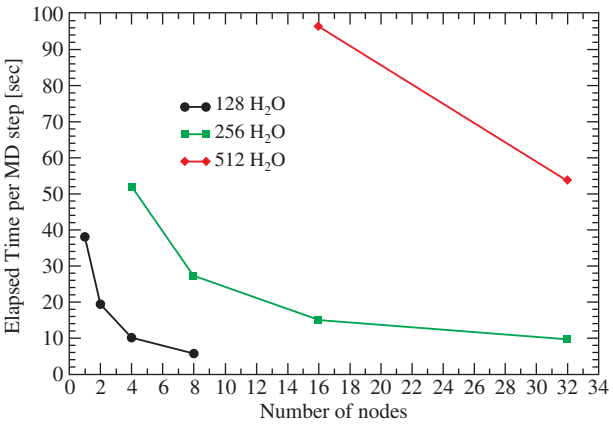

Fig. 3 Elapsed time per MD step as a function of the number of nodes for ambient water. The electronic structure was obtained using the plane-wave pseudopotential approach to density functional theory in the Perdew-Burke-Ernzerhof approximation. The valence-core interaction was described by the Troullier-Martins and the Car-von Barth pseudopotentials for oxygen and hydrogen, respectively. The valence orbitals were expanded in plane waves with an energy cutoff of 70 Ry.

溶液の第一原理分子動力学シミュレーション

利用責任者
平田　勝　　日本原子力研究所 東海研究所 副主任研究員
著者
池田 隆司　　日本原子力研究所 東海研究所 研究員

本研究課題は、第一回課題選定において留保になり、再申請にて採択された課題のひとつである。このため、実質2ヶ月間の最適化作業内容と最適化の結果について報告する。本研究テーマで使用するプログラム(コード名CPMD)は、平面波基底を用いた第一原理電子状態計算により各原子に働く力をHellmann-Feynman力として求め、古典ニュートン方程式に基づき原子の動力学を求めるものである。計算コストの高いルーチンとしては、1電子軌道の直交化、重なり行列要素の計算、 およびFFTを挙げることができる。前者2ルーチンについてはNetlibで公開されているBLASレベル3ルーチンであるDGEMMなどを多用しており、FFTについてはStefan Goedecker により開発され、GPLライセンス下で配布されている並列化されたFFTパッケージを用いている。よって、実施した最適化作業としては、BLASレベル3ルーチンのベクトル化とOpenMPによるノード内並列化、FFTと関連ルーチンのOpenMPによるノード内並列化を重点的に行った。最適化作業終了後、水分子512個の系を用いて性能測定を実施し、ベクトル化率99.31 %、並列化率99.8835 %に達したことを確認した。また、常温常圧下の水を例に、 原子数とMDの1ステップあたりの経過時間から、およそ1000原子までの系について長時間の分子動力学シミュレーションが32ノードを用いて可能であると結論した。

キーワード：溶液、第一原理分子動力学法、BLAS、OpenMP、水