

面向异构计算的高性能计算算法与软件*

徐 顺^{1,2}, 王 武^{1,2}, 张 鉴^{1,2}, 姜金荣^{1,2}, 金 钟^{1,2,3}, 迟学斌^{1,2}



¹(中国科学院 计算机网络信息中心, 北京 100190)

²(中国科学院 计算科学应用研究中心, 北京 100190)

³(广东省生物医药计算重点实验室(中国科学院 广州生物医药与健康研究院), 广东 广州 510530)

通讯作者: 金钟, E-mail: zjin@sccas.cn

摘 要: 研发适应国产异构计算环境的高性能计算算法与软件是非常重要的课题, 对我国高性能计算软件研发匹配高性能计算硬件高水平发展的速度具有重要意义. 首先, 简要介绍高性能计算应用软件的现状、趋势和面临挑战, 并对几类典型高性能计算应用软件开展并行计算算法特征分析, 涵盖了宇宙 N 体模拟、地球系统模式、计算材料相场动力学、分子动力学、量子计算化学和格点量子色力学等多个问题、尺度和领域. 其次, 讨论了面向国产异构计算系统的对策, 提炼出若干典型应用算法和软件的共性问题, 涉及核心算法、算法发展、优化策略等. 最后, 面向异构计算体系结构, 对高性能计算算法与软件进行了总结.

关键词: 高性能计算; 异构计算; 并行算法; 科学计算软件

中图法分类号: TP301

中文引用格式: 徐顺, 王武, 张鉴, 姜金荣, 金钟, 迟学斌. 面向异构计算的高性能计算算法与软件. 软件学报, 2021, 32(8): 2365–2376. <http://www.jos.org.cn/1000-9825/6008.htm>

英文引用格式: Xu S, Wang W, Zhang J, Jiang JR, Jin Z, Chi XB. High performance computing algorithm and software for heterogeneous computing. Ruan Jian Xue Bao/Journal of Software, 2021, 32(8): 2365–2376 (in Chinese). <http://www.jos.org.cn/1000-9825/6008.htm>

High Performance Computing Algorithm and Software for Heterogeneous Computing

XU Shun^{1,2}, WANG Wu^{1,2}, ZHANG Jian^{1,2}, JIANG Jin-Rong^{1,2}, JIN Zhong^{1,2,3}, CHI Xue-Bin^{1,2}

¹(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

²(Center of Scientific Computing Applications and Research, Chinese Academy of Sciences, Beijing 100190, China)

³(Guangdong Provincial Key Laboratory of Biocomputing (Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences), Guangzhou 510530, China)

Abstract: It is very important to develop high performance computing algorithm and software adapting China heterogeneous supercomputer. This has also great significance in minimizing the gap between China's HPC hardware and HPC software. Firstly, the article briefly introduces current trend and challenges of high-performance computing application software and analyzes computing algorithm characteristics of various typical high performance computing applications including *N*-body simulation in computing

* 基金项目: 国家重点研发计划(2016YFB0201701); 国家自然科学基金(41931183, 11871454); 中国科学院战略性先导科技专项(C类)(XDC01030200); 中国科学院“十三五”信息化专项(XXH13506-404, XXH13506-405); 广东省生物医药计算重点实验室资助项目(2016B030301007)

Foundation item: National Key Research and Development Program of China (2016YFB0201701); National Natural Science Foundation of China (41931183, 11871454); Strategic Priority Research Program of the Chinese Academy of Sciences (Category C) (XDC01030200); Special Fund of the Chinese Academy of Sciences for Informatization Construction in 13th “Five-Year Plan” (XXH13506-404, XXH13506-405); Guangdong Provincial Key Laboratory of Biocomputing (2016B030301007)

本文由“国产复杂异构高性能数值软件的研制与测试”专题特约编辑孙家昶研究员、李会元研究员推荐.

收稿时间: 2019-09-01; 修改时间: 2019-12-05; 定稿时间: 2020-01-22

cosmology, earth system model, computational material phase field dynamics, molecular dynamics, quantum chemistry, and lattice QCD. Secondly, solution to use domestic heterogeneous computing system has been discussed and typical application algorithms, common questions of software including core algorithm, development of algorithm, strategies of optimizing codes have been summarized as well. Finally, a summary of high-performance computing algorithms and software for heterogeneous computing is given.

Key words: high performance computing; heterogeneous computing; parallel algorithm; scientific computing software

高性能计算应用对国家工业、国家安全、生命健康和科学发现均具有重要的作用.作为信息科技政策制定咨询权威组织,美国信息技术与创新基金会(Information Technology and Innovation Foundation,简称 ITIF)早在 2016 年就作了题为《高性能计算对美国竞争力至关重要》^[1]的专题分析报告.近年来,我国也探索出了“天河”、“神威”和“曙光”三大国产高性能计算机的发展之路,均具备自主研发高性能处理器和加速器的能力;同时也计划在 2020~2021 年前后设计实现 E 级(每秒百亿亿次浮点运算)计算机,在全球高性能计算性能 Top500 排行榜上^[2],可看到国产超级计算机性能处于全球前列水平.但不难看出,我国超算系统软件和应用软件的研发步调相对落后,国产超算上的软件生态没有同步形成^[3].一个客观的困难是:国际上主流基于 Intel x86 或 Nvidia GPU 的科学计算软件无法直接在部分国产超算上编译运行,特别是面向异构计算加速这方面,需要有大量的移植适配工作.对于神威超算这种特殊的主从核架构,大部分软件需要修改甚至重写以实现从核加速计算.

