
基于集成众核的 3D 蒙特卡罗半导体器件模拟*

方民权¹⁺, 张卫民¹, 张理论¹², 曾琅³⁴, 刘晓彦⁵, 尹龙祥⁵

¹(国防科学技术大学 计算机学院,长沙 410073)

²(国家超级计算广州中心,广州 510006)

³(北京航空航天大学 电子信息工程系,北京 100091)

⁴(北京航空航天大学 自旋电子交叉研究中心,北京 100091)

⁵(北京大学 微纳电子学研究院,北京 100871)

Monte Carlo Simulator For Semiconductor Devices on Many Integrated Core*

FANG Min-Quan¹⁺, ZHANG Wei-Min¹, ZHANG Li-Lun¹², ZENG Lang³⁴, LIU Xiao-yan⁵, YIN Long-xiang⁵

¹(School of Computer Science, National University of Defence Technology, Changsha 410073, China)

²(National Supercomputing Center in GuangZhou, GuangZhou 510006, China)

³(School of Electrical and Information Engineering, Beihang University, Beijing 100091, China)

⁴(Spintronics Interdisciplinary Center, Beihang University, Beijing 100091, China)

⁵(Institute of Microelectronics, Peking University, Beijing 100871, China)

+ Corresponding author: Phn: +86-15874293052, E-mail: fmq@hpc6.com

Received 2014-07-30; Accepted 2014-XX-XX

Abstract: 3D Monte Carlo simulation for semiconductor devices consumes long time. Especially, when grids are growing and particles are increasing, computing scale becomes very large. By analysing the hotspots and the second level parallelism, the parallel schemes on Many Integrated Core for Effective Potential Method is presented; Parallelization of particle free fighting, statistics of simulation information and surface roughness scattering rate are researched. 3-level parallel Monte Carlo simulator for semiconductor devices is implemented and validated. Result shows that the 3-level parallel program can get better performance than original version. When accuracy is improved, the 3-level parallel program can get greater speed-up.

Key words: Monte Carlo; Device Simulating for semiconductor; Many Integrated Core; Effective Potential Method; particle free flight

摘 要: 3 维蒙特卡罗器件模拟计算量大,计算量随网格与粒子数增加而急剧增加.通过分析 3D 蒙卡模拟加速热点和进一步可并行性,研究有效电势方法的集成众核并行方案;研究粒子自由飞行、统计模拟信息、计算表面粗糙散射等热点并行方案,最终实现基于 CPU/MIC 的三级并行 3 维蒙特卡罗器件模拟软件.实验结果显示三级并行比单

* Supported by the National Natural Science Foundation of China under Grant No. 41375113 (国家自然科学基金); the National Natural Science Foundation of China under Grant No.61272146 (国家自然科学基金).

作者简介: 方民权(1989—),男,浙江东阳人,博士研究生,主要研究领域为大规模科学与工程计算;张卫民(1966—),男,浙江诸暨人,研究员,博士生导师,主要研究领域为数值天气预报,大规模科学与工程计算.

级并行获得更好的性能;当提高模拟精度时,三级并行蒙特卡罗模拟对单级并行加速比增加。

关键词: 蒙特卡罗;半导体器件模拟;集成众核;有效电势方法;粒子自由飞行

中图分类号: TP391

文献标识码: A

传统的半导体器件开发方法以试验为基础,即通过建立工艺程序,试制并完成器件的检测,再对比初始指标是否符合,若不符合则须重新修改工艺程序和器件结构参数,进行新的试制过程^[1]。试制无法一次成功,往往需要重复非常多次;根据摩尔定律,基层电路复杂性急剧增加,试制复杂度也随之增加;以上因素导致传统试制方法成本巨大,已不适合现代半导体器件开发。而利用计算机辅助设计技术的半导体器件模拟由于其效果理想、成本低廉等优点被广泛使用。半导体器件模拟从一维、二维到三维,其计算量急剧增加,巨大的计算时间消耗严重影响了半导体器件的开发速度。因此需要引入高性能并行计算技术,加速半导体器件模拟过程,降低半导体器件开发周期,最终促进微电子产业发展。

