“天河一号”超级计算机系统简介

国防科学技术大学，2010年11月

“天河一号”是在国家863计划支持下，通过与天津市滨海新区密切合作，由国防科学技术大学研制成功的国家超级计算天津中心业务主机，并作为计算主结点加入中国国家网格，面向国内外提供超级计算服务。

“天河一号”研制工作从2005年起步，按预先研究、方案设计、工程研制和推广应用四个阶段依次实施，其中工程研制阶段从2008年开始，按两期工程实施。一期系统于2009年9月研制成功，峰值性能为每秒1206万亿次双精度浮点运算（TFlops），持续性能为563.1TFlops（LINPACK实测值），是我国首台千万亿次超级计算机系统，参加2009年11月世界超级计算机TOP 500排名，位列亚洲第一、世界第五，实现了我国自主研制超级计算机能力从百万亿次到千万亿次的跨越，使我国成为继美国之后世界上第二个能够研制千万亿次超级计算机的国家。二期系统于2010年8月在国家超级计算天津中心升级完成，峰值性能提升为 4700TFlops，持续性能提升为 2507TFlops（LINPACK实测值），部分采用了自主研制的飞腾－1000中央处理器，参加2010年11月世界超级计算机TOP 500排名，位列世界第一，实现了从亚洲第一向世界第一的重大跨越，取得了我国自主研制超级计算机综合技术水平进入世界领先行列的历史性突破。

“天河一号”采用 CPU 和 GPU 相结合的异构融合计算体系结构，硬件系统主要由计算处理系统、互连通信系统、输入输出系统、监控诊断系统与基础架构系统组成，软件系统主要由操作系统、编译系统、并行程序开发环境与科学计算可视化系统组成。总体技术指标如下：

（1）峰值性能 4700TFlops，持续性能 2507TFlops（LINPACK 实测值），内存总容量 262TB，存储总容量 2PB，满负荷运行最大功耗为 4.04MW。

（2）计算处理系统：包含 7168 个计算结点和 1024 个服务结点。每个计算结点包含 2 路英特尔 CPU 和一路英伟达 GPU，每个服务结点包含 2 路飞腾 CPU。全系统共计 23552 个微处理器，其中英特尔至强 X5670 CPU（2.93GHz、6 核）14336

个、飞腾－1000 CPU（1.0GHz、8 核）2048 个、英伟达 M2050 GPU（1.15GHz、14 核/448 个 CUDA 核）7168 个，CPU 核共计 102400 个，GPU 核共计 100352 个。

（3）互连通信系统：采用自主设计的高阶路由芯片 NRC 和高速网络接口芯片 NIC，实现光电混合的胖树结构高阶路由网络，链路双向带宽 160Gbps，延迟 1.57us，单背板交换密度 61.44Tbps。

（4）输入输出系统：采用 Lustre 全局分布共享并行 I/O 结构，2 个元数据管理结点，128 个对象存储结点，总容量 2PB。

（5）监控诊断系统：采用分布式集中管理结构，实现系统实时安全监测、控制和调试诊断。

（6）基础架构系统：采用高密度双面对插组装结构，冷冻水空调密闭风冷散热。全系统包含 140 个机柜，其中计算机柜 112 个、服务机柜 8 个、通信机柜 6 个、I/O 存储机柜 14 个，占地总面积 700 平方米，总重量 160 吨，环境温度 10℃～35℃，湿度 10％～90％。

