

# 我国高端计算之路与“神威”的探索

陈左宁

(国家并行计算机工程技术研究中心, 北京 100080)

**[摘要]** 21世纪初,受应用需求的驱动,国内外对高端计算的研究与开发掀起了新一轮的热潮。分析了国外高端计算的发展现状及趋势,描述我国高端计算发展的总体状况及应用开发情况,介绍神威系列高性能计算机的主要技术特点和应用成果,探讨当前高端计算技术发展所面临的问题,并提出了发展我国高端计算事业的几点建议。

**[关键词]** 高端计算; 网格计算; 高性能服务; 高效能计算; 总体拥有成本; 神威系列计算机

**[中图分类号]** TP338.4 **[文献标识码]** A **[文章编号]** 1009-1742(2004)09-0023-06

## 1 国外高端计算发展现状及趋势

### 1.1 国外高端计算发展现状

21世纪,高端计算(high end computing)技术已成为衡量一个国家经济技术综合实力的重要标志,它对国民经济、社会发展、国家安全和国防现代化建设具有重要意义。以美国和日本为代表的发达国家十分重视高端计算机系统的研制及其应用技术的开发。综观当前国际高端计算的发展形势,一方面,在应用需求和市场因素的驱动下,高端计算正从科学计算向经济和商业各个领域扩展,人们更看重高端计算机系统的服务能力,以计算为中心转向以数据和应用为中心;另一方面,受新的科学发现、国防建设、国家安全等应急性任务需求的牵引,高端计算的研究、开发以及对高端计算系统的采购出现了新的热潮。

根据全球实用超级计算机500强最新排行榜的统计分析,目前国际上已经有242台系统的Linpack实测性能超过1万亿次/秒(2003年12月前只有131台);500强系统的总性能为813 TF/s(2003年12月前为528 TF/s);排行榜中“最慢”系统(第500台机器)的速度为624 GF/s(2003

年12月前为403.4 GF/s);现在500强系统的主流结构是Cluster, Constellations和MPP三种结构类型。所有系统分布在世界上35个国家和地区,美、日、德、英等发达国家占了80%的计算资源,其中仅美国就安装了255台,占总性能的56%;并且500台系统中的91%是由美国制造的,所有这些数据均表明美国在高端计算机的使用和生产方面仍然保持着绝对的领先优势。TOP10超级计算机见表1。

### 1.2 国外高端计算发展趋势

国外高端计算系统今后的开发热点是计算速度为十万亿次/秒左右的系统,中期目标是百万亿次/秒,长期目标是千万亿次/秒甚至更高。未来国际高端计算的发展将呈现以下趋势。

1) 超级计算机和网格计算2种实现形态共存  
高端计算应用的多样性导致了高端计算实现形态的多样性。未来的高端计算将朝着2个方向发展:一是受“9.11”事件的影响和日本成功推出“地球模拟器”的触动,居高端计算霸主地位的美国加大投入,不断研制面向科学研究应用的尖端超级计算机;二是受网络技术迅猛发展的推动,网格计算作为无缝集成的协同计算环境,又称“虚拟超级计算

机”，会得到进一步重视。目前，美、日等国政府和研究单位在这两方向都有相应的资金投入和具体规划。未来高端计算系统一定是网格中的重要计算

资源，它需要适应网格环境的计算模式、编程模型、开发运行和管理系统等；网格技术也将是未来获取高端计算能力的优良环境，二者将互为补充。

表1 国际最新 TOP10 超级计算机排名<sup>[1]</sup>

Table 1 A list of the most advanced TOP10 supercomputers in the world

排位	制造商/系统名称/处理器数	实测速度/峰值速度/GF·s <sup>-1</sup>	效率/%	年代
1	NEC/Earth-Simulator /5120	35 860.00 /40 960.00	88	2002
2	California Digital Corporation/Thunder Tiger4 1.4 GHz/4096	19 940.00/22 938.00	87	2004
3	HP/ASCI Q-AlphaServer SC ES45, 1.25 GHz/8192	13 880.00/24 480.00	68	2002
4	IBM/BlueGene/L DD1 Prototype (0, 5 GHz PowerPC440) /8192	11 680.00/16 384.00	71	2004
5	Dell/Tungsten PowerEdge 1750, P4 Xeon 3.06 GHz/2500	9 819.00/15 300.00	64	2003
6	IBM/eServer pSeries 690 (1.9 GHz Power4+) /2112	8 955.00/16 051.00	56	2004
7	Fujitsu/RIKEN Super Combined Custer/2048	8 728.00/12 534.00	70	2004
8	IBM/BlueGene/L DD2 Prototype (0.7 GHz PowerPC440) /4096	8 655.00/11 469.00	75	2004
9	HP/Mpp2 Integrity rx2600 Itanium2 1.5 GHz, Quadrics/1936	8 633.00/11 616.00	74	2003
10	Dawning/Dawning 4000A Opteron 2.2 GHz, Myrinet/2560	8 061.00/264.00	72	2004

