高性能计算机研究的现状与展望

樊建平　陈明宇

摘要　高性能计算机的研制受永无止境的探求复杂的物理世界与人类社会本身的应用计算需求的驱动及研制者所处环境（人员、经验、经费等）及当时的可选择的实现使能技术的影响。回顾历史，任何时刻研制的最高性能的计算机总是服务于当时的科学计算的需求（材料模型、药物设计、气候模拟、核武器模拟、电磁学等）或者称是以科学计算为最初应用的靶子进行设计的（当前最快的日本 Earth Simulation 与 IBM BLUE/Gene 两个项目是很好的例子），而高性能计算机使用与发明的技术逐步向商用领域转移与转化（SMP、MPP、Cluster 等），计算性能（当前的设计目标是 Petaflops）及与其相匹配的存储、带宽等指标成为高性能计算机设计者追求的持续性关键指标。高性能计算机的实现使能技术包括计算数学（计算模型与算法）、计算机体系结构与部件构成技术三部分，为保持每十年性能增加 700-1000 倍左右的速度（远高于摩尔定律单芯片的发展速度）及高性能计算能力 70%的年增长需求，高性能计算机设计师仅仅考虑体系结构与部件构成两部分已不能满足现实的需求，对计算数学有相当的了解已成为必然。本文以性能为叙述主线，介绍高性能计算机研制的历史、现状与未来展望。

1　高性能计算机研制的历史回顾

1.1 1950-2002 历史的简单回顾

电子计算机在诞生之初主要就是为科学计算服务的。到 1960 年代，随着技术的成熟，计算机开始走向各种商业领域的应用，并且应用范围越来越广泛。因此，为了有别于“通用计算机”，专门针对科学计算进行优化设计的计算机开始被称为“高性能计算机”，或简称 HPC。

可以把 1970 年代出现的向量计算机看作是第一代的高性能计算机。通过在计算机中加入向量流水部件，可以大大提高科学计算中向量运算的速度，其中比较著名的有 CDC 系列、CRAY 系列、NEC 的 SX 系列向量机。中国有代表性的是银河一号及中科院计算所的 757 计算机。

80 年代初期，随着 VLSI 技术和微处理器的技术的发展，向量机一统天下的格局逐渐被打破。通过多个廉价的微处理器构建的并行化超级计算机首先从成本上具有了无可比拟的优势。“性能/价格比”而非单一性能成为衡量高性能计算机系统的重要指标。按照摩尔定律速度发展的微处理器的性能快速超越传统向量机。1990 年代初期，大规模并行处理（MPP）系统已经开始成为高性能计算机发展的主流。

MPP 主要由多个微处理器通过高速互联网络构成，每个处理器之间通过消息传递的方式进行通讯和协调。比较有代表性的有 TMC 的 CM-5, Intel Paragon 等。中国的第一个 MPP 系统是计算所国家智能机中心的曙光 1000 计算机。

较 MPP 早几年问世的对称多处理机 SMP 系统，是由数目相对较少的微处理器共享物理

内存和 I/O 总线形成的计算机系统（国内最早基于微处理器的 SMP 为曙光 1 号）。和 MPP 相比，早期的 SMP 扩展能力有限，并不具有很强的计算能力。但由于 SMP 与单机系统兼容性好，是单机系统的升级与增强，被广泛应用于商业计算领域。

1990 年代中后期的一种趋势是将 SMP 的优点和 MPP 的扩展能力结合起来，这一趋势发展成后来的 CC-NUMA 结构，即分布式共享内存。每个处理器节点都可以访问到所有其它节点的内存，但访问远程内存需要的延迟相对较大。代表性的系统有 Sequent NUMA-Q, SGI-Cray Origin 等,国内的神威与银河系列等。CC-NUMA 本身没有在提高性能的角度上进行较大的创新，主要优点是便于程序的开发和与 SMP 的兼容性。而对科学计算任务 CC-NUMA 结构是否优于 MPP 系统仍存在争议。

