

DOI: 10.7652/xjtuxb201407011

GPU 加速窦房结计算机仿真的实现及优化

张虹, 郑霄, 赵丹

(西安交通大学电气工程学院, 710049, 西安)

摘要: 针对窦房结电生理计算机仿真运算量巨大、耗时长的问题,提出了基于高性能图形处理单元(GPU)实现并行计算及优化的方法。首先考虑窦房结细胞中央和边缘的差异,构建了一维非匀质窦房结组织模型;利用算子分裂方法使模型的解算任务具备并行性。根据具体解算过程提出了三种并行化策略,并对其中耗时最短的策略从线程块设置、数据交换频率以及存储模式等方面进行了进一步优化。结果表明:对于 500 个细胞的仿真,CUDA 程序较串行程序的执行时间下降了 60%,进一步优化后,CUDA 程序的执行时间可下降 84%;窦房结组织越大,GPU 的加速效果越明显。结果验证了 GPU 加速解算方法可显著提高窦房结模型的解算速度,降低实际执行时间。

关键词: 图形处理单元;并行计算;窦房结;计算机仿真

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 0253-987X(2014)07-0060-05

Implementation and Optimization of Computer Simulations with GPU Acceleration for Sino-Atrial Nodes

ZHANG Hong, ZHENG Xiao, ZHAO Dan

(School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A parallel computation method with its optimization based on graphic processing unit (GPU) is proposed to improve the problem that the electrophysiological simulations for sino-atrial nodes (SAN) needs large amount of calculation and time consumption. A one-dimensional inhomogeneous model for tissue of SAN is established by considering the difference of properties in the central and the edge of SAN cell. Then, the calculation of the model is parallelly performed by using the operator splitting method. Three parallelization strategies are put forward based on the calculation process. The strategy with the shortest running time is further optimized by considering the block size, the data transfer and data partitioning across various memories. Simulation results with 500 cells show that, compared with a serial program, the execution time of the non-optimized CUDA program decreases by about 60%, and the execution time further reduces to 84% when the CUDA program is optimized. It is also observed that the larger the tissue is, the more significant the acceleration effect of GPU is. It can be concluded that the proposed GPU accelerating method significantly speeds up the computation of SAN tissue model and decreases the execution time.

Keywords: graphic processing unit; parallel computation; sino-atrial node; computer simulation

窦房结为心脏的起搏点,研究其电活动规律可 为揭示心律失常,尤其是房性心律失常的内在机制

收稿日期: 2014-01-15。 作者简介: 张虹(1969—),女,副教授。 基金项目: 国家自然科学基金资助项目(81271661);教育部回国留学基金资助项目(第 40 批);中央高校基本科研业务费专项资金项目(xj2011087)。

网络出版时间: 2014-04-25 网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20140425.0930.005.html>

提供帮助。然而,描述窦房结细胞电活动行为的数学模型十分复杂,涉及的方程和参数众多。对于构成的组织模型,在细胞数量较多时,冗长的仿真时间往往成为制约研究的瓶颈。

近年来,随着图形处理单元(Graphic Process Unit,GPU)的发展,利用显卡上的处理器完成通用计算以加快模型的解算速度受到人们的关注^[1-2]。GPU 内往往包含数百个流处理器,可提供一个理想的并行体系结构,同时处理数以千计并发的线程。此外,GPU 全局和共享存储器的内存带宽都很大,因此读写速度快,处理效率高^[3]。这些特性为 GPU 应用于心脏计算机模型的解算提供了可能。但是,不同于以往无须关心计算机底层硬件的编程模式,要想提高 GPU 的加速效果,程序员往往需要针对不同的处理对象和不同的任务量从线程块的大小、CPU 主机和 GPU 设备之间的通信、整个数据在 GPU 上存储器的分配等诸多方面综合考虑后提出具体的解决方案^[4]。