2) 从高性能计算走向高效能计算 随着高性能计算向高效能服务转变，超级计算机系统追求的目标也将从“高性能” (high performance) 走向“高效能” (high productivity)。按美国 DARPA HPCS 计划说明，High productivity 的综合含义是指提高超级计算机系统的计算性能、可编程性、可移植性和鲁棒性，同时努力降低系统的开发、运行及维护成本。HPCS 计划表明，超级计算机要想保持快速发展势头，必须要有本质的变化，即必须采用先进技术，平衡各项设计指标，实现系统的高可靠性、高可用性、高可维性、高安全性和低功耗。

3) 进一步发展“提速”技术 开发深度并行性，增强单位面积的处理能力，是未来提高超级计算机性能的一条重要措施。目前，国外正在加紧研发的“提速”技术包括：深度并行体系结构（例如 SMT, IBM ViVA, Cray locale）；单芯片多处理器（CMP）技术；存储器内处理器（PIM）技术，可重构多处理器结构，以及光连接技术等等。

国外计划研制的超级计算机系统见表 2。

表 2 国外计划研制的超级计算机系统\*

Table 2 The supercomputers under development in the world

系统名称	峰值性能 /TF·s <sup>-1</sup>	处理器	处理器数	完成年代	投资方
Cray 红色风暴	40	AMD Opteron	10368	2004	Sandia
CrayX1 (原 SV2)	52.4	定制 MSP	4096	2003	DOD
富士通 HPC2500	85.1	SPARC64 V	16384	未确定	自筹
IBM ASCI 紫色	100	Power5	12544	2004	DOE
IBM 蓝色行星	150	Power5	16384	2005	DOE
IBM 蓝色基因/L	360	PowerPC440	65536	2005	DOE

## 2 中国高端计算发展概况

### 2.1 总体发展情况

我国高端计算机系统研制开始于 20 世纪 70 年代中后期，大体经历了 3 个主要发展阶段：第一阶段从 70 年代中后期到 80 年代中期，主要以研制大型向量系统为主（以银河 I 为代表）；第二阶段从 80 年代中后期到 90 年代末，主要以研制大规模并行系统为主（以神威 I 为代表）；第三阶段从 90 年代中期起，主要以研制大规模机群系统为主（以曙光机为代表）。目前，参与高端计算机研制的单位已经从科研院所发展到企业界。

进入新世纪，随着研制高端计算机系统的诸多关键技术被攻克（尤其是机群技术），我国自行研制的高端计算机系统已开始形成自己的品牌系列和一定的市场规模，其发展呈现星火燎原之势头。近两年，随着“神威”、“银河”、“曙光”、“深腾”、“天梭”等一批知名产品的出现，使我国成为继美、日之后第三个具备高端计算机系统研制能力的国家，被誉为世界未来高端计算市场的“第三股力量”。

根据中国软件行业协会数学软件分会 2003 年 11 月份公开发布的 2003 年中国高性能计算机 TOP100 排行榜最新统计，我国高端计算机系统的总计算能力在 19.56 TF/s 峰值左右，其中国产机器有 23 台，引进系统 73 台，具体情况见表 3。在

\* 国外超级计算机发展现状及趋势研究报告

体系结构上，呈现出多样化的局面，从 SMP, CC-NUMA, Cluster 到 MPP，其中 Cluster 结构占很大比例。另外，目前国产系统已开始进入世界超级计

