

数字岩心中的 Boltzmann 并行算法优化*

宋文博,滕奇志,吴小强

(四川大学 电子信息学院 图像信息研究所,四川 成都 610064)

摘要: 针对 LBM 在应用过程中计算规模过大的问题,提出了一种优化算法。该算法明显改善了计算时间过长的问题;对访问的边界数据进行特殊处理,减少了分支判断逻辑;根据三维岩石微观图像的特点进行了空间结构优化,避免了无效格子点在线程中的运行,提高了 GPU 运算效率。实验结果表明,优化后的计算速度有明显提升。

关键词: Boltzmann;并行计算;GPU;CUDA

中图分类号: TP391.4

文献标识码: A

文章编号: 1674-7720(2014)03-0027-04

Boltzmann parallel algorithm optimization in digital cores

Song Wenbo, Teng Qizhi, Wu Xiaoqiang

(Research of Image Information, School of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China)

Abstract: In this paper, an optimization algorithm is proposed for the huge amount of calculation of Lattice Boltzmann method in the application process. It significantly improves the problem that the computation time is too long. Special handling to boundary data accessed reduces the branching decision logic. According to three-dimensional microscopic images of rock characteristics, spatial structure is optimized to avoid invalid grid point in the thread to run, which improves the efficiency of GPU computing. The experimental results show that the speed of the optimized calculation has improved significantly.

Key words: Boltzmann; parallel computing; GPU; CUDA

研究岩石的渗透特性、孔隙特性以及变化规律是预测油气形成、分布规律与探索提高采收率的主要方法^[1]。渗透特性反映流体通过岩石空间的能力,以往大多通过实验来定性研究,但近年来,为从微观尺度对渗流特性进行定量分析,国内外学者在该领域展开了大量研究。为了使研究的渗流结果具有实际意义,建立一个能够真实反映岩心孔隙特征的数字岩心往往是开展研究的第一步。目前,国内外学者能够借助数值重建方法或物理实验手段建立数字岩心^[2]。利用数字岩心进行渗透率分析的方法主要有两种:孔隙网络模型和格子 Boltzmann 方法 LBM (Lattice Boltzmann Method)。孔隙网络模型是以孔隙和喉道为基本组成单元的模型,它对不规则的孔隙喉道进行规则化抽象处理,将其等效为具有任意横截面的管道,并且采用形状因子记录真实的形状信息,使几何计算得以简化,从而使得更多的物理公式能直接用于渗流模拟分析。孔隙网络模型是根据研究的问题定义具

体的规则进行流动模拟的^[3]。LBM 把模拟介质看作由细微的“格子”状的基本单元组成,通过模拟微观粒子的运动过程和分布规律来模拟渗流过程。LBM 是一种成熟的算法,被广泛应用于多种领域的流体运动模拟^[4-5]。但是传统算法的计算规模过大,运算时间太长,在实际应用中极为耗时。为了解决模拟时间过长的问题,国内外学者都在研究一些基于并行技术的加速方案,这些方案多采用集群并行计算机来加速 LBM 模拟的方法。

本文主要以 D3Q19 为模型来模拟流体,并针对 LBM 的算法特点对访问的边界数据进行特殊处理,减少了访问边界数据时的分支跳转;同时,根据三维微观岩石的空间结构特点,剔除了无效点,减少了运算时间。在国家自然科学基金项目中,参考文献[5]在计算结构和粒子空间上进行了优化,本文在此基础上进一步优化,取得了一定的效果。

1 LBM 和 CUDA 平台

1.1 LBM 的基本方法

在 LBM 中,复杂流体运动被转化为规则、离散的格

* 基金项目:国家自然科学基金(60972130、61372174)

子空间中的运动,一般采用 DdQm 模型来表示粒子在离散空间中的运动模型。其中, d 代表空间的维数, m 代表离散速度的方向数。与 Navier-Stockes 方程求解非线性偏微分方程不同^[6], LBM 仅用离散格子的迁移与碰撞来表述流体的运动。

模拟三维空间中流体的运动,一般采用 D3Q13、D3Q15、D3Q19 或 D3Q27 模型。在 DdQm 模型中, m 越大时, LBM 的精度越高,收敛速度越慢。

