

# 应当重视高性能计算的应用和基础性研究

桂文庄  
中国科学院

## Should Pay More Attention to Applied and Fundamental Study for High-Performance Computing

Gui Wenzhuang  
Chinese Academy of Sciences

2008年11月1日，在江苏无锡召开的中国2008高性能计算学会年会上，公布了2008年中国高性能计算机性能TOP100排行榜<sup>[1]</sup>。上榜的第一名是由中国科学院计算技术研究所和曙光公司研制的曙光5000A，它采用了30720个AMD四核1.9MHz处理器，峰值速度达到233.5万亿次/秒，LINPACK值达到180.6万亿次/秒，是我国首台LINPACK值超过百万亿次的超级计算机；排名第二的是联想公司研制的深腾7000，它采用了12216个Intel Xeon四核3.0MHz处理器，峰值速度为146.0万亿次/秒，LINPACK值为102.8万亿次/秒<sup>[2]</sup>。这两

台计算机分别安装在上海超级计算中心和中国科学院计算机网络信息中心，在11月17日公布的世界超级计算机TOP500<sup>[3]</sup>中分别位于第10名和第19名。它们都是在国家863计划支持下研制的，其投入运行必将对我国高性能计算的应用起到重大推动作用。据悉，我国正在着手千万亿次计算机系统的研制，可能将于2010-2011年问世。

我国高性能计算机速度的增长令世人瞩目，但是从应用的角度来看，我们与国际先进水平相差还是很大的。根据最新公布的数据，我们得到了下列两个统计表。表1是世界TOP500计算机和我国10万

亿次/秒以上计算机应用领域的比较。统计表明，我国用于科学计算研究的计算机比例偏少，世界TOP500计算机中37%安装在国家实验室、大学、公共超级计算中心用于科学计算，并且几乎囊括了最强大的计算机系统，而我国用于科学计算的高端计算机不到20%；另一情况是国际上金融业用了58台TOP500计算机，70台用于工业制造，包括航天航空、汽车、建筑、医药、半导体工业、计算机硬件和系统设计等，我国仅1台超过10万亿次计算机用于金融业，制造业几乎是空白；我国虽有11台超级计算机用于信息服务业，但其中却有6台在游戏公司，而世界上主要用于信息处理与服务和企业管

另一方面，在应用高性能计算求解问题的规模上，我国和世界先进水平也有较大差距。我国科研工作中利用超级计算机的情况近年来虽然有了很大发展，但是尚未有十万亿次以上规模应用问题的计算实践，而美国和欧洲已经有许多数十万亿次量级的应用。当前所有超级计算机都是依靠大量处理器并行计算工作的，百万亿次计算机会有上万个处理器并行工作。据了解，我国目前的并行计算规模，绝大多数只用到几个到几十个处理器，只有极少数应用用到几百个到上千处理器。因此，在我们刚刚有了百万亿次的计算机，已经开始研制千万亿次计算机时，我们是能否用好这样的机器，这些机器能

表1：世界TOP500计算机和我国10万亿次/秒以上计算机应用领域的比较

类别	世界TOP500		我国（10T以上）	
	数量	占比	数量	占比
科学计算	185	37.00%	10	18.50%
能源/地球物理	56	11.20%	21	38.90%
气象	10	2.00%	3	5.60%
电信	10	2.00%	3	5.60%
金融	58	11.60%	1	1.90%
政府应用	8	1.60%	5	9.30%
工业制造	70	14.00%	0	0.00%
信息服务	103	20.60%	11	20.40%
合计	500	100.00%	54	100.00%

表2：拥有世界TOP500计算机最多的国家分应用领域的比较

类别	美国	英国	法国	德国	日本	中国	意大利	印度	俄国
科学计算	84	12	8	14	14	3	7	6	7
能源/地球物理	46		2			4	1		
气象	5		1				1		1
电信	1		1		1	5	2		
金融	35	8	7	3	1				
政府应用	6	1							
工业制造	57	6		2	2		1		
信息服务	57	18	7	5		3		2	
合计	291	45	26	24	18	15	12	8	8

表2是拥有世界TOP500计算机最多的9个国家分应用领域的比较。我国拥有TOP500计算机排名第7，应属于世界前列。但是从应用领域来看，仍然表明在科学计算和工业制造应用领域相对薄弱。

否充分发挥它们应有的作用，显然是一个大问题。

上述严峻的事实说明我们的应用领域分布的不合理，应用水平与国际相比有较大差距。我们决不能仅仅看到我国研制的计算机性能的进步，更重要

► 的是应用水平和对科研、经济发展和国防建设的作用。

科学计算应用问题对计算机能力的需求是无止境的，我们不怀疑我国已经有用到数百万亿次、上千万亿次的运算能力的计算需求，但是必须看到在2~3年后要实现与此能力相应的实际应用，我们还面临着巨大的困难，因为我们的准备实在不足。近年来，应当说我国在超级计算硬件环境方面已经取得长足进步，已经挤入国际前列；但是在适应于高度并行的算法、数值模拟与高端科学计算应用软件方面研究远远不够，与发达国家相比差距很大，这是我国应用水平落后的根本原因。特别是要用到数百万亿次以上运算能力的时候，更是有许多基础性的问题尚待深入研究，如不给予足够重视，将严重影响我国高端计算的应用与发展。

