

基于 OpenMP 的 Winograd 并行矩阵乘算法应用研究*

阳光亮, 李鸿健, 豆育升, 唐 红

(重庆邮电大学 高性能计算与应用研究所, 重庆 400065)

摘要: 为了提高半经典分子动力学模拟中矩阵乘法效率, 通过一种稀疏矩阵分解方法化简矩阵乘法, 基于 OpenMP 实现矩阵相乘的 Winograd 并行算法。该算法将 Winograd 算法中各部分依次采用 OpenMP 并行计算, 降低了数据通信。在 16 核服务器上测试表明, 该方法能够显著提高半经典分子动力学模拟中矩阵乘法效率, 并行加速比能够达到 9.47, 并具有良好的可扩展性, 为大分子体系的模拟提供了可能。

关键词: 分子动力学; 矩阵乘; Winograd; 并行计算; 加速比; OpenMP

中图分类号: TP391.9

文献标志码: A

文章编号: 1001-3695(2012)07-2435-03

doi:10.3969/j.issn.1001-3695.2012.07.009

Application and research on Winograd parallel algorithm of matrix multiplication based on OpenMP

YANG Guang-liang, LI Hong-jian, DOU Yu-sheng, TANG Hong

(Institute of High Performance Computing & Application, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)

Abstract: In order to improve the efficiency of the matrices multiplication in the semi classical molecular dynamics simulation, simplified through a the matrix multiplication sparse matrix decomposition method. And then, achieved the matrix multiplication by Winograd parallel algorithm based on OpenMP. This algorithm calculated parallelly each part of Winograd algorithm in order by OpenMP to reduce the data communication. Tested on the server with 16 cores, this method could significantly improve the efficiency of the matrix multiplication in the semi classical molecular dynamics simulation. Its parallel speedup ratio can reach 9.47 and has good scalability. It provided probability for the simulation in large molecular system.

Key words: molecular dynamics(MD); matrix multiplication; Winograd; parallel computing; speedup ratio; OpenMP

分子动力学模拟是固体、液体和气体的各种微观性质的常用研究手段, 广泛应用于物理、化学、生物、材料等多个学科领域^[1-3]。半经典分子动力学模拟是分子动力学(MD)的一个分支, 是从原子、电子水平对分子体系进行精细的理论研究, 目前已成为分子动力学理论研究中的一个热门课题, 受到越来越多计算化学研究者的青睐^[4]。

半经典分子动力学的基本原理是原子核的运动轨迹由经典力学计算, 而电子的运动则通过量子力学处理。半经典分子动力学广泛应用于分子动力学过程研究, 其中典型的模型包括美国伊利诺斯大学的 Martinez 研究组提出的 AIMS (ab initio multiple spawning) 模型^[5]和英国帝国理工学院的 Robb 研究组提出的 MM/QM 模型^[6-7]。半经典分子动力学模拟过程是时间步循环, 单步计算包含核运动的力 F 计算、哈密顿矩阵元 H 和重叠矩阵 S 的计算、哈密顿矩阵元 H 和重叠矩阵 S 的特征值计算。当原子和价电子数量增加时, 半经典分子动力学计算量巨大, 其中核动力 F 的计算占 90%^[4]。在半经典分子动力学模拟计算中, 为了使模拟结果更加精确, 就会用尽可能多的粒子来进行模拟。随着模拟计算体系的不断扩大, 矩阵乘法计算十分耗时。因此, 优化并提高矩阵相乘运算对于提高半经典分子动力学的计算速度十分必要。

本文提出了半经典分子动力学特殊稀疏矩阵计算的优化

方法, 基于计算矩阵特征对矩阵乘法进行化简, 在半经典分子动力学矩阵乘法计算中引入 Winograd 算法^[8,9], 降低矩阵乘法时间复杂度, 并基于 OpenMP 实现了 Winograd 并行矩阵乘法。

1 特殊稀疏矩阵乘法分解和优化

在半经典分子动力学模拟中, 核动力 F 的计算是对每个原子进行一次矩阵 $\partial S / \partial X$ 和矩阵 HS^{-1} 的乘法, 其中 H 为哈密顿矩阵, S 为重叠矩阵, X 为坐标位置, 相乘的两个矩阵均为方阵, 其行列数为模拟分子的所有原子外层电子数总和^[4]。当两个原子间距离达到一定范围时, 粒子间的相互作用力极其微弱, 可以忽略不计。由此产生的 $\partial S / \partial X$ 矩阵具有图 1 中 A 矩阵的稀疏特征, 同时该矩阵是关于对角线对称的, 而相乘的另一个矩阵 HS^{-1} 是普通方阵, 不具有对称性。为方便表示, 设 $A = \frac{\partial S}{\partial X}$, $B = S^{-1}H$, 则 $A \cdot B = \frac{\partial S}{\partial X} \cdot S^{-1}H$ 。