（7）操作系统：64 位麒麟 Linux，面向高性能并行计算优化，支持能耗管理、高性能虚拟计算域等，可广泛支持第三方应用软件。

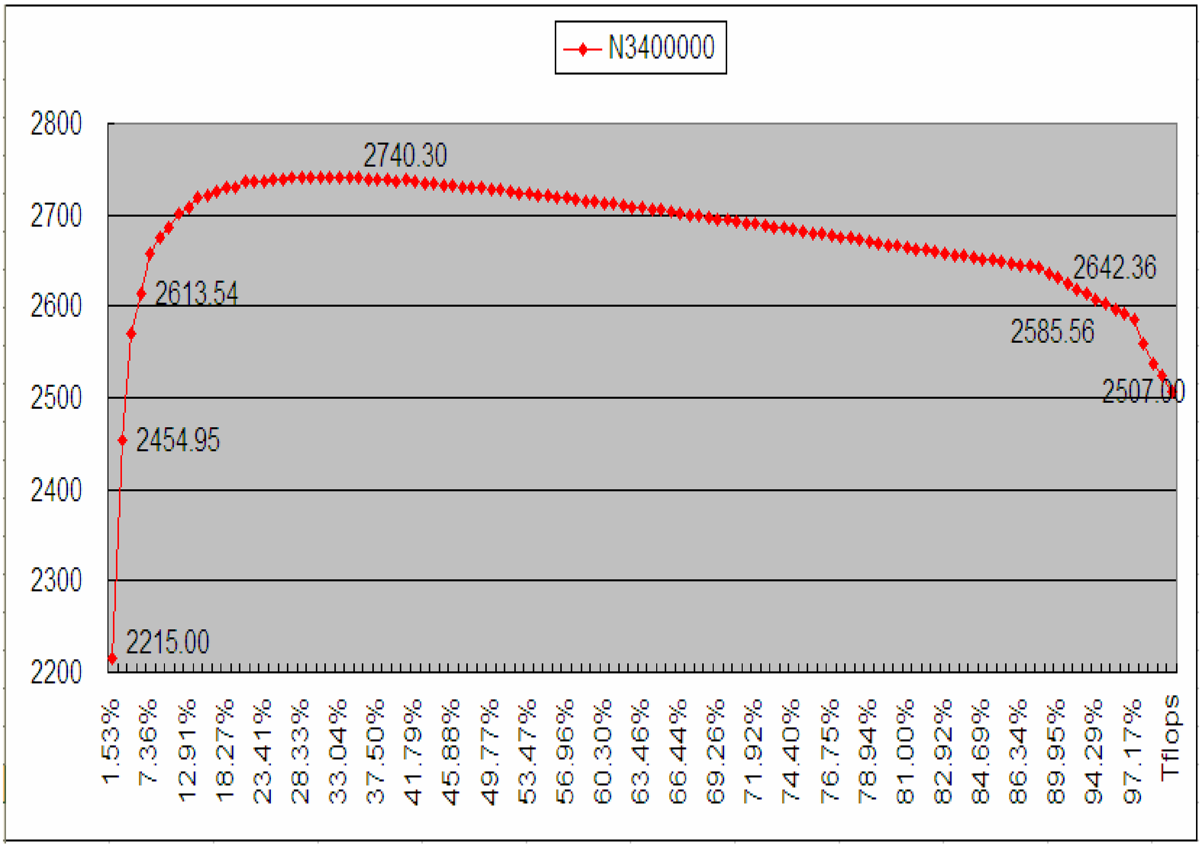
（8）编译系统：支持 C、C++、Fortran77/90/95、Java 语言，支持 OpenMP、MPI 并行编程，支持异构协同编程框架，

高效发挥 CPU 和 GPU 的协同计算能力。

“天河一号”立足自主创新、坚持实用好用，具有高自主、高性能、高能效、高安全和易使用等显著特点。

（1）高自主：自主研制了 3 款大规模集成电路芯片（飞腾－1000 CPU、高阶路由芯片 NRC 和高速网络接口芯片 NIC），4 类结点（计算结点、服务结点、I/O 管理结点和 I/O 存储结点），2 套网络（互连通信网、维护诊断网）以及 15 种 PCB 电路板；自主研制了操作系统、编译系统、并行程序开发环境与科学计算可视化系统。

（2）高性能：峰值性能 4700TFlops，持续性能 2507TFlops（由中国高性能计算机 TOP100 组织权威评测的 LINPACK 实测值），位列 2010 年 11 月世界超级计算机 TOP 500 第一位。



（3）高能效：系统能效为 620.54MFlops/W，可列 2010 年世界 Green 500 第四位，仅次于并列第一名的 3 台 IBM BlueGene（PowerXCell 8i）系统。

（4）高安全：麒麟操作系统符合 B2 级安全标准，提供基于隔离的用户安全工作环境。

（5）易使用：提供统一的全系统资源管理视图、友好的系统管理使用界面、一体化的并行应用集成开发环境和虚拟化的网络计算环境。

“天河一号”研制过程中取得了一系列核心关键技术突破，综合技术水平进入世界领先行列。主要技术创新点有：

（1）异构融合体系结构。国际上首创 CPU 和 GPU 相融

合的高效并行计算体系结构，突破了资源管理异构协同技术和混合编程技术，充分发挥 CPU 和 GPU 的协同计算能力，满足应用对计算资源的不同需求。

（2）64 位多核多线程自主 CPU。突破了多核多线程体系结构与片上并行系统（PSoC）设计技术，自主研制成功 8 核 64 线程飞腾－1000 CPU。片内集成 DDR3 存控、PCIe 2.0 I/O 接口和 CPU 直连接口，主频 1GHz，峰值性能 8GFlops，SPECint_rate2000 值为 37.7，SPECfp_rate2000 值为 41.6。