当前高性能计算机快速发展,世界超级计算机 TOP500 更新换代,搭载集成众核(Many Integrated Core, MIC)结构产品 Intel Xeon Phi 协处理器的中国天河 2 号超级计算机已蝉联 3 届榜首^[2]。相对于超级计算机的快速发展,配套的高性能应用软件相对较弱^[3]。本文利用高性能计算研究蒙特卡罗器件模拟,在一定程度上填补高性能应用软件领域的空缺。

国内外也有很多专家学者对器件模拟进行并行化研究,比如 Donald 等^[4]研究了三维器件模拟的大规模并行算法;Chow 等^[5]采用 MIMD 实现了砷化镓 MESFET 器件并行电热模拟,加速 14 倍;Zhang^[6]等研究了 MPI 并行蒙特卡罗器件模拟器。在 CPU/GPU 异构系统上,Faria 等^[7]基于 GPU 实现 1 维和 2 维器件电磁模拟,加速 30 倍;Tsuyoshi 等^[8]基于 GPU 研究了 PIC-FDTD 算法,加速 4 倍。在最新的 CPU/MIC 高性能计算架构上,还未有人做过相应的研究,本文基于 CPU/MIC 异构架构研究蒙特卡罗器件模拟。

蒙特卡罗(Monte Carlo, MC)^[9-10]模拟方法是研究半导体器件模拟的经典方法,相对于其他方法,比如流体动力学模型(Hydrodynamic, HD)^[11]、漂移扩散模型(Drift Diffusion, DD)^[10,12]等,MC 方法能更细致地模拟粒子运动,结果也更准确。但 MC 方法是随机模拟方法,要获得精确的模拟结果需要大量的迭代计算,特别是在模拟 3 维器件时,大量的网格和粒子数将导致计算规模急剧增长^[6,9]。基于巨大的计算需求以及模拟过程中存在的可并行性,本文对蒙特卡罗模拟方法进行并行化研究。

本文结构安排如下,第 1 部分简单介绍蒙特卡罗器件模拟方法,分析加速热点和进一步可并行性;第 2 节研究有效电势方法的集成众核并行化;第 3 部分研究粒子并行自由飞行、统计模拟信息、计算表面粗糙散射等热点的并行方案;第 4 节分析实验结果;最后是总结与展望。

1 蒙特卡罗器件模拟与热点分析

1.1 蒙特卡罗器件模拟方法

蒙特卡罗方法是用概率解决物理和数学问题的统计数值方法,通过直接求解玻尔兹曼方程(Boltzmann Transport Equation, BTE),对分布不做任何假设,适用于模拟和随机过程相联系的物理现象^[10]。蒙特卡罗模拟可以正确描述器件中复杂的非局域输运现象。

玻尔兹曼方程:

$$\frac{\partial f}{\partial t} + \vec{u} \cdot \nabla_r f - \frac{q\vec{E}}{h} \cdot \nabla_k f = \frac{\partial f}{\partial t_{coll}} \quad (1)$$

$$\frac{\partial f}{\partial t_{coll}} = \frac{V}{2\pi} \int [W(k',k)(1-f(k))f(k') - W(k,k')(1-f(k'))f(k)]dk \quad (2)$$

其中 $f=f(x_j, u_j, t)$ 为分布函数, x_j 、 u_j 是广义坐标 x 和群速度 u 的分量, coll 表示碰撞;蒙特卡罗方法利用随机过程求解 BTE, 方程左边载流子运动方程:

$$\frac{d\hbar k}{dt} = q\vec{E} \quad (3)$$

$$\frac{d\vec{r}}{dt} = \vec{u} \quad (4)$$

其中 t_f 是自由飞行时间, 满足 $\int_0^{t_f} \lambda(t) dt = -\ln(1-r_t)$, r_t 是随机数, 满足 0-1 均匀分布, $\lambda(t)$ 是散射率函数。

