

## 基于GPU的蒙特卡洛模拟在辐射剂量计算中的研究进展及应用

黄莹, 刘海宽

复旦大学放射医学研究所, 上海 200032

**【摘要】**蒙特卡洛(MC)算法在辐射剂量计算中发挥着重要作用,但是计算速度限制了其在临床中的应用。随着图形处理器(GPU)技术的发展,GPU并行加速方法被越来越多地应用到MC计算中。本文主要介绍基于GPU的MC在光子、电子和质子输运模拟过程在辐射剂量计算方面的研究进展及其在医学物理上的应用。

**【关键词】**图形处理器;蒙特卡洛

**【中图分类号】**R144.1;R312

**【文献标志码】**A

**【文章编号】**1005-202X(2017)10-0973-04

### Development and application of graphics processor units-based Monte Carlo simulation in radiation dose calculation

HUANG Ying, LIU Haikuan

Institute of Radiation Medicine, Fudan University, Shanghai 200032, China

**Abstract:** Monte Carlo calculation plays an important role in medical physics, but the application in routine clinical use is limited by its computing speed. With the development of graphics processor units (GPU), GPU parallel speedup was increasingly used for MC simulation. Herein, we report on the implementation of photon, electron and proton on a card equipped with GPU and its development in radiation dose calculation and application in medical physics.

**Keywords:** graphics processor units; Monte Carlo

### 前言

蒙特卡洛(MC)模拟方法凭借其粒子输运过程和几何建模的精确性,在医学物理与辐射剂量学领域发挥着重要作用。剂量计算在放疗计划的制定中有着重要作用,MC模拟凭借其准确性被称为剂量计算的金标准<sup>[1]</sup>。常用的MC计算软件主要包括EGS4/5<sup>[2-4]</sup>、EGSnrc<sup>[5]</sup>、MCNP<sup>[6]</sup>、PENELOPE<sup>[7-8]</sup>、GEANT4<sup>[9]</sup>等,但是MC模拟的速度和耗时问题成为限制其临床应用的主要因素,如在实际临床靶区剂量计算中,要在满足2.5%不确定程度下,常规基于CPU的MC剂量模拟计算需花费6 h<sup>[10]</sup>。因此采用多核CPU并行计算的方法分别加速模拟不同能量范围的光子、电子、质子的输运,成为MC剂量模拟计算的一个重要

研究方向。同时开发出了一些通过简化物理过程进行运算加速的MC程序,如VMC++<sup>[11]</sup>、MCDOSE/MCSIM<sup>[2, 12-13]</sup>、剂量计划方法(Dose Planning Method, DPM)<sup>[14]</sup>等。尽管上述方法使得MC计算的速度不断提升,但是距离临床所需的近似于实时MC计算还有很大的发展空间。图形处理器(GPU)由于其众多运算单元并行、高内存带宽、支持浮点数算法、单元计算的成本低、程序接口开放等优点,在MC模拟技术领域具有更为广阔的应用前景。

### 1 基于GPU光子、电子剂量的MC模拟

DPM是基于CPU的MC剂量加速包。Jia等<sup>[15]</sup>在DPM基础上开发了基于GPU的MC剂量计算包gDPM v1.0, gDPM将GPU所有的线程看作独立的计算单元,追踪源粒子和次级粒子的所有输运过程。使用20 MV的单能电子点源和6 MV的光子点源进行验证,对临床放疗能量范围的光子束和电子束均有较好的准确性,与2.27 GHz Intel Xeon CPU相比,在NVIDIA Tesla C1060 GPU卡上可以实现5.0~6.6倍的加速。

**【收稿日期】**2017-05-29

**【基金项目】**国家自然科学基金(11475047)

**【作者简介】**黄莹,硕士研究生,主要研究方向:医学物理, E-mail: 1621-1140003@fudan.edu.cn

**【通信作者】**刘海宽,副研究员, E-mail: liuhk@fudan.edu.cn

尽管gDPM v1.0与CPU模拟计算相比有较好的剂量特性,但是加速比相对较低(5.0~6.6倍)。Hissoiny等<sup>[16]</sup>开发了MC剂量计算包GPUMCD。GPUMCD是基于已经存在的MC软件包的物理过程,将所有粒子的输运过程当成一个整体,重写针对GPU的代码,通过与EGSnrc比较来验证该软件包模拟计算的准确性<sup>[5]</sup>。该软件包通过产生一个粒子队列来存放新产生的次级粒子,并顺次处理电子和光子输运过程。GPUMCD通过单独模拟电子和光子来消除不同粒子同时输运产生的线程分散,电子和光子分别放在两个队列中模拟,使得GPU一次仅模拟一种粒子。就准确率而言,98%甚至更多体素的 $\gamma$ 通过率2%~2 mm。在NVIDIA GTX 480卡上模拟15 MV 100万电子和400万光子所需时间少于0.3 s(包括CPU和GPU之间内存转换的时间)。GPUMCD的运行速度比EGSnrc快900倍,比DPM快200倍,另外与单个C1060卡相比,使用两个NVIDIA C1060卡有1.9倍的加速。