1.2 CUDA 平台

在 20 世纪 90 年代, NVIDIA 公司提出了图形处理器 GPU (Graphics Processing Unit), 并且实现了硬件加速的三维坐标变换和光源计算 T&L (Transformation and Lighting) 技术。在数字图像应用 (游戏、工程制图等) 的强烈刺激下, GPU 的性能不断进步。在 2001 年, NVIDIA 推出了一款支持可编程图形流水线 (Programmable Graphics Pipeline) 的标志性产品 GeForce 3。随后, 人们开始注意到 GPU 的高度并行化架构的优势, 通用计算图形处理器 GPGPU 技术逐步发展起来, 并且在科学计算领域得到了应用^[7]。在 2006 年, NVIDIA 公司发布了 GeForce 8 系列 GPU, 同时发布了 GPU 运用于通用计算的统一计算设备架构 CUDA (Compute Unified Device Architecture)^[8]。从此, GPU 技术的发展进入了全新的时代。目前, GPU 通用计算技术已广泛应用于流体渗流模拟、天文计算、实时数字图像处理与医疗诊断等众多科学计算领域。在全球超级计算机前 10 排行榜中, 有 3 个采用了 CPU 加 GPU 的异构方式^[9]。

2 CUDA 平台上 LBM 并行算法优化

本文主要从两个方面对 CUDA 平台上 LBM 并行算法进行优化: 一是针对边界数据进行处理, 减少访问存储器次数; 二是针对岩石的多孔介质的空间特点, 从实质上提高运算过程中 Wrap 内有效点的比例, 从而达到提高运算速度的目的。首先, 在 CUDA 平台上对 LBM 进行层次化划分。CUDA 线程划分方式如图 1 所示。

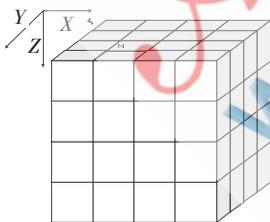


图 1 LBM 的 CUDA 线程划分方式

本文选择每个 CUDA 线程处理一个 LBM 的离散格子。GPU 中采用网格分层的方式来组织线程, 合理地划分 Grid 与 Block 才能充分发挥 GPU 的计算能力。

2.1 边界数据处理

在图 1 所示的 CUDA 线程划分方式下, 分析粒子空间中格子数据的具体处理过程。考虑 \vec{x} 与 \vec{y} 方向的边界格子中粒子的碰撞与迁移过程。因为在粒子沿坐标轴正方

向迁移时, 粒子分布函数将会离开格子空间发生越界, 所以在对边界数据进行特殊处理。

针对格子数据越界访问的问题, 首先考虑一般的处理方式, 即在越界访问数据时对数据进行一个分支判断, 然后查找下一个数据所在位置, 并将数据读入进行处理。然而对于 GPU 而言, 它属于单指令单数据流 SIMD (Single Instruction Single Data Stream) 结构, 执行分支条件语句的效率较低。当 GPU 执行一条指令时, 对应的多个线程会同时执行, 一个 Wrap 内的线程会串行执行。如果 Wrap 内的线程因为分支判断而走向了不同分支 True 或者 False, 那么无论执行 True 或者 False 任意一个分支时, 另一条分支的线程则会执行空操作, 在 CUDA 程序中称为 Divergent^[10]。例如, 在执行 True 分支时, 所有 False 分支内线程的内存操作会无效。为了减少分支判断的产生, 针对这一类边界数据进行特殊处理, 本文选用了与参考文献 [11]~[14] 中提到的处理手法类似的方式。为 \vec{x} 与 \vec{y} 方向数据的每一层的最后一个数据分别增加一个额外的节点数据, 即虚拟层, 如图 2 所示。在 \vec{x} 与 \vec{y} 方向的边界格子发生迁移时, 直接迁移到节点数据中, 省去了分支判断。对 Boltzmann 方法的计算过程而言, 节点数据没有任何意义。通过这样的方式, 分支判断的数目明显下降。

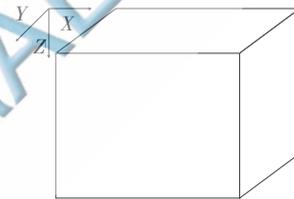


图 2 LBM 中添加的虚拟层

2.2 空间结构优化

对于三维岩石微观图像而言, 其孔隙空间具有多孔介质形态的特点, 并且大多数情况下其孔隙度不超过 20%。在渗流分析中, LBM 的离散格子空间中有超过 80% 的粒子点不参与计算, 如果这些格子点在参与计算, 那么 GPU 中的 Wrap 有效率会大大降低。针对这种情况, 可以在进行 GPU 执行语句前, 在 CPU 程序中通过分支语句跳过。然而, 引入分支条件一般会降低程序的运行效率 (Wrap 内的分支条件会被串行执行)。