在发展 CC-NUMA 同时，机群系统（Cluster）也迅速发展起来。类似 MPP 结构，机群系统是由多个微处理器构成的计算机节点通过高速网络互连而成。节点一般是可以单独运行的商品化计算机。由于规模经济成本低的原因，机群系统具有比 MPP 更高的性能/价格比优势。机群系统还继承 MPP 系统的编程模型，更进一步加强其竞争优势。代表性的系统是 IBM SP2，国内有曙光 3000，4000 等系列。到 2000 年初机群实际上已经构成了高性能计算机系统的主流。据 2003 年的统计，TOP500 中的 MPP(含 CC-NUMA)占 42%,Cluster 占 29.8%。

Rank	Manufacturer Computer / Procs	R_{max} R_{peak}	Installation Site Country / Year
1	**NEC** Earth-Simulator/ 5120	**35860.00** 40960.00	**Earth Simulator Center** Japan/2002
2	**Hewlett-Packard** ASCI Q - AlphaServer SC ES45/1.25 GHz/ 8192	**13880.00** 20480.00	**Los Alamos National Laboratory** USA/2002
3	**Linux Networx** MCR Linux Cluster Xeon 2.4 GHz - Quadrics/ 2304	**7634.00** 11060.00	**Lawrence Livermore National Laboratory** USA/2002
4	**IBM** ASCI White, SP Power3 375 MHz/ 8192	**7304.00** 12288.00	**Lawrence Livermore National Laboratory** USA/2000
5	**IBM** SP Power3 375 MHz 16 way/ 6656	**7304.00** 9984.00	**NERSC/LBNL** USA/2002
6	**IBM** xSeries Cluster Xeon 2.4 GHz - Quadrics/ 1920	**6586.00** 9216.00	**Lawrence Livermore National Laboratory** USA/2003
7	**Fujitsu** PRIMEPOWER HPC2500 (1.3 GHz)/ 2304	**5406.00** 11980.00	**National Aerospace Laboratory of Japan** Japan/2002
8	**Hewlett-Packard** rx2600 Itanium2 1 GHz Cluster – Quadrics/ 1540	**4881.00** 6160.00	**Pacific Northwest National Laboratory** USA/2003

《图一：2003 年 6 月 TOP500 List 前 8 位》

MPP 取代向量机和机群逐步替代 MPP 这两个进程的背后都是摩尔定律在起作用。高性能计算机体系结构的创新必须与半导体技术和产业发展相结合，否则很难变成主流技术，这也是 SIMD 系统、阵列机、数据流等新型体系结构没有流行起来的主要原因。

1.2 高性能计算机面临的主要问题

For every 1 gigaflop peak performance, we need

Capability (Flops)	1990 (10^9)	2000 (10^{13})	2020 (10^{17})	2050 (10^{23})
1 GB memory size	10^9	10^{13} (10^{12})	10^{17}	10^{23}
50 GB disk storage	$5x10^{10}$	$5x10^{14}$ (10^{14})	$5x10^{18}$	$5x10^{24}$
10 TB archival storage	10^{13}	10^{17} (10^{14})	10^{21}	10^{27}
16 GB/s cache bandwidth	$1.6x10^{10}$	$1.6x10^{14}$ (10^{11})	$1.6x10^{18}$	$1.6x10^{24}$
3 GB/s memory bandwidth	$3x10^9$	$3x10^{13}$ (10^{10})	$3x10^{17}$	$3x10^{23}$
0.1 GB/s I/O bandwidth	10^8	10^{12} (10^{10})	10^{16}	10^{22}
0.01 GB/s disk bandwidth	10^7	10^{11} (10^9)	10^{15}	10^{21}
1 MB/s archival storage band.	10^6	10^{10} (10^8)	10^{14}	10^{20}

Where 10^6 = mega, 10^9 = giga, 10^{12} = tera, 10^{15} = peta, 10^{18} = exa, 10^{21} = zetta, and 10^{24} = yotta

(蓝色数字为实际系统达到的量级)