数百万亿到上千万亿次的高性能计算机应用的困难是需要成千上万个处理器并行计算带来的（我们姑且把这种并行度称为“海量并行”）。现有的应用软件无法直接应用到这种计算机上去，我们首先面临的是需要研制适用于海量并行的应用软件。经验证明，基于现有算法的软件，并非都可以直接放大到海量并行的情形。许多对于串行计算或是少量并行（例如几十个处理器）计算很好的算法，到了海量并行效率很差，甚至不能工作。因此，适用于海量并行的算法研究就特别重要。在海量并行和极大规模计算量的条件下会有什么问题发生，需要什么样的算法模型，如何保证计算的精度与可靠性，现在还没有得到足够深入的研究。必须从应用问题的物理模型和数学模型出发，研究与机器体系结构相适应的算法，而不是简单地从原有算法出发进行移植和改造。总之，海量并行和极大规模计算量会带来许多新问题。要高效、可靠地算出结果来，对算法研究是一个很大的挑战。

除了算法问题，机器要能够方便使用，必须高度重视系统管理和用户编程支持环境的研究开发。高度并行的计算机系统使用困难，不能想象全靠程

序员手工编制和优化一个使用成千上万个处理器的程序，目前更做不到靠机器系统本身自动并行化，这就对系统从并行编译到并行优化支持以及对系统的调度管理提出了更为迫切和更高的要求，这与一个小规模的并行系统区别很大。如此大量的处理器进行并行运算，特别是处理器拥有许多核的情形，目前尚没有成熟的编译、调度和编程方法，这些都需要大量的基础性研究。如果计算机系统不能很好解决以上这些问题，不能给用户海量并行处理条件下的稳定与可靠运行，不能确保系统高效、灵活和优良的运行环境，高性能计算应用将难以实现。

此外，海量并行的体系结构、系统的平衡设计、数据结构与通讯调度、可靠性、稳定性和功耗，对海量并行的计算系统都带来新的问题。为了向更高的计算能力发展，对于那些需要千万亿次计算能力的应用问题，研究面向应用问题的新的计算机体系结构，实现更加高效、成本更为低廉的计算，也是重要的研究方向。

事实表明，我国有关科技计划在高性能计算机发展的问题上，对上述基础性问题重视和安排不够，高性能计算机本身的发展与高性能计算的应用环境、算法和软件的研究开发脱节。正如中国科学院学部咨询报告“加速发展我国高性能计算的若干建议”<sup>[4]</sup>中指出的：“长期以来，（我国）对高性能计算存在认识上的误区。将高性能计算等价于高性能计算机的研制，并以高性能计算机的发展战略代替高性能计算的发展战略。”这对于我国高性能计算的应用和发展十分不利，如不及时解决这些问题，我们花巨资投入研制的海量并行计算机系统将会成为摆设。

从我国现有基础到使用数百万亿次到上千万亿次的应用实现需要一个过程。试想一下，如果不能在我国的千万亿次计算机问世的时候，哪怕是初步解决上述问题，这些斥资巨大的计算机系统能否有效发挥其投资效益？会不会成为一种浪费？诚然，

需要用到数百万亿次到千万亿次的应用在一段时间内不会很多，这些机器除了支持这些应用外还应当支持很多规模较小的应用。但是如果都是小型的应用，千万亿次机器还有什么意义？应用的计算规模反映了人们对客观世界认识的深入程度和理解的精细水平，这对于科研和应用都是十分重要的事。高性能计算已成为科学技术发展和重大工程设计中具有战略意义的研究手段，对于提升我国自主创新能力、增强国家竞争力、保障国家安全、促进经济发展、建设创新性国家有着十分重要的战略意义。当前要特别重视高性能计算全面、协调和可持续发展的问題，按照科学发展观的要求，统筹安排高性能计算机的硬件、体系结构、基础软件和应用

环境、计算方法和数值模型、重大应用及其软件的研究和开发，特别重视应用和高性能计算的基础性研究，使我国高性能计算又好又快地发展。

现在我国百万亿次系统即将投入使用，这为千万亿次规模的应用和相关各种问题的研究以及用户的培养提供了必要的基础条件。有关部门应当充分抓住这个时机，尽快部署和加强对数百万亿次到千万亿次高性能计算的应用和基础性研究，利用现有的百万亿次计算机系统，遴选一批具有重大科学意义与应用价值的问题，开展针对更大规模计算应用的算法、软件的研究，并全面安排和带动数百万亿次到千万亿次的高性能计算各方面的基础性研究，为未来更大规模计算的应用和持续发展打好基础。

## 参考文献

- [1] 孙家昶等，《2008年中国高性能计算机性能TOP100排行榜》，2008年11月，（<http://www.samss.org.cn>）
- [2] 深腾7000的性能数值根据11月17日公布的世界TOP500进行了修正。
- [3] TOP500 Supercomputer sites, List-Nov 2008,（<http://www.top500.org/>）
- [4] 《中国科学院关于呈报“加速发展我国高性能计算的若干建议”的报告》，（科发学部字[2006]304号）

来稿时间:2008年11月22日