在 GPU 中, 计算与访存均以 Wrap (即 32 个线程) 为单位。经过统计, 在格子空间规模为 $128 \times 128 \times 128$, 孔隙度为 12.42% 的岩石孔隙结构中, 有效点数为 0 的 Wrap 占总 Wrap 数的 30% 左右, 有效点数越多的 Wrap 所占的百分比越小, 如图 3 所示。有效点不为 0 或者 32 的 Wrap 内会出现分支条件, 且有效点数为 0 的 Wrap 是无效的, 其内部所有线程均不作任何有效计算。并且, GPU 平台最优的是共享存储器, 而不是全局存储器, 故访问全局存储器相对耗时。如果能避免这些 Wrap 的无效点,

那么访问全局存储器的数目就会大大降低,从而提高运算速度。为此,本文引入了分支条件,对于无效格子点,既不读取任何数据也不执行任何计算。该分支条件对有效点数为 32 的 Wrap 不会产生任何影响。

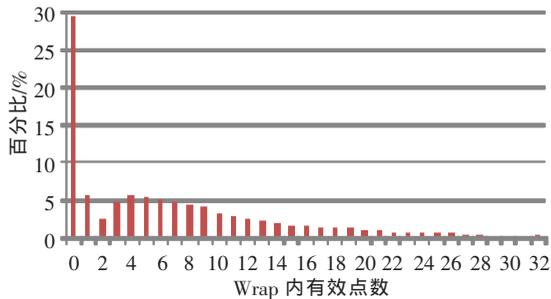


图3 Wrap内有效点数分配情况

3 实验结果与分析

本文的实验是在一台 64 位的台式计算机上完成的。计算机的处理器是英特尔第三代酷睿 i7-3770K, 主频为 3.50 GHz, 4 核 8 线程, 一级缓存 4×32 KB, 二级缓存 4×256 KB, 三级缓存 8 MB, 8 GB 内存。

3.1 不同孔隙度的运行结果

为了研究孔隙度参数对优化结果影响, 本文分别采用了孔隙度为 0.11、0.26 和 0.48 共 3 组岩石微观图像进行实验。实验在 9600GT、GT240 和 GTX560Ti 3 种不同性能的显卡上进行。图 4 所示为经过三维重建后的三维岩石图像。图 5 为 3 种显卡上的运行结果, 其中横坐标为不同数据规模, 纵坐标为碰撞迁移过程迭代 1000 次所消耗的时间。



图4 重建后的三维岩石图像

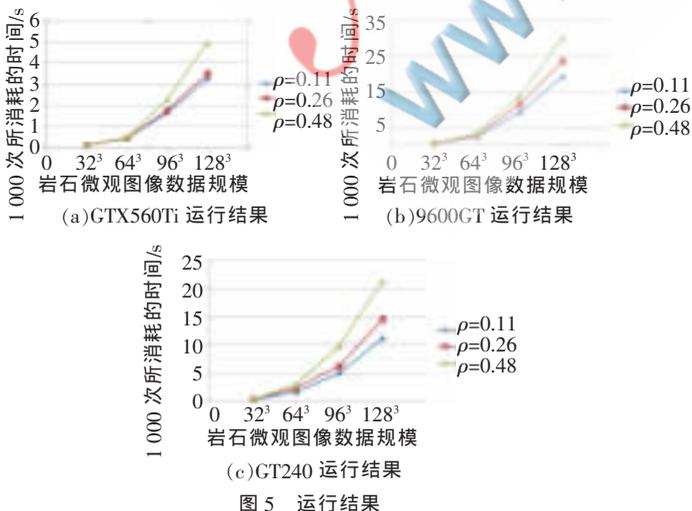


图5 运行结果

图 5 中数据表明, 经过 3 种显卡测试, 随着孔隙度的减小, 优化后均得到了明显的加速效果。孔隙度越小, Wrap 中无效点的数目就越多, 经过判断跳过这些大量的无效点, 运算时间大量减少。因为空间结构优化会受孔隙度的影响, 所以不同孔隙度下的运行结果会有较大差距。

