

编者按 科研装备创新是科技创新的重要基础与保障,是衡量一个国家科技水平与综合竞争能力的重要标志。为推动我国科研装备的创新,“十一五”期间,财政部设立了国家重大科研装备研制专项,在中科院进行试点。根据财政部的要求,中科院从我国前瞻性基础研究和战略性高技术发展需求出发,结合近年来各研究所在仪器设备相关技术方面取得的重要突破,认真组织策划,先后部署了 15 个项目,在管理模式、支持方式等方面进行了有益探索,并取得了良好进展。为使读者更好地了解中科院重大科研装备自主研制方面的工作和进展情况,本刊推出系列仪器设备研制介绍,希望能够引起广大科技工作者的关注和支持,以更好地推动科研装备的自主研制工作。

高效能低成本多尺度离散模拟 超级计算应用系统

葛 蔚

曹 凝*

(中国科学院过程工程研究所 北京 100190) (中国科学院计划财务局 北京 100864)

关键词 科研装备,多尺度离散模拟,超级计算

1 科学背景

随着现代科学技术的飞速发展,传统的科学实验和理论研究方法已不能完全满足当代科学研究与技术进步的需求,计算机模拟作为 20 世纪发展起来的新型科研手段被视为现代科学技术进步的“加速器”而受到越来越多的重视。大型计算机模拟是以计算能力为基础,依据研究目标的理论模型,运用数值计算方法,在计算机上进行虚拟实验。其特点是费用低、周期短、方便灵活、应用广泛,甚至可以模拟人类现代技术无法实现的科学实验,如星系演化和宇宙形成过程等。因此,超级计算系统已成为世界诸强竞相研制的国之重器。

超级计算系统有通用和专用之分。前者能较均衡地满足各领域的计算需求,而后者是为特定算法专门优化设计的。主频和集成度等元器件技术的进步能整体提升各种计算机对各种算法的计算速度。但近年来,以集成电路的线宽为标志,人类正在逼近当前技术的极限,因此速度的提升已越来越多地依靠多核并行、异构和流处理等需要与软件和算法更紧密配合的途径。另外,随着性能的飞速提升,个人计算机已进入了很多原本需要超级计算的领域,使超级计算的需求逐渐集中于若干专门领域。因此,过分追求通用,会使超级计算系统如同自行车与跑车混行的高速公路,在效率和经济性上都面临挑

* 中国科学院计划财务局副局长。E-mail:zbc@cashq.ac.cn
收稿日期 2011 年 6 月 20 日

战。而如果对各种算法分别研制专用系统,计算速度虽高,但开发、生产、运行维护和应用成本同样提高,也缺乏普遍意义。

高效能低成本多尺度离散模拟超级计算应用系统则提出了能够涵盖若干领域、适合大量超级计算问题的“多尺度离散模拟”通用算法框架。针对该框架设计高效的软件和优化的硬件系统,实现低成本、低能耗、高效率和高性能的计算,兼备通用与专用系统的优势,其具体表现是:

(1)适用范围广。利用此框架,代表各种应用的各种单元间的作用方式可模块化地嵌入通用的总体算法和数据结构中,而无需独立编写相应的计算软件。

(2)可扩展性强。通用超级计算系统为适应各种不同的算法和应用问题,需要任何一对处理器间能快速交换数据,多尺度离散模拟的特点决定了每个处理器只需和特定的极少数相邻处理器交换或共享数据,只要可靠性允许,这样的系统中处理器数量可任意扩展而保持相对的成本和使用效率不变。

(3)并行效率高。各离散单元间作用的计算可在大量处理器上同时进行,而不必采用传统的中央处理器(CPU)顺序处理,可极大地提高处于计算操作中的元器件的比例、减少存储硬件的开销,从而降低同样计算能力下硬件制造难度、成本及运行功耗。

