

基于 GPU 加速的定向凝固相场模拟计算研究

胡延苏¹ 高 昂² 王志军³ 慕德俊⁴

(长安大学电子与控制工程学院 西安 710064)¹ (西北工业大学电子信息学院 西安 710072)²

(西北工业大学材料学院 西安 710072)³ (西北工业大学自动化学院 西安 710072)⁴

摘 要 相场法作为一种极具优势的微观组织数值模拟方法,已经在凝固微观组织演化机制的研究中得到了广泛应用。然而无论是从计算尺度还是微观组织演化时间上考虑,相场模拟计算量均非常大,对计算机有着非常高的要求。相对于传统的中央处理器(CPU)计算,图形处理器(GPU)计算是最近发展的一种高效计算手段。提出了一种基于 GPU 加速的定向凝固相场模拟计算策略,实现了大尺度条件下的定向凝固界面形态演化的加速计算。计算结果表明,对于单个计算机,GPU 计算与 CPU 计算的加速比可以高达 30 余倍。GPU 加速将为相场模拟的发展及应用带来新的契机。

关键词 相场,定向凝固,图像处理单元,统一计算设备架构

中图法分类号 TP301.6 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.7.005

GPU Application on the Phase-field Simulation

HU Yan-su¹ GAO Ang² WANG Zhi-jun³ MU De-jun⁴

(School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China)¹

(School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China)²

(School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China)³

(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)⁴

Abstract As a very preponderant microstructure numerical simulation method, the phase field method has been widely used in the research of solidification microstructure evolutionary mechanism. But whether the simulation scale or the evolution time is considered, there is huge computation which sets a high demand for the computer. Compared with the traditional CPU, GPU is much more efficient for the parallel computing. The paper proposed a GPU-based acceleration strategy to complete the directional solidification phase field simulation in large scale. The results show that the computation time can reduce 30 times compared with a single CPU, which brings a new opportunity for the phase field simulation in large scale.

Keywords Phase field, Directional solidification, Graphics processing unit (GPU), Compute unified device architecture (CUDA)

1 引言

利用相场法数值模拟定量再现实验中观察到的凝固微观组织演化过程,是近年来微观组织演化研究的重要方法之一。相场法作为一种极具优势的微观组织数值模拟方法,已经在凝固微观组织演化机制的研究中得到了广泛应用^[1-5]。以定向凝固界面形态演化为例,Boettinger 和 Warren^[1]首次采用相场法研究了高抽拉速度条件下,定向凝固胞状组织向平界面的转变,再现了高速生长下溶质截留现象,证实了线性稳定性理论预测的高速平界面状态的存在。Wang 等^[2]利用 B-S 热力学一致相场模型,预测了两种绝对稳定性出现的条件,再

现了随抽拉速度变化凝固界面形态发生深胞-浅胞-平界面的演化过程。Lan 等^[3]采用相场法研究了平界面初始失稳过程以及失稳后胞晶的淘汰、横向调整过程,得到了稳态胞晶形貌及一次间距随控制参数的变化规律。Gurevich 等^[4]采用相场法对三维胞晶的生长进行了细致研究,发现随着生长速度的变化,胞晶生长存在 3 种生长模式:低速下指状、高速下针状及中速下两者共存。综上所述,相场法已广泛应用于定向凝固微观组织形态演化和组织选择的研究,这为凝固组织演化机制的研究及凝固理论的发展提供了强有力的支持。

然而,虽然相场法在材料微观组织演化的研究中得到了广泛应用,但是由于其计算量巨大,目前大多数研究仍局限于

到稿日期:2014-12-25 返修日期:2015-04-18 本文受中国博士后科学基金资助项目(2013M540772),国家自然科学基金(61203233, 51101124, 51101125)资助。

胡延苏(1985—),女,博士,讲师,主要研究领域为并行计算、云资源调度, E-mail: huyansu@163.com; 高 昂(1984—),男,博士,讲师,主要研究领域为计算机网络、物联网应用; 王志军(1984—),男,副教授,硕士生导师,主要研究领域为凝固理论与凝固技术; 慕德俊(1963—),男,教授,博士生导师,主要研究领域为网络化控制、网络与信息安全。