针对心脏的计算机仿真,Yu 等人给出了 GPU 加速心脏解剖模型解算的方法^[5],主要研究对象是心脏的机械活动而非电活动。Sato 等人报道了 GPU 加速心室组织电活动仿真的可行性^[6],但并未对程序的优化进行探讨,而且所用的单细胞模型属于第一代,需要解算的方程和涉及的参数相对第二代动作电位动力学模型要简单得多。目前尚未见到 GPU 在窦房结仿真中应用的报道,因此本文以第二代窦房结单细胞模型为基础建立一维非匀质组织,提出并比较不同的并行化策略对解算速度的影响,并对 GPU 加速模型仿真的实现及优化方法进行了探讨。

文中配置的计算机显卡为 NVIDIA GeForce GTX550Ti。它由 4 个多处理器组成,每一个处理器包含 48 个流处理器。整个程序的实现基于统一计算设备架构 CUDA 4.2(The Compute Unified Device Architecture)。

1 计算及优化方法

1.1 模型解算的并行化策略

窦房结组织由许多细胞通过缝隙连接电导彼此耦合而成。组织中的膜电位变化遵循如下方程^[7]

$$\nabla \cdot (\sigma V) = A_m \left(C_m \frac{\partial V}{\partial t} + I_{ion} \right) \quad (1)$$

式中: V 是细胞膜电位; C_m 是膜电容; A_m 是细胞的表面积; t 是时间; I_{ion} 是流经细胞膜的各种离子

电流、交换电流以及泵电流的总和; σ 是细胞间缝隙连接电导。

为使模型的解算具备并行性,本文首先采用算子分裂法^[8]将式(1)分成描述单细胞电活动行为的常微分方程(2)以及细胞间电紧张扩步的偏微分方程(3)

$$\frac{dV}{dt} = -\frac{1}{C_m} I_{ion} \quad (2)$$

$$\frac{\partial V}{\partial t} = \frac{1}{A_m C_m} \nabla \cdot (\sigma V) \quad (3)$$

算子分裂使得解算时可以先计算每个细胞的动作电位,然后再考虑细胞间电耦合的相互作用。由于单细胞模型的解算实际是耗时最多的,尤其对组织细胞数量多的情况,因此算子分裂使不同单细胞的解算可以交给不同的线程同时处理,从而使解算任务具备了可并行性。

本文构建了一维非匀质窦房结组织。由于中心到边缘细胞形态的差异,因此每个细胞的膜电容和细胞大小的指数式计算系数^[9]

$$F = 1.07 \times 29.0 \frac{1.0n}{N} / (30.0(1.0 + 0.7745 \exp(-(29.0 \times \frac{1.0n}{N} - 24.5)/1.95))) \quad (4)$$

式中: F 代表第 n 个细胞对应的系数; N 代表总细胞数。

细胞自身电活动采用第二代动作电位动力学模型描述^[10],流经细胞膜的不同电流受各自通道门控因子的调节,门控因子又与膜电位呈函数关系。门控因子 s 与膜电位 V 之间的关系如下

$$\frac{ds}{dt} = F(V, s) \quad (5)$$

因此,单细胞的解算首先需要计算参数,即利用式(4)得到各细胞的实际膜电容和细胞大小,然后求解常微分方程(5)得到门控因子,再计算相应的离子电流,最后根据总的离子电流,求解常微分方程(2)得到单个细胞的膜电位。此处,所有常微分方程均采用欧拉法求解,时间步长取 0.03 ms。

由于一个细胞膜电位的求解涉及 20 多个常微分方程,运算量大,因此本文将这部分的计算全部交由 GPU 完成。根据解算过程,设计了三个核函数依次完成变量的初始化(核函数 1),电容、细胞大小等参数和门控因子的求解(核函数 2),离子电流和膜电位的计算(核函数 3)。其中,核函数 1、2 和 3 需要的寄存器数分别为 19、32 和 25,每个多处理器中的线程数分别为 256、1 024 和 1 024。