2 装置综述

解决自然界很多复杂问题的瓶颈在于缺乏对其时空多尺度结构的认识,这也是复杂性科学

研究的焦点问题。中科院过程所从1984年开始就致力于用多尺度方法研究气固两相系统,逐步发展成极值型多尺度方法。在推广应用极值型多尺度方法的思想 and 用离散方法证明不同系统稳定性条件的过程中,逐步认识到多尺度和离散化是很多工程问题的共性,自此致力于建立针对这一共性的计算模拟方法和软件,并设计相应计算机系统。

2007年6月,Nvidia(英伟达)公司发布了CUDA(Compute United Device Architecture)1.0,中科院过程所意识到可以借用“CPU+GPU”方案来实现多尺度离散模拟。为此,该所仅用4个月时间就建立了由126台HPUxw8600工作站组成,单精度峰值超过100Tflops的Mole-9.7集群并行计算系统。利用该系统,该所成功开展了多相流动直接数值模拟、材料和纳微系统微观模拟和生物大分子动态行为模拟等应用,证明了多尺度离散化并行计算模式的优势和前景。

2009年1月,高效能低成本多尺度离散模拟超级计算应用系统项目正式启动。项目组首先对Mole-9.7的应用情况进行分析,通过采用Nvidia GTX295显卡,于2009年



多尺度离散模拟 Mole-8.5 系统

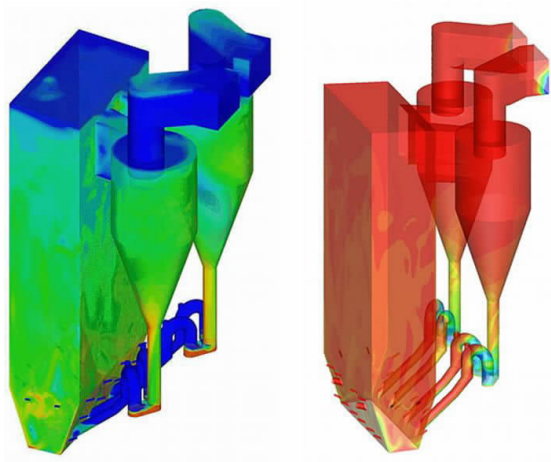
初将 Mole-9.7 系统升级到了单精度峰值 450Tflops,同时选用 AMD 的 HD4870x2 显卡构建了另一套单精度 150Tflops 单元系统。这两套系统分别采用 CUDA 和 Brook+ 编程,为了实现耦合计算,项目组采用了基于相同数据接口的多道程序 MPI 并行模式,成功进行了计算流体力学模型的尝试。这一成功为更方便灵活地利用多种 GPU 资源提供了保证,也形成了系统建设上的新思路:通过与合作公司的合作,提高系统的产品化水平,力争推广到更多用户;同时,将这些系统通过既有网络连接起来,开展异构资源的分布式计算。为此,中科院计财局等主管部门积极组织协调,促成过程所与联想、曙光公司各建立一套单精度峰值 200 万亿次的单元系统的技术方案,并于 2009 年 2 月安装到位。两家公司分别采用了 Nvidia 的 GTX280、Tesla C1060 和 AMD 的 HD4870 x2 显卡。它们与过程所研制的两套系统联网共同形成了 Mole-8.7 系统。全系统共 1 000 多块 GPU 卡。采用多程序 MPI 的方式成功在 953 块 GPU 上实现了凹纹槽流的格子玻尔兹曼模拟,并且相对效率与 CPU 系统相当。2009 年 4 月 20 日,中科院正式对外宣布了这一成果。这不仅是该项目的阶段性成果,很好地检验了预定的整体设计思路,而且是国内首套单精度峰值超过 1 000 万亿次的超级计算系统。