较小尺度下微观组织演化的过程再现与机理分析,距离实际条件下(空间尺度、时间尺度)微观组织演化的真实定量预测还存在较大距离。多年来,研究者们一直在努力试图通过模型改进、算法应用及计算硬件性能提高等提高相场模型的求解速度,但效果仍十分有限。目前,大尺度相场模拟均采用 CPU 并行计算,其对计算机的要求和计算效率本身均极大地限制了相场模拟的应用和发展。近年来 GPU 计算技术的快速发展,使得大规模科学计算有了新的发展动力,这也为个人电脑上实现大尺度相场模拟计算提供了可能。但目前针对 GPU 加速相场模拟计算的研究并不多见。Yamanaka 等对枝晶生长的相场模拟进行了 GPU 加速实验^[6,7],结果证实 GPU 加速计算可极大地提高计算速度,这表明 GPU 加速在相场模拟计算方面有着重要的应用前景。

本文将定向凝固界面形态演化为例,对 GPU 加速计算大尺度相场模拟做深入研究。基于图形处理器计算技术及相场法数值模拟计算的特点,提出一种 GPU 加速相场模拟计算的策略,并对计算网格及相场模拟时间对计算加速比的影响进行分析,最终给出 GPU 加速在定向凝固胞/枝晶列间距调整过程中的应用。

2 相场模型

定向凝固中温度场采用线性假设 $T = T_M + Gz$, 界面上的 Gibbs-Thomson 关系为:

$$U + [z + (mc_\infty/k)/G]/l_T = -d_0 \mathcal{K} - \beta_0 V_n$$

以固定温度区间为参考系的定向凝固定量相场模型为:

$$\tau_0 (1 - (1-k) \frac{z + (mc_\infty/k)/G}{l_T}) (\partial_t \phi - V_p \partial_z \phi) = \nabla \cdot (W(\theta)^2 \nabla \phi) - \partial_x [W(\theta) W'(\theta) \partial_y \phi] + \partial_y [W(\theta) W'(\theta) \partial_x \phi] + \phi - \phi^3 + f_n - \lambda (1 - \phi^2)^2 (U + \frac{z + (mc_\infty/k)/G}{l_T}) \quad (1)$$

$$\partial_t c - V_p \partial_z c = \nabla \cdot [D_L \frac{1 - \phi}{1 + k - (1-k)\phi} \nabla c + (D_L \frac{1 - \phi}{1 + k - (1-k)\phi} + \frac{W_0 (\partial_t \phi - V_p \partial_z \phi)}{\sqrt{2} |\nabla \phi|}) \frac{c(1-k)}{1 + k - (1-k)\phi} \nabla \phi] \quad (2)$$

其中:

$$W(\theta) = W_0 (1 + \gamma_4 \cos 4(\theta - \theta_0))$$

$$U = (\frac{2kc/c_\infty}{1 + k - (1-k)\phi} - 1) / (1-k)$$

$$f_n = \alpha_n r_g (1 - \phi^2)$$

$$l_T = |m| c_\infty (1-k) / (kG)$$

其中, W_0 为界面厚度参数, τ_0 为界面驰豫时间参数, λ 为耦合常数, γ_4 为界面能各向异性强度系数, θ 为界面法线方向与界面能择优取向间的夹角, θ_0 为界面能择优取向与热流方向间的夹角, r_g 为在 $(-1, +1)$ 之间呈高斯分布的随机变量, α_n 为与界面厚度有关的噪声强度系数, $(1 - \phi^2)$ 保证了噪声只在固/液界面上存在。模型参数与实际物性参数的对应关系为: $d_0 = a_1 W_0 / \lambda$, $\tau_0 = a_2 \lambda W_0^2 / D$, 其中 $d_0 = \Gamma / (mc_\infty (1-k) / k)$ 为溶质毛细长度。

界面厚度是相场模型中一个重要的特征参量,在计算求解时需确定界面厚度 W_0 ,且界面厚度与计算尺度及计算工

作量密切相关,小的界面厚度会导致计算量太大,从而限制模拟尺度;大的界面厚度会人为引起界面的非平衡现象,从而导致模拟的失真及非定量化。通常在保证计算结果收敛的前提下,选取的界面厚度比描述的微观组织特征尺度小一个数量级左右。界面厚度取值的局限性使大尺度的相场模拟需要更多的计算网格。