尽管高性能计算可应用于工业、国防、生命健康等诸多领域,但这些应用主要以科学计算为基础.随着计算机软硬件技术的高速发展,特别是对称处理器核数、内存容量、网络带宽的提升引发(并行)计算算法需要考虑新的平衡因素,促使在科学计算与数值模拟应用方面提出了大量新的软件架构和理论算法.比如在高能物理领域,形成了美国 USQCD^[4]和 MILC^[5]、英国 Grid^[6]和日本 Bridge++软件等诸多代表性的格点量子色动力学(quantum chromodynamics,简称 QCD)模拟软件,这些软件形成了自身特色的软件架构,或者是面向特定体系结构而优化设计的.特别是 USQCD 的 Chroma 软件构建了一套多层次、松耦合的软件架构体系,取得了有效整合软件模块和增强硬件适应性的显著效果;他们也开发了基于 GPU 加速计算的软件包 QUDA^[7],充分利用了 GPU 异构计算的优势.在分子模拟应用领域,LAMMPS^[8]和 NAMD^[9]软件为代表的的应用软件在高性能计算方面具有显著的性能优势,但它们的软件架构完成不同,而两者都具有代表性.LAMMPS 是典型的 MPI 分布式并行程序,具有非常强的软件功能和计算性能的可扩展性;NAMD 更是采用 Charm++并行计算框架支持新的并行编程模式,数据通信采用远程对象的函数调用方式实现.此外,兼备量子化学研究的 NWChem^[10]软件引入了 Global Arrays 并行计算框架来支撑计算化学类的特殊问题,更多的是考虑高可伸缩的计算化学算法使用传统的消息传递通信模型很繁琐,而使用可移植共享内存编程模型更简便^[11].以上这些高性能计算典型应用软件不仅着眼于科学问题的研究,而且通过其成功的应用大大推动了高性能计算应用技术的发展,在高性能异构计算方面取得了诸多理论和应用成果.比如格点 QCD 应用和 NAMD 软件均获得过高性能计算应用领域重要的“戈登·贝尔”奖(Gordon Bell Prize,简称 GB 奖).美国的 E 级计算计划(Exascale Computing Project,简称 ECP)^[12]中均毫无悬念地将这些软件作为重点发展的对象.

相形之下,我国在相关的科学计算应用领域的自研程序代码普遍依托于相应领域的课题组开发和继承、大多是以科学问题研究为主,缺乏专业的开发人员来维护软件.更是缺乏前瞻性的软件顶层设计,也难以形成规模化的可推广使用的软件成果,软件的并行计算性能更是难以保证.近年来,GPU 异构计算和国产自主处理器的兴起使得软件框架和软件模块不断推陈出新,导致软件程序无法跟上硬件架构和软件应用环境快速发展的脚步,并行计算设计优化加大了困难.庆幸的是,莫则尧团队针对以上问题探索出了一条创新的技术路径.其基本方法是凝练数值模拟应用领域的共性需求来建模,并设计模型驱动的并行算法和采用了并行编程技术^[13].他们提出了面向结构网格和非结构网格的标准化网格数据模型,分别开发了结构网格的编程框架 JASMIN^[14]和非结构网格的编程框架 JAUMIN^[15],均一定程度支持并行算法和并行编程的软件复用.虽然我国科学计算软件的开发积累了不少经验,但从串行程序到大规模并行程序、从算法程序到软件产品还有相当长的道路要走.从《国家高性能计算环境发展报告》^[16]可得,在我国各大超级计算中心的超级计算系统上所运行应用软件大部分来自国外,一旦出现“断供”,我国计算科学研究即显现出被“卡脖子”的被动局面.

因此,为了发展国产超级计算系统的应用生态并适应未来 E 级计算的需求,融合国际主流软件的优点并实现支撑科学计算应用研究的最终目的,非常有必要梳理适配国产高性能计算系统的科学计算应用软件的发展情况.本文将从高性能计算应用算法和软件的角度,阐述一些典型应用软件核心算法的并行计算特征,并讨论面对国产异构系统的算法对策及其发展思路.

1 典型高性能计算应用的核心算法与软件特征

高性能计算应用软件种类繁多,不可能一一罗列.现以美国国家能源研究科学计算中心(National Energy Research Scientific Computing Center,简称NERSC)的计算资源使用情况作为代表来分类.NERSC是美国能源部科学局的旗舰科学计算单位,在通过计算加速科学发现的领域处于世界领先地位.图1是2018年NERSC超级计算机时使用量按学科分布图(数据来源:<https://www.nersc.gov/assets/Uploads/2018NERSCAnnualReport.pdf>),可见其超算应用主要集中在地球与环境、化学、材料科学、高能物理、天文/宇宙学等学科领域.

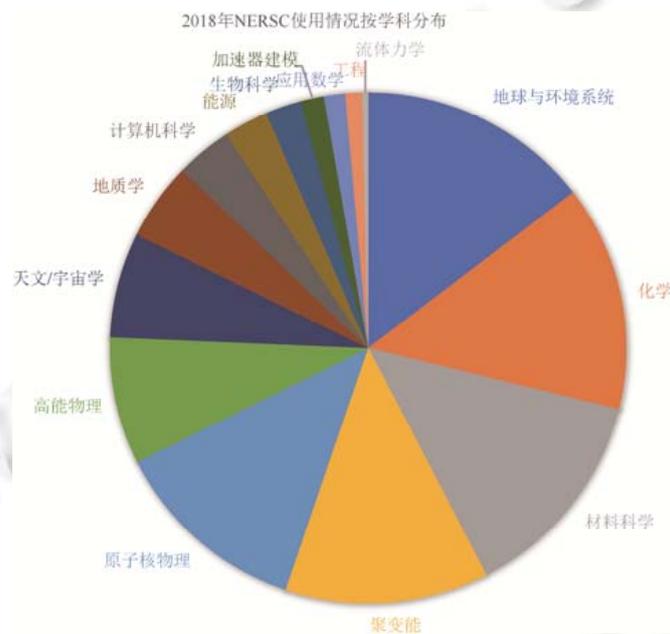


Fig.1 Usage amount in subjects on NERSC supercomputer in 2018

图1 2018年NERSC超级计算机按学科分的使用量

本文也大致选择以上主要计算应用领域来讨论其中的应用软件核心算法和软件特征.这些领域的研究对象具有不同的计算问题研究尺度,按尺度量级从大到小排序见表1.“计算模拟”较“实验”和“理论”两大科学研究手段有很多优势:可以对不同尺度问题建模,可以对不同数据维度建模,也可以对确定性问题(如偏微分方程)或非确定性问题(如概率统计问题)建模.以下列举的几大类领域应用将说明相关的建模方式.