GPUMCD实现了较好的加速效果,重写的代码中缺少实际临床计划剂量计算的必要特征。Jia等<sup>[17]</sup>从两个方面优化gDPM代码,开发了gDPM v2.0。首先,为保证模拟的准确性,gDPM v2.0保持与DPM相同的粒子输运过程和不确度;再者,gDPM v2.0增加了各种临床常用关键内容(如计划中的能量密度图)以提高其临床适用性。与2.27 GHz Intel Xeon CPU处理器相比,gDPM v2.0使用NVIDIA Tesla C2050 GPU卡实现69.1~87.2倍的加速。gDPM是首个基于GPU可用于病人临床治疗剂量计算的MC剂量包,对于一个临床IMRT和VMAT计划,可以在36.1~39.6 s内完成MC剂量计算,标准差小于1%。

以上所有基于GPU的MC剂量计算软件均没有具体描述源信息的相空间函数。Townson等<sup>[18]</sup>在基于GPU的MC剂量包gDPM3.0上使用3种不同的相空间技术。第1种相继从患者独立的相空间读取粒子文件,在传输过程中根据粒子种类和能量进行分类;第2种方法是建立一个简单的二级准直器模型和运行的通量图保证病人独立的相空间可以被使用;第3种是PSL(phase-space-let)技术,PSL技术首先将源相空间根据粒子种类、能量和位置分类,然后使用新的代码运行。Townson将基于3种技术的结果与BEAMnrc/DOSXYZnrc进行相互验证,结果发现PSL技术能在计算效率和准确率之间建立最大的优化,从而将PSL作为gDPM3.0默认的相空间技术。对一个IMRT放疗计划使用PSL技术,97%通过率的情况下,单个GPU的计算时间为50 s, BEAMnrc/DOSXYZnrc的计算时间为8.4 h。

大多数的GPU-MC是在Nvidia CUDA平台上运行,限制了代码的可移植性。基于OpenCL(Open Computing Language)平台的MC计算工具goMC对MV级外照射光子和电子进行耦合光电子的模拟。采用EGSnrc MC包的电子输运过程和PENELOPE中随机铰链的方法来替代gDPM中剂量规划的方法。Tian等<sup>[19]</sup>计算6 MV光子线和15 MV电子线在均匀水模、水-骨组织-肺-水平板模体、半平板模体中的剂量分布,所有的情形都实现了与gDPM较好的一致性。由于不同的电子输运过程和开发环境,各种模拟情况下goMC的效率比gDPM低4%~16%。

基于个体化的CT或者锥形束CT(CBCT)剂量是目前备受关注的问题,MC计算可以实现较高的准确性,但是却受到计算效率的限制。Jia等<sup>[20]</sup>利用病人扫描的CT/CBCT图像得到基于GPU的光子输运代码gCTD,对源谱和通量图进行建模,在均匀水模中与EGSnrc相比可以实现400倍的加速,Zubal模体中实现76.6倍的加速,通过简单的几何调整可以评估CT扫描的图像剂量。在Trilogy直线加速器(瓦里安)OBI系统上对gTCD的X射线源进行验证,对于X射线点源,在Nvidia GTX 590上评估CBCT扫描的辐射剂量,对于 $10^9$ 个粒子的计算时间少于5 min。计算结果与之前测量和MC计算数据相比在量级上相同,但是绝对值上有差异。