在以上成果的基础上,中科院组织协调了 10 个研究所来推广部署由曙光和联想生产的上述单元系统。历经半年多时间的设计、安装、调试,至 2010 年初,各家单位都成功开展了 GPU 计算和应用研究,并在地质勘探数据处理和天体物理模拟等方面取得了国际前沿水平的应用和理论结果。与此同时,过程所密切关注 GPU 计算技术的最新发展,与 Nvidia 和 Tyan 等公司紧密合作,

形成了最终目标系统的设计方案。通过分析应用特性,其核心系统 Mole-8.5 主要采用 Tyan 的 S7015 主板,最多可安装 8 块 Nvidia 公司最新的 Tesla C2050 GPU 卡(Fermi),从而使单机点执行离散模拟的性能价格比能得到最充分的发挥,并使过程所成为了 Nvidia C2050 GPU 卡在全球的首个批量用户。

2010 年 4 月 24 日,Mole-8.5 系统初步建成,并实现了与 Mole-8.7 以及 10 个研究所的 10 套系统的联网计算,形成了单精度峰值近 5 000 万亿次的分布式超级计算环境。2010 年 5 月,中科院过程所在软件所的通力合作下,及时提交了 Linpack 测试结果,并在 2010 年 6 月 Top500 排名中名列第 19 位,而在稍后的 Green500 排名中更进入了前 10(列第 8 位)。2010 年 7 月,采用离散单元法实现了工业规模装置准实时模拟,并开展了实时模拟和虚拟过程示范系统的前期设计。为适应实时模拟的需求,项目组还提出了计算与显示耦合的在线可视化方式并在 Mole-8.5 系统上成功应用。后期,过程所通过与中科院计算所的密切合作,实现了三维并行显示计算耦合,并应用于海量计算数据的离线可视化。

2010 年 10 月,系统硬件、软件和应用



超临界锅炉模拟

等各方面顺利达到了验收要求:理论峰值计算速度为每秒 2 224.8 万亿次单精度浮点运算,实例测试中达到的最高计算速度为每秒 1 432 万亿次单精度浮点运算,系统 Linpack 测试峰值 207.3 Tflops; 系统能耗 563 KW, 系统总能耗 (含冷却系统 200KW)763KW, 占地面积 145m², 系统内存容量 17.792TB, GPU 显存容量 6.48TB, 共计 24.272 TB; 计算系统重量 12.6 吨, 磁盘总容量 720.584TB; 系统软件主要包括结点操作系统 CentOS 5.4、GCC/G++-4.1.2 编译器、MPI/OpenMP/CUDA 编程环境、Ganglia 和 MoleMonitor 监控软件等, 实现了远程系统访问和作业管理。2010 年 11 月 8 日, 系统正式通过验收, 标志着我国在高性能计算应用系统领域又取得重要进展。

3 组织与管理

在高效能低成本多尺度离散模拟超级计算应用系统研制阶段, 采取工程管理方式组织实施, 设立工程领导小组、总体组、咨询委员会、监理委员会和用户联盟等 5 个组织单元。领导小组由中科院主管院领导任组长, 财政部及中科院主管司局领导、项目承担单位领导任副组长, 主要负责工程总体目标的设置、任务的审定和工程实施过程中重大事项的决策, 与国家相关部门之间沟通协调等。总体组主要负责工程的全过程管理, 成员由项目牵头单位中科院过程所聘任, 报领导小组批准, 对单位法定代表人负责。咨询委员会由相关行业专家组成, 对领导小组负责决策咨询, 对总体组负责技

术咨询。监理委员会由中科院计财局聘任, 负责对工程质量、经费管理及工程进度等进行全程目标控制、跟踪和监督。用户联盟由研制系统现有目标用户单位和潜在用户单位组成, 负责参与项目的市场调研和相关讨论, 是研制系统技术需求和市场需求的来源。

