

基于 GPU 的蒙特卡洛放疗剂量模拟

司 森, 吴鑫全, 包尚联, 黄斐增

北京大学医学物理和工程北京市重点实验室, 北京 100871

【摘要】目的:蒙特卡洛模拟被认为是目前剂量计算方面最为精确的算法,但是因为其模拟时间过长,在临床应用上受到限制。EGSnrc 作为目前在医学物理领域应用最为广泛的蒙特卡洛模拟软件,因为其过长的执行时间,其在临床方面的应用受到很大限制。为了克服这一障碍,我们开发了一个基于 GPU 的蒙特卡洛模拟程序,以期放疗剂量提供一个高效和低成本的成本卡程序。**方法:**本文给出了一种基于 GPU (Graphic Processing Unit) 的蒙特卡洛模拟的新方法,开发语言是 CUDA 5.0,将目前最为通用的蒙特卡洛程序 EGSnrc 移植到 GPU 平台,保留 EGSnrc 的核心物理过程以及输运过程的算法,这可以在最大限度保持原来 EGSnrc 模拟精度的前提下,极大地提高蒙特卡洛模拟的效率。GPU 版本的蒙特卡洛模拟程序运行在一块英伟达 Tesla C2050 显卡上。GPU 版本的 EGSnrc 精度的验证采用了纯水模体,同时,入射的射线我们选择为 6 MV 的光子。为了进一步检验 GPU 版本的 EGSnrc 的精度,我们进行了一个逐体素的检验,检验结果显示, GPU 版本的 EGSnrc 和 EGSnrc 符合的很好。**结果:**最终实验结果表明,在模拟 20 亿个相空间事例的情况下,使用 NVIDIA Tesla C2050 显卡,新的基于 GPU 的蒙特卡洛程序的速度比在单核的 Intel Xeon 2.0 GHz CPU 上的模拟速度提高了 43 倍,且其精度与 EGSnrc 的精度相当。计算结果的方差在高剂量区域 ($D > D_{\max}$) 小于 0.5%,剂量误差经过 D_{\max} 归一化之后,其和 EGSnrc 的误差小于 1% 的比率在占整个区域的 90% 以上。**结论:**通过此新程序表明,基于 GPU 的蒙特卡洛算法可以极大地提高蒙特卡洛程序的运行效率,与此同时, GPU 版本的 EGSnrc 在最大程度上保持了 EGSnrc 的模拟精度。考虑到 GPU 版本的 EGSnrc 程序的速度以及精度优势,其在未来的临床应用中有巨大的前景。

【关键词】放射治疗; 蒙特卡洛模拟; GPU; EGSnrc

【DOI 编码】doi:10.3969/j.issn.1005-202X.2015.01.001

【中图分类号】TP391.9

【文献标识码】A

【文章编号】1005-202X(2015)01-01-03

GPU-Based Monte Carlo Simulation for Radiation Therapy Dose Simulation

SI Sen, WU Xin-quan, BAO Shang-lian, HUANG Fei-zeng

The Beijing City Key Lab of Medical Physics and Engineering, Peking University, Beijing 100871, China

Abstract: Objective Monte Carlo simulation is generally considered to be the most accurate algorithm for radiation therapy dose calculation. However, it is very time consuming for Monte Carlo simulation to get a desirable accuracy, which makes it difficult to use in routine clinical application. EGSnrc is considered to be the most widely used Monte Carlo Simulation code in Medical Physics, the long execution time is the main barrier for its routine clinical application. To overcome this obstacle, a graphics processing units (GPU) based parallel computing version of EGSnrc was developed to provide a fast and low cost solution for accurate Monte Carlo simulation. **Methods** In this paper, we give a GPU-based Monte Carlo simulation program package using the CUDA 5.0 environment. We implement the EGSnrc to the GPU platform, the EGSnrc's core physics interaction and particle transportation mechanism were maintained in the GPU version of EGSnrc, which will maintain the accuracy of EGSnrc, at the same time we accelerate its execution speed. The GPU based Monte Carlo simulation code is run on a NVidia Tesla C2050 card. And the test phantom we use in this experiment is a pure water phantom, and the incident source we choose is a 6 MV photon. **Results** Our research results show that using a NVIDIA Tesla C2050 GPU card against an Intel Xeon CPU card, both simulates 2 billion phase-space histories, our GPU-based Monte Carlo program package will get a 43 speed-up factor at the same time maintain the accuracy of EGSnrc. To further test the agreement of the calculation result, we run a voxel by voxel test,

