

基于 GPU 的生物大分子计算平台的构建与优化

俞 伟¹, 宁璐璐², 许 菲^{*2}

(1. 江南大学, 信息化建设与管理中心, 江苏 无锡 214122; 2. 江南大学, 生物工程学院, 江苏 无锡 214122)

摘要: 通过生物大分子计算平台对分子动力学的模拟运算, 计算值可以描述分子的运动轨迹, 从而揭示原子间的内在关系, 但这一过程较为复杂且昂贵。作者研究了较低投入下大分子计算平台的构建与优化, 在并行 GPU 条件和 CUDA 体系下使用 AMBER 软件包完成分子动力学的模拟运算。通过对两个大分子体系的运算, 实验结果表明计算平台达到了 60 倍的计算加速, 实现了较高的加速性能, 完成了高性价比的高性能计算平台的建设。

关键字: 生物大分子, 分子动力学, 计算加速, 图形处理单元

中图分类号: TP 302.7 **文献标志码:** A **文章编号:** 1673-1689(2017)10-1101-05

Construction and Optimization of GPU-Based Computing Platform for Biological Macromolecules

YU Wei¹, NING Lulu², XU Fei^{*2}

(1. NIC, Jiangnan University, Wuxi, 214122, China; 2. School of Biotechnology, Jiangnan University, Wuxi 214122, China)

Abstract: Biological macromolecules by molecular dynamics computing platform for analog operation, calculated value can describe the trajectory of molecules, which reveals the intrinsic relationship between atoms, but the process is more complicated and expensive. In this paper, the construction and optimization to achieve lower investment macromolecules computing platforms, using a software package AMBER molecular dynamics simulation runs in parallel GPU and CUDA system conditions. Through the operation of two large molecules, experimental results show that the computing platform reached 60 times the computing speed up to achieve a high acceleration performance, completed the construction of cost-effective, high-performance computing platform.

Keywords: biological macromolecules, molecular dynamics (MD), accelerated computing, graphic processing unit (GPU)

收稿日期: 2015-10-08

基金项目: 江苏省自然科学基金项目(BK20151126)。

作者简介: 俞 伟(1975—), 男, 江苏无锡人, 工学硕士, 讲师, 主要从事计算机工程研究。E-mail: yuwei@jiangnan.edu.cn

* 通信作者: 许 菲(1978—), 女, 黑龙江哈尔滨人, 工学博士, 教授, 博士研究生导师, 主要从事生物化学、生物大分子研究。

E-mail: feixu@jiangnan.edu.cn

引用本文: 俞伟, 宁璐璐, 许菲. 基于 GPU 的生物大分子计算平台的构建与优化[J]. 食品与生物技术学报, 2017, 36(10): 1101-1105.

随着计算机计算能力的快速提升,生物大分子的动力学模拟计算受到了科研人员的重视。分子动力学(Molecular Dynamics,MD)通过软件模拟计算出生物大分子在原子水平上的相互作用力,从而展示生物大分子运动的微观过程,得到生物大分子结构变化、系统静态和动态的动力学性质和热力学性质等详细信息。

2007年,马国正等^[1]实现了双CPU下生物大分子分子动力学的模拟。2008年,Yasuda^[2]最早实现了使用GPU来提高计算电子排斥积分的速度并加以改进。文献^[3-5]均提及GPU对于分子动力学的模拟计算有显著的加速效果。2007年,CUDA被用于对分子动力学的计算进行改进,在GPU上计算简化的范德华势。上述研究表明多种分子计算程序包伴随着GPU的加入,计算速度得到显著的提升^[6]。

综上所述,GPU的引入为计算化学带来了快速发展,基于GPU的开源软件也有了长足的进步。但是习惯于传统CPU计算化学的国内用户对于GPU还是不甚了解,甚至认为GPU是高贵神秘的技术。因此GPU在计算化学中的使用需要宣传、推广和普及。研究人员提出了使用PDB代码、Web程序等生物信息学方法描述源蛋白大分子中连接肽的位置、长度、溶剂可及性和二级结构等信息^[6]。部分学科需要通过生物大分子计算平台的计算值来描述分子的运动轨迹,从而揭示原子间的内在关系,但是巨大的计算量延缓了科研的进展。作为一名计算机工作人员,协助这类学科构建大分子计算平台是作者的一项研究课题。通过实践,作者证明了可以通过较低的投入完成高性能大分子计算平台的构建。

如何构建和优化基于GPU的大分子计算平台是作者研究的重点和核心,通过对分子动力学的模拟计算,证明并行效果良好,GPU环境下计算速度较CPU环境下有近60倍的提高,为构建分子动力学计算平台提供了性价比高的解决方案。