Table 1 Scale size of different scientific subjects

表1 不同科学研究对象的尺度大小

研究对象	尺度量级	量级单位
宇宙	光年	10^{15} m
地球	千米	10^3 m
材料组织	微米	10^{-6} m
分子原子	纳米	10^{-9} m
原子核	飞米	$<10^{-15}$ m
基本粒子	阿米	$<10^{-18}$ m

1.1 宇宙N体模拟

N 体问题又称“多体问题”,兼具数据密集和计算密集的特点,是物理科学高端模拟中 7 个主要问题之一^[17]. 此类问题在天体物理学、分子动力学等领域处于核心的地位.天文学的一些关键现象(如星系成团及演化、暗物质晕等)需采用大尺度、高解析度的大规模 N 体模拟结果来解释和指导天文观测;宇宙模型和理论也通过计算结果得以验证和完善.随着先进快速算法和高性能计算技术发展,万亿量级规模 N 体问题的高效大规模并行计算模拟已成为可能.

基于树遍历的算法如快速多极子方法(fast multipole method,简称 FMM),不仅具有近似线性的复杂度和较高的精度,而且适合并行.其中短程力的求解具有计算密集的特征,可有效利用高性能异构平台加速器件(如 GPU、MIC 或众核处理器等)的超强计算能力.基于快速傅里叶变换(fast Fourier transform,简称 FFT)的粒子网格方法(particle-mesh method,简称 PM)虽速度更快,但由于受限于精度,只适用于计算长程力,而且将带来全局通信.为兼顾计算速度和精度,重叠短程力的计算与长程力的通信,混合算法成为高性能并行 N 体模拟的主流^[18,19].如图 2 所示,虚线内为 FMM 计算区域,其中粒子间的短程力相互作用(P2P)占 FMM 计算时间的 90%以上,适合在加速卡上实现;虚线外的区域采用 PM 计算长程力,可在 CPU 端调用 FFT 实现.

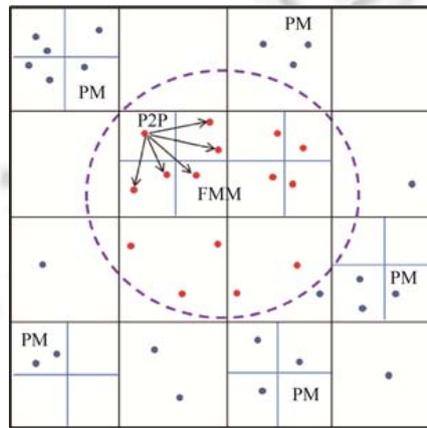


Fig.2 A schematic diagram of FMM-PM algorithm

图 2 FMM-PM 算法示意图

中国科学院国家天文台和计算机网络信息中心团队合作,在国产高性能异构计算平台上采用 FMM 与 PM 耦合算法实现了大规模的 N 体问题高效并行模拟.计算的并行效率达 80%,浮点峰值效率达 30%以上,为实现更高分辨率的超大规模天文应用模拟奠定了基础.

1.2 大气与地球系统模式

大气与地球系统模式中,Stencil 计算是核心算法,也是物理科学高端模拟中 7 个主要问题之一^[17]并广泛存在与科学与工程计算中.Stencil 计算(如图 3 所示)绝大部分是访存密集型计算,计算访存比低,主存带宽受限严重^[20].已经有很多 Stencil 计算优化方面的工作:如 Strumpfen 等人^[21]利用时间局部性优化三级存储系统;Mickevicus 等提出了 GPU 平台上的高阶 Stencil 优化方法.此外,也有一些基于 Stencil 计算的应用在国产机器上的移植优化工作,如 Fu 等人^[22]将大气模式 CAM 和大型非线性地震模拟系统移植到“神威·太湖之光”超级计算机上.

与大气模式 CAM 类似,中国科学院地球系统模式中的有限差分 Stencil 计算的运行时间占比超过 90%,但是其浮点计算效率仅有不到 5%.而且这些 Stencil 计算包含更多更复杂的数据结构,Stencil 的种类和差分格式、算法求解过程和特点也各不相同.其主要特点有:通信和同步多、复杂计算多、变量多而计算网格点少、计算访存比和主存带宽受限问题更突出.这对当前的国产并行计算平台的计算和访存能力提出了新需求.而在高性

能计算机普遍采用异构架构的当下,性能调优显得尤其重要.代码的移植优化工作是一项具有挑战性的工作.主要的优化策略有:通过通信优化手段减少通信占比,提高程序的可扩展性;通过计算子区域规模划分,调节计算资源和高速寄存器及缓存使用率,达到较好的均衡;通过改进算法或算法实现,减少数据依赖和分支计算,提高数据访问连续性和 cache 命中率;提高程序的向量化率等.

