

* 学科发展 *

高性能计算技术展望

夏培肃* 胡伟武

(计算技术研究所 北京 100080)

摘要 文章概述了高性能计算机系统的主要类型与发展现状,对 21 世纪超高性能计算机系统及其新技术作了展望,并提出了我国应采取的对策。

关键词 高性能计算技术

人类对计算技术的需求是没有止境的。高性能计算技术在航空航天技术、石油勘探和开发、大范围气象预报、核爆炸模拟、材料设计、药物设计、基因信息学、密码学、人工智能、经济模型、数字电影等领域起着重要的作用。高性能计算机可以对所研究的对象进行数值模拟和动态显示,可以模拟没有重力的宇宙空间、天体运动、核聚变中的等离子、酶的活性、人体器官和骨骼、大气环流等,还可以用来研究车辆的碰撞过程、引擎内部的燃烧机制、物体表面的原子排列、抗流星碰撞的宇宙飞行器设计、大规模数据库的同时搜索等。它在当前的科学发展中起着不可替代的重要作用。在国际高科技竞争日益激烈的今天,高性能计算技术及其应用的水平已成为显示国家综合实力的一个标志。

1 高性能计算机的主要类型与发展现状

美国是高性能计算技术最先进的国家,目前世界上运算速度最快的 50 台计算机中有 37 台是美国研制和生产的。美国已于 1996 年研制出持续运算速度超过每秒 1 万亿次浮点运算的 ASCI Red 计算机,并将在明年之内完成每秒运算 3 万亿次的 ASCI Blue 计算机。

高性能计算机系统都是并行处理系统,它有 5 种主要类型:

- (1) 传统的向量多处理机系统。
- (2) 可伸缩的大规模并行处理(MPP)系统。

MPP 系统是在同一地点由大量(数百或数千)处理单元构成的高性能并行处理系统。从应用程序界面和存储管理的角度看,多机系统可分为共享存储系统和消息传递系统两类。在共享存储系统中,所有处理机共享主存储器,处理机之间的通信通过访问共享存储器来实现;在消息传递系统中,每个处理机都有自己的局部存储器,处理机之间的通信通过显式的消息传递来进行。与消息传递系统相比,共享存储系统具有较强的通用性,且可以方便地移植现有的应

* 中国科学院院士
收稿日期:1998 年 6 月 5 日

用软件,但可伸缩性较差。近年来的发展表明,消息传递系统与共享存储系统日趋统一。一些研制 MPP 系统的公司都推出了分布式共享存储 MPP 系统。另外,在消息传递的硬件机制上用软件提供共享存储的编程界面(即所谓的共享虚拟存储)也是近年来研究的热点。

目前的 MPP 系统主要存在以下两个问题,即运行效率低和可编程性差。运行效率低指系统实际运算速度远低于峰值运算速度;可编程性差指并行程序的开发比较困难,串行程序向并行程序自动转换的效果不好,且并行程序在不同平台之间的有效移植也有一定的难度等。这些问题的原因是多方面的,涉及到计算机体系结构、系统软件、并行计算方法等。为了解决这些问题,各个方面正在进行积极的研究,包括使用分布式共享存储系统、利用多线程技术以隐藏访存延迟、开发高带宽的互连网络、提高访存的局部性、提高算法的并行度、利用预取和后存技术、开发支持数据并行和任务并行的 HPC++ 语言、建立可伸缩的软件库等。

(3) 工作站网。

工作站网是将大量同一品种或不同品种的工作站(或微机)通过高速网络互连,以构成廉价的高性能计算系统。

(4) 可伸缩的、由高速网络互连的异构计算机系统。

将分布在不同地点、不同类型的计算系统(包括 MPP 系统、工作站等)通过高速网络互连,可以实现比 MPP 系统性能更高、功能更强的计算系统。这种异构计算机系统内的不同计算机可以执行所计算问题的不同内容,是高性能计算系统的一个重要发展方向。

(5) 高性能专用计算系统。

高性能计算机系统一般采用 COTS 技术,即用现成的商品硬件(例如微处理器)和商品软件制作计算机系统。COTS 技术使计算机的研制周期短,性能价格比高,但计算机系统性能的进一步提高和实现新的设计思想受到限制。