3 GPU 加速定向凝固相场模拟计算策略

计算统一设备构架 Computer Unified Device Architecture (CUDA) 是 NVIDIA 公司于 2007 年推出的一种全新的处理和管理 GPU 的硬件和软件构架,通过将传统的顶点着色器和像素着色器合二为一,由具备可编程特性的 CUDA 内核 (CUDA Core) 统一处理,更加适用于结构化数据的并行计算。

3.1 多流处理单元

基于 Fermi 构架的 GPU 中最基本的处理单元是 CUDA 核心 (在 Fermi 之前的构架中也称为流处理器 Stream Process)。如图 1 所示,每 32 个 CUDA 核心加上寄存器、共享内存等组成一个多流处理单元 Streaming Multiprocessors (SMP)。

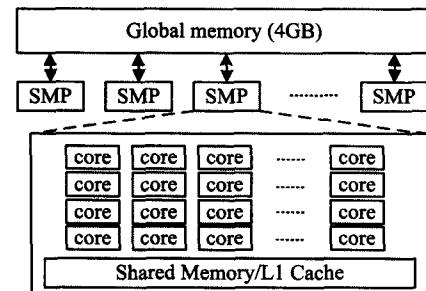


图 1 Fermi 构架中多流处理单元结构

以本文实验采用的 Quadro 6000 显示芯片为例,448 个 CUDA 核心组成了 14 个多流处理单元,每个多流处理单元内部集成 64kB 共享存储器 (On-Chip DRAM),同时片外集成了 6GB 全局显存 (Off-Chip DRAM)。所有的 CUDA 核心都可以访问全局显存,但只有同一个多流处理单元中的 CUDA 核心可以通过共享存储器同步数据。相对于全局显存约数百个时钟周期的延迟,对共享存储器的访问是 GPU 原子操作指令,延迟几乎只有全局显存的 1/100,存取速度与寄存器相当^[8]。

3.2 线程结构设计

在 CUDA 的架构下,并行程序在结构上分为两部分:运行于 CPU 上的 host 程序和运行于 GPU 上的 kernel 程序。Host 程序负责将规格化的待处理数据载入显存中,由 GPU 执行 kernel 程序进行密集计算并返回处理结果。

如图 2 所示, kernel 程序在 GPU 中是以线程 (Thread) 为最小的运行单元,由 core 进行处理;多个线程组成一个线程块 (Thread Block),通过共享内存进行同步,线程块所包含的线程数因 GPU 的硬件构架而异,在硬件上由多流处理单元 (SMPs) 处理,执行同一个 kernel 程序的线程块可以合成一个线程块栅格 (Grid of Thread Blocks)。因此, GPU 加速实际上是通过将单一 kernel 程序由多流处理单元执行,从而实现大量线程并行运行实现的。

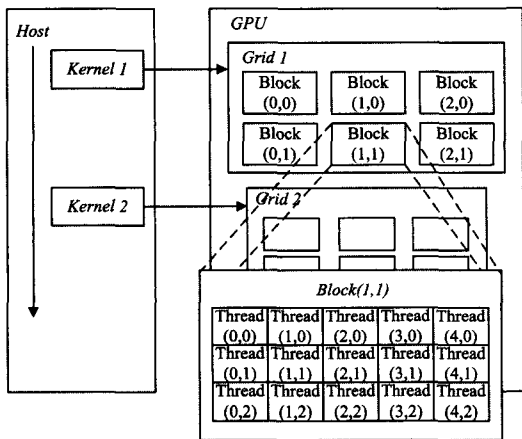


图2 CUDA编程模型

图3所示为GPU加速定向凝固相场模拟计算的流程。host程序根据问题规模为相场矩阵 ϕ 和溶质场 c 分配内存空间,完成边界条件的设定、相关矩阵在内存和显存之间的拷贝并启动kernel程序。对host程序而言,当启动kernel程序时,线程块栅格中的各个线程块会被分配给GPU中的各个多流处理单元,以并行的方式执行。其中kernel1和kernel2函数分别对应式(1)求解晶体凝固过程的相场 ϕ 和式(2)求解溶质场 c ,并根据演化次数控制迭代深度。