（3）基于高阶路由的高速互连通信。突破了片上高阶网络体系结构技术，自主设计了高效通信协议、高阶瓦片式（Tile）片上交换网络和高密度片间互连网络，使链路双向通信带宽达到 160Gbps、单背板交换密度达到 61.44Tbps，分别为当前国际主流商用互连 IB QDR 的 2 倍和 2.37 倍。

（4）多级并行编译优化。设计了优化资源利用的多核多线调度机制、多级并行动态负载平衡算法、全程序过程间分析等编译算法，高效支撑 JASMIN 编程框架，实现易用高效的应用编程与运行。

（5）高性能虚拟计算域。突破了高效用户容器技术、负载均衡技术和虚拟化网络终端技术，创新地在高性能计算机系统中实现了安全隔离和可定制用户环境功能，有效提升了安全性和易用性。

（6）软硬一体的低功耗控制。设计了一体化能耗管理框

架，通过监控系统自反馈冷却调节、处理器调频调压和自适应结点能耗状态转换等方法，有效降低了系统运行功耗。

“天河一号”已在国家超级计算天津中心投入业务运行，已在石油勘探、高端装备研制、生物医药、动漫设计、新能源、新材料、工程设计与仿真分析、气象预报、遥感数据处理、金融风险分析等方面获得成功应用。“天河一号”将为国家重大科技进步，为国家战略性新兴产业技术创新，为国民经济建设又好又快发展，为创新型高科技人才培养作出重大贡献。

国家超级计算天津中心简介

NSCC－TJ，2010年11月

国家超级计算天津中心（简称“天津超算中心”）2009年5月经国家科技部批准，由天津市滨海新区和国防科学技术大学合作建设。共建天津超算中心是双方贯彻落实胡主席关于走军民融合式发展道路指示精神的重大举措，是充分发挥国防科技成果对国家战略性新兴产业发展的牵引作用，为国民经济建设服务的成功实践。

天津超算中心建设于2009年6月启动。中心是天津市滨海新区的事业单位，由滨海新区科学技术委员会主管，编制40人。中心位于天津经济技术开发区滨海服务外包产业园5号楼，建筑面积约6500平方米。

天津超算中心装备的“天河一号”超级计算机系统是在国家863计划重大项目“千万亿次高效能计算机系统”和天津市滨海新区的支持下，由国防科学技术大学研制成功的。“天河一号”一期系统于2009年9月研制成功，峰值性能为每秒 1206 万亿次双精度浮点运算（TFlops），持续性能为563.1TFlops（LINPACK 实测值），是我国首台千万亿次超级计算机系统，参加2009年11月世界超级计算机TOP 500排名，位列亚洲第一、世界第五，实现了我国自主研制超级计算机能力从百万亿次到千万亿次的跨越，使我国成为继美国之后

世界上第二个能够研制千万亿次超级计算机的国家。二期系统于 2010 年 8 月在天津超算中心升级完成，峰值性能提升为 4700TFlops，持续性能提升为 2507TFlops（LINPACK 实测值），部分采用了自主研制的飞腾－1000 中央处理器，参加 2010 年 11 月世界超级计算机 TOP 500 排名，位列世界第一，实现了从亚洲第一向世界第一的重大跨越，取得了我国自主研制超级计算机综合技术水平进入世界领先行列的历史性突破。“天河一号”研制成功使得天津超算中心成为我国首个拥有千万亿次计算机的超算中心，也是世界上少数几个拥有千万亿次计算机的科研和公共服务机构之一。

天津超算中心的建设得到了国家科技部、天津市委、市政府、滨海新区和国防科学技术大学的高度重视和大力支持。国家 863 计划投入 2 亿元、天津市滨海新区配套 2 亿元研制“天河一号”超级计算机系统，天津经济技术开发区投入 1 亿元作为中心配套建设和前期运营费用。国防科学技术大学依托高性能计算机研制方面的长期技术积累，组织高水平研制团队，在较短的时间内研制成功了“天河一号”超级计算机系统。

为大力推动超算应用工作，滨海新区科委在 2010 年“自主创新平台与环境建设”专项中划拨出 3000 万元作为超算应用推广经费，滨海新区经信委在 2010 年度和 2011 年度第一批信息化项目中设立“超级计算应用”专题予以支持。天津