【收稿日期】2014-08-28

【基金项目】国家自然科学基金(11175007)

【作者简介】司森(1986-),男(汉族),北京大学物理学院核技术与应用专业研究生。

【通讯作者】黄斐增, E-mail: huangfz@pku.edu.cn。

which shows that the GPU based EGSnrc agree well with the EGSnrc. When the uncertainty is less than 0.5% everywhere in the high dose region ($D > 0.5D_{\max}$), the dose difference normalized to D_{\max} between GEGS and EGSnrc is less than 1% for more than 90% of the voxels. **Conclusion** Our new GPU-based Monte Carlo program package could accelerate the Monte Carlo simulation in a great degree, at the same time, maintains the accuracy of EGSnrc. The GPU based EGSnrc code has great potential in routine clinical application of dose calculation for radiation therapy, considering its speed and accuracy.

Key words: radiation therapy; Monte Carlo simulation; GPU; EGSnrc

前言

作为一种被广泛认为是目前最为精确的剂量算法,蒙特卡洛模拟在放疗剂量计算方面有着广阔的应用前景^[1]。在过去几十年中,随着 EGS 系列和 MCNP 系列等模拟粒子输运的蒙特卡洛软件的陆续开发,蒙特卡洛模拟在放疗剂量计算方面也有了广泛的应用。但是,因为蒙特卡洛模拟的精度和模拟粒子事例数的平方成正比,所以当需要达到一定精度的时候,蒙特卡洛程序需要模拟大量粒子,会耗费大量时间。这个特性成为蒙特卡洛模拟在临床中应用的最大障碍。如果在 CPU 平台上面解决这一问题,主要的思路有两个:第一,不断提高 CPU 的频率,从而使得模拟速度不断提高,但随着摩尔定律的失败,计算机 CPU 的频率提高也已遇到了瓶颈。第二,进行多 CPU 的并行计算,此种方法面临的问题是多 CPU 系统的能耗问题。上述两个解决问题的思路都因为 CPU 计算自身的缺陷而面临种种困难。

自从 2006 年 NVIDIA 公司推出了基于 GPU 的并行计算平台 CUDA 之后,基于 GPU 的蒙特卡洛模拟日益成为蒙特卡洛剂量计算研究中的热点领域^[2],相较于 CPU 计算, GPU 计算单位计算能力的价格和能耗都大大降低^[3],这些特征都展现了 GPU 计算在医学物理领域的广阔应用前景。在这一时期,世界范围内有多个研究组报告了他们的基于 GPU 的蒙特卡洛模拟程序,例如 gDPM^[4], GPUMCD, GMC^[5], SMC^[6] 等等,所有这些程序都在保持了蒙特卡洛模拟精度的前提下极大的提高了模拟的速度。但是基于 GPU 的蒙特卡洛剂量计算软件在国内尚处于空白阶段。

作为目前在医学物理领域应用最为广泛的蒙特卡洛程序^[7], EGSnrc^[8] 可以精确的模拟电子 - 光子的输运过程,并且可以准确计算放疗射线能量范围内的剂量分布情况^[9]。但因为其复杂的物理过程模型, EGSnrc 为了保证上述的精确性而牺牲了效率,从而使得其在临床中的应用受到了很大的限制。为了提高 EGSnrc 程序的运行效率,我们开发了一种基于 GPU 的蒙特卡洛模拟程序 (GEGS), 其在保留了 EGSnrc 主要的电子 - 光子输运算法的同时可以利用 GPU 的并行计算能力进行并行计算,从而大大提高了蒙特卡洛程序的运行效率。

为了验证 GEGS 的性能,我们在相同的条件下分别用 GEGS 和 EGSnrc 进行了剂量计算,同时给出了

两种软件计算得到的剂量深度分布曲线 (PDD) 和不同深度的离轴比 (OAR)。所选择的硬件分别是 NVIDIA Tesla C2050 和 Intel Xeon 2.0 GHz。

1 方法与材料

为了保证 GEGS 的精度,我们保留了 EGSnrc 电子 - 光子输运的主要物理过程和算法,考虑到放疗常用的射线的能量范围主要集中在中低能, GEGS 中没有将三光子效应这一物理过程考虑进去,我们将在后面看到舍去这个过程并没有降低 GEGS 的精度。