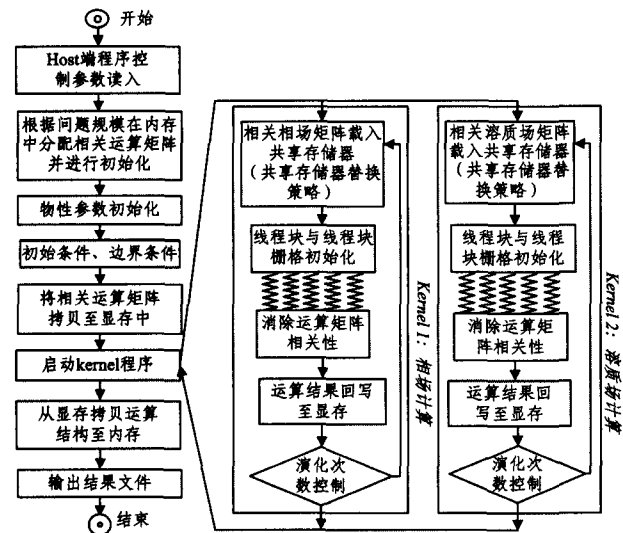


图3 GPU加速定向凝固相场模拟计算流程

3.3 共享存储器访问优化

每个计算网格单元的相场和溶质场与其相邻的4个单元相关,因此kernel程序的每个线程块在求解过程中至少应该从全局显存向每个多流处理单元中共享存储器读入 $(Num_{Block,x}+2) \times (Num_{Block,y}+2) \times sizeof(double) \times (n+1)$ 个字节,其中 n 为中间矩阵变量的个数。

Fermi 构架中每个多流处理单元中的64kB片上DRAM可以按照两种方式工作:48kB共享存储器+16kB L1Cache或者16kB共享存储器+48kB L1Cache。然而,随着问题规模的增大以及线程块大小的增加,共享存储器有可能无法一次从显存中载入当前线程块所需的所有数据。因此,本文将共享存储器划分为 M 块,按照各多流处理单元中线程块 y 方向分组计算,即线程块中每 $Num_{Block,x} \times M$ 个线程为一组,在CUDA内核上并发执行。如图4所示,假设每个多流处理单

元中的共享缓存被分割为3块($M=3$),分别称为SM0、SM1和SM2。当线程块中第1列线程组在求解相场 ϕ 时,与之相邻的第0列和第2列相场数据也相应地从全局显存中被载入共享存储器中参与计算;同理,当第2列线程组在求解时,应载入第1列和第3列相场数据,此时从全局显存中读入第3列相场数据,覆盖共享存储器SM0。依次类推,经过 $(Num_{Block,y}-1)/M$ 次显存访问操作,该线程块中对应的所有相场数据都会被依次载入共享存储器参与运算。

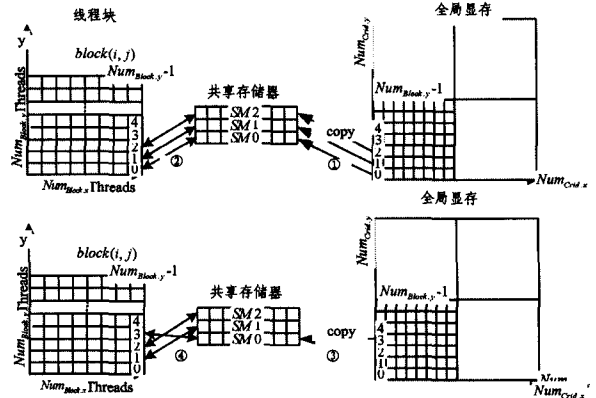


图4 共享存储器访问流程示意图

4 实验结果及分析

4.1 实验配置

由于实验中通常采用类金属透明系合金来研究凝固过程,为方便与实验结果及理论分析结果进行对比,本文主要对透明有机物的定向凝固过程进行分析,选取的二元单相合金为丁二腈-香豆素(SCN-C152)^[9]。这种合金已在凝固微观组织演化原位观察实验中得到了较深入的研究,合金物性参数如表1所列。采用的控制参数为 $V_p=50\mu\text{m/s}$, $G=20.2\text{K/cm}$,界面能各向异性参数为 $\gamma_4=0.005$,噪声强度系数 $a_n=0.0001$ 。