算机 500 强排行榜。2004 年曙光、联想、浪潮等公司产品均在最新 TOP500 中榜上有名，其中曙光 4000A 系统跃居第 10 名。

表 3 中国 TOP100 制造商分析<sup>[2]</sup>

Table 3 The analysis of Chinese TOP100 manufactures

产地	厂商	台数	份额/%	$R_{max}/GF \cdot s^{-1}$	$R_{peak}/GF \cdot s^{-1}$	处理器数
国产机器	联想	9	9.00	7 194.04	11 224.0	2 360
	神威	7	7.00	1 660.73	4 187.86	1 504
	自制	4	4.00	295.96	597.20	408
	清华	2	2.00	866.44	1 600.16	400
	浪潮	2	2.00	727.4	1 238.4	228
	华云/奥利金	2	2.00	155.5	265.6	50
	曙光	1	1.00	39.00	111.70	160
	小计	27	27.00	10 943.07	19 224.92	5 110
引进机器	HP	57	57.00	6 594.08	10 032.3	3 196
	SGI	8	8.00	504.46	614.4	576
	IBM	7	7.00	1 585.68	4 039.4	944
	SUN	1	1.00	52.7	64.00	80
	小计	73	73.00	8 732.92	14 750.1	4 796
总计		100	100.00	19 675.99	33 975.02	9 906

2.2 高端计算应用状况

高端计算机系统的发展有力地带动了应用的开发，在石油行业、气象气候预报、核能模拟、生物工程、生命科学等实际使用中已取得显著成效，拓

宽了我国科学研究和工程应用的深度和广度。表 4 描绘了中国 TOP100 高性能计算系统的应用领域分布情况。

2.2.1 石油行业应用现状

表 4 中国 TOP100 行业领域分析<sup>[3]</sup>

Table 4 The analysis of Chinese TOP100 application areas

应用领域	台数	份额/%	$R_{max}/GF \cdot s^{-1}$	$R_{peak}/GF \cdot s^{-1}$	处理器数
科学计算	29	29.00	10 240.18	15 020.76	4 912
金融保险	21	21.00	2 383.98	3 671.30	1 168
电信	16	16.00	2 097.89	3 363.10	1 012
税务	8	8.00	1 208.70	1 904.00	544
能源	7	7.00	2 250.54	5 810.48	1 232
工业	7	7.00	759.52	1 139.40	400
生物信息	3	3.00	1 112.30	1 724.88	356
交通运输	3	3.00	223.18	365.90	160
大气气象	2	2.00	155.50	265.60	50
邮政	2	2.00	142.2	224.0	64
研究与开发	1	1.00	196.90	307.20	64
政府	1	1.00	71.10	112.00	32
总计	100	100.00	19 675.99	33 975.02	9 906

1) 主要软件依赖进口，每年花费人民币数亿元。该类软件的针对性不强，不能完全满足我国石油勘探开发工作的需要；

2) 国内石油软件研发投入不断加强，已逐步应用到油藏模拟、石油勘探等领域。

2.2.2 气象气候应用现状

1) 从银河 II 开始一直走移植加改造的道路，

气象模式从 T63, T106 到 T213，已在国产高端计算机上建立了业务运行系统；

2) 应用软件很难从国外得到完全的版本，通常不能跨平台使用，可移植性差；

3) 已起步国产化软件的开发，建立统一区域气象模式和中国气候系统模式。

2.2.3 生物工程生命科学领域应用现状

1) 近几年得到快速发展,已使用峰值性能达万亿次/秒量级的机器,主要应用于分子动力学模拟、基因比对和药物筛选等;

2) 所有应用软件来自国外,需要移植改造。

### 3 神威系列高性能计算机发展

#### 3.1 国家并行计算机工程技术研究中心概况

国家并行计算机工程技术研究中心(以下简称中心)主要从事高性能计算机系统的研发及推广工作。在多处处理机系统、并行处理技术、系统软件、超精细多层印制电路板设计与生产、高密度电子组装技术等诸多方面,中心有很强的技术开发和工艺制造实力,同时中心还有一支实力很强的应用开发队伍,可为用户提供应用软件的并行开发移植、并行算法性能优化等全方位技术服务。

国家并行中心自成立以来取得了诸多成就:1999年10月,中心研制的第一台神威I型机进入国家气象局运行;2001年1月,中心研制的第二台神威I型机进入上海超级计算中心运行;中心还拥有P系列、A系列、I系列、O系列4个系列的集群系统;先后在不同单位安装近30台套神威“新世纪”集群系统。