3.2 优化结果

实验采用孔隙度为 0.26 的一组微观岩石图像, 在 9600GT、GT240 和 GTX560Ti 3 种不同性能显卡上进行实验。图 6 为 CPU 在不同图片规模下的运算结果。图 7 为加速比, 可以看出, 经过优化后的计算速度有了显著提升。



图6 CPU运行结果

图7 加速比

从图 7 可以看出, 随着图像规模的增大, 加速比有明显的提升。因为随着图像规模增大, 在格子迁移到边界的数目增多, 经过边界数据处理后粒子直接命中下一个, 分支判断大量减少, 提高了 GPU 中线程的使用效率, 提高了运算速度。在图片规模为 128³、显卡为 GTX560Ti 的情况下, 优化加速比可以达到 12 左右。与参考文献[5]的实验结果对比, 本文的实验结果加速比提高了 1.9 倍以上。

本文旨在提高 Boltzmann 算法在三维岩石微观图像上的计算速度, 减少计算时间。对边界数据进行处理, 在边界迁移时减少了分支判断条件, 减少了访问全局存储器的数目; 针对三维岩石微观图像的空间结构特点, 减少了 Wrap 内无效点数目, 提升了 GPU 的线程利用率。三维岩石微观图像的流体渗透率的计算规模过大, 计算时间较长的问题得到一定的改善。这扩大了 Boltzmann 算法在计算流体渗透率中的应用, 使 Boltzmann 算法在较大规模的岩石微观图像的应用上更加具有实际意义。预计将来 GPU 计算将会代替 CPU 计算成为岩石微观图像渗流分析的主要手段。

参考文献

- [1] 陈杰, 周改英, 赵喜亮, 等. 储层岩石孔隙结构特征研究方法综述[J]. 特种油气藏, 2005, 12(4): 11-14.
- [2] 赵秀才. 数字岩心及孔隙网络模型重构方法研究[D]. 北京: 石油工程学院, 2009.
- [3] MANWART C, AALTOSALMI U, KOPONEN A, et al. Lattice-Boltzmann and finite-difference simulations for the permeability for three-dimensional porous media[J]. Physical Review E, 2002, 66(1): 16702.

- [4] GLADROW W. Lattice-gas cellular automata and Lattice Boltzmann models[D]. Springer, 2005.
- [5] Zeng Fansong, Teng Qizhi. Application and optimization for Lattice Boltzmann method in calculation of fluid permeability[J]. Computer Engineering and Applications, 2011,47 (22): 156-159.
- [6] Ming Pingjian. Numerical simulation of sloshing in rectangular tank with vof based on unstructured grids[J]. Journal of Hydrodynamics, 2010, 22(6): 856-864.
- [7] 仇得元.GPU 编程技术——从 GLSL、CUDA 到 OpenGL[M]. 北京:机械工业出版社,2011.
- [8] 张舒,褚艳利,赵开勇,等.GPGPU 高性能运算之 CUDA[M]. 北京:中国水利水电出版社,2009.
- [9] SRIVASTAVA N. Top 10 super computer in the world [EB/OL].<http://www.top10gallery.com/2012/02/top-10-super-computer-in-world-2012.html>, 2012-2.
- [10] NVIDIA CUDA C Best Practices Guide[Z]. Version 4.0. NVIDIA Corporation, 2011:103-105.
- [11] VALDERHAUG T K. The Lattice Boltzmann simulation on multi-GPU systems[D]. Trondheim: Norwegian University of Science and Technology, 2011.
- [12] HABICH J. Performance evaluation of numeric compute kernels on NVIDIA GPUs[M]. Master's Thesis of Friedrich-Alexander-Universit Erlangen-Nurnberg, 2008.
- [13] 李博,李曦鹏,张云,等.耦合 Nvidia/AMD 两类 GPU 的格子玻尔兹曼模拟[J].科学通报,2009,54(20):3177-3184.
- [14] 李森,李新亮,王龙,等.基于 OpenGL 的并行方腔流加速性能分析[J].计算机应用,2011,28(4):1401-1403,1421.
(收稿日期:2013-11-11)

作者简介:

宋文博,男,1989年生,硕士研究生,主要研究方向:图像处理与模式识别。

滕奇志,女,1961年生,教授,硕士生导师,主要研究方向:图像处理与模式识别。

吴小强,男,1969年生,高级工程师,主要研究方向:图像处理与数据库系统方面。