《图二：应用期望的高性能计算机系统相关指标》

通用高性能计算机系统必须在计算能力、存储能力、带宽、延迟及 I/O（输出/输入）等五个方面均衡发展，才能保证实际应用系统性能的提升。从 1976 年 Cray-1 计算机以后的系统设计开始失去平衡，特别表现在 CPU（中央处理器）与 Memory（内存）之间数据存取的时间延迟，cache （缓存）带宽、内存带宽、I/O 带宽等方面发展的滞后（详见图一）。同时为增强系统的可靠性、可用性、可维护性及性能价格比等特征，高性能计算机的设计师需要对电耗性、冷却系统、机器体积与运行环境、造价等综合考虑。

目前基于商用 SMP 节点构造机群系统的主要问题包括：SMP 节点本身并非为构造大规模并行处理系统而设计、连接 SMP 节点的高效互连网络价格高、高耗电（单位电耗产生的计算性能低）、解决存储部件之间（cache 之间、cache 与内存之间、内存与外存、内存与通讯接口之间）的延迟花费大量的资源。在追求性能/$(单位资金)、性能/单位空间、性能/单位电耗的比赛中越来越吃力。要实现 Petaflops 以上的 HPC 仅仅通过扩展目前系统的规模很难达到，仍然需要在体系结构上有所突破。

1.3 设计高性能计算机需要考虑的三个层次（作者称为 ASC Model —Application System and Components 模型）

高性能计算机的设计应从构成部件技术、系统结构与应用系统三个层面来考虑。回顾历史，1950（Univac-1）—1976(Cray-1) 设计师的注意力主要集中于构成部件技术的创新与优化（单机 RAM 模型）；1982(Cray X-MP)—2002(Earth Simulator) 设计师的注意力集中于体

系结构层面，各种并行体系结构（SMP, MPP, Cluster）与学术界在 1998-2000 年鼓吹的“Connection （连接）是一切”反映了这一时期的特征；2003（IBM Blue Gene）以后，高性能计算机的设计开始重视应用系统本身的特点，从应用角度出发来选择与改进体系结构与构成部件技术。这一趋势发展的基础是构成部件技术发展迅速（商用 CPU 与商用节点机、FPGA、SOC、光互连技术等）、体系结构的研究多样化（多层次并行、网格、可重构等）。目前我们非常缺乏对应用系统的系统研究，性能测试(各种 Benchmark）只能对已完成机器的性能进行测试，对机器本身的设计（如：选择体系结构与部件技术）帮助不大。我们已有各类的并行程序设计模型（PRAM, LogP 等）帮助并行算法设计师建模与进行算法分析，机器设计师也需要相关模型（我们称为 ASC 模型）帮助其从应用系统角度指导体系结构及部件的选择与创新。例如当机器设计师构造 100 TFLOPS 的计算机时，假设有以下三种选择：光互连+商业 PC 主板、FPGA 定制芯片+3D mesh、高性能 SOC+2D Mesh，根据 ASC 模型应该可以定量分析每一个方案并给出选择的理由。

2 现状

目前，世界上最快的超级计算机 Earth Simulator 的实际计算速度是 35T flops，即 35 万亿次。对高性能计算机研究的下一个挑战是 1Petaflops，即千万亿次计算。

预计 Petaflops 计算机将由 10000~1000000 个处理器、10TB~1PB 主存、1PB~100PB 在线存储、100PB 以上离线存储构成。第一个系统实现将在 2010 年前完成。如何达到千万亿级，是继续依靠摩尔定律的发展，还是在体系结构或者构件技术上找到新的突破，现在还是一个悬念。

2.1 高性能计算机体系结构的研究

目前高性能计算机体系结构的主流仍然是 MPP 和机群技术的进一步发展，通过将更多的处理器连接起来构建更大规模的并行系统。其中最具代表性的就是美国能源部的 ASCI 计划，分别由 Intel、SGI、IBM、HP 等构建超大规模的机群系统，其中 HP ASCI Q 共有 8192 个处理器，20Tflops 的峰值。ASCI 计划原计划 2004 年达到 100Tflops。