3 维蒙特卡罗模拟中每个时间步主要包括 Poisson 方程求解和模拟粒子自由飞行两个关键部分。求解 Poisson 方程前需要统计模拟过程中的信息, 如声子散射次数, 电子密度、电流大小等, 还要更新电子浓度信息。通过 Poisson 方程的求解获得电势, 计算有效电势后转换为电场, 计算网格电荷、表面粗糙散射率等。粒子自由飞行中假设粒子间没有相互作用, 粒子仅在电场作用下飞行。

1.2 加速热点与进一步可并行性分析

3D 蒙特卡罗器件模拟是一个弱并行^[13]问题, 其模拟过程是一个严格迭代的过程, 只能从每次迭代(时间步)中寻找可并行的热点。运行 3D 蒙特卡罗程序模拟 20nm 双栅晶体管 FinFET, 通过测试获得模拟中每次迭代过程各函数的时间消耗, 如表 1 所示。表中显示, 解 Poisson 方程获得电势并转换为电场(compute_field)过程耗时最大, 占了近 80%; 其次是粒子自由飞行(update_particle)、模拟粒子后统计(statistic)等过程。求解电场、粒子自由飞行、模拟后统计等过程都是并行化考虑的热点。

Table 1 Time-consuming functions of 3DMC simulation

表 1 3DMC 模拟耗时函数

Functions	Every Iteration Time(ms)	ratio
compute_field	3609.64	80.40%
update_particle	636.37	14.17%
statistic	146.43	3.26%
particle_to_density	37.93	0.84%
GetSurfRoughnessPhononScRate	35.01	0.78%
compute_cell_charge	24.35	0.54%
...	0.14	0.00%

在不同的 FinFET 模拟精度情况下, 即对 3 维网格进行细分以提高模拟精度, 分别测试 1 维、2 维、3 维网格细分情况下蒙特卡罗模拟中每个时间步中各函数的时间, 计算其相对原始网格的时间变化比率(图 1)。图中数据显示, 电场计算过程变化率最大, 说明该过程随精度提高耗时增长最快, 另表 1 显示其耗时最长, 根据 Amdahl 定律^[14]对耗时多的部分做并行可获得较好的加速效果, 因此将计算电场过程作为本文考虑的首要加速热点。其他函数的变化率均小于总变化率, 说明提高精度时, 这些函数耗时比率会有所下降, 总和必小于 20%, 因此是次要加速热点。

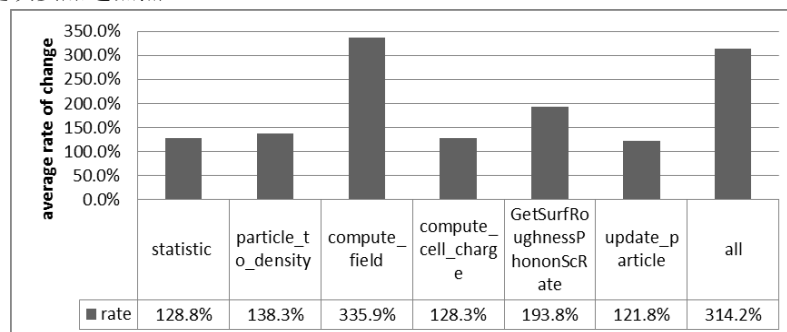


Fig.1 Change rate for different functions when accuracy improving

图 1 提高精度时各函数耗时变化率

在基于 MPI 并行 3D 蒙特卡罗模拟器中, 进程间的网格划分是对粒子传输方向(y 方向)进行划分的, 网格中 y 方向数量有限, 因此进程数量无法无限制扩展。在 3 维器件网格中, 还有 x 与 z 两个方向的并行性

能没有挖掘, 本文基于此进一步挖掘 3 维蒙特卡罗器件模拟的并行性。

2 基于 MIC 的有效电势方法并行化

2.1 有效电势方法

当器件尺寸缩小至纳米级, 载流子的量子效应增强, 并最终影响器件特性。处理量子效应的关键思路是用量子电势纠正传统 Poisson 电势。比如在 Wigner 修正中, 用 Wigner 变换方程计算量子电势, 薛定谔修正中求解薛定谔方程获得量子电势。有效电势方法(Effective Potential Method, EPM)^[15]通过求解高斯分布的 Poisson 电势卷积来描绘有效电势。该方法直截了当, 避免了薛定谔方程的求解(薛定谔量子修正法)和不确定载流子局部浓度二级派生物的计算(Wigner 量子修正法)。