在获得单个细胞的膜电位后,按照算子分裂的思想,考虑细胞间耦合对电位的作用,求解偏微分方程(3)。此处,以单细胞电位计算的结果作为初始值,利用三点差分法解算式(3)。空间步长取0.1 mm。 σ 按照式(4)从中心到边缘呈指数形式变化。组织边界处采用无通量边界条件进行处理,即

$$(\nabla V)|_{x=0} = (\nabla V)|_{x=l_s} = 0 \quad (6)$$

式中: x 代表组织中的空间坐标; l_s 代表窦房结组织的总长度。

细胞间耦合部分的计算采用了两种策略,在并行策略1中由CPU处理,在并行策略2中由GPU完成。因此,策略2除了单细胞解算时设计的三个核函数外,还增加了核函数4用于扩散和边界条件的计算。核函数4用到的寄存器数为26,每个多处理器中的线程数为1024。可见,策略1和策略2是将整个解算任务细分为数个模块,交给不同的核函数分别完成。

为比较不同任务划分对运算速度的影响,在策略3中将核函数进行整合,函数数量由策略2中的4个缩减为2个。除初始化任务交给一个核函数完成外,其余计算任务均在一个核函数中完成。两个核函数需要的寄存器数分别为19和32,每个多处理器中的线程数分别为256和1024。

1.2 CUDA程序的优化

为提高性能,本文利用NVIDIA可视化性能分析器Visual Profiler对程序进行分析。根据分析结果,从线程块大小、数据交换以及存储器分配三方面对CUDA程序进行优化。文中选择三种并行策略中执行时间最短的策略作为优化的起点。

NVIDIA GeForce GTX550Ti的计算能力满足CUDA规格,因此每个多处理器最多可容纳1536个线程、8个线程块。为了使处于工作状态的线程占满整个多处理器,线程块中的线程总数应大于每个多处理器可容纳的线程数除以线程块数的总数(192)。根据多处理器的寄存器总数和每个线程需要的寄存器数可知,一个多处理器上可以容纳1024个线程。因此,针对本文的计算任务,将线程数设计为512,共2个线程块。

如果每执行一个时间步长的运算,都将数据由显存复制到主存,则大量的时间消耗在频繁的数据交换上。而一次尽可能多的在GPU上完成计算并将结果暂存,待一批计算任务完成后回传数据到主存,这样便可减小数据交换的频率。受限于GPU共享存储器的容量,本文以500个时间步长的计算

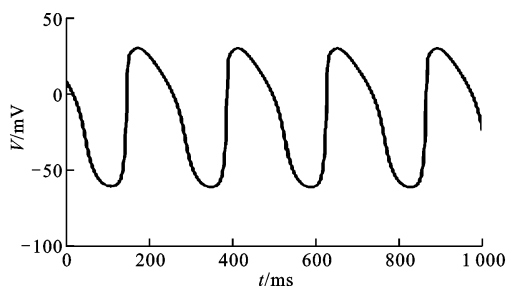
为一组安排一次数据交换,通过这种方式缩减主机与GPU间数据交换的次数。

由于GPU设备的全局存储器容量大且可被所有线程访问,所以往往作为数据的主要存储器,但其读写速度慢则降低了执行效率。本文通过调整存储结构,将计算中使用频率较高的数据(如膜电位、膜电容等)放入片上存取速度极快的共享存储器中以加快访问速度;局部变量尽量保存在寄存器或共享存储器中以提高访问速度。

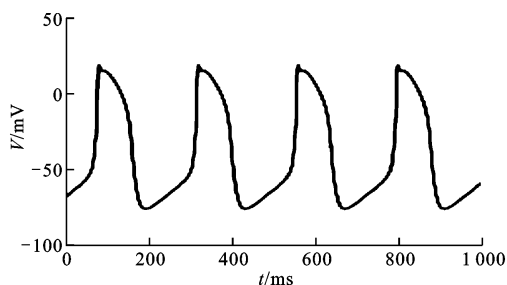
2 结果与讨论

2.1 组织细胞的动作电位

为确保文中GPU解算方法的正确性,首先计算得到了窦房结组织模型细胞的动作电位。图1a和图1b分别为中心细胞和边缘细胞的动作电位。



(a) 中心细胞



(b) 边缘细胞

图1 窦房结组织模型细胞的动作电位

可以看出,细胞具有在每一次复极化后静息电位自动抬升、逐渐达到阈值并触发又一次激动的自律特性。同边缘细胞相比,中心细胞除极速率慢、动作电位时程长。窦房结组织中心和边缘细胞的最大舒张电位分别为-57 mV、-76 mV,极上升速率分别为7.9 V/s、12.5 V/s。这些结果与已有研究^[9],以及利用CPU计算所得的仿真结果一致^[11],说明了本文GPU解算的准确性。