高性能计算机系统常用的一些系统软件有:可管理成千的并行处理单元、控制各种外围设备并具有连网功能的可扩展的微内核操作系统;可直接支持数据并行的 HPF 语言及其编译器,并行 C、并行 C++ 语言及编译器,将串行程序自动并行化的 SUIF;以及可移植的异构编程环境 PVM、适用于消息传递的标准平台 MPI、分布式共享存储编程环境 Tread Marks 等软件平台。

高性能计算机系统常用的一些测试程序和程序库有:LINPACK、NAS、适用于测试可伸缩的分布式存储计算机系统的 PARKBENCH、可伸缩的线性代数程序库 SCALAPACK、可移植的不规则复杂问题程序库 SCPLIB 等。

高性能计算应用有计算密集型应用、数据密集型应用和通信密集型应用三个方面。计算密集型应用包括大规模科学和工程计算、数值模拟等;数据密集型应用包括数字图书馆、数据仓库、数据挖掘和分析、可视化等;通讯密集型应用包括协同工作、遥控、远程医疗诊断等。

2 面向 21 世纪的高性能计算系统

目前,以美国为代表的发达国家已经制定或正在制定发展高性能计算技术的战略计划。美国为了保持其高性能计算技术在世界上的领先地位和解决对美国至关重要的挑战性问题,1991 年美国国会根据戈尔副总统的建议通过了高性能计算法,接着先后实施了“高性能计算和通信”(HPCC)、“加速战略计算倡议”(ASCI)及“计算、信息和通信”(CIC)等多项国家级的战略性计划。前面提到的每秒运算 1 万亿次和 3 万亿次的计算机都是 ASCI 计划的一部分。根

据 ASCI 和 CIC 计划,美国将在 2000 年完成每秒运算 10 万亿次的计算机,2001 年完成 30 万亿次的计算机,2003 年完成 100 万亿次的计算机。另外,还将每秒运算 1 000 万亿次浮点运算的计算机硬件和软件的研究工作正式列入 1998 年度的 CIC 计划中。预计美国在 5 年之内将完成 1 000 万亿次的专用计算机 GRAPE-6,10 年之内将完成 1 000 万亿次的通用计算机。在 1998 年度的 CIC 计划中,还包括“先进计算基础设施伙伴”(PACI)计划,这是将分布在不同地点的计算资源和数据资源通过高速网络互连,使用户可以把网络上的资源作为一个整体来看待。

如上所述,在下世纪初,超高性能计算机系统可有三种形式,即:每秒运算 10 万亿次至 1 000 万亿次的 MPP 系统;针对某种应用而设计的高性能专用计算机系统;由高速网络互连的元计算(Metacomputing)系统。

每秒运算 1 000 万亿次的 MPP 系统如果采用目前的每秒运算 10 亿次的微处理器,则需要 100 万个这样的处理器,处理器数量的增加会带来很多困难,因此必须提高微处理器的运算速度。Intel 公司预计在 2006 年将生产出每秒运算 200 亿次的微处理器,实现 1 000 万亿次的 MPP 系统需要 50 000 个这样的微处理器。由于存储器速度的增长不如微处理器快,处理器的运算速度提高后,它与存储器之间的速度差距会更大,从而导致访存延迟大幅度增加,特别是远程访问的延迟更为严重。1 000 万亿次的 MPP 系统的系统软件必须管理成千上万台处理机、几十万个线程,还必须面对多层次(7—10)的存储系统,这无疑是对系统软件的严重挑战。对于算法,要求能提供 100 000—1M 个并发线程,并能够容忍访存延迟。所以,为了解决或缓解上述困难,研制 1 000 万亿次的 MPP 系统应有新的体系结构、新的硬件技术、新的软件思想和新的算法,而更重要的是它们之间的有机结合。

由于实现超高性能的通用计算机的难度很大,对于求解某些只需对少数算法进行大量反复运算的问题(例如天体物理中的多体问题),如果使用高性能专用计算机,不失为一种较好的选择。1 000 万亿次的专用计算机的研制费用只有相应的通用计算机的 1%。

元计算系统将各种不同的计算机通过高速网络互连,网络资源对用户透明。用户在自己的工作stations上工作时,好象在使用一台性能很高的计算机。这种系统利用现成的计算资源,可以以较少的投资实现高性能计算,但它的系统效率低。