表1 丁二腈-香豆素(SCN-C152)^[9]合金的物性参数

Parameters	Name	SCN-C152
$T_m(\text{K})$	熔点 Melting point of solvent	331.24
$C_{\infty}(\text{mol}\%)$	初始浓度 Initial concentrate	0.0135 mol%
K	分配系数 Partition coefficient	0.05
$m(\text{K/mol})$	液相线斜率 Slope of liquidus	-542
$D_L(\text{m}^2/\text{s})$	液相扩散系数 Diffusion coefficient in liquid	0.45×10^{-9}
$\Gamma(\text{K/m})$	G-T系数 Gibbs-Thomson coefficient	6.48×10^{-8}

实验采用的硬件配置如表2所列。为了比较CPU和GPU在相场模拟过程中的运算差异,求解过程同时设计了基于C语言的串行程序和基于CUDA的并行程序。当迭代10万次时,两者的数值计算误差可以控制在 10^{-12} 以内。因此,串行程序和并行程序在精度上是保持一致的。

表2 系统硬件配置

Name	Configuration
CPU	Intel Xeon X5690(3.46GHz,6Core)×1
内存	24GB(DDR3)
GPU	NVIDIA Quadro 6000×1
显存	6GB(DDR3)
操作系统	Windows 7 Profession Edition CUDA5.0

(下转第56页)

[19] Chen Wei-zhu, Yan Jun, et al. Document transformation for multi-label feature selection in text categorization[C]//Seventh IEEE International Conference on IEEE, 2007;451-456

[20] Zhang Min-ling. ML-RBF: RBF neural networks for multi-label learning[J]. Neural Processing Letters, 2009, 29(2): 61-74

[21] Zhang Min-ling, Zhou Zhi-hua. Multi-label learning by instance differentiation[C]//AAAI, 2007, 7: 669-674

[22] Zhang Min-ling, Zhou Zhi-hua. ML-kNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038-2048

(上接第 21 页)

4.2 实验分析

在对定向凝固过程的模拟中,模拟对象的复杂程度决定了总的计算量。对于不同的计算量, GPU 的加速效果会有所差异。为了比较 GPU 相对于 CPU 在定向凝固相场模拟中的优势,本文分别从演化次数和问题规模¹⁾方面具体说明,其中 t_{CPU}/t_{GPU} 称为 GPU 相对于 CPU 的加速比,简称加速比。

首先,考察演化次数对加速比的影响。如图 5 所示,当问题规模固定为 $N_x \times N_y = (250 \times 1500)$ 时,随着演化次数的增加,加速比逐渐增加。当演化次数从 100 步逐步增加到 100 万步时,加速比逐步稳定在 36.5 左右;当演化次数大于 20 万步后,演化次数对加速比的影响逐步减弱。

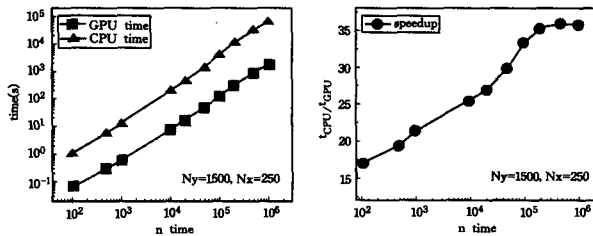


图 5 加速比随演化次数的变化:当问题规模固定时,随着演化次数的增加,加速比逐步稳定在 36.5 左右

其次,考察问题规模对加速比的影响。如图 6 所示,当演化次数固定为 10 万时,随着问题规模从 (250×1500) 逐步增加到 (5000×1500) ,加速比逐步稳定到 36.5 左右;当问题规模大于 (2000×1500) 时,加速比基本保持不变。