其中载流子不仅是传统粒子, 还有类波性质, 可通过有效电势方法来捕获这种粒子类波行为。在 3 维模拟中, 有效电势定义为:

$$V_{\text{eff}}(x, y, z) = \iiint V(x', y', z') G(x-x', y-y', z-z') dx' dy' dz' \quad (5)$$

$$G(x, y, z) = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right) \quad (6)$$

其中 $V(x, y, z)$ 是求解 Poisson 方程获得的电势, $G(x, y, z)$ 是 3 维高斯分布函数, $\sigma_x \sigma_y \sigma_z$ 分别是 x, y, z 方向上的标准差。使用有效电势 $V_{\text{eff}}(x, y, z)$ 替代 Poisson 电势来计算粒子模拟的电场。

在 $i(j, k)$ 方向上, 有效电势:

$$V_{\text{eff}}(i) = \frac{1}{8} \sum_j [V(j) - E_B(j \in S_i O_2)] \times [\text{erf}(x_j'' - x_i) - \text{erf}(x_j' - x_i)] \times [\text{erf}(y_j'' - y_i) - \text{erf}(y_j' - y_i)] \times [\text{erf}(z_j'' - z_i) - \text{erf}(z_j' - z_i)] \quad (7)$$

其中 $V(j)$ 是在 j 网格上的 Poisson 电势, $[x_j', x_j''] \times [y_j', y_j''] \times [z_j', z_j'']$ 是计算卷积时 j 网格的环绕立方体。氧化场中 $E_B = 3.1 \text{ V}$, 变量 $\sigma_x = \sigma_y = \sigma_z = 0.5 \text{ nm}$, 积分范围 $|x-x'| < 4\sigma_x$ (y, z 同理)。3 维模拟中, 增加积分范围将导致大量计算。

2.2 集成众核上有效电势并行计算

在进程级, 真实空间在粒子传输方向 (y 方向) 上进行任务分割, 每个进程只需计算属于自己的网格的有效电势。每个网格上的电势计算都是独立的过程, 不同格点上的有效电势计算可以并发进行。在 3D 网格中, x 和 z 方向(甚至 y 方向)上仍存在可并行空间, 因此可开发线程级并行。3 维空间上任务划分有 3 种模式(图 2): 1. 按面划分, 即根据线程数量对 x 或 z 方向分割, 每个线程处理自身分配到的网格面上格点的有效电势计算; 2. 按线划分, 即同时对 x, z 两个方向进行分割, 每个线程分配的是器件网格化后一条竖线上的所有网格格点; 3. 按点划分, 即以网格格点作为基本分割单位, 将所有网格点平均分配给所有线程。集成众核中线程数量众多, 为充分利用资源, 应使子任务数量尽可能多, 因此理论上按点划分最为合适。

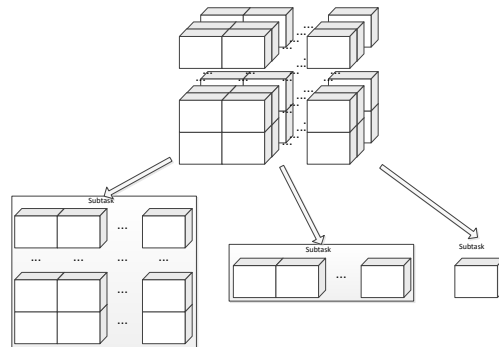


Fig.2 Task partitioning for effective potential in 3D grid

图 2 3D 网格有效电势任务划分

在上述并行任务划分方案基础上, 要利用集成众核计算有效电势, 需要将任务映射到集成众核上进行计算。具体实现时, 利用 ICC 编译器提供的 offload 功能, 将需要 MIC 计算的任务 offload 到 Intel Xeon Phi 上,