矩阵 A 的大小可以平面延伸, 图 1 中矩阵 A 的阴影区由非零元素占据, 表示粒子间有相互作用力; 非阴影区元素值为零, 表示粒子间没有相互作用力。图中的阴影宽度为外层电子数, 如 C 原子最外层有 4 个电子。该矩阵在半经典分子动力学模拟的矩阵乘法中是乘矩阵, 根据该矩阵的稀疏特征, 可以对非零元素按行、列分解为两个子矩阵, 再分别进行矩阵相乘。

收稿日期: 2011-12-05; 修回日期: 2012-01-06 基金项目: 国家自然科学基金资助项目(21073242)

作者简介: 阳光亮(1986-), 男, 重庆铜梁人, 硕士, 主要研究方向为矩阵并行计算(yangglzh@163.com); 李鸿健(1981-), 男, 重庆铜梁人, 博士研究生, 主要研究方向为并行算法设计与应用; 豆育升(1953-), 男, 陕西西安人, 教授, 博导, 主要研究方向为量化计算; 唐红(1957-), 女, 教授, 博士, 主要研究方向为网络计算。

由于矩阵中绝大部分零元素不用参与计算,因此缩短了运算时间。特殊稀疏矩阵相乘的分解优化流程如图 1 所示。

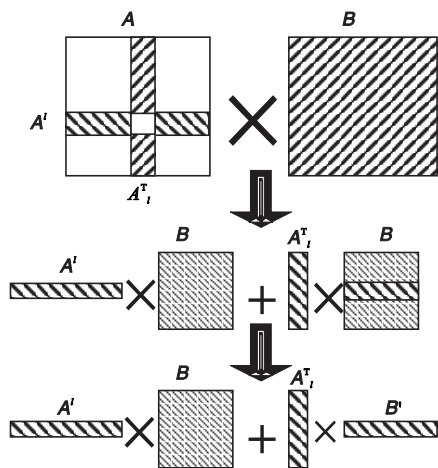


图1 特殊稀疏矩阵相乘分解优化流程

因此,优化该矩阵乘法主要是对 $\partial S/\partial X$ 矩阵进行分解。将矩阵 A 分解为两个子矩阵,按式(1)分别相乘并对应累加得到最终结果为

$$\begin{aligned} A \cdot B &= A_1 \cdot B + A_2^T B = \\ &A_1 \cdot B + A_2^T B_1 \end{aligned} \quad (1)$$

如果将所有行与列直接对应相乘,则两个方阵相乘的时间复杂度为 $O(N^3)$,计算比较复杂。通过上述分解后,时间复杂度降低到 $O(N^2)$,有效地缩短了计算时间,为大分子体系在短时间内模拟提供了现实的可能性。

2 基于 OpenMP 的 Winograd 算法并行

2.1 Winograd 算法介绍

Winograd 算法是一种广泛使用的矩阵串行乘法^[8-10]。两个矩阵相乘,结果矩阵元素值为对应乘矩阵的行与被乘矩阵的列的点积。设有两个向量: $V(v_1, v_2, v_3, \dots, v_n)$ 和 $M(m_1, m_2, m_3, \dots, m_n)$,它们的点积为

$$V \cdot M = v_1 \cdot m_1 + v_2 \cdot m_2 + v_3 \cdot m_3 + \dots + v_n \cdot m_n \quad (2)$$

普通的矩阵乘法按照结果式(2)直接运算。Winograd 算法是一种有效的矩阵乘法^[9],其具体思想是增加了少量加法运算而减少了乘法运算。Winograd 算法描述如下:令 $k = \lfloor n/2 \rfloor$,当 n 为偶数时,

$$V \cdot M = \sum_{i=1}^k (v_{2i-1} + m_{2i})(v_{2i} + m_{2i-1}) - \sum_{i=1}^k v_{2i-1} v_{2i} - \sum_{i=1}^k m_{2i-1} m_{2i} \quad (3)$$