因为 CUDA 的计算是以线程 (Thread) 为单位来进行^[10], 类似于一个 CPU, 每一个线程可以模拟一个粒子从进入模体到消失的全过程, 但是在 CUDA 中, 每 32 个线程组成一个线程束 (Warp)^[11], 这 32 个线程需要同步进行内存的读写操作。以蒙特卡洛剂量模拟为例, 32 个线程中的粒子需要同时将能量沉积写入内存当中, 所以就造成了模拟速度快的粒子去等模拟速度慢的粒子这样一个问题, 从而大大降低 CUDA 程序的运行效率。这一问题在 CUDA 计算中叫做线程发散 (Thread Divergence)^[12]。为了解决这一问题, Hissoiny^[13] 于 2011 年提出了 CUDA 粒子输运并行计算的一个基本原理, 即每次并行模拟的粒子应该大体上具有相似的物理性质, 这样, 这样可以保证每个线程束中模拟的粒子可以经历大体相同的物理过程, 其模拟的时间也会相近, 这样会在最大程度上降低线程发散的问题, 从而提高并行计算的效率。在 GEGS 中, 我们也沿着这一思路来安排程序的框架, 保证每次进行并行计算的粒子具有相同的物理性质。

作为对比, 我们用 GEGS 和 EGSnrc 分别模拟 20 亿个相空间粒子, 选用 30 cm×30 cm×30 cm 的纯水模作为试验模体, 每个体素的大小为 0.5 cm×0.5 cm×0.2 cm, 射野大小为 10 cm×10 cm。为了进行精度验证, 我们给出了两种软件计算纯水模的剂量深度分布曲线 (PDD) 和 2.5 cm, 7.5 cm 以及 12.5 cm 深度出的离轴比 (OAR), 光子的截断能量设置为 0.01 MeV, 电子的截断能量设置为 0.7 MeV。相空间粒子为 6 MV 的光子, 通过模拟一台 Varian CLINAC 2100C 得到。

2 模拟结果

图 1 和图 2 分别给出了在纯水模中的 PDD 曲线和不同深度的 OAR 曲线, 可以看到, 即使在不考虑三光子效应的条件下, GEGS 和 EGSnrc 的结果有着很

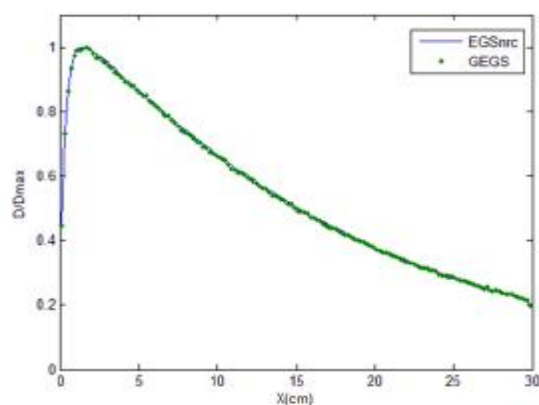


图 1 6 MeV 入射光子条件下的 GEGS 和 EGSnrc 剂量分布曲线
Fig.1 PDD Curves With 6 MeV Photon of GEGS and EGSnrc

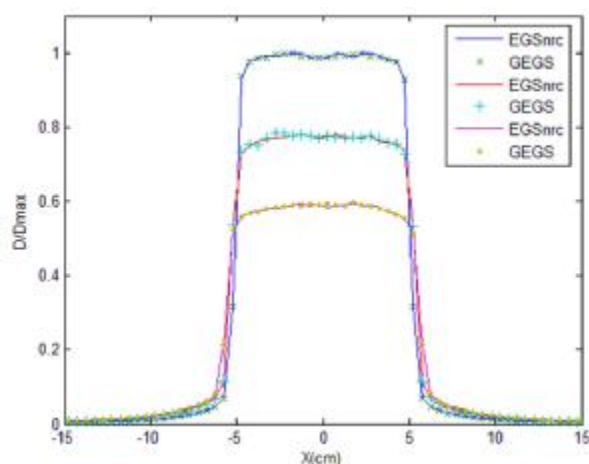


图 2 6 MeV 入射光子条件下 GEGS 和 EGSnrc 的离轴比曲线
Fig.2 OAR Curves With 6 MeV Photon of GEGS And EGSnrc