日本 NEC 的 Earth Simulator 结合了向量处理技术和 MPP 的技术，利用带向量部件的节点构建 MPP 系统，不但取得了 Top500 峰值第一位，而且实际应用运行效率也比较高。美国的 Cray X1 系列也采用了类似的结构。

IBM 计划在 2000 年~2005 年，每年花费 1 亿美圆研究经费，以便最终建造出用于生物计算的 petaflops 级机器。2002 年该项目中的 Bluegene/L 结构设计已经确定，预计峰值计算速度 360Tflops。Bluegene/L 的设计中没有采用传统的高功耗的高端处理器，而是采用了低功耗的 SOC 芯片。IBM 将这种技术称为 cellular 结构。虽然每个处理器性能并不很高，但是可以通过增加系统中的处理器数量来达到高的峰值计算能力。Bluegene/L 共有 65536 个节点，计划中的 BlueGene/C 处理器个数可能达到 100 万个。

美国 Stanford 大学的 Streaming SuperComputer 计划，尝试采用专门设计的 streaming 处理器来构建超级计算机。一个 Streaming 节点中集成了 128 个 1GHz 的 FPU，一个背板支持 32 个节点，32 个背板就可以达到 PetaFlops，而预计成本只有 4 千万美元。Streaming 技术最初的思想来源于专用游戏机的设计中。现有科学计算应用是否能有效移植还有待研究。

美国 NASA 支持的 HTMT(Hybrid Technology Multi-threaded)采用了另一条不同的路线。HTMT 试图避开摩尔定律，采用了超导逻辑、光交换、全息存储，PIM 等等全新的技术，其核心处理单元 SPELL 的频率可达 100GHz，而主要设计挑战是能够满足这样高速处理器的存储体系。HTMT 原计划在 2006 年左右达到 1Pflops，但此计划研制费用极高，技术风险大，因此很有可能再次让位于摩尔定律。

网格技术有可能成为实现 PetaFlops 的另一条途径。网格是近年来计算机体系结构发展的一个重要方向，其基本思想是通过 Internet 进行资源共享和协同工作。目前连接到 Internet 的计算机已经达到 1 亿台以上，通过互联网可能达到的聚合计算潜力是不可估量的。国际上已经有 Globus 等组织为网格环境制定标准和参考实现。但是用网格技术实现 petaflops 仍需要关键技术上的突破：一方面互联网连接的速度和带宽仍有待提高，近年网络通信技术以超摩尔定律的速度高速增长已经为此提供了可能，达到实用阶段只是时间问题。另一方面是有效的网格体系模型和计算模型还没有建立。网格的资源是分散和动态的，计算也是一种分散的、动态的过程，传统的并行共享内存或消息传递程序模式不能直接有效的利用。如何使科学计算高效使用网格的计算能力是当前一个主要研究方向。

2.2 增强高性能计算机功能与特征的研究

Berkeley 的 ROC(Recovery Oriented Computing)项目中提出未来峰值不是高性能计算机面临的主要问题，相反，如何将硬件、软件故障，包括人为失误考虑在内，真正提高系统的可用性是一个挑战。为此 ROC 项目研究了一系列通过硬件和软件的进行故障监测、故障屏蔽、故障注入、故障恢复等技术方法。

USC 的 PAMA (Power Aware MultiProcessor Architecture) 则关注高性能计算中的功耗问题，其开发的实验平台可以监测到系统中实际的功率消耗，并通过结合应用程序动态调整处理器的功率(可在 10^2 的范围内内调整)，从而达到减少总功耗的目的。

美国的 LANL 实验室在高密度计算研究项目中，设计了一个可以在 1 立方米放下 240 个处理器的 beowulf 机群系统。其主要技术是刀片式 (blade) 结构，通过简化处理器主板的设计，可以在更少的空间内放入更多的处理器，从而达到高的性能/空间比。