以上研究表明,通过不同优化方法可实现基于GPU的光子、电子的输运过程MC剂量计算,并实现了不同程度的加速比,但是距离实际临床应用还有一定的发展空间。

## 2 基于GPU质子剂量的MC模拟

随着质子放疗的发展,其优越性也逐渐突出,基于GPU的MC在质子的输运也是当前研究的热门问题。Yepes等<sup>[21]</sup>用geant4代码在均匀的水模体中产生质子输运历程的数据库,在快速MC计算器(Fast Dose Calculator, FDC)基础上通过重复追踪的方法对基于GPU的质子剂量进行加速计算。在病人案例的剂量计算中,基于重复追踪的MC根据患者身体的材料在数据库中选择合适的质子径迹,根据散射角和轨迹长度的适当比例计算剂量分布。这种方法在计算上有效避免了采样过程中的物理相互作用,实现了计算模型与SIMD结构的兼容,每一个线程总是执行同一个操作,如轨迹重复。在配备Geforce GTX 295 GPUs的双GPU系统中1 min内可以实现1%的精确度,与基于CPU的模拟相比提速75.5倍。另一个基于GPU的质子剂量的计算是在简化蒙卡(Simplified Monte Carlo Method, SMC)上



进行的。Kohn等<sup>[22-23]</sup>在进行质子沉积能量的模拟计算中,沉积的能量由已知水的深度剂量分布等效模型决定,同时建立了多次库伦散射模型。这个简单的模型使得它与GPUSIMD结构相兼容,每一个线程根据目前质子的地位独立使用不同的数据执行相同的指令,在实现过程中使用高速的共享内存。在临床案例中较CPU实现了12.3~16.0倍的加速,在NVIDIA Tesla C2050 GPU仅花了9~67 s计算临床剂量,而且不确定度在可以接受的范围。

Yepes<sup>[21]</sup>和Kohn<sup>[22]</sup>等的努力缩短了质子剂量计算时间,但是从某种程度上讲都是在类MC的简化模型中进行的。Jia等<sup>[24]</sup>开发了兼顾计算效率和准确性的质子放疗剂量计算的全MC包(gPMC),每一个线程追踪质子的实际物理过程,其剂量计算的准确性与质子MC模拟剂量计算的金标准TOPAS/Geant4结果进行比较<sup>[25]</sup>。在NVIDIA C2050 GPU卡上,模拟1 000万质子花费时间6~22 s,相对不确定度1%左右。当两个线程同时在一个体素沉积能量时,基于GPU的MC剂量计算会出现内存写入冲突,能量沉积连续进行来获得正确的结果。这种内存写入冲突在光子剂量计算中也会出现<sup>[15]</sup>,但是在质子中更严重,因为质子的轨迹几乎是直线而且以同步的形式前进,导致内存写入冲突的可能性更大。Jia等<sup>[26]</sup>采用多剂量计数器技巧来减缓这个问题。使用多剂量计数器,随机调用能量沉积事件到一个计数器,减少了写入冲突的可能性,提高了计算效率。

虽然已有在GPU上实现快速MC的质子输运,但是对于质子核非弹性碰撞都是简化的MC过程。Tseung等<sup>[27]</sup>研究质子输运过程的加速,采用详细的质子核的弹性和非弹性碰撞。将与放疗相关的粒子考虑到质子碰撞的作用过程中,计算次级粒子产生的剂量、非均匀介质模体中剂量以及商业的放疗计划系统上复杂的头颈部计划剂量,并与geant4.9.6p2/TOPAS进行比较,剂量计算结果基本一致。在NVIDIA GTX680卡上模拟1 000万个粒子的时间大约20 s(包括CPU和GPU之间的数据转换)。

为更好发挥gPMC的功能,提高准确性和高效性,Qin等<sup>[28]</sup>在gPMC v1.0基础上进行改进,在OpenCL环境下开发了基于GPU的质子输运程序gPMC v2.0。gPMC v2.0更新了核相互作用的物理模型,扩展了计数函数使得不同类型粒子通量、沉积剂量都可以得到,同时可以得到线性传递能量(LETd),并且使用多计数器来减少计数过程中出现的内存写入冲突问题。在AMD Radeon R9 290x GPU模拟 $10^7$ 个粒子花费时间8~17 s,平均数据不确定度小于1%。