#### 3.2 神威系列高性能计算机技术特点

神威系列高性能计算机的主要技术特点包括:高可扩展性和高效性技术;高可靠性和高可用性技术;友善环境技术;总体工程设计与制造技术。

##### 3.2.1 高可扩展性和高效性技术

1) 研制开发了高速互连网络:高带宽,低延迟,低阻塞;全定制,工业标准(IBA),硬件支持消息-旁路OS,全局消息,多层次深度流水;

2) 低延迟存储结构:NUMA, NEAR-SMP结构,新型目录协议,OpenMP支持全局共享;

3) 系统软件可扩展性:多层次分布并行操作系统(可扩展OS体系结构,亲缘性资源分配算法等),并行文件系统(支持MPI-2);

4) 并行编译:静态/动态优化措施(延迟分布、页面分布、页面迁移、反馈式编译等);

5) 采用高密度组装,系统物理规模大大缩小。

##### 3.2.2 高可靠和高可用技术

1) 降低CPU和其他芯片结温:多种冷却措施;

2) 校错、纠错、冗余、作业分区的隔离与保护;

3) 系统级容错体系,系统自愈技术,故障恢复技术,容错技术,保留恢复功能(全透明的checkpoint)。

##### 3.2.3 友善环境技术

1) 大规模并行系统的单一映像;

2) 多种编译优化技术,实现最佳程序运行;

3) 大规模并行调试、性能监测和自动性能分析;

4) 多种并行支撑库,高效编写/运行程序;

5) 面向问题的并行开发环境提升了系统好用性;

6) 分布式环境下的大规模并行科学计算可视化。

##### 3.2.4 总体工程设计与制造技术

1) 深亚微米工艺专用集成电路物理设计技术;

2) 多层印制板设计与制造;

3) 高密度电子组装工艺;

4) 大功率器件、部件散热及冷却技术。

#### 3.3 神威系列高性能计算机应用情况

神威系列机在我国航天、气象、石油、生物、国防和核工业等领域均取得了一批具有重大价值的应用成果,如神舟飞船返回舱飞行过程中其周围空气流动情况的模拟(见封三彩图1),轴对称射流流场的直接数值模拟(见封三彩图2),气象模拟(见封三彩图3),药物筛选试验(见封三彩图4),基因序列对比(见封三彩图5),石油勘探模拟(见封三彩图6)等,描绘了神威系列机在相关应用领域所取得的丰硕成果。

### 4 高端计算发展面临的问题

长期以来,高性能计算界一直以“峰值运算速度”作为超级计算机技术水平的评价标准。为实现足够高的峰值性能,业界不断推出物理上可扩展性较好的、“堆积”了越来越多通用微处理器的MPP系统。但是,这些MPP系统的实际应用性能与其峰值速度往往相差甚远,而且系统规模也越来越大,耗电越来越多,编程困难,应用效率低。20世纪90年代成熟并流行的HPC技术,如今已遇到音障,归纳起来主要有以下问题。

#### 4.1 实际应用的持续性能低

TOP500 采用的 Linpack 标准以规则的计算密集型模型为背景，反映的是经过软件专家“精心雕琢”的以 CPU 为核心的硬件系统性能。当前，绝大多数高端计算系统实际使用的持续性能只有其峰值性能的 5%~10%。造成这一现象的主要原因包括：

1) 存储器墙 (memory wall) 问题——DRAM 性能每年增加 7%，而微处理器性能每年增加 60%，它们之间的性能差每年扩大 50%；

2) 主机性能与外围性能不匹配——主机性能每年增加 73% (4 年增加 10 倍)，而外围系统 (外存储系统) 性能每年只增加 13% (9 年增加 3 倍)；

3) 缺少高效并行应用软件开发环境，高端计算应用系统开发周期长；

4) 系统资源难以有效调度以获得负载均衡；

5) 应用问题的物理模型及并行算法与现有并行系统体系结构的不适应。

#### 4.2 复杂性挑战

摩尔定律只解决计算能力问题，未解决复杂性问题。当前，高端计算系统所面临的复杂性挑战包括：

1) 结构的复杂性，大规模，多层次；

2) 使用的复杂性，缺乏统一的应用接口，移植难，互操作难，用户更多精力放在解决可计算性问题，而不是应用问题本身；

3) 管理的复杂性，资源的配置、优化、管理等。

#### 4.3 可靠性与可用性问题

目前，采用堆积商用微处理器构造的 MPP 系统当规模进一步扩大时，全系统的平均故障间隔时间 (MTBF) 明显地下降。

#### 4.4 功耗问题

在摩尔定律的推动下，通用高性能微处理器的速度提高很快，但其功耗也越来越大。据 ITRS 估计，到 2009 年 COMS 高性能 CPU 的功耗将达到 120~200 W。由此导致了随着传统高性能计算机的堆积规模越来越大，其总功耗的增加幅度逐渐逼近人们可承受的极限。当前，10 TF/s 量级超级计算机系统的功耗都已超过了 1 MW，相当于一个中小城市的用电总量。功耗过高会影响到整机系统的稳定性和可靠性，同时如此高的用电成本也严重地制约了高端计算系统的普及性。