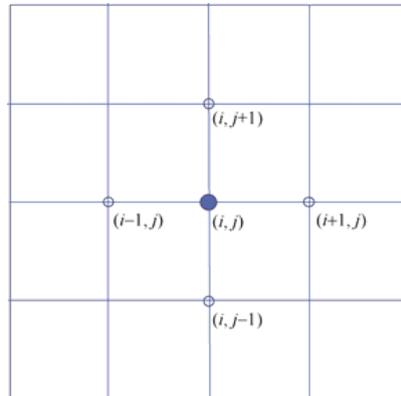


Fig.3 Discrete structure in Stencil computation

图 3 Stencil 计算的离散结构

1.3 相场模拟微结构演化

合金材料的微观组织对材料的诸多性能起到决定性的影响,因此深入研究其微观组织的演化机理具有重要的意义.相场方法是模拟合金材料微观组织演化的主要手段,主要通过求解关于相场、成分场等变量的一组偏微分方程来进行微观组织演化过程的模拟.方程具有强刚性和强非线性的特点,大体系长时间的模拟是非常具有挑战性的课题.传统的有限差分显式算法虽可在当代高性能计算平台上有较高效的实现^[23],但由于时间步长受到算法稳定性条件的限制而无法满足长时间模拟的需求.而隐格式的设计则面临着相场模型方程强非线性以及相应的大规模线性和非线性代数方程组并行求解等问题.中国科学院计算机网络信息中心和金属研究所团队合作,针对国产异构高性能计算平台设计了可扩展指数时间差分算法,可高效地进行相场模型的大体系长时间模拟.2016年在“神威·太湖之光”超级计算机上使用全机约 1 000 万计算核心实现了持续性能为 50 PFLOPS (每秒 5 亿亿次浮点运算)的超大规模相场模拟^[24].基于“神威”超级计算机的“申威”处理器的主从核架构,数据通信使用主核 DMA 方式,而七点 Stencil 计算则在从核上计算.最终应用测试(如图 4 所示)的并行效率达到 70%,浮点性能达到峰值的 41%,远超同类的计算软件.

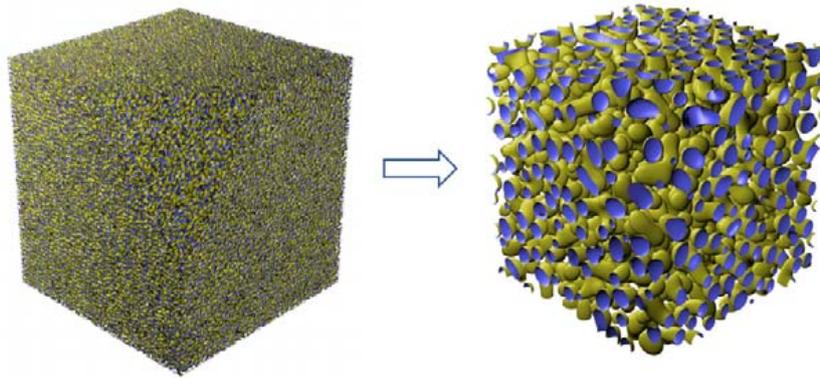


Fig.4 Configuration coarsening evolution of two-phase system governed by the Cahn-Hilliard equation

图 4 两相体系由 Cahn-Hilliard 方程控制在粗化过程中的构型演化

1.4 分子动力学模拟

分子动力学模拟(molecular dynamics,简称 MD)是在原子水平上利用经典牛顿力学方程研究分子状态的计算机模拟方法^[25].随着计算机计算能力的提高,可实现全原子模型的动力学模拟,在原子结构层次揭示某些现象随时间的发展过程,已广泛应用于计算化学、生物化学、纳米材料和药物分子设计等领域.

分子动力学模拟的计算核心在于力场的计算,具体可细分为长程作用力计算、短程力计算与近邻列表的计算、成键作用力计算等多种形式;除了核心的力场计算,还存在着一些物理系统约束需要实现,比如分子结构约束方法 SHAKE 和 LINCS、周期性边界条件和统计系综理论约束等.加速分子动力学模拟,在方法上可采用增强抽样算法,在实现上则可借助大规模并行计算.在并行计算时任务划分常采用体系空间分解法,可采用 MPI+X 的多级并行计算设计,其中 X 可为多核 CPU 计算或 GPU 异构计算等形式.比如美国能源部主导开发的 MD 软件 LAMMPS(large-scale atomic/molecular massively parallel simulator),可利用多核 CPU、GPU 和 Intel Xeon Phi 等多种方式加速其计算性能.该软件提供了 GPU package 和 KOKKOS package 可利用 GPU 部件加速计算,使用 MPI 计算任务下的 GPU 任务 offload 工作方式^[26].这种方式将计算密集型任务交给 GPU 来计算完成,并获得返回结果,充分利用了异构计算环境来提高并行效率.结合区域空间分解和 MPI 并行技术,可进行数万核以上的超大规模并行计算.GPU 作为异构计算设备用来加速分子动力学模拟的一种实现流程如图 5 所示,不同软件在实现上可能存在差异,但基本用于加速分子作用力计算以及相关的近邻列表构建等计算密集型任务.

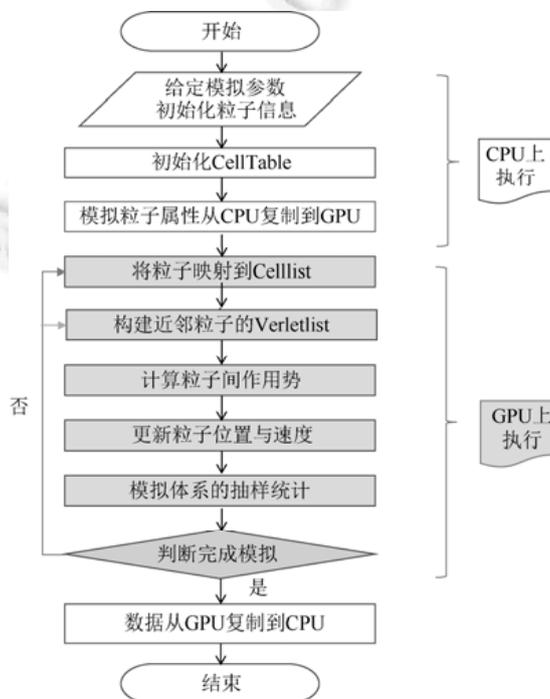


Fig.5 A flowchart for an implementation of GPU accelerated molecular dynamics simulation^[27]

图 5 一种实现 GPU 加速分子动力学模拟的流程图^[27]

1.5 量子化学

量子化学源自理论化学,应用量子力学的基本原理和方法来研究化学问题.经过多年的发展,已经有多种量子化学计算软件被开发出来.科学家们用这些工具计算和求解分子结构问题,不断地扩大其可解问题的深度和广度.