当 n 为奇数时,

$$V \cdot M = \sum_{i=1}^k (v_{2i-1} + m_{2i})(v_{2i} + m_{2i-1}) - \sum_{i=1}^k v_{2i-1} v_{2i} - \sum_{i=1}^k m_{2i-1} m_{2i} + v_n m_n \quad (4)$$

对比式(3)与(4)可以发现,当 n 为奇数时,计算与偶数类似,仅多一个修正项。由式(3)可以发现,在 Winograd 算法中的 $\sum_{i=1}^k v_{2i-1} v_{2i}$ 与 $\sum_{i=1}^k m_{2i-1} m_{2i}$ 均是计算一次,而可重复使用。

对于 n 阶方阵相乘,普通算法的运行时间包含 $n^2(n-1)$ 次加法、 n^3 次乘法,其计算复杂度为 $T_1(n) = n^2(2n-1)$;而 Winograd 算法中包含 $n^2(n+2)$ 次加法、 $n^2(n/2+1)$ 次乘法,计

算复杂度为 $T_2(n) = n^2(3n/2+3)$ 。当 $n > 8$ 时, $T_1(n) > T_2(n)$,Winograd 算法比普通串行乘法的计算复杂度低。

2.2 Winograd 算法并行分析

由于 Winograd 算法与矩阵普通乘法相比有显著的优越性,将半经典分子动力学模拟中分解优化后的矩阵相乘用 Winograd 算法实现可以降低计算复杂度。为了进一步加快计算,应该采用并行技术来提高计算速度。目前的并行技术主要分为进程级并行、线程级并行和进程级与线程级混合并行。

OpenMP 是一种线程级的并行编程模型,适合共享内存多处理系统(SMP)和多核处理器体系结构(CMP)^[11-13]。由于其编程简单且易于实现,OpenMP 在并行编程中得到了广泛的应用。OpenMP 采用 Fork-Join 的并行执行方式,其编译指导语句包含制导标志符、制导名称、子句三部分^[14-16]。

很多矩阵相乘研究者认为 Winograd 算法中有公用的行列处理部分($\sum_{i=1}^k v_{2i-1} v_{2i}$, $\sum_{i=1}^k m_{2i-1} m_{2i}$),不易进行并行化。有部分学者针对 Winograd 算法这一特点,将高阶矩阵分解到各个节点计算,在具体子矩阵相乘时再采用 Winograd 算法计算^[8,9]。这种进程级的并行存在固有的通信、同步和等待问题,有一定的通信开销^[12,13]。当其传递子矩阵的公用行列处理值时,会占用大量的内存,增加了通信量;若不传递,又存在一定的重复计算,增加了计算量。所以这种任务划分的并行效率比较低下,一般在 50% 以下^[8,9]。

半经典分子动力学模拟中,Winograd 算法的 OpenMP 并行是各计算部分依次采用 OpenMP 并行,没有重复计算与数据传递,占用的内存总量也相对较少,因而具有更高的并行效率。

2.3 基于 OpenMP 的 Winograd 矩阵乘并行算法

基于 OpenMP 的 Winograd 算法并行的基本思想是对以下三部分分别进行并行: $\sum_{i=1}^k (v_{2i-1} + m_{2i})(v_{2i} + m_{2i-1})$ 、 $\sum_{i=1}^k v_{2i-1} v_{2i}$ 、 $\sum_{i=1}^k m_{2i-1} m_{2i}$ 。计算时先计算出后两部分的值并存储在数组中,以备重复使用;然后在第一部分计算时,减去相应的后两部分计算值,得到最终结果,以减少计算过程中不必要的数据传递与存储。该算法的类 Fortran 代码描述如下:

```
use omp_lib
implicit none
integer n,i,j,k,m
! 定义乘矩阵 a、被乘矩阵 b、结果矩阵 C
real(kind=8) :: a(n,n),b(n,n),c(n,n)
real(kind=8) :: e(n),d(n)
设定 OpenMP 并行调用的线程数
设置时间函数 secnds(0.0)
call omp_set_num_threads(num_threads)
!! 对矩阵行与列的处理,并分别存入数组 e(n),d(n)中
置 m 的初值为 n/2
! $ OMP parallel do private(i,j) schedule(static)
do i = 1,n
    置 e(i)、d(i)的初值均为 0
    do j = 1,m
        让 e(i)累加上 a 矩阵中 a(i,2*j-1)与 a(i,2*j)的积;
        让 d(i)累加上 b 矩阵中 b(2*j-1,i)与 b(2*j,i)的积
    enddo
enddo
! $ OMP END parallel do
!! 计算第三部分并加入前两部分的修正值得到最终结果
! $ OMP parallel do private(j,i,k) schedule(static)
do j = 1,n
    do i = 1,n
        置 c(i,j)的初值为 e(i)与 d(j)和的反数
```

```

    enddo
    do k = 1, m
        do i = 1, n
            把 a(i, 2 * k - 1) 与 b(2 * k, j) 的和、a(i, 2 * k) 与 b(2 * k
            - 1, j) 的和, 的乘积累加到 c(i, j)
        enddo
    enddo
    enddo
    ! $ OMP END parallel do
    设置时间函数 secnds(0.0)
    打印时间函数之差作为并行计算时间