3 高性能计算机的新技术

在未来的高性能计算机中,半导体技术将继续获得发展,一些非半导体技术,如光计算技术、超导体计算技术、量子计算技术以及生物分子计算技术也将得到发展。

在半导体技术方面,除了微处理器将提高集成度和提高工作频率,从而提高运算速度外,PIM (Processor In Memory) 技术也是近年来研究得较多的半导体技术。其主要思想是将 CPU、寄存器和存储器交错地集成在同一芯片上,使得处理机和存储器之间的通信带宽有数量级的增加,从而大大缓解 CPU 的访存瓶颈。如将一个 SMP 系统集成在一个芯片上,用很多这样的芯片构成分布式系统,可以实现性能很高的计算机。这种系统的主要问题是散热和芯片引脚的限制。

光计算技术是近年来的研究热点之一。目前全光计算机在技术上尚未成熟,但光互连和全息存储技术的研究取得了较大进展,尤其是点-点的光纤互连技术已趋于成熟。光纤互连有频带宽、无电磁干扰、可高密度并行连接、多信号源和多扇出、传输速度快、衰减小、不需接地等优

点。光纤的波分光交换技术在 MPP 系统的互连网络中有自动寻径功能,具有诱人的前景。全息存储将信息以光栅的形式存储于晶体中,光栅是两束激光(信号光束和参考光束)在光敏材料中相互干涉时形成的。用参考光束再次照射晶体,发生衍射,就重现原来的信息。全息存储的带宽可达 100 Gbps,存储密度可达每立方厘米 1TB。

超导技术又称为 RSFQ (Rapid Single Flux Quantum) 技术,被认为是实现 1 000 万亿次计算机的最具吸引力的技术之一。它具有速度快、功耗小、工艺简单等优点。它利用超导状态下的 SQUID (Superconducting Quantum Interface Device) 在外加电流时可以储存或释放一份磁通量的原理,用 SQUID 中磁通量的有无表示“1”和“0”。由于工作在超导状态,速度快,超导处理机的工作频率可达 100 GHz。由于外加电流很小,功耗也很小,每个门的功率只有 0.1 μ W。其缺点是负载驱动能力差,需要低温(4—5K)环境等。超导技术可以和半导体技术及光技术一起构成高性能计算机。

量子计算机由利用量子效应作为工作基础的量子器件组成。量子器件的工作原理和常规计算机所用的器件不同。常规器件的信息位不是处于“0”态就是处于“1”态,而量子器件的信息位除了“0”态和“1”态以外,还有相干态。相干态可以是“0”也可以是“1”,它们出现的概率相等。量子计算机比常规计算机具有大得多的状态空间。例如常规计算机的 5 位寄存器只能存放一个 5 位数,而 5 位量子寄存器可以存放 2^5 个数。而且,量子计算机可以对这些数同时进行运算。量子计算机适合于模拟量子物理系统和进行大数分解。实现量子计算机需要解决两个难题:为了维持量子逻辑的一致性,量子系统和环境的隔离问题;设备缺陷所引起的逻辑错误问题。

分子计算机的原理是利用 DNA 链保存信息,并通过 DNA 链之间的生化反应完成运算,产生表示结果的 DNA 分子。其优点是高度并行(所有 DNA 分子同时参加运算)、功耗小(半导体计算机的十亿分之一)、存储密度大(磁存储器的一万亿倍)。DNA 计算机的高度并行性适合于解决半导体计算机难以解决的问题,如可满足性问题等。利用 DNA 计算机存储密度大的特性为半导体计算机提供巨大的存储空间。实现分子计算机需要解决的问题包括生化反应慢、DNA 分子容易水解、操作不可靠、DNA 分子之间难以通信等。

4 结束语

高性能计算技术对促进科技进步、推动经济发展和保障国家安全有十分重要的作用。美国和高性能计算有关的计划由计算、信息和通信委员会(CCIC)负责,该委员会直属以克林顿总统为首的国家科技领导小组(NSTC)。高性能计算技术不是由市场驱动,而是由对国家至关重要的挑战性问题所驱动。它们的科研工作不是企业行为,而是国家行为。

在我国,如何发展面向 21 世纪的高性能计算技术,在国际高性能计算领域占有一席之地,不仅需要专业科技人员的积极努力,更需要各级领导的高度重视。当务之急是及早制定我国发展面向 21 世纪高性能计算技术的战略规划并积极组织实施。

参考文献

- 1 K. Hwang and Z. Xu. Scalable Parallel Computing. McGraw-Hill, 1998.
- 2 美国 CIC 计划 1998 蓝皮书。