自洽场(self-consistent field,简称 SCF)迭代计算是量子化学计算软件的核心方法.它主要包括计算双电子

积分、构造 Fock 矩阵和求解广义特征值方程,兼具计算密集型和访存密集型的特点.计算双电子积分一般采用解析的方法,是计算密集型的,计算量与体系大小的 4 次方成正比.在引入了 prescreening 方法后,可以有效降低积分计算的 scaling,同时利用分子划片的方法,可以得到近似线性的 scaling 结果.构造 Fock 矩阵是访存密集型的,有大量的随机不连续访存.如何协调处理计算积分和构造矩阵是异构优化自洽场计算的难点.2008 年,Yasuda^[28]首先提出利用 GPU 计算双电子积分.同年利用积分公式自动展开的方法将积分的计算转化为适合 GPU 运行的方式,取得了较好的结果.后来,Ufimtsev 等人^[29,30]提出了预先排序的方法,实现了在 GPU 上构造 Fock 矩阵(如图 6 所示).尽管算法相对于传统的 CPU 算法在计算量上有一定的损失,但仍然取得了近百倍的加速效果(相对于 1 个 CPU 核).

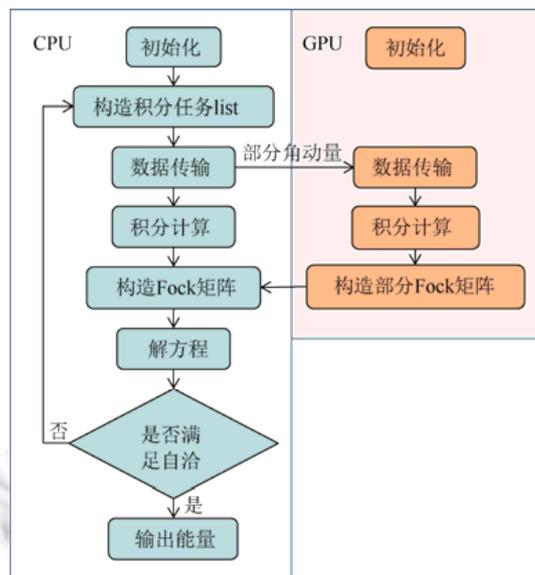


Fig.6 A flowchart of GPU accelerated SCF calculation

图 6 GPU 加速 SCF 计算流程图

后自洽场(post-SCF)计算在提高量子化学计算结果精度和研究前沿科学问题上有着非常重要的作用.作为多参考方法起点的多组态(multi-configurational)自洽场的计算热点非常类似 Hartree-Fock 的计算,区别之处是在多组态自洽场使用的积分常为分子轨道下的积分,自洽迭代过程中积分的多个指标需要进行持续的变换操作.此部分计算同样兼具访存密集型和计算密集型的特点.基于多行列式的微扰计算则是成功地实现了 GPU 加速,极具代表性的有 NWchem 软件中的 CCSD(T)方法.该法中的 CCSD 计算复杂,由 CPU 处理;后续基于 CCSD 波函数的微扰计算(即 T 部分)则交由计算能力较强的 GPU 来实现,需针对偶合簇波函数中的每一个行列式表示的系数进行微扰校正,计算热点为密度矩阵同电子积分的乘积运算.

1.6 格点量子色动力学

量子色动力学(即 QCD)是一种强相互作用的规范理论,它和电弱统一理论(EW)通称为标准模型,是高能物理研究的基本理论体系.QCD 描述了组成强作用粒子(强子)的夸克和与色量子数相联系的规范场的相互作用,可统一地描述强子的结构和它们之间的强相互作用,被认为是有希望的强作用基本理论.格点量子色动力学(格点 QCD)是从第一性原理出发的研究 QCD 的非微扰方法^[31],对于高能物理中的重大物理问题,如夸克禁闭、宇宙早期演化、标准模型精确检验及超出标准模型新物理需寻找等研究有不可替代的作用.格点 QCD 方法的计算模拟过程对高性能计算资源需求巨大,其计算过程体现出计算密集型特征,而其数据分析则体现出数据密集型特征.计算精度和数据分析效率直接反映了格点 QCD 计算方法的应用水平.大规模计算应用中通常采用空间

区域分解的并行计算方式(如图7所示),格点区域上的计算任务分配到不同的处理器上计算,各处理器核负责格点边界上的数据通信,总体上使得数值模拟计算具有高可并行性和高可扩展性等特点,是高性能科学计算的典型应用领域^[32],曾获得1988、1998、2006年的Gordon Bell奖,2018年也入围该奖项的决赛^[33].