#### 4.5 应用研究滞后

高端计算应用水平低，应用的研究与开发滞后于高端计算机的发展：

1) 应用的并行度普遍在百十量级，应用到更大规模的很少 (并非没有需求)；

2) 高端应用软件的开发、高效并行算法的研究尚不能与高端计算机发展同步；

3) 应用投入不够，应用研发力量薄弱且分散，企业基本未介入；

4) 应用范围和并行计算技术普及与研究尚有很大的潜力；

5) 缺乏跨学科的综合型人才。

这种严重的不平衡已经制约了高端计算的发展。

### 5 发展我国高端计算的几点建议

美国和日本政府历来将高端计算技术看作是 21 世纪经济技术领域占据世界领先地位的关键因素，双方都把高端计算技术列入科技发展长远规划，投入大量的人力和物力，加强研究开发高端计算系统及相关的应用技术。针对国内高端计算的发展现状及国际高端计算的未来走势，特提出以下几点发展我国高端计算事业的宏观建议。

1) 政府投入，政策倾斜 高端计算技术是 21 世纪科技创新的重要工具，是衡量一个国家综合技术实力的重要标志。因此，政府部门应把高端计算技术列入科技发展长远规划；加强研究我国的高端计算发展战略；结合重大应用项目，设立国家专项计划，研制高端计算机及应用系统；实行政策倾斜，鼓励使用国产高端计算机系统。

2) 企业介入，产学研结合 国际高端计算的发展实践证明，振兴我国的高端计算事业必须要有企业介入，特别是希望有一批知名企业介入行业应用软件平台的研发；有必要成立行业协会，加强高端计算机研发单位 (包括院校、研究所等) 与生产厂家及应用单位之间的合作，规范行业行为。

3) 培育应用群体，重视应用软件的研发 高端计算事业的发展离不开高水平的应用群体，因此必须加强技术培训和交流，普及并行计算基础知识，拓展用户应用水平；以应用为牵引，加强与其他应用学科的共同攻关，研制国民经济和国防建设急需的高端应用系统平台，特别强调与高端计算机同步研发。



4) 抓住机遇, 自主创新 目前, 国内高端应用市场逐渐成熟, 国际高端计算的发展也出现了新的热潮。我国应抓住这一良好的机遇, 自主创新, 打破国外封锁与垄断, 坚持“提速”技术是基础、“服务”技术是关键、“高效能”技术是目标的设计理念, 加强对创新性体系结构、器部件技术和软件系统的研究, 力争研制出满足用户需求的国产高端计算系统, 使我国的高端计算技术赶超世界先进水平, 在世界高端计算领域占有一席之地。

#### 参考文献

- [ 1 ] TOP500 List for June 2004[EB/OL]. <http://www.top500.org/lists/>, 2004-06
- [ 2 ] 中国软件行业协会数学软件分会. 中国 TOP100 制造商分析[EB/OL]. <http://www.samss.org.cn/>, 2003-11
- [ 3 ] 中国软件行业协会数学软件分会. 中国 TOP100 行业领域分析[EB/OL]. <http://www.samss.org.cn/>, 2003-11

## High End Computing in China and the Exploration of “Sunway” Computers

Chen Zuoning

(*National Research Centre of Parallel Computer Engineering & Technology, Beijing 100080, China*)

**[Abstract]** In the beginning of the 21st century, driven by application requirements, the research and development of HEC are getting a new upsurge both in China and abroad. This paper analyses the development status and trends of international HEC, describes the overall development and application status of HEC in China, introduces the major technical features and application achievements of Sunway series high-performance computers; discusses the problems faced by today's HEC technologies, and finally makes some suggestions on how to extend progress in the HEC industry in China.

**[Key words]** high-end computing; grid computing; high performance service; high productivity computing; total cost of ownership; Sunway series computers