4 应用

由于应用需求迫切, 高效能低成本多尺度离散模拟超级计算应用系统在建设过程中就开始承担“大型油气田及煤层气开发”等国家重大专项、国家科技支撑计划及国家自然科学基金重大基金项目中的重要计算服务, 并为中石化、中石油、宝钢、兖矿以及通用电气、阿尔斯通、必和必拓、联合利华等国内外大型企业提供模拟计算任务, 成功应用于化工、冶金、能源、生物和材料等领域的过程模拟与优化设计, 以及物理、力学、化学和生命科学领域的若干基础研究问题。在气固系统多尺度模拟、复杂生物分子模拟和金属晶体材料模拟等领域成功实施了世界前沿水平的大规模并行计算。该系统的硬件成本和运行费用较传统的 CPU 超级计算系统显著降低, 提高了实际应用效能, 有力支持了国家重大项目的研究和国际合作, 有力推



项目验收会

动了我国超级计算应用水平的跨越式发展。

该系统主要针对过程工程中大型反应器的优化和改造开展了模拟研究,包括流化床、搅拌槽、锅炉和冶金炉等,模拟尺度达到米级而分辨率达到毫米级,采用的粒子或网格数最多超过了 10 亿。通过材料表面结构的分子动力学模拟与气体剪切流动的拟颗粒模拟的耦合实现了对高温高速下气动加热和材料变形破坏等问题的研究,具备了在航空航天、能源动力和材料力学等领域实际应用的能力。实现了缝洞型油藏中油水驱替过程的介观模拟,实现了微米级裂缝与厘米级孔洞的耦合模拟,整体计算规模可超过米级,具备了部分代替油藏物理模拟的能力。针对钢铁冶炼新工艺的主反应器和钢渣处理等工艺过程实现了全系统、全尺寸的模拟优化,并在计算速度上接近实时模拟。

另外,该系统开展的应用还包括天体演化的物理学计算、金属合金材料设计、集成电路的电磁场分析、高能物理实验数据分析、高能粒子与材料间相互作用、油藏勘探地震波数据的反演、医学影像三维重构、移动空间物体跟踪分析、基因比对等方面。如在多尺度模拟计算系统上进行大规模计算模拟三维流感病毒在细胞液中的动态结构,体系中包含的原子个数约 3 亿,模拟使用了多达 1 728 个 Tesla C2050 GPU,是目前已知的对该类型体系的最大规模的模拟,模拟结果可用于研究流感病毒的结构,探索抗流感药物同流感病毒间的作用。

5 发展展望

作为我国超级计算应用水平跨越式发展的典型标志,高效能低成本多尺度离散模拟超级计算应用系统在向我们展示其广阔

应用前景的同时,也给我们带来了许多启示,并为未来发展指明了方向。

其一,为我国超级计算系统的应用发展探索了新的模式,即通过软件的通用化实现广泛的应用,而利用硬件的专门化提高计算效率、降低制造和运行成本,突破百万亿次级多尺度离散模拟超级计算系统的产品化技术,促进超级计算的普及化;

其二,开辟了应用牵引、软件主导、系统创新的计算机和模拟技术发展道路,扭转了开发与应用的脱节,并利用软硬件优势的集成与互补,巧妙地突破了国外模拟软件和硬件的垄断;

其三,为产业界和科技界提供了强有力的技术支撑平台,形成若干具有自主知识产权的应用软件,解决若干重要工业过程开发和基础科学研究中的计算难题,走出一条具有特色的科学和工程计算道路,并在多相复杂系统、多尺度模拟等领域进入国际领先行列;

其四,有力促进计算机模拟在过程研发中的应用,推动过程工程领域的科技水平和自主创新能力的整体进步;

其五,为我国培养了超级计算应用系统研发的新型团队,及覆盖多个领域的新型模拟应用人才,形成研发与应用的可持续发展能力;

最后,高效能低成本多尺度离散模拟超级计算应用系统项目的研制成功也启示我们,在开放的环境下,只要我们解放思想,不断从创新实践中提出新概念,提出新思想,通过体制机制创新,有效整合优势技术资源联合攻关,我们完全有能力做出世界一流的仪器设备。



中国科学院