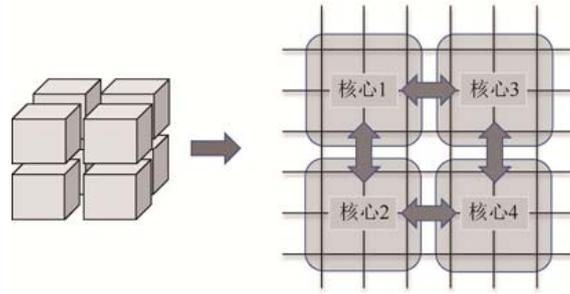


Fig.7 A schematic diagram of domain decomposition in parallel computing lattice QCD

图7 格点 QCD 区域划分并行计算示意图

在格点 QCD 软件设计上,抽象分层软件架构的设计方法得到了重要应用,比如美国的 USQCD 软件框架分为 4 层:最上层应用层包括 Chroma,之下 3 层为支撑库层,分别进行消息通信、数据结构定义和基本算法实现.再如英国的 Grid 软件^[6]也是抽象了数据并行计算模型,针对格点 QCD 的热点计算抽象了统一的数据并行模型,支持多种平台的 SIMD 矢量化计算,如 AVX2 和 AVX512 等形式.这种数据并行模型可以有效支撑上层应用的开发,保证较高的高性能计算效率.软件框架的这种分层抽象的设计方法极大地提高了代码的开发效率和运行效率,大大地促进了整个格点 QCD 研究领域的高速发展.针对异构计算环境,美国为主的团队开发了针对 GPU 加速的格点 QCD 软件包 QUDA,软件包主要采用 C++ 模板类支持不同数据类型的加速计算,已在美国 Summit 超级计算机上实现大规模 GPU 加速的格点 QCD 计算,取得了显著的加速效果.基于 CUDA 设计的 QUDA 软件如要移植到曙光国产超算平台,需要经过转码为 ROCm 支持的 HIP 代码形式.

2 共性问题的提炼及解决对策

2.1 共性算法

2004 年著名高性能计算专家 Colella^[17]从计算物理问题中提出了科学计算的 7 种算法模型,即结构性网格、非结构性网格、快速傅里叶变换(FFT)、稠密线性代数(dense linear algebra)、稀疏线性代数(sparse linear algebra)、粒子动力学(particles)、蒙特卡洛(Monte Carlo).这 7 种算法模型还可以进一步归类为求解数值微分方程类和求解数值线性代数类,实现形式上有显性计算方式和隐性计算方式.这些核心的共性计算模型对高性能计算机提出了不同方面和层次的需求,也基本覆盖了常见高性能计算应用.如 N 粒子相互作用中等设计 FFT 计算长程相互作用,粒子动力学是一个微分方程的求解,格点 QCD 计算也会涉及大量的线性系统求解,通常为千万到亿级矩阵维度的巨大计算量,同时存在海量 3x3 低维向量矩阵乘法.提炼以上共性算法特征,有利于集中研发力量开发可共享使用的软件算法包;也有利于突破计算关键技术问题,提高高性能计算软件研发的效率.

国产软件编程框架 JASMIN^[14]和 JAUMIN^[15]核心正是从共性算法提炼出发,对大量物理模型计算问题进行抽象和提炼共性特征,形成编程框架的核心.编程框架底层以网格数据模型为基础,提供数值算法构件,中间层实现并行计算模式和负载均衡建模,顶层为数值代数解法器库或者独立于编程框架的第三方解法器库接口.

2.2 浮点计算精度

不论何种科学计算应用,浮点计算精度的问题均扮演着重要的角色.从应用问题的角度来看,数值结果精度往往越高越好,但高精度浮点计算较低精度浮点计算需要更多的时间.一般 64 位的双精度计算所需时间是 32 位的单精度浮点计算时间的两倍,因此常常遇到在高精度和高计算性能之间的取舍问题.高低精度混合计算是

一般的处理方式,但大型科学计算模拟中浮点运算过程非常复杂,很难从理论上推导出合理的精度位,不得不凭经验判断,但这不可靠更不科学.一个相对科学的方式是先评估科学计算应用算法中最少需要多少位数值有效位数,而这个有效位数能满足对结果需求精度.这可以借助软件工具 Verificarlo^[34]来推算出计算算法中保持结果所需的最低有效位,然后选择相匹配的浮点数表示精度.另外,在某些体系结构支持多个低精度浮点拼凑为高精度浮点的运算,但这增加了算法实现的复杂度.在通用处理器中一般只提供浮点精度计算,对应于 C 语言中的 double 和 float 精度;而专用 FPGA 加速模块却可以自定义浮点计算精度;最近的 Nvidia GPU 也支持半精度的运算.科学计算中浮点运算追求的目标之一是所谓的“高准确率低精度(high-accuracy low-precision,简称 HALP)”^[35].

2.3 海量数据存储及可视化

高性能计算应用在大规模模拟中,往往会遇到大量数据存储的问题.比如分子模拟常规 100 万原子体系中,每个原子单精度浮点数据仅存储三维的速度和位移数据,大约需要 24MB 存储空间,如果需要 1 万帧轨迹数据则需要 240GB 文件,如果再需要不同温度和压强的组合条件下的轨迹数据,对存储空间需求剧增.另外,数据格式在异构系统上不兼容,直接数据传输存储会产生问题.当前可以借助专门的 HDF(hierarchical data format)文件格式接口来避免,它可存储不同类型的图像和数码数据的文件格式,并可在不同类型的机器上传输,同时还有统一处理这种文件格式的函数库.

数值模拟数据的可视化特点见表 2,数据源可以按网格特征进行划分,数据的可视化也具有一般的流程.大规模数值模拟数据对其可视化也带来了困难,主要体现在空间表示和空间存储方面.比如粒子数超过了 32 位整数可表示的范围可能导致软件崩溃,渲染时 GPU 显存一次性无法容纳一个数据文件时,导致数据分批导入显存,增加了算法的复杂度也往往降低了计算性能.