根据并行子任务数量、相应的 Phi 核数等参数, 合理设置协处理器上的线程数量, 以获得最大性能。针对有效电势计算过程, 在 CPU 求解 Poisson 方程获得电势后, 通过 offload 将有效电势计算分载到协处理器 MIC 上, MIC 完成有效电势的并行计算并返回结果, CPU 继续下一步, 即将电势转换为电场, 见图 3。

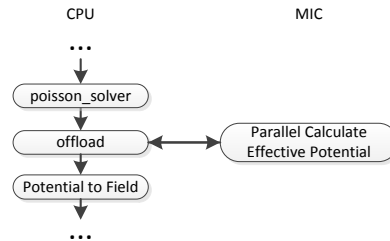


Fig.3 Flow chart for effective potential calculating on CPU/MIC

图 3 CPU/MIC 协作计算有效电势简单流程图

3 其他热点并行化

3.1 粒子并行自由飞行

模拟粒子自由飞行时, 假设粒子间没有相互作用, 仅在电场作用下飞行。给定一个时间步 dt , 假设时间步内电场不变。粒子飞行过程会遇到散射和飞越网格问题。电子散射主要包括声子散射、杂质散射和表面散射。N 个粒子的自由飞行过程如下:

```

for i= 1 to N do
|   step1  $t=dt$ 
|   step2 预计每个散射事件发生时间
|   step3 选择最先发生事件和时间  $dt'$ 
|   step4 求粒子自由飞行  $dt'$  时间的状态 (速度、位置、k 矢量等)
|   step5 处理散射事件
|   step6  $t=t- dt'$ 
|   step7 if  $t>0$  then goto step2
end for
  
```

在 3 维模拟过程中, 将器件网格分为很多块, 位于不同进程中, 粒子可能在一个时间步内飞出网格。各进程通过添加属于其他进程的“ghost”网格来追溯飞出网格的粒子运动, 在时间步结束前, 将“ghost”信息传输到目标进程。由于运动的不确定性, 粒子有可能飞出“ghost”块, 即“fly too far”。此时, 将“fly too far”的粒子标记为“unfinished”并停止该粒子模拟; 当存在“unfinished”粒子时, 启动新循环模拟这些粒子飞行。

在线程级, 由于同一进程的不同线程共享存储, 因此线程间不需要类似“ghost”的处理。在粒子飞行模拟假设中, 粒子间无相互作用, 因此不同粒子可以同时飞行, 即多个线程同时模拟不同的粒子飞行。由于粒子分布不均匀, 每个网格中的粒子数量不一。当并行任务区域划分比较集中时, 各线程所在区域内粒子数量可能相差较大, 导致负载不均衡。为了达到较好的负载均衡, 缩小划分单位, 避免集中分配, 以获得更均衡的负载; 避免集中不均匀分配还可采用乱序分配、动态分配等方法。

3.2 并行统计模拟信息

在上一个时间步模拟粒子自由飞行结束后, 下一个时间步求解 Poisson 方程前, 需要统计模拟过程中的信息, 如声子散射次数, 电子密度、电流大小、电子浓度信息等。网格格点上的信息通过统计网格内粒子对网格的贡献获得的, 由于粒子与网格距离的不同, 其贡献也有差距, 简单地说, 粒子与格点距离越近, 其贡献也越大。比如粒子电荷对格点的贡献:

$$\begin{aligned}
Q_{i,j,k}^{par} = & \sum_{(ii) \vec{r}_{ii} \in \Omega_{i-1,j-1,k-1}} Q_{ii}^{par} \frac{x_{ii}-x_{i-1}}{x_i-x_{i-1}} \frac{y_{ii}-y_{j-1}}{y_j-y_{j-1}} \frac{z_{ii}-z_{k-1}}{z_k-z_{k-1}} + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i,j-1,k-1}} Q_{ii}^{par} \frac{x_{i+1}-x_{ii}}{x_{i+1}-x_i} \frac{y_{ii}-y_{j-1}}{y_j-y_{j-1}} \frac{z_{ii}-z_{k-1}}{z_k-z_{k-1}} \\
& + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i-1,j,k-1}} Q_{ii}^{par} \frac{x_{ii}-x_{i-1}}{x_i-x_{i-1}} \frac{y_{j+1}-y_{ii}}{y_{j+1}-y_j} \frac{z_{ii}-z_{k-1}}{z_k-z_{k-1}} + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i,j,k-1}} Q_{ii}^{par} \frac{x_{i+1}-x_{ii}}{x_{i+1}-x_i} \frac{y_{j+1}-y_{ii}}{y_{j+1}-y_j} \frac{z_{ii}-z_{k-1}}{z_k-z_{k-1}} \\
& + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i-1,j-1,k}} Q_{ii}^{par} \frac{x_{ii}-x_{i-1}}{x_i-x_{i-1}} \frac{y_{ii}-y_{j-1}}{y_j-y_{j-1}} \frac{z_{k+1}-z_{ii}}{z_{k+1}-z_k} + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i,j-1,k}} Q_{ii}^{par} \frac{x_{i+1}-x_{ii}}{x_{i+1}-x_i} \frac{y_{ii}-y_{j-1}}{y_j-y_{j-1}} \frac{z_{k+1}-z_{ii}}{z_{k+1}-z_k} \\
& + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i-1,j,k}} Q_{ii}^{par} \frac{x_{ii}-x_{i-1}}{x_i-x_{i-1}} \frac{y_{j+1}-y_{ii}}{y_{j+1}-y_j} \frac{z_{k+1}-z_{ii}}{z_{k+1}-z_k} + \sum_{(ii) \vec{r}_{ii} \in \Omega_{i,j,k}} Q_{ii}^{par} \frac{x_{i+1}-x_{ii}}{x_{i+1}-x_i} \frac{y_{j+1}-y_{ii}}{y_{j+1}-y_j} \frac{z_{k+1}-z_{ii}}{z_{k+1}-z_k}
\end{aligned} \quad (8)$$

计算时, 一个粒子会对 3 维网格的 8 个格点产生贡献, 将改变 8 个格点的对应值。如果同时计算两个相邻网格上粒子对网格的贡献或同一网格内粒子对网格的贡献时, 可能需要同时修改某些或全部网格格点的值, 导致出错。为了解决这一问题, 本文在并行设计时, 设计了跳格的方法, 即每次跳过 1 个网格, 两个不相邻的网格中粒子的对格点的贡献不会产生冲突, 可以同时计算, 如图 4 所示。

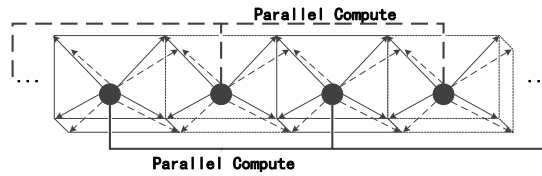


Fig.4 Parallel simulating information statistics on 1 dimension
图 4 单维度上模拟信息并行统计

3.3 并行计算表面粗糙散射

在求解 Poisson 方程并计算获得电场后, 需要统计网格电荷和计算表面粗糙散射。其中统计网格电荷只需要将网格内粒子电荷累加。由于实际工艺总会使栅氧和沟道间的界面存在粗糙度, 载流子与界面碰撞时将发生镜面反射或漫反射, 漫反射会对载流子沿沟道方向的运动分量产生影响, 最终将影响器件特性, 因此在模拟过程中需要考虑表面粗糙散射。表面粗糙散射率计算公式:

$$S_{sr}(E_{eff}) = \frac{\pi m^* e^2}{h^3} (\Delta \cdot L)^2 \cdot E_{eff}^2 \quad (9)$$

其中表面有效横向电场 $E_{eff} = \frac{\int E_T(z)n(z)dz}{\int n(z)dz}$, Δ 是表面粗糙幅度, L 是相关长度。在上述两个计算过程中, 不同网格的计算各自独立无相关, 不同网格上的计算可以并发进行。

4 实验结果

4.1 各热点线程级并行效果

本文在天河 2 号超级计算机上模拟 3 维 FinFET, 其中网格为 50*64*28。实现集成众核有效电势方法并行方案, 测试性能并与原始程序进行对比。表 2 统计了不同节点执行有效电势计算的耗时对比, 表中数据显示, 集成众核并行计算对比原始版本加速 66-82 倍不等, 说明有效电势方法非常适合在集成众核计算。