### 3 讨论

上述基于GPU加速的MC模拟,在光子、电子和质子输运方面均取得了不同程度的运算加速,这表明GPU加速的MC模拟在光子、电子、质子的输运方面具有广阔的应用前景,但也面临着GPU的硬件结构限制以及MC模拟线程分配等多方面的挑战。第一,GPU的调用结构单指令多线程,GPU的多处理器使用32个并行的线程运行程序,如果线程中使用if-else声明,程序一次占用一个线程,而其他的线程处于空闲期。实现高速运算的关键在于所有线程在同一个进程中,但是MC计算不同路径的数据是独立的;第二,GPU的内存速度与CPU相比是很慢的,获取随机内存的代价很高。在MC模拟中,所有的线程使用GPU的全局内存,每个线程随机访问不同的内存地址,极大减慢了MC计算速度。高能MC计算中,粒子经历很多物理过程,这些物理过程产生的次级粒子更增加了其复杂性。在GPU MC计算中,每种粒子占据一个线程<sup>[29]</sup>,需要在每一个线程之内完成粒子所有相关物理过程的计算,这些分歧线程由调用程序排序。所以,在任何时候GPU可能仅计算一些粒子历程;第三,GPU计算另一个具有挑战性的工作是计数,由于要从成千上万个粒子历程中收集数据,需要较慢的原子操作来避免数据写入的混乱。

现有的基于GPU的MC模拟多是针对特定应用的物理过程、应用范围、特定能量范围,或是仅仅涉及光子、电子、质子的模拟。另外,现有的研究多采用物理过程近似的方法进行GPU加速MC计算,一定程度上降低了MC计算结果的准确度。而建立基于GPU加速的GEANT的MC模拟平台,适用于质子、电子和光子在所有医学物理上的应用过程,包括医学影像、放疗、内照射治疗等。在GEANT平台上可以采用不经过近似物理过程的方法快速计算截面数据、散射角和能量损失等,增加了结果的可信度。另外,GEANT平台也可以解决GATE(Geant4 Application for Tomographic Emission)运行时间较长的问题,为医学影像和放疗最受欢迎的MC模拟平台GATE的广泛使用奠定了基础<sup>[30]</sup>。

MC模拟要同时兼顾准确性和效率的问题。不同的计算单元计算不同的粒子输运过程,这对于基于CPU的MC计算适用,加速比和处理器的个数成线性关系<sup>[31]</sup>。根据Shen等<sup>[32]</sup>提出的多核CPU和GPU结合的可能性,Kalantzis等<sup>[33]</sup>利用OpenMP将GPU和CPU同时使用加速MC模拟,这种混合方法减少了每一个粒子的模拟时间,在基于GPU的MC模拟中实现了67.2倍的加速,在不影响准确度的情