Table 2 Visualization features in numerical simulation

表 2 数值模拟数据可视化特点

分类	特点
网格特征	均匀网格、正交网格、曲线网格、无结构网格、离散点等分类
维度场	标量场、矢量场
可视化流程	数据预处理、几何映射、绘制、显示

2.4 软件框架设计

软件框架决定了软件的生命力,其主要特征是软件功能的适应性,具体包括理论建模支撑、计算模块、逻辑流程控制 and 数据分析等模块.随着计算机硬件特性和软件编程语言的发展,许多科学计算软件已不能很好地适应应用需求,大部分问题是软件的原框架受到冲击导致的.如原有程序采用 Fortran 语言编写,后来为了适应新的运行环境改用 C++ 实现大规模设计,又需要新的 C++11 改进编程效率,导致软件框架不稳定.其次早期采用 OpenMP 为主的多 CPU 核计算模型,需要转为异构 OpenCL 计算模式;另外由于新的加速设备,原来的 CUDA 程序可能需要改造为 OpenACC 并行编程方式.以上问题在高性能计算软件的设计中普遍存在,内在原因为软件并行计算模式较多,硬件体系结构差异大.

对于高性能计算软件来说,优秀的软件框架可提高编程开发效率,适应新的软件需求和用户服务推广.如何以不变的设计原则应对多变的软件需求是一个关键问题.对于软件框架来说,分层式设计常用方式如通用并行软件 JASMIN 和格点 QCD 应用软件 Chroma.但从异构计算这方面考虑,面向异构计算环境有机统一计算模拟的各方面是根本性解决问题,如统一处理器核计算、主存和加速内存管理和网络通信等三大模块,屏蔽异构计算带来的差异性可以使上层软件设计保持不变性;当前 Khronos 组织发表的 SYCL 规范^[36]和 Intel oneAPI^[37]正在建立面向异构计算环境的统一编程接口,SYCL 底层基于 OpenCL 跨平台异构 API 规范,基于 SYCL 的 DPC++ 作为 Intel oneAPI 的主体,已经实现了 Intel GPU 和 Nvidia GPU 的异构加速计算;针对 AMD GPU 常用选择 ROCm 框架下的 HIP 编程方式,而 HIP 实际上能够兼容 CUDA 在 Nvidia GPU 上编译运行,另外,值得注意的是,

基于 HIP 也可以实现 SYCL 统一规范^[38].

3 总结

本文基于高性能计算应用软件的现状趋势,对宇宙 N 体模拟、大气与地球系统模式、计算材料相场动力学、分子动力学、量子计算化学和格点量子色力学等应用开展了并行计算特征分析,提炼了应用的典型算法和软件的共性问题,并开展了解决的对策讨论.

在国产异构计算高性能计算环境下,由于国产架构不同于主流的体系架构,其算法实现和软件设计有很大差异.另外我国自主的硬件体系架构设计起步较晚,所构建的软件生态环境包括系统工具和程序调试工具等等还远不成熟,给应用程序开发团队开发程序增加了现实难度.为了有效开展基于国产高性能异构计算算法和软件设计,现给出两点发展建议:

(1) 尽快建立我国高性能计算的统一软硬件标准.可从指令集、API 接口、软件架构、网格环境对接等方面建立统一规范.

(2) 兼容已有国际主流标准并积极参与国际规范标准的制定,提高国产超级计算平台与开源软件的兼容性.

References:

- [1] Ezell SJ, Atkinson RD. The vital importance of high performance computing to U.S. competitiveness. 2016. <http://www2.itif.org/2016-high-performance-computing.pdf>
- [2] Dongarra J, *et al.* Supercomputer Top500 list. 2019. <https://www.top500.org/lists>
- [3] Jin Z, Lu ZH, Li HY, Chi XB, Sun JC. Origin of high performance computing—Current status and developments of scientific computing applications. *Bulletin of Chinese Academy of Sciences*, 2019, 34(6):625–639 (in Chinese with English abstract).
- [4] USQCD Collaboration. US lattice quantum chromodynamics. 2019. <https://www.usqcd.org>
- [5] DeTar Carleton. MILC code. 2019. http://www.physics.utah.edu/~detar/milc/milc_qcd.html
- [6] Boyle P, Cossu G, Yamaguchi A, Portelli A. Grid: A next generation data parallel C++ QCD library. arXiv:1512.03487v1, 2015.
- [7] Clark MA, Babich R, Barros K, Brower R, Rebbi C. Solving lattice QCD systems of equations using mixed precision solvers on GPUs. *Computer Physics Communication*, 2010,181:1517–1528.
- [8] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 1995,117:1–19.
- [9] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 2005,26(16):1781–1802.
- [10] Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, de Jong WA. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications*, 2010,181(9):1477–1489.
- [11] Dongarra J, Foster IAN, Fox G, Gropp W, Kennedy K, Torczon L, White A. *Source Book of Parallel Computing*. Morgan Kaufmann Publishers, 2003.
- [12] Kothe DB, Diachin L. The exascale computing project. 2019. <https://www.exascaleproject.org>
- [13] Mo ZY, Zhang AQ, Liu QK, Cao XL. Parallel algorithm and parallel programming: From specialty to generality as well as software reuse. *Science in China (Information Sciences)*, 2016,46(10):1392–1410 (in Chinese with English abstract).
- [14] Mo ZY, Zhang AQ, Cao XL, Liu QK, Xu XW, An HB, Pei WB, Zhu S. JASMIN: A parallel software infrastructure for scientific computing. *Frontiers Computer Science in China*, 2010,4(4):480–488.
- [15] Liu QK, Zhao WB, Cheng J, Mo ZY, Zhang AQ, Liu JJ. A programming framework for large scale numerical simulations on unstructured mesh. In: *Proc. of the IEEE Int'l Conf. on High Performance and Smart Computing (IEEE HPSC)*. New York, 2016. 310–315.
- [16] Chi XB, *et al.* *National High Performance Computing Environment Development Report*. Beijing: Science Press, 2018 (in Chinese).
- [17] Colella P. Defining software requirements for scientific computing. 2004. <http://www.lanl.gov/orgs/hpc/salishan/salishan2005/davidpatterson.pdf>