Table 2 Speedup of effective potential parallel calculating on MIC

表 2 集成众核上有效电势并行计算加速效果

Nodes	Original (ms)	MIC (ms)	Speedup
1	3522.68	53.38	66
2	1740.42	21.21	82
4	849.32	11.18	76
8	392.46	5.72	69
16	174.60	3.07	57

分别实现粒子自由飞行、统计模拟信息和计算表面粗糙散射等热点并行方案, 测试时间并与原始程序执行时间进行对比, 见表 3。表中数据显示, 添加了线程级并行后, 各热点的计算时间有了明显的缩减, 各热点加速 1.5-20.5 倍不等, 说明本文针对各个热点设计的并行方案是有效的。

Table 3 Parallel preference of thread-level for different hotspot

表 3 各热点线程级并行性能

Function	Original (ms)	OMP (ms)	speedup
statistic	146.43	14.14	10.4
particle_to_density	37.93	5.59	6.8
compute_cell_charge	24.35	4.15	5.9
GetSurfRoughnessPhononScRate	35.01	1.71	20.5
update_particle	636.37	422.76	1.5

4.2 三级并行优势

在 3D 蒙特卡罗半导体器件模拟软件包中实现上述集成众核并行方案和线程级并行方案, 新的蒙特卡罗模拟程序包含进程级、线程级、集成众核三个层次并行。分别用原始和三级并行 3DMC 软件包模拟 1000 个时间步的 3 维 FinFET, 通过测试获得平均每个时间步所消耗的时间 (图 5)。图中数据显示, 三级并行可以大幅减少半导体器件蒙特卡罗模拟时间, 加速 3.0~6.7 倍不等。进程数量增加时, 三级并行蒙特卡罗模拟对比单 MPI 的蒙特卡罗模拟加速比有所下降, 主要是增加进程时, 相同任务规模中每个进程分配到的任务减少, 进程中可并行分量的比率下降。

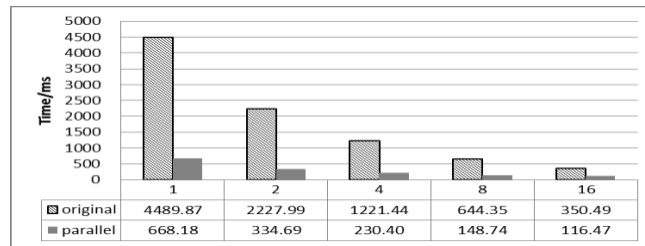


Fig.5 average time of every step for Monte Carlo simulation

图 5 单个时间步蒙特卡罗平均模拟时间

在真实模拟中, 利用 16 个节点并行模拟 30000 个时间步的 3 维 FinFET, 原始蒙特卡罗模拟软件包需要 3.2 小时, 而本文开发的三级并行蒙特卡罗模拟软件包仅需 1.1 小时, 明显减少了器件模拟时间。

4.3 提高精度两级并行性能分析

扩大网格尺寸, 提高模拟精度时, 对比三级并行和原始软件包进行半导体蒙特卡罗模拟时每个时间步的时间消耗, 表 3 统计了分别将网格放大 1、2、4、8 倍情况下使用 1 个和 16 个节点时的加速比。从表中数据不难发现, 随着网格的扩展, 模拟精度的提高, 两级并行蒙特卡罗器件模拟对比单级并行的加速比是增加的。说明本文开发的三级并行蒙特卡罗器件模拟软件包在模拟更高精度器件时更占优势。

Table 4 Change of speedup when grid changed

表 4 加速比随网格扩展变化

Nodes	original	2X	4X	8X
1	6.7	14.9	25.0	32.0
16	3.0	4.3	6.2	6.9

5 总结与展望

本文基于 3 维蒙特卡罗半导体器件模拟软件, 分析加速热点和进一步可并行性, 研究了有效电势方法的

集成众核并行方案以及其他热点的线程级并行方案, 最终实现 MPI+openMP+MIC 三级并行 3 维蒙特卡罗半导体器件模拟。实验结果显示, 三级并行软件包能获得比原始单级并行软件包更好的性能, 在精度扩展情况下, 三级并行蒙特卡罗模拟更具优势。