1 相关技术

1.1 AMBER

目前分子动力学计算模拟软件中,常用的包括AMBER、CHARMM和GROMOS等。其中AMBER分子力场由加利福尼亚大学P.A. Kollman课题组最初为计算蛋白质和核酸体系而开发,计算参数数据均来源于实验值。随着Kollman课题组和其他课题

组对AMBER内容的不断丰富,逐渐形成了一个具备用于生物大分子、有机小分子和高分子模拟计算功能的力场体系。AMBER包含了分子动力学模拟计算的源代码和演示程序,是一个在生物大分子模拟计算领域被广泛应用的分子力场软件。AMBER的优势在于对生物大分子的计算,对小分子体系的计算还需借助其他模拟计算体系。

AMBER力场的势能函数形式简单,所需参数少,计算量比较小,但也在一定程度上限制了这个力场的扩展性。AMBER力场用傅立叶级数的形式描述二面角扭转能,用谐振子模型计算键长伸缩能和键角弯转能,用Lennard-Jones势模拟范德华力,其势能表达式(1)为:

$$V_{(r)} = \sum_{\text{bonds}} \frac{1}{2} k_b (l - l_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N-1} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \right\}$$

表达式(1) r_{ij} 为两个原子之间的距离, σ 为碰撞常数(collision parameter), ϵ_{ij} 为势阱深度(well-depth), q_i 和 q_j 分别为两个原子的电荷, ϵ_0 为常数, ϵ 为介电常数。

作者使用的计算软件AMBER 14由AMBER 12升级而来。通过升级,实现了对GPU的完全支持,将注意力集中于扩展封装性能、提升单GPU运行性能30%、支持CUDA(5、6和7)、支持多GPU有效运行、进一步提升对分子动力学的加速等方面。

1.2 CUDA在GPU计算中的使用

GPU不仅用于图形的渲染,由于其具备万亿次浮点运算的性能导致从金融到医药产生了爆发式的应用。使用高层语言,可以在CPU上通过使用GPU加速应用的运行,既优化了CPU的单线程性能,又加速了GPU的并行处理。

CUDA是由NVIDIA公司开发的编程模型和并行计算平台,通过GPU的使用实现计算能力的大幅提高。CUDA被广泛应用于各类计算应用和研究论文,目前至少有3亿个支持CUDA的GPU被用于笔记本、工作站、计算机集群和超级计算机。通过CUDA实现了使用C、C++和Fortran等高级语言直接操作GPU,而不再需要汇编语言。

1.3 MPI的使用

在多处理器前提下,计算机指令的执行分为串行和并行两种方式。并行执行指令的方式(MPI)在

完善的资源调度和通信机制下可以取得更高的性能,通过 MPI 可以将计算任务分摊给多个处理单元实现处理单元的负载均衡,最终实现各处理单元性能的最大发挥。MPI 不是一种语言,是一种并行执行的思想,是一种规范和模型。课题使用的服务器配置了 2 块 GPU 卡,因此需要通过 MPI 的引入来实现 2 块 GPU 卡的并行工作,在较短的时间内完成更大的计算量。课题通过 OpenMPI 开源工具来实现 MPI。

2 大分子计算平台的构建

针对课题计算的需求,构建了如表 1 的硬件平台。

表 1 计算平台硬件配置

Table 1 Hardware configuration of computing platform

硬件类型	具体配置
CPU	Intel Xeon E5-2650v2 2.6G/八核/L3 20M/22NM(2 颗)
内存	128G DDR3 ECC 1600MHZ(16G*8)
硬盘	2 块 4T SATA 盘,1 块 480G 固态硬盘
GPU	NVIDIA Tesla K40M 12GB/2880 核心(2 块)

课题构建的生物大分子计算平台使用了 NVIDIA 的高性能 GPU 加速器 TeslaK40,该加速器峰值双精度浮点性能为 1.43 Tflops,峰值单精度浮点性能为 4.29 Tflops,存储器带宽(ECC 关闭)为 288 GB/秒,存储器容量(GDDR5)达到 12 GB,在桌面获得集群的性能。在表 1 的硬件条件下在 CentOS 6 操作系统中首先要完成基础运行环境的构建,主要包括如下两个步骤:

1.安装 Intel CPU 环境下的 C++环境

```
cd /home/tony
tar zxvf l_ccomp_xe_2013_sp1.0.080.tgz
cd ./intel/l_ccomp_xe_2013_sp1.0.080
su
setenforce 0
getenforce
yum install libstdc++.so.5
exit
./install.sh
cd
vim .bashrc
```

```
export MKL_HOME=/home/tony/intel/mkl
source/ home/tony/intel/composer_xe_2013_sp
1.0.080/bin/compilervars.sh intel64
source .bashrc
su
yum install gcc flex tcsh zlib-devel bzip2-devel
libXt-devel libXext-devel libXdmcp-devel
2. 安装分子计算平台 AMBER 14 和
AMBERtools 14
```

```
cd /home/tony
tar xvfj AmberTools15.tar.bz2
tar xvfj Amber14.tar.bz2
vim .bashrc
export AMBERHOME=/home/tony/amber14
source .bashrc
cd $AMBERHOME
./configure intel
make install
make test
```

上述环境搭建后已可以进行分子动力学的计算,但为了实现高效和并行,需对平台做如下优化,主要包括如下两个步骤:

1.编译安装 openmpi-1.6.tar.bz2

```
tar xvfj openmpi-1.6.tar.bz2
cd /home/tony/openmpi-1.6
./configure --prefix=/home/tony/openmpi -intel
FC=ifort CC=icc CXX=icpc F77=ifort
make
make install
vim .bashrc
export MPI_HOME=/home/tony/openmpi-intel
export PATH = $MPI_HOME/bin:
$AMBERHOME/bin:$PATH
export LD_LIBRARY_PATH=$MPI_HOME/lib:
$AMBERHOME/lib:$LD_LIBRARY_PATH
export MKL_HOME=/home/tony/intel/mkl
2.安装并行
cd $AMBERHOME
./configure-mpi intel
make install
```

3 实验与仿真

3.1 测试体系说明

平台对两个体系进行了测试，体系参数如表 2 所示。

表 2 测试体系
Table 2 Test system

研究目标	体系原子个数	水分子个数	水盒子形状	抗衡离子数
DNA	9 788	3 043	截断正八面体	18 (Na ⁺)
蛋白质	12 673	3 812	六面体	3 (Na ⁺)

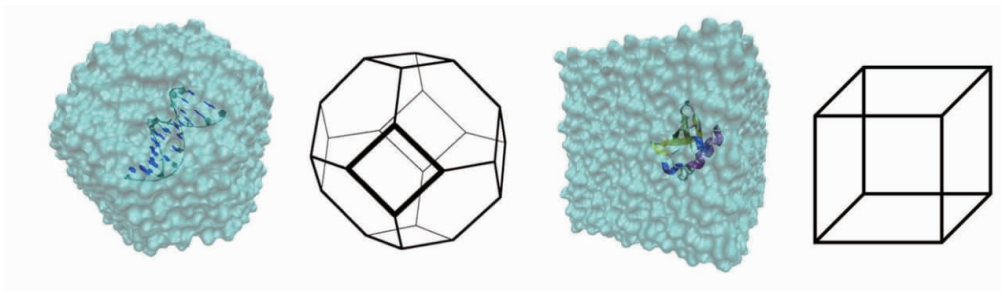


图 1 DNA 与蛋白质测试体系的模拟形状

Fig. 1 Shape of DNA and protein about simulation test system

表 3 模拟参数

Table 3 Simulation parameters

体系	力场	系综	控压方法	控温方法	积分步长	总步长
DNA	AMBER ff14SB	NPT	Berendsen barostat	Langevindynamics	2 fs	5 000 000
蛋白质	AMBER Ff14SB	NPT	Berendsen barostat	Langevin dynamics	2 fs	2 500 000

表 4 力场参数

Table 4 Force field parameters

体系	Lenard-jones 势能截断值	Columb 势	氢原子	边界条件
DNA	12 Å	Particle Mesh Ewald	SHAKE	周期性
蛋白质	12 Å	Particle Mesh Ewald	SHAKE	周期性