```

3 实验结果及分析

选取包含 100 ~ 500 个原子以内分子体系进行半经典分子动力学实验测试。采用了各种不同的策略和算法对矩阵相乘进行计算,并比较和分析了计算结果。

3.1 串行矩阵乘的性能测试

为了测试矩阵乘法分解算法与 Winograd 算法在矩阵乘法中的影响,采用一个 2.0 GHz 单核处理器计算不同体系的矩阵乘法。对普通矩阵乘法、Winograd 算法、分解算法、分解 Winograd 算法的计算时间测试结果如表 1 所示。

表 1 四种串行矩阵相乘算法的计算时间/s

矩阵乘策略与算法	atoms				
	100	200	300	400	500
普通矩阵乘法	0.048 9	0.953 2	4.371	12.97	20.48
Winograd 算法	0.041 4	0.762 4	3.942	12.24	19.44
分解算法	0.004 5	0.026 2	0.038 2	0.078 6	0.113 5
分解 Winograd 算法	0.002 8	0.015 4	0.028 8	0.053 4	0.083 2

由表 1 可知,在不同原子规模体系计算中,分解算法比普通矩阵乘法显著减少了计算时间;分解 Winograd 算法同样比 Winograd 算法显著减少了计算时间。由此说明,对特殊矩阵相乘的分解,其时间复杂度从 $O(N^3)$ 减少到 $O(N^2)$,能够有效提高半经典分子动力学矩阵乘法的计算性能。

同时,在不同原子规模体系计算中,Winograd 算法比普通矩阵乘法有更好的计算性能;分解 Winograd 算法比分解算法也有更好的计算性能。这是由于 Winograd 算法用增加加法次数来减少乘法次数,能显著地加快矩阵相乘运算。

3.2 基于 OpenMP 的 Winograd 并行乘法分析

定义加速比公式为 $S_N = T(1)/T(N)$,当模拟特定的分子体系时, $T(1)$ 表示一个处理器的运行时间, $T(N)$ 表示 N 个处理器参与的并行时间。为了比较特殊稀疏矩阵分解后,基于 OpenMP 的 Winograd 并行算法与 OpenMP 并行算法的并行加速比,定义 $T(1)$ 为单处理器计算普通矩阵乘法的时间。系统配置为 16 核 2.0 GHz 处理器,4 GB 内存和 Linux 操作系统。

选取 100 个和 400 个原子的体系进行测试,对比基于 OpenMP 普通并行算法和基于 OpenMP 的 Winograd 并行算法的加速比。计算 100 个原子体系的加速比如图 2 所示。

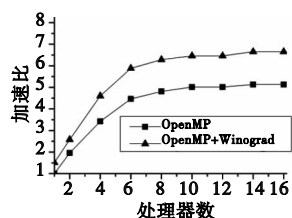


图2 100个原子体系的加速比

在图 2 中,随着处理器数的增加,OpenMP 普通算法和基于 OpenMP 的 Winograd 算法的加速比均呈增长趋势。由于该计算体系仅仅包含 100 个原子,计算体系较小,当处理器数量超过 6 个时,两种算法的加速比增长都变得缓慢。在采用相同的处理器计算时,基于 OpenMP 的 Winograd 并行算法的加速比大于 OpenMP 普通并行算法的加速比,表明基于 OpenMP 的 Winograd 并行乘法相对于 OpenMP 普通算法具有显著的优越性。