本文研究了高性能多级并行器件模拟的可行性, 下一步工作是基于 CPU/GPU 高性能计算架构研究加速蒙特卡罗半导体器件模拟过程, 或研究其他半导体器件模拟方法的高性能多级并行加速。

References:

- [1] Gong Lei. Designing of analysis software for semiconductor device simulation[D]. Hunan:Hunan University,2002.
- [2] Hans Meuer, Horst Simon, et al. TOP 500 Supercomputer Sites[EB/OL]. <http://www.top500.org/lists/>
- [3] Hong Wu,Xuebian Chi. Development Situation Analysis of High performance computer system and platform[EB/OL]. The 2012 Annual Report for Computer Science and Technology of China ,volume B.
- [4] Donald M.W.,Eric T.,et al. A Massively Parallel Algorithm for Three-Dimensional Device Simulation[J]. IEEE Transactions on Computer-Aided Design.1991(10):1201-1209.
- [5] Chow S.T.P.,Christopher M., et al. A Parallel Implementation of an Electrothermal Simulation for GaAs MESFET Devices[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,1996(15):308-316.
- [6] Wei Zhang, Gang Du, et al. A 3D Parallel Monte Carlo Simulator For Semiconductor Devices[C]. IEEE:Computational Electronics.beijing:2009:1-4.
- [7] Faria M.,Ferrera A.,et al. One and Two Dimensional Devices Electromagnetic Simulation Using Parallelism on GPUs[C]. Microwave & Optoelectronics Conference(IMOC) 2011 SBMO/IEEE MTT-S international.Natal:2011:924-927.
- [8] Tsuyoshi I.,Akiko O.,et al. GPGPU-based parallel computing of PIC-FDTD simulation for the development of novel terahertz radiation devices[J]. Physics and Simulation of Optoelectronic Devices XXII,SPIE.2014(8980):89801V_1-89801V_7.
- [9] Liangxiu Ye. Monte Carlo Simulation of Small Size Semiconductor Devices[M]. Beijing:Science press,1997.
- [10] Zhiliang Xia. Investigation of Transport in Nano-scale MOS Devices[D]. Beijing:Peking University,2007.
- [11] S.M.Sze. Physics of semiconductor devices[M]. JOHN WILEY&SONS,1981.
- [12] A.Forghieri,R.Guerreri,P.Ciampolini,et al. A New Discretization Strategy of the Semiconductor Equations Comprising Momentum and Energy Balance[J]. IEEE Trans. Computer-Aided Design,vol.7(2),1988:231-242.
- [13] Weiming Zhou. Multi_core Computing and Programming[M]. Wuhan:Huazhong university of science and technology press,2009.
- [14] Guoliang Chen. Parallel computing——Structure,algorithm,programming[M].third edition. Beijing:Higher education press,2011.
- [15] Wei Zhang, Gang Du, et al. 3D Full Band Monte Carlo Simulation of Nanoscale FinFETs Structures using Effective Potential Method[J]. Sci China Inf Sci.

附中文参考文献:

- [1]龚磊. 半导体器件模拟分析软件设计[D]. 湖南:湖南大学,2002.
- [3]武虹, 迟学斌. 高性能计算机系统及平台发展状况分析[EB/OL]. 2012 年度中国计算机科学技术年度报告 B 卷. <http://www.ccf.org.cn/resources/1190201776262/2013/12/30/2012baogaob.pdf>
- [9]叶良修. 小尺寸半导体器件的蒙特卡罗模拟[M]. 北京:科学出版社,1997.
- [10]夏志良. 纳米尺度 MOS 器件输运特性研究[D]. 北京:北京大学,2007.
- [13]周伟明. 多核计算与程序设计[M]. 武汉:华中科技大学出版社,2009.
- [14]陈国良. 并行计算——结构·算法·编程[M].第三版. 北京:高等教育出版社,2011.