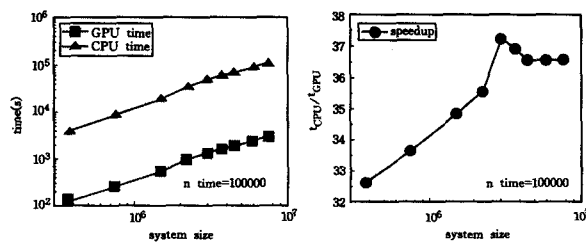


图 6 加速比随问题规模的变化:当演化次数固定时,随着问题规模的增加,加速比逐步稳定在 36.5 左右

通过两组对比实验,可以得出:一方面,随着演化次数和问题规模的增加, GPU 并行计算显著地加快了凝固相场模拟的速度;另一方面,可以明显看出,加速比最终收敛于 36.5 左右。由此可见,只有在大规模计算的时候, GPU 加速的能力才能最大限度地发挥出来。

产生这种现象可能是因为受限于实验的硬件平台。在计算规模较小的时候, GPU 的计算能力部分处于空闲状态,不能完全发挥其优势;随着计算规模的增加, GPU 的能力被充分使用,与 CPU 的串行计算相比,其优势越来越明显;但是,当计算规模达到一定程度时,无论是 GPU 并发执行 kernel 线程块栅格还是 CPU 单线程串行求解, GPU 的硬件资源(主

要是 CUDA 核心和共享存储器空间)和 CPU 的硬件资源(主要是 CPU 计算能力)都消耗完全,即硬件的计算能力达到饱和,此时加速比收敛于一个反映了 GPU 和 CPU 固有处理速度的比值。

结束语 本文针对液/固相场模拟过程中,由于大计算尺度和长演化时间导致的模拟计算量巨大的问题,采用 CUDA 计算模型,对相场模拟进行并行化改造,使其适应于 GPU 上的高效并行处理。实验结果表明,在相同的计算尺度和演化时间条件下, GPU 计算与 CPU 计算的加速比可以高达 36 倍。

虽然目前国内类似方法尚处于研究阶段,但本文的研究结果初步表明, GPU 强大的浮点计算能力以及高效的并行处理机制,使得相场理论研究向着大尺度仿真、实时/近实时模拟成为可能;并且,通过采用多核 GPU 的并行协同工作,加速比的理论值还可以进一步提高。

参考文献

[1] Boettinger W J, Warren J A. Simulation of the Cell to Plane Front Transition during Directional Solidification at High Velocity[J]. J. Cryst. Growth, 1999, 200(3/4): 583-591

[2] Wang J C, Zhang Y X, Yang Y J, et al. Phase field modeling for dendritic morphology transition and micro-segregation in multi-component alloys[J]. Science in China Series E: Technological Sciences, 2009, 52(2): 344-351

[3] Lan C W, Shih C J, Lee M H. Quantitative Phase Field Simulation of Deep Cells in Directional Solidification of an Alloy [J]. Acta Mater, 2005, 53(8): 2285-2294

[4] Gurevich S, Karma A, Plapp M, et al. Trivedi. Phase-Field Study of Three-Dimensional Steady-State Growth Shapes in Directional Solidification [J]. Phys. Rev. E, 2010, 81(1)

[5] Wang Y B, Wang Y X, Cheng Z H, et al. Phase-field Simulation of Interface Effect during Grain Nucleation of Solidification Processing [J]. Rare Metal Materials and Engineering, 2012, 41(6): 1045-1048

[6] Yamanaka A, Aoko T, Ogawa S, et al. GPU-accelerated phase-field simulation of dendritic solidification in a binary alloy [J]. Journal of Crystal Growth, 2010, 318(1): 40-45

[7] Takaki T, Yamanaka A, Nukada A, et al. Peta-scale phase-field simulation for dendritic solidification on the TSUBAME 2.0 supercomputer [C]// Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis(SC '11), 2011

[8] Glaskowsky P N. NVIDIA's Fermi: The First Complete GPU Computing Architecture[M]// NVIDIA Corporation White Paper. 2009

[9] Losert W, Shi B Q, Cummins H Z. Evolution of Dendritic Patterns during Alloy Solidification: Onset of the Initial Instability [J]. Proc. Nat. Acad. Sci. USA, 1998, 95(2): 431-438

¹⁾ 为了实现合金定向凝固枝晶列生长过程的模拟加速,本文模拟了 $N_x \times N_y$ 大小的二维剖面, $N_x \times N_y$ 称为问题规模。