计算 400 个原子体系的加速比如图 3 所示。随着处理器的增加,基于 OpenMP 普通并行算法和基于 OpenMP 的 Winograd 并行算法同样都呈增长趋势。400 个原子体系加速比的增长趋势比 100 个原子体系加速的比增长趋势更加明显。基于 OpenMP 普通并行算法,当处理器数超过 8 后,OpenMP 并行的加速比增长缓慢,而基于 OpenMP 的 Winograd 并行随着处理器的增加其加速比不断增加,当处理器达到 16 时,加速比为 9.47,并行效率为 59.2%。该算法在分子体系规模增大时具有良好的可扩展性,为大分子体系的半经典分子动力学模拟提供了可能。

通过以上分析可以发现,该算法实现了 Winograd 算法的良好并行化,与普通的 OpenMP 并行算法相比,具有更高的并行效率,且当分子体系规模增大时其加速比具有良好的可扩展性;和以往 Winograd 算法的进程级并行相比也具有更高的并行效率^[8,9],在 16 核时其能达到 59.2%。

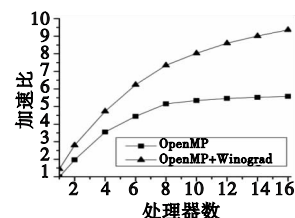


图3 400个原子体系的加速比

4 结束语

对半经典分子动力学模拟中特殊稀疏矩阵相乘进行分解计算,其计算时间能显著降低;在此基础上运用 Winograd 算法的 OpenMP 并行能得到更好的计算效果,进一步加速了计算,为实现大分子体系模拟提供了现实的可能性。

参考文献:

- [1] 朱林山,金石声,苟富均,等.低能入射 Si 与 Si(001)2 × 1 重构表面相互作用过程的分子动力学模拟[J].半导体学报,2007,28(11):1748-1755.
- [2] 陈敏,肖定全.单壁碳纳米管中稳态机械波的激发条件研究[J].人工晶体学报,2010,39(2):355-360.
- [3] 胡晓君.金刚石振动特性的分子动力学模拟[J].人工晶体学报,2005,34(6):1109-1112.
- [4] 李鸿健,白明泽,唐红,等.混合并行技术在激光化学反应模拟中的应用[J].计算机应用,2010,30(6):1687-1689.
- [5] MICHAL B N, MARTI'NEZ T J. Ab initio quantum molecular dynamics [J]. Advances in Chemical Physics, 2002, 121: 439-512.
- [6] BERARDI F, OLIVUCCI M, ROBB M A. Simulation of MC-SCF results on covalent organic multi-bond reactions: molecular mechanics with valence bond (MM-VB) [J]. Journal the American Chemical Society, 1992, 114(5): 1606-1616.

(上接第 2437 页)

- [7] BERPARK M J, BERNARDI F, OLIVUCCI M, *et al.* Molecular mechanics valence bond methods for large active spaces-application to conjugated polycyclic hydrocarbons[J]. *Chemical Physics Letters*, 1994, 217(5-6):513-519.
- [8] 谭福平,刘洪刚. Winograd 矩阵乘法算法用于任意阶矩阵时的一种新处理方法[J]. *应用数学与计算数学学报*, 2004, 18(1):94-98.
- [9] 周德俊,邢永丽. S-W 型二阶矩阵快速乘法的最少乘法运算次数[J]. *工程数学学报*, 1999, 16(1):90-94, 23.
- [10] 周德俊,林彦芬,田增保. 二阶矩阵快速乘法的一个新的算法集合[J]. *高等学校计算数学学报*, 1996, 18(04):380-383.
- [11] 古志民,孙贤和. 并行计算机系统结构与可扩展计算[M]. 北京:清华大学出版社,2009.
- [12] 章隆兵,吴少刚,蔡飞,等. PC 机群上共享存储与消息传递的比较[J]. *软件学报*, 2004, 15(6):842-849.
- [13] 陈华,范宜仁,邓少贵,等. 基于多核微机的微粒群并行算法[J]. *计算机工程与应用*, 2010, 46(13):37-39.
- [14] 章隆兵,吴少刚,蔡飞,等. 适合机群 OpenMP 系统的制导扩展[J]. *计算机学报*, 2004, 27(8):1129-1136.
- [15] CHEN Yong-jian, SHU Ji-wu, LI Jian-jiang, *et al.* Static analysis of OpenMP directive nesting types and its application[J]. *Journal of Software*, 2005, 16(2):194-204.
- [16] 顾丽红,吴少刚,章隆兵,等. 针对非规则应用的 OpenMP 制导扩展[J]. *小型微型计算机系统*, 2005, 26(1):124-128.