Processor in Memory（PIM）也是近年来研究比较多的一个方向。其基本思想是一方面处理器主频提高和内存访问速度之间的差距不断增大，另一方面芯片内计算逻辑相比于存储占用的晶体管面积只有很小一部分，因此将部分处理功能集成到 Memory 中，可以提高存储器的利用效率，同时增加并行处理的能力。这方面的研究有 IRAM、Imagine、FlexRAM、DIVA 等项目。

MIT 的 RAW 项目与 PIM 的思想有些相通，通过在一个芯片中加入多个嵌入式处理器和互联网络，来更有效利用芯片内部的功能单元，并且可以通过动态调整改变功能单元、通道和输出管脚的分配和联结，最大限度发挥单位面积硅的计算能力。有观点认为 RAW 将是未来处理器芯片的主要模式。

与 RAW 类似的是可重构计算的研究。通过利用 FPGA 等复杂的现场可编程逻辑器件，可以根据应用的特点动态改变芯片的内部结构，从而得到较高的性能。通常把这种可重构的单元称为 RPU。 RPU 的有效性在很多实际应用中得到验证。一些研究项目如 DISC、MATRIX、BRASS 等集中于探讨如何结合 CPU 和 RPU 功能的新型体系结构。随着芯片集成度的进一步提高，可编程逻辑器件计算的能力也将不断加强，可重构计算有可能最终打破原有高性能计算中硬件/软件的分界线。

随着 ASIC 和复杂可编程逻辑器件技术的普及，专用计算机(special purpose computer)的研制也逐渐在高性能计算领域占据重要的地位。如日本 RIKEN 高性能计算中心研制的分子动力学模拟专用计算机 MD-GRAPE 系列的峰值速度甚至超过同时代最快的通用计算机，其中的 MDM 在 2001 年就达到 78Tetaflops 的峰值。而研制中的"Protein Explorer"很可能会成为世界上第一个 Petaflops 的系统。

2.3 高性能计算机构成器件的研究

微处理器仍是高性能计算机的核心技术。目前的微处理器技术已经开始向单芯片多核心（如 IBM Power4）和单芯片多线程（如 Intel Pentium4）以及 SoC 等方向发展。单个芯片的处理能力还会进一步提高。SONY 公司的计划中 PS3 单芯片到 2005 年将达到 1Tflops 的处理能力。但由于 10 年之内摩尔定律继续有效，芯片集成度和频率每 18 月翻番，导致芯片的功耗问题日益显著。最高端的微处理器功耗可达几十甚至上百瓦，使得系统散热成为不可忽视的问题。

芯片之间的互联也因为信号频率的提高成为一个突出的问题。使用铜线连接的脉冲信号传输受寄生电阻、电容、电感的影响，而且频率越高这种影响越大。尽管当前主流 CPU 主频已经达到 2Ghz，但板级的并行总线互联仍限制在 800Mhz 以内。

光互联有可能成为最终的互联解决方案。相比于铜线连接光互联具有高带宽、长距离、低损耗等特点。而基于表面发射激光的 VCSEL 技术已经成功的将半导体技术和激光技术结合起来。主机之间的光互联已经广泛应用，主板之间光互联也发展成熟，基于光波导的板内互联技术也在实验之中。

光互联应用到高性能计算机的主要问题是成本，这是因为 VCSEL 技术还只能用于 GaAs 等半导体工艺，不能和 CMOS 直接结合。一旦这方面技术取得突破，光互联必将进入计算机系统的内部。

全光交换技术近年来在骨干通信网络中已经开始采用。传统的集中式电路交换技术受电信号之间交叉干扰和电信号频率的限制，其进一步大容量扩展受到限制，目前主流技术在 Tb/s 的量级已经很难提高。而光传输没有串扰和带宽限制，因此全光交换的潜力远远超过电交换的极限。目前已经有基于微机电系统(MEMS)技术的 1000 路自由空间光交换实验系统。全光交换系统的发展将进一步为高性能计算机的动态系统互连提供支持。