好的相合度。同时,我们给出了 GEGS 和 EGSnrc 各自的模拟速度,其中 GEGS 为每毫秒 722 个粒子,EGSnrc 速度为每毫秒 17 个粒子。可以看到,相比于 EGSnrc,我们 GEGS 在保留了模拟精度的前提下,极大的提高了模拟的效率。为了进一步验证 GEGS 的精度,我们对 GEGS 和 EGSnrc 的剂量结果进行了逐体素的检验,检验结果显示,整个模体所有体素中,两者计算误差小于最大剂量 1% 的体素占有所有体素的 99.47%, 误差小于最大剂量 2% 的体素占有所有体素的 99.99%。这进一步显示出了 GEGS 极大程度的保留了 EGSnrc 精度。

3 结论与讨论

在本文中,我们给出了我们最新开发的基于 GPU 的蒙特卡洛模拟软件包,同时对其速度与精度进行了测试,结果显示,相比于 CPU 平台下的蒙特卡洛模拟程序来说,我们的程序可以加速达到 43 倍,与此同时很好的保持了蒙特卡洛模拟的精度,为以后的日常临床应用打下了一个很好的基础。

但同时,我们可以预见到,我们的程序仍旧有很大的加速空间,这主要基于以下几点。第一,可以考虑

把部分截面数据存储到纹理内存之中,进一步提高数据读取的效率。第二,类似于多核 CPU 计算,我们可以考虑进行多 GPU 的并行计算,在 Jia^[14]的报告中,多 GPU 并行计算可以得到线性的加速表现。第三,进一步降低线程发散的情况,这个工作可以通过在读取相空间的过程中,对相空间粒子进行按照能量排序的预处理,这样保证每次并行模拟的粒子都有相近的能量,这就使其经历更加相近的物理过程,从而达到减小线程发散的目的。

【参考文献】

- [1] Andreo P. Monte Carlo techniques in medical radiation physics [J]. Phys Med Biol, 1991, 36: 861-920.
- [2] Prax G, Xing L. GPU computing in medical physics: A review[J]. Med Phys, 2011, 38: 2685-2697.
- [3] 甘阳谷,黄斐增. 基于 GPU 的蒙特卡洛放疗剂量并行计算[J]. 中国医学物理学杂志, 2012, 29(6): 3715-3717.
Gan YG, Huang FZ. GPU-based Parallel Monte Carlo Simulation for Radiotherapy Dose Calculation[J]. Chinese Journal of Medical Physics, 2012, 29(6): 3715-3717.
- [4] Jia X, Gu X, Sempau J, Choi D, et al. Development of a GPU-based Monte Carlo dose calculation code for coupled electron-photon transport[J]. Phys Med Biol, 2010, 55(11): 3077-3086.
- [5] Jahnke L, Fleckenstein J, Wenz F, et al. GMC: a GPU implementation of a Monte Carlo dose calculation based on Geant4[J]. Phys Med Biol, 2012, 57(5): 1217-1229.
- [6] Kohno R, Hotta K, Nishioka S, Matsubara K, et al. Clinical implementation of a GPU-based simplified Monte Carlo method for a treatment planning system of proton beam therapy[J]. Phys Med Biol, 2011, 56(22): 287-294.
- [7] Rogers DW. Fifty years of Monte Carlo simulations for medical physics[J]. Phys Med Biol, 2006, 51(13): R287-301.
- [8] Kawrakow I. Accurate condensed history Monte Carlo simulation of electron transport I[J]. Med Phys, 2000, 27: 485-498.
- [9] Ali ES, Rogers DW. Benchmarking EGSnrc in the kilovoltage energy range against experimental measurements of charged particle backscatter coefficients[J]. Phys Med Biol, 2008, 53(6): 1527-1543.
- [10] Farber R. CUDA application design and development[M]. Morgan Kaufmann, 2011.
- [11] Sanders J, Kandrot E. CUDA by example and introduction to general-purpose GPU programming [M]. Addison-Wesley, 2010.
- [12] NVIDIA, Cuda c best practices guide (R)(2012).
- [13] Hissoiny S, Ozell B, Bouchard H, et al. GPUMCD: A new GPU-oriented Monte Carlo dose calculation platform[J]. Med Phys, 2011, 38: 754-764.
- [14] Jia X, Gu X, Graves YJ, et al. GPU-based fast Monte Carlo simulation for radiotherapy dose calculation [J]. Phys Med Biol, 2011, 56: 7017-7031.