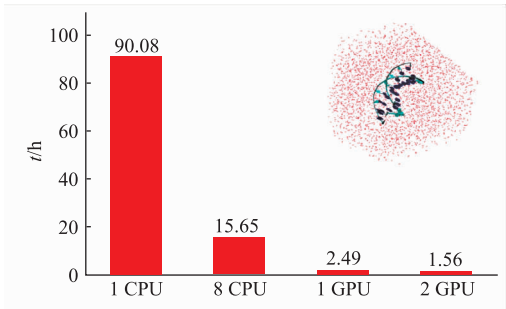


图 2 DNA 体系测试结果

Fig. 2 Result of DNA system

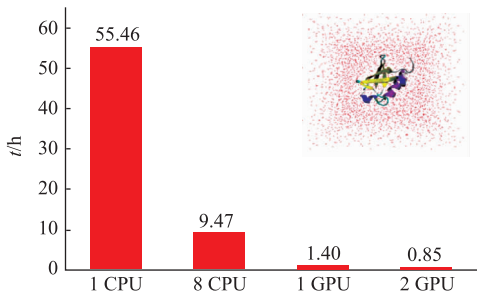


图 3 蛋白质体系测试结果

Fig. 3 Result of protein system

表5 不同运算单元的加速比
Table 5 Speedup different process unit

体系	1CPU	8CPU	1GPU	2GPU
DNA	1	5.76	36.18	57.74
蛋白质	1	5.86	39.61	65.25

4 结 语

实现了低投入高性价比的生物大分子计算平

台的构建与优化。证明了 GPU 对于加速分子动力学的模拟运算起着至关重要的作用。

后续课题将从事如下两方面的工作:目前 2GPU 较 1GPU 的加速比分别是 1.60 和 1.65,需要优化 GPU 环境下的计算平台,实现 2GPU 较 1GPU 的加速比接近或达到 2;构建低成本的计算集群,通过多机的并行运行实现更高的加速比,为更高级别的原子数量的模拟计算提供计算平台。

参考文献:

- [1] MA Guozheng, NAN Junmin. Molecular dynamic simulations of biomacromolecule on dual Process-based system using MPICH technology[J]. **Computers and Applied Chemistry**, 2007, 24(8): 852-855. (in Chinese)
- [2] YASUDA K. Two-electron integral evaluation on the Graphics[J]. **J Comput Chem**, 2008, 29: 334-342.
- [3] FEI Hui, ZHANG Yunquan, WANG Ke, et al. Parallel algorithm and implementation for molecular dynamics simulation based on GPU[J]. **Computer Science**, 2011, 38(9): 1056-1058. (in Chinese)
- [4] LIN Jianghong, LIN Jinxian, LU Tun. Accelerated molecular dynamics simulation using multi-core CPU and GPU[J]. **Journal of Computer Applications**, 2011, 31(3): 367-381. (in Chinese)
- [5] LI Jiangyu, ZHAO Dongsheng, WANG Yumin. GPU computing and its application in biomedical research [J]. **Mil Med Sci**, 2011, 35(8): 1024-1028. (in Chinese)
- [6] LI Jianfang, WANG Chunjuan, WU Minchen. Design of linker peptides and its application in fusion protein [J]. **Journal of Food Science and Biotechnology**, 2015, 34(11). (in Chinese)

会 议 消 息

会议名称(中文):第四届中国休闲食品产业发展高峰论坛

所属学科:动物食品科学

开始日期:2017-11-01

结束日期:2017-11-03

所在城市:北京市 东城区 具体地点:中国国际展览中心(三元桥静安庄馆)

主办单位:中国食品科学技术学会休闲食品加工技术分会

协办单位:中国农业大学食品科学与营养工程学院功能食品配料研发中心 中国林业产业联合会流通分会 好想你枣业股份有限公司

承办单位:中国农业科学院农产品加工研究所果蔬食品制造与营养科学团队 国家粮食局科学研究院粮食加工研究组 中国食品工业(集团)公司 中国包装和食品机械有限公司

议题:天然、营养、健康、方便、多样。

联系电话:15210811896; 010-62816487

E-MAIL:zgxxspfh@163.com

会议注册费:¥1200

会议网站: <http://www.cifst.org.cn/a/665.html>

会议背景介绍:当前,中国休闲食品产业已经成为食品工业中快速发展的朝阳产业之一。国家高度重视和支持休闲食品产业发展。国务院办公厅关于进一步促进农产品加工业发展的意见(国办发〔2016〕93号)中明确提出支持大中城市郊区重点发展主食、方便食品、休闲食品等。2017年中央一号文件明确提出大力发展方便食品、休闲食品等。随着人们生活水平的提高和生活节奏的加快,对营养、健康、方便、安全、多样的休闲食品需求越来越高,产业发展空间巨大。为进一步促进我国休闲食品产业持续稳定健康发展,不断提高行业的科技创新能力、装备制造水平和产品竞争力,搭建休闲食品政府部门、产业界、学术界、商界、流通界等新政策、新技术、新装备、新产品、新模式和新业态共享平台,定于2017年11月1-3日在“第十五届中国国际食品加工和包装机械展览会”期间,同期举办第四届中国休闲食品产业发展高峰论坛。