随着网络技术的发展，网络化的器件也成为高性能计算机的一部分。突出的例子就是网络存储的发展。一个计算机系统中不需要有专用的存储设备，只要拥有网络接口，就可以通过网络访问远程的共享存储服务。网络存储服务把一类功能相同的器件集中管理起来并通过

网络对外提供服务。这是一种网格化的方式。构成计算机系统的其他器件如 CPU、Memory 等是否也可以采用同样的方式分解、集中和重组，是当前研究的一个新的方向。

3 展望

3.1 机群系统的应用面扩大、形成产业规模市场

基于 Linux 的机群系统在下五年的使用面将进一步扩大，高性能计算机产业前景更加光明，个人用高性能计算机时代将出现。以 Beowulf PC Linux 机群为标志，高性能计算机的门槛现在已经不再是高不可攀了。高性能计算机的普及也将使其应用面进一步扩大。除了从前的国家级战略单位以外，产业界和地方部门也可能逐步利用高性能计算机，而随着网格技术的发展和应用软件的进一步丰富，可以预见不远的将来会出现个人高性能计算的时代。由此高性能计算的战略意义和产业前景更加重要。

3.2 网格不仅影响各种应用，对计算技术也将产生巨大影响

网格作为下一代 Internet 的应用不仅影响最终用户，同时对其它技术的发展产生巨大的影响，高性能计算机领域将首当其冲。网格作为下一代 Internet 应用，其特征是以资源共享为目标，同类资源集中，异种资源分离，资源的调用服务化，资源的使用按需分配。对于高性能计算机系统来说，资源的网格化是一种使能技术，为更大尺度的高性能计算机系统的设计提供了支持。但是不能简单的将网格和未来的高性能计算机系统等同起来。作为一种共享技术，网格只是使现有的高性能计算资源更好的被共享使用，而并不能增加高性能计算资源。我国的高科技发展计划一度用网格发展专项取代了新一代高性能计算机的研制，是走入了一个误区。

网格化技术真正对高性能计算机的推动是提供了一种动态的、网格化的资源作为高性能计算机的新的组成部件，如网络 CPU、网络 RAM、网络磁盘等等。网格将不仅仅是“计算机通过网络连接起来”，而是成为真正意义上的“网络连接起来的计算机”。

如何利用这些部件构造更高性能的计算机系统仍然需要更多的体系结构的研究。中科院计算所智能中心提出的“Dagger (De-coupled Architecture with Grid-key and Grid Enabling Resource) 网格计算机体系结构”并应用于曙光 4000 系列计算机系统设计中，代表了对这一趋势的积极探讨。

3.3 光电结合是未来计算机制造技术的发展趋势

光电技术的结合是下二十年计算机制造技术的发展趋势，PCB（印刷电路板） 板间光互连将成为未来高性能计算机的标准互连方式。如果按照摩尔定律继续发展，下二十年后，半导体技术将进入 THz 时代。而 THz 是电信号和光信号交叉之处。可以预见光电技术的结合将是未来物理学和技术发展的前沿。据估计，未来 3 年内，光底板内互联将发展成熟，而 5-10 年内，芯片之间的自由空间互联将成为可能，15-20 年内芯片内部也有可能采用光互联技术。

计算所智能中心从事的“网格化动态自组织体系结构 DSAG 及光互连高性能计算机的实现”项目（HPC-OG）结合光互连技术、网格技术及可重构计算，通过功能部件的分离和重组构造具有动态性、时效性、按需资源分配和共享、高性能与高可靠性的超级计算机系统。

是一种集成创新的尝试。

3.4 并行体系结构向多层次、多粒度方向发展

并行计算机体系结构向多层次、多粒度方向发展，使能技术多样化和 SOC/FPGA 可能带来较大发展机遇。为解决大规模系统并行（连接上万个 CPU）及处理器与内存之间的存取速度慢（包括 CPU 与 CACHE 之间）的技术挑战，采用多层并行体系结构成为高性能计算机设计师的选择。应用程序的程序设计模式同时要支持多粒度的并行模式（细、中、大粒度）。构造高性能计算机的部件已有较多的选择，除直接用传统的 PC 及服务器节点外，由各类 SOC、FPGA、DSP 芯片制造的主板在处理某一类科学问题时，其性能价格比较传统 PC 节点提高几倍或十几倍。基于新型节点构成的万亿次计算机以低于 10 万美圆价格销售的日子可能很快到来。