2.2 执行时间和加速比

为了评估GPU解算的效果,文中利用NVIDIA Visual Profiler测量了三种策略的实际执行时间。

图 2a 给出了仿真 1.6 s、200 个细胞的情况。结果表明,策略 3 的执行时间最短,策略 1 所需时间最长。这是因为,同其他策略相比,策略 1 的偏微分方程由 CPU 解算,无形中降低了程序的并行度,同时增加了主机和 GPU 设备之间数据交换的频率。策略 3 相对于策略 2,将多个核函数整合后可以减少因函数间调用和存储器频繁访问所耗费的时间,从而提高了整体的执行效率。由此可见,尽可能地提高任务的并行度且减少核函数的数量有利于提高运算速度。

图 2b 给出了策略 3 不同任务。可见,GPU 上计算门控因子、离子电流、膜电位以及电扩布 SAN_computing 耗时占比 72.68%,在 CPU 上执行任务的时间占 21.31%;变量初始化的执行时间占 1%;CPU 和 GPU 之间通过内存复制交换数据的时间占 4.01%;为了等待线程同步所耗费的时间占 1%。作为主机,CPU 负责内核函数的调用、内存的分配和数据交换等任务。

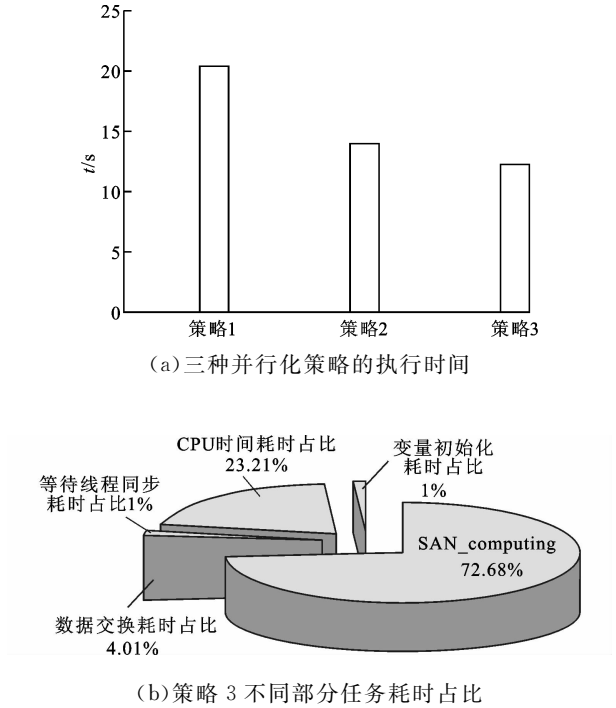


图 2 三种并行策略的执行时间及策略 3 不同任务耗时占比情况

在三种并行策略中,由于策略 3 的执行时间最短、效率最高,因此本文对策略 3 进行了优化。图 3a 给出了不同细胞数量的组织模型采用串行(标记*)、策略 3 CUDA 并行(标记○)及策略 3 优化(标记△)程序的执行时间。图 3b 为策略 3 的 CUDA 并行程序优化前后的加速比。图 3a 反映出,对

于 200 和 500 个细胞的组织,CUDA 并程序的执行时间远少于串行程序,而且组织越大,细胞数量越多,加速效果越显著,优化后可使执行时间进一步缩短。例如,500 个细胞的组织仿真,相对于串行程序,CUDA 并行程序在优化前执行时间减少了 60%,优化后执行时间减少了 84%。但是,对 30 个细胞而言,CUDA 并行程序优化前执行时间为 8.92 s,反而比串行程序的执行时间长,优化后执行时间减少到 5.86 s,但仍大于串行程序的执行时间。这可能是由于,在细胞数量少、模型组织小的情况下,并行计算节约的时间小于 CPU 和 GPU 间数据交换所耗费的时间,造成总体执行时间的延长。这说明 GPU 加速仿真的方法对于小规模运算不但无法发挥其优势,反而会降低执行效率。从图 3b 也可看出,组织越大,加速比越大,优化后效果显著。