况下改进的混合方法提速20%左右。未来CPU和GPU结合为MC模拟提供了更好的前景。

## 【参考文献】

- [1] ROGERS D. Fifty years of Monte Carlo simulations for medical physics[J]. *Phys Med Biol*, 2006, 51(13): R287-301.
- [2] MA C, MOK E A, PAWLICKI T, et al. Clinical implementation of a Monte Carlo treatment planning system[J]. *Med Phys*, 1999, 26(10): 2133-2143.
- [3] BIELAJEW A F, HIRAYAMA H, NELSON W R, et al. History, overview and recent improvements of EGS4NRCC Report PIRS-0436 National Research Council[C]. Ottawa, Canada.
- [4] HIRAYAMA H, NAMITO Y, BIELAJEW A F, et al. The EGS5 Code System SLAC-R-730, KEK 2005-8 (Menlo Park, CA: Stanford Linear Accelerator Center)[C].
- [5] KAWRAKOW I. Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version[J]. *Med Phys*, 2000, 27(3): 485-498.
- [6] BRIEEMEISTER J F. MCNP-A General Monte Carlo N-Particle Transport Code[C]. 2010.
- [7] BARÓ J, SEMPÁU J, FERNÁNDEZ-VAREA J M, et al. PENELOPE: An algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter[J]. *Nucl Instrum Methods Phys Res*, 1995, 100(1): 31-46.
- [8] SEMPÁU J, ACOSTA E, BARO J, et al. An algorithm for Monte Carlo simulation of coupled electron-photon transport[J]. *Nucl Instrum Methods Phys Res*, 1997, 132(3): 377-390.
- [9] STÉZOWSKI O. Geant4: a simulation toolkit[J]. *Nucl Instrum Methods Phys Res Sect A*, 2013, 506(3): 250-303.
- [10] PAGANETTI H, JIANG H, PARODI K, et al. Clinical implementation of full Monte Carlo dose calculation in proton beam therapy[J]. *Phys Med Biol*, 2008, 53(17): 4825-4853.
- [11] KAWRAKOW I, FIPPEL M, FRIEDRICH K. 3D electron dose calculation using a voxel based Monte Carlo algorithm (VMC)[J]. *Med Phys*, 1996, 23(4): 445.
- [12] LI J S, PAWLICKI T, DENG J, et al. Validation of a Monte Carlo dose calculation tool for radiotherapy treatment planning[J]. *Phys Med Biol*, 2000, 45(10): 2969-2985.
- [13] MA C M, LI J S, PAWLICKI T, et al. MCDOS: a Monte Carlo dose calculation tool for radiation therapy treatment planning[M]//The Use of Computers in Radiation Therapy. Berlin: Springer, 2000.
- [14] SEMPÁU J, WILDERMAN S J, BIELAJEW A F. DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations[J]. *Phys Med Biol*, 2000, 45(8): 2263-2291.
- [15] JIA X, GU X, SEMPÁU J, 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.
- [16] HISSOINY S, OZELL B, BOUCHARD H, et al. GPUMCD: a new GPU-oriented Monte Carlo dose calculation platform[J]. *Med Phys*, 2011, 38(2): 754-764.
- [17] JIA X, GU X, GRAVES Y J, et al. GPU-based fast Monte Carlo simulation for radiotherapy dose calculation[J]. *Phys Med Biol*, 2011, 56(22): 7017-7031.
- [18] TOWNSON R W, JIA X, TIAN Z, et al. GPU-based Monte Carlo radiotherapy dose calculation using phase-space sources[J]. *Phys Med Biol*, 2013, 58(12): 4341-4356.
- [19] TIAN Z, SHI F, FOLKERTS M, et al. A GPU OpenCL based cross-platform Monte Carlo dose calculation engine (goMC)[J]. *Phys Med Biol*, 2015, 60(19): 7419-7435.
- [20] JIA X, YAN H, GU X, et al. Fast Monte Carlo simulation for patient-specific CT/CBCT imaging dose calculation[J]. *Phys Med Biol*, 2011, 57(3): 577.
- [21] YEPES P P, MIRKOVIC D, TADDEI P J. A GPU implementation of a track-repeating algorithm for proton radiotherapy dose calculations[J]. *Phys Med Biol*, 2010, 55(23): 7107-7120.
- [22] KOHNO R, HOTTA K, NISHIOKA S, 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): N287.
- [23] CHEN G T, SINGH R P, CASTRO J R, et al. Treatment planning for heavy ion radiotherapy[J]. *Int J Radiat Oncol Biol Phys*, 1979, 5(10): 1809-1819.
- [24] JIA X, SCHÜMANN J, PAGANETTI H, et al. GPU-based fast Monte Carlo dose calculation for proton therapy[J]. *Phys Med Biol*, 2012, 57(23): 7783-7797.
- [25] PERL J, SHIN J, SCHUMANN J, et al. TOPAS: an innovative proton Monte Carlo platform for research and clinical applications[J]. *Med Phys*, 2012, 39(11): 6818-6837.
- [26] JIA X, SCHUEMANN J, PAGANETTI H, et al. Development of GPMC V2.0, a GPU-based Monte Carlo dose calculation package for proton radiotherapy[J]. *Med Phys*, 2012, 39(6): 3945.
- [27] TSEUNG H W, MA J, BELTRAN C. A fast GPU-based Monte Carlo simulation of proton transport with detailed modeling of nonelastic interactions[J]. *Med Phys*, 2015, 42(6): 2967-2978.
- [28] QIN N, BOTAS P, GIANTSOURI D, et al. Recent developments and comprehensive evaluations of a GPU-based Monte Carlo package for proton therapy[J]. *Phys Med Biol*, 2016, 61(20): 7347.
- [29] BADAL A, BADANO A. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit[J]. *Med Phys*, 2009, 36(11): 4878-4880.
- [30] XIAO S, YANG X, SZTEJNBERG M, et al. Geant4 based Monte Carlo dose calculation engine for radiation therapy[J]. *IEEE Trans Nucl Sci*, 2010, 57(2): 775-781.
- [31] TYAGI N, BOSE A, CHETTY I J. Implementation of the DPM Monte Carlo code on a parallel architecture for treatment planning applications[J]. *Med Phys*, 2004, 31(9): 2721-2725.
- [32] SHEN W, WEI D, XU W, et al. Parallelized computation for computer simulation of electrocardiograms using personal computers with multi-core CPU and general-purpose GPU[J]. *Comput Methods Programs Biomed*, 2010, 100(1): 87-96.
- [33] KALANTZIS G, TACHIBANA H. Accelerated event-by-event Monte Carlo microdosimetric calculations of electrons and protons tracks on a multi-core CPU and a CUDA-enabled GPU[J]. *Comput Methods Programs Biomed*, 2014, 113(1): 116.

(编辑:黄开颜)