3.5 高性能计算机硬件发展逐步面向应用

减少用户使用高性能计算机复杂性的努力将有回报，体系结构创新与自动并行程序设计工具的发展可能是关键因素。针对高性能计算机体系结构来设计高性能算法依然是应用科学家今天必须面临的问题。对体系结构及系统软件详细的了解与理解是写出高效程序的关键，也有用户为提高应用程序效率自己重新开发操作系统的事例。如何摆脱应用系统随一代一代机器的研制而重复设计一遍的局面成为计算机设计师追求的目标和研究的方向之一。目前体系结构的可重构计算（如计算所从事的网格化动态自组织体系结构及基于光互联的实现(HPC-OG)）以及并行编译的识辨与串行程序并行化辅助工具的发展有力推动这一趋势的形成。

4 机遇与挑战

高性能计算机的研制已走入发展的十字路口。美国 Illinois 大学计算机系 Daniel A. Reed 教授曾指出：“在美国目前还没有积极的大规模体系结构和原型研究项目。简单的说，我们目前正面临着体系结构的危机，包括软件和硬件。”这一十字路口很像 1990 年国内有关使用微处理器构成并行系统还是用大规模集成电路自己设计向量计算机的选择。较 1990 年更困难的是当时并行机在国外已有商品化的机器供我们发展参考。今天我们与美国人处于同一十字路口，我们的技术选择及产业化方面的努力有可能对国际高性能计算机的技术走向产生较大的影响，从长远看还可能影响到低端的服务器市场。“十年磨一剑”，我们应庆幸十年一遇的创新机会，争取在技术与产业化两方面都有所突破。

2002 年日本 Earth Simulator 系统取代美国获得 Top 500 第一位之后，已经促使美国认识到传统的机群/MPP 结构并不是 HPC 体系结构的终结。美国 DARPA 在 2002 年启动了 HPCS(High Productivity Computing Systems)计划，其主要目标是为了“填补当前基于 80 年代后期技术的 HPC 和未来的量子计算技术之间的高端计算”，并强调性能(Performance)、可编程性（Programmability)、可移植性（Portability）和稳固性（Robustness)。整个计划分三个阶段，第一阶段为概念评估阶段，第二阶段到 2006 年为系统和关键技术评估阶段，第三阶段到 2010 年为研发和系统实现阶段。到 2003 年 7 月该计划的第一阶段已经基本完成，Cray、IBM、Sun 正式入选第二阶段。

机遇和挑战同时存在，未来五到十年也将是中国高性能计算机技术和产业发展至关重要

的时期。我国在高性能计算机方面的研究与产业化已有相当的基础，有进行重大技术创新的条件。我国高性能计算机的市场已进入高速发展期，需求牵引将逐渐表现出对技术创新的拉动作用。中科院计算所、国防科大及江南所已有相当的技术储备与人力资源。曙光公司、联想集团、浪潮集团已建立有一定规模的产业化基础。国家设立新的高性能计算机发展专项不仅可行与必要，同时时机也已经成熟。高性能计算机已进入国际的新一轮竞争，目前处于各种新思想与新方法产生的活跃期，学术界争论很大。企业界在沿用过去学术成果不断推出低成本的 Cluster 系统的同时，不断参与尝试用新的构成部件建立的并行计算系统，同时对目前系统的可用性、耗电性、可管理性等进行持续性的改进与改良。设立高性能计算机专项，抓住创新期，从计算模型与算法、部件技术与体系结构三个层次及其相互联系研究新一代的高性能计算机系统，其收获与意义将巨大。

作者：樊建平 中国科学院计算技术研究所 副所长、研究员、博导

陈明宇 中国科学院计算技术研究所 副研究员、博士