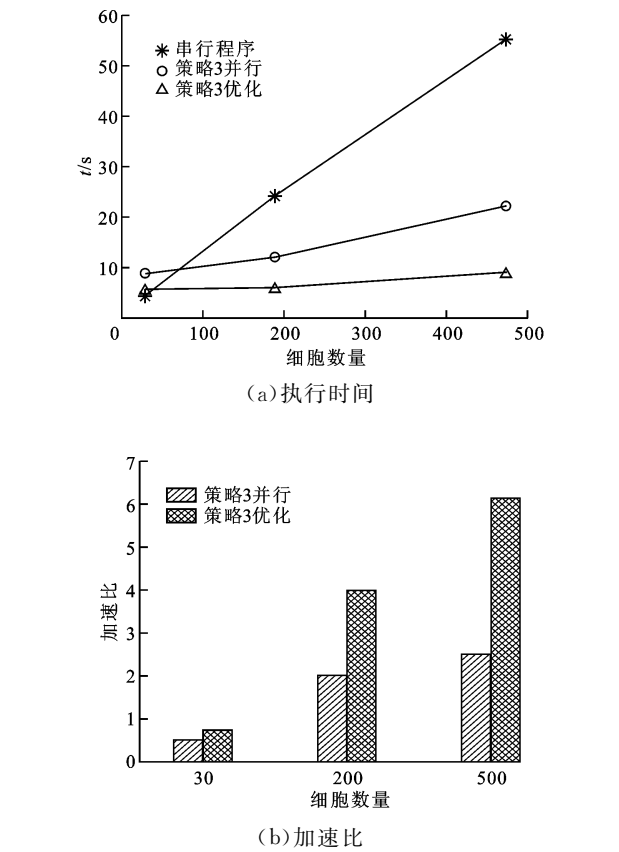


图 3 不同细胞数量和编程策略对执行时间和加速比的影响

图 4 为策略 3 CUDA 并行优化后各任务耗时占比。与图 2b 相比,GPU 上的计算时间增加了 6.08%,CPU 上任务的执行时间减少了 16%,CPU 与 GPU 间数据交换时间减少了 25.19%。可见,通过分组缩减、批量传输的方式可显著降低数据交换耗时。

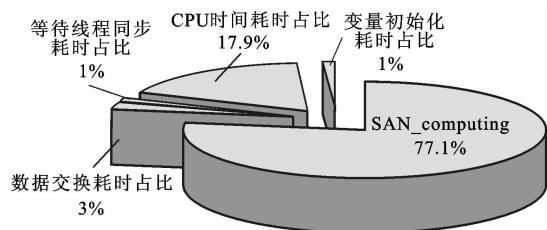


图4 策略3优化后不同任务耗时占比

3 结 论

本文对利用 GPU 加速窦房结组织电活动仿真的方法进行了探索。结果表明:采用算子分裂后,窦房结组织模型解算任务本身具备的可并行性为利用 GPU 实现加速提供了可能。将单细胞及扩散部分的求解全部交给 GPU 尽量增大并行度,通过批量传输减少数据交换频率,以及充分利用共享存储器和寄存器等措施,可使加速效果明显。本文提出的并行化策略和优化方法为未来在个人电脑上进行更为复杂的窦房结组织仿真研究奠定了基础。

参考文献:

- [1] 赵丽丽, 张盛兵, 张萌, 等. 基于 CUDA 的高速 FFT 计算 [J]. 计算机应用研究, 2011, 28(4): 1556-1559. ZHAO Lili, ZHANG Shengbin, ZHANG Meng, et al. High performance FFT computation based on CUDA [J]. Application Research of Computers, 2011, 28(4): 1556-1559.
- [2] JAYSHREE P, JITENDRA K, MADHURA B A. GPGPU processing in CUDA architecture [J]. Advanced Computing: An International Journal, 2012, 3(1): 105-105.
- [3] MACEDONIA M. The GPU enters computing's mainstream [J]. Computer, 2003, 36(10): 106-108.
- [4] HOANGTRONG T M, WILLIAMS G S B, LEDERER J W, et al. GPU-enabled stochastic spatiotemporal model of rat ventricular myocyte calcium dynamics [J]. Biophysical Journal, 2011, 100(3): 557.
- [5] YU Rongdong, ZHANG Yin, ZHANG Sanyuan, et al. GPU accelerated simulation of cardiac activities [J]. Journal of Computers, 2010, 5(11): 1700-1704.
- [6] SATO D, XIE Yuanfang, WEISS J N, et al. Acceleration of cardiac tissue simulation with graphic processing units [J]. Medical & Biological Engineering & Computing, 2009, 47(9): 1011-1015.
- [7] POTSE M, DUBE B, RICHER J, et al. A comparison of monodomain and bidomain reaction-diffusion models for action potential propagation in the human

heart [J]. IEEE Transactions on Biomedical Engineering, 2006, 53(12): 2425-2435.

- [8] 金印彬, 杨琳, 张虹, 等. 二维心室肌中动作电位传导的数值算法研究 [J]. 西安交通大学学报, 2004, 38(8): 851-854. JIN Yinbin, YANG Lin, ZHANG Hong, et al. Numerical algorithm for conduction of action potential in two dimensional cardiac ventricle tissue [J]. Journal of Xi'an Jiaotong University, 2004, 38(8): 851-854.
- [9] GARNY A, KOHL P, HUNTER P J, et al. One-dimensional rabbit sinoatrial node models: benefits and limitations [J]. Journal of Cardiovascular Electrophysiology, 2003, 14(10): 121-132.
- [10] ZHANG Henggui, HOLDEN A V, KODAMA I, et al. Mathematical models of action potentials in the periphery and center of the rabbit sinoatrial node [J]. American Journal of Physiology: Heart and Circulatory Physiology, 2000, 279(1): 397-421.
- [11] ZHANG Hong, WANG Juan, WU Ruijuan, et al. Effects of atrium on the electrical activity of sino-atrial node [J]. Key Engineering Materials, 2011, 467: 742-747.

[本刊相关文献链接]

- 唐铁轩, 吴俊敏, 陈国良, 等. 并行片上网络仿真器 ParaNSim 的设计及性能分析. 2012, 46(2): 24-30. [doi: 10. 7652/xjtuxb201202005]
- 张学一, 付晓梅, 周化锰. 一种格约减辅助的协作多址接入方案. 2012, 46(4): 100-105. [doi: 10. 7652/xjtuxb201204017]
- 刘海涛, 王磊, 赵万华. 考虑模态特性的高速机床进给系统刚度匹配研究. 2014, 48(1): 90-95. [doi: 10. 7652/xjtuxb201401015]
- 丁军平, 蔡皖东. 面向对等网络 SEInR 传播模型的理论分析. 2012, 46(2): 31-37. [doi: 10. 7652/xjtuxb201202006]
- 刘学, 焦淑红, 司锡才. 观测域求积分卡尔曼滤波的机载无源定位算法. 2011, 45(5): 137. [doi: 10. 7652/xjtuxb201105025]
- 张保, 曹海军, 董小社, 等. 面向图形处理器重叠通信与计算的数据划分方法. 2011, 45(4): 1-4. [doi: 10. 7652/xjtuxb201104001]
- 康雨, 闫相国, 郑崇勋, 等. 任意可分负载的多轮调度算法. 2009, 43(8): 125-129. [doi: 10. 7652/xjtuxb200908025]
- 郭振宇, 程博, 叶敏, 等. 一种并行混沌差异演化算法. 2007, 41(3): 299-302. [doi: 10. 7652/xjtuxb200703008]
- 王润新, 刘进军. 无须考虑电源接法和性质的三相电压型整流器平均模型. 2009, 43(12): 81-85. [doi: 10. 7652/xjtuxb200912017]

(编辑 赵炜)