- [18] Bagla JS. TreePM: A code for cosmological N-body simulations. *Journal of Astrophysics Astronomy*, 2002,23(3):185–196.
- [19] Ishiyama T, Nitadori K, Makino J. 4.45 Pflops astrophysical N-body simulation on K computer—The gravitational trillion-body problem. In: *Proc. of the Int'l Conf. on High Performance Computing, Networking, Storage and Analysis (SC)*, 2012. 1–10.
- [20] Wang YZ, Jiang JR, Zhang H, Dong X, Wang LZ, Ranjan R, Zomaya AY. A scalable parallel algorithm for atmospheric general circulation models on a multi-core cluster. *Future Generation Computer Systems*, 2017,72:1–10. [doi: 10.1016/j.future.2017.02.008]
- [21] Strumpen V, Frigo M. IBM research report: Software engineering aspects of cache oblivious stencil computations. Research Report, RC24035 (W0608-077), IBM, 2006.
- [22] Fu HH, Liao JF, Ding N, Duan XH, Gan L, Liang YS, Wang XL, Yang JZ, Zheng Y, Liu WG, Wang LN, Yang GW. Redesigning CAM-SE for peta-scale climate modeling performance and ultra-high resolution on Sunway TaihuLight. In: *Proc. of the Int'l Conf. on High Performance Computing, Networking, Storage and Analysis*. 2017. Article No.12 [doi: 10.1145/3126908.3126909]
- [23] Shimokawabe T, Aoki T, Takaki T, Yamanaka A, Nukada A, Endo T, Maruyama N, Matsuoka S. Peta-scale phase-field simulation for dendritic solidification on the TSUBAME 2.0 supercomputer. In: *Proc. of the Int'l Conf. on High Performance Computing, Networking, Storage and Analysis (SC 2011)*. 2011. Article No.3.
- [24] Zhang J, Zhou CB, Wang YG, Ju LL, Du Q, Chi XB, Xu DS, Chen DX, Liu Y, Liu Z. Extreme-scale phase field simulations of coarsening dynamics on the Sunway TaihuLight supercomputer. In: *Proc. of the Int'l Conf. on High Performance Computing, Networking, Storage and Analysis (SC 2016)*. 2016. Article No.4.
- [25] Allen MP, Tildesley DJ. *Computer Simulation of Liquids*. 2nd ed., Oxford University Press, 2017.
- [26] Brown WM, Wang P, Plimpton SJ, Tharrington AN. Implementing molecular dynamics on hybrid high performance computers—Short range forces. *Computer Physics Communications*, 2011,182(4):898–911.
- [27] Zhang S, Xu S, Liu Q, Jin Z. Cell Verlet algorithm of molecular dynamics simulation based on GPU and its parallel performance analysis. *Computer Science*, 2018,45(10):298–301,306 (in Chinese with English abstract).
- [28] Yasuda K. Two-electron integral evaluation on the graphics processor unit. *Journal Computational Chemistry*, 2008,29(3):334–342.
- [29] Ufimtsev IS, Martinez TJ. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *Journal of Chemical Theory and Computation*, 2008,4(2):222–231.
- [30] Ufimtsev IS, Martinez TJ. Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. *Journal of Chemical Theory and Computation*, 2009,5(4):1004–1015.
- [31] Wilson KG. Confinement of quarks. *Physical Review D*, 1974(10):2445–2459.
- [32] Brower R, Christ N, DeTar C, Edwards R, Mackenzie P. Lattice QCD application development within the US DOE exascale computing project. In: *Proc. of the EPJ Web Conf*. 2018. Article No.09010.
- [33] Hines J. Five Gordon bell finalists credit summit for vanguard computational science. 2019. <https://www.olcf.ornl.gov/2018/09/17/uncharted-territory/>
- [34] Denis C, de Oliveira Castro P, Petit E. Verificarlo: Checking floating point accuracy through Monte Carlo Arithmetic. In: *Proc. of the IEEE 23rd Symp. on Computer Arithmetic*. 2016. 55–62.
- [35] Aberger CR, De Sa C, Leszczynski M, *et al.* High-accuracy low-precision training. arXiv:1803.03383, 2018.
- [36] Wong M. C++ single-source heterogeneous programming for OpenCL. 2020. <https://www.khronos.org/sycl/>
- [37] Smith J. Intel® oneAPI toolkits^(Beta). 2020. <https://spec.oneapi.io/versions/latest/introduction.html>
- [38] Alpay A. HIPSYCL: SYCL 1.2.1 over AMD HIP/NVIDIA CUDA. 2020. <https://github.com/illuhad/hipsycl>

附中文参考文献:

- [3] 金钟,陆忠华,李会元,迟学斌,孙家昶.高性能计算之源起——科学计算的应用现状及发展思考. *中国科学院院刊*,2019,34(6):625–639.
- [13] 莫则尧,张爱清,刘青凯,曹小林.并行算法与并行编程:从个性、共性到软件复用. *中国科学:信息科学*,2016,46(10):1392–1410.
- [16] 迟学斌,等.国家高性能计算环境发展报告.北京:科学出版社,2018.
- [27] 张帅,徐顺,刘倩,金钟.基于 GPU 的分子动力学模拟 Cell Verlet 算法实现及其并行性能分析. *计算机科学*,2018,45(10):298–301,306.



徐顺(1980—),男,博士,副研究员,CCF 专业会员,主要研究领域为并行算法和代码优化,计算物理.



姜金荣(1977—),男,博士,研究员,CCF 专业会员,主要研究领域为并行计算,计算地球科学.



王武(1982—),男,博士,副研究员,CCF 专业会员,主要研究领域为并行算法,计算电磁学,计算宇宙学.



金钟(1974—),男,博士,研究员,博士生导师,CCF 专业会员,主要研究领域为并行计算,生物医药计算.



张鉴(1972—),男,博士,研究员,博士生导师,CCF 专业会员,主要研究领域为高性能计算,科学计算.



迟学斌(1963—),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为并行计算,计算数学.

www.jos.org.cn

www.jos.org.cn