市科委通过科技支撑计划重点项目支持超算应用，天津市发改委、经信委将超算应用推广工作纳入了“十二五”发展规划。

天津超算中心应用领域已涵盖石油勘探、高端装备研制、生物医药、动漫设计、新能源、新材料、工程设计与仿真分析、气象预报、遥感数据处理、金融风险分析等方面。天津超算中心与企业、科研院所和高校开展合作，现已为20多个用户单位提供了高性能计算服务。通过构建企业创新研发平台，为天津一汽、天津汽车模具公司、三星电机等企业提供计算服务，提高企业设计研发能力和水平；通过构建科研合作机制，与中国科学院、中新生态城国家动漫产业示范园、天津国际生物医药联合研究院、天津市气象局、天津大学、南开大学、天津科技大学、北京大学、清华大学等单位合作开展高性能应用及研究工作。

国家超级计算天津中心将进一步夯实基础，强化能力，以应用需求为牵引，立足天津，面向全国，按照“高科技公共服务、信息产业发展、信息专业人才聚集培养”三位一体的思路进行建设和运营，为提高国家科技创新能力和促进战略新型产业技术创新提供强大的高性能计算支撑。

Introduction of Tianhe-1 Supercomputer System

NUDT, November 2010

Tianhe-1 supercomputer was built by National University of Defense Technology of China (NUDT), which has been supported by National 863 Hi-Tech Program and Tianjin Binhai New Area. Tianhe-1 is the host computer system of National Supercomputer Center in Tianjin (NSCC-TJ), and also works as a backbone node of Chinese National Grid. It can provide high-performance computing services for both domestic and foreign customers.

The research and development of Tianhe-1 system began in 2005 and it consisted of four stages: preliminary research, scheme design, system implementation, and applications exploitation. The stage of system implementation began in 2008 as a two-phase project. The phase I system was built successfully in September 2009, and the theoretical peak performance of this system is 1206TFlops while the result of its LINPACK test is 563.1TFlops. It was the first petaflop supercomputer in China. On the biannual TOP 500 supercomputer list released in November 2009, it was ranked the fastest supercomputer in Asia, and the fifth fastest computer in the world. The completion of phase I system indicated that it was a breakthrough for China to develop supercomputers from one hundred trillion flops per second to quadrillion flops per second. China has become the second country with the petaflop supercomputer R&D capability in the world besides the United States. The phase II system had been upgraded and enhanced in NSCC-TJ in August 2010. The theoretical peak performance of the enhanced system was 4700TFlops, and its

LINPACK test reached 2507TFlops. The enhanced system partly adopted the domestic FeiTeng-1000 CPUs, and was ranked the first on the TOP 500 list released in November 2010. The system has made a significant progress from being the first in Asia to being the first in the world. This indicates that China's overall technology of supercomputer development with self-intellectual property has reached the international advanced level, which is a historic breakthrough.

Tianhe-1 system adopts a hybrid architecture of heterogeneous integration of CPUs and GPUs. The hardware of the system consists of 5 components: the computing system, the interconnected communication system, the I/O system, the monitoring and diagnostic system, and the infrastructure system. The software of Tianhe-1 system is composed of the operating system, the compiling system, the parallel programming environment, and the scientific visualization system. Tianhe-1 system's overall technical indexes are as follows:

(1) It has the theoretical peak performance of 4700TFlops, the LINPACK test performance of 2507TFlops, the total memory of 262TB, and the disk capacity of 2PB. The power dissipation at full load is 4.04MW.

(2) The computing system includes 7,168 computing nodes and 1,024 service nodes. Each computing node is configured with two Intel CPUs and one NVIDIA GPU. Each service node has two FeiTeng-1000 CPUs. There are totally 23,552 microprocessors, including 14,336 Intel Xeon X5670 CPUs (2.93GHz, six-core), 2,048 FeiTeng-1000 CPUs (1.0GHz, eight-core), and 7,168 NVIDIA M2050 GPUs (1.15GHz, 14 cores/448 CUDA cores). There are totally 102,400 CPU cores and 100,352 GPU cores.

(3) The interconnected communication system uses high-radix Network Routing Chips (NRC) and high-speed Network Interface Chips (NIC) to construct the communication network. Both NRC and NIC chips are designed with self-intellectual property. The interconnection topology is an optic-electronic hybrid fat-tree structure with the bi-directional bandwidth of 160Gbps, the latency of 1.57μs, and the switch density of 61.44Tbps in a single backboard.

(4) The I/O system uses the Luster file system, which is a distributed global shared file system. It has two I/O management nodes, 128 I/O storage nodes, and 2PB disk capacity.

(5) The monitoring and diagnostic system adopts a distributed centralized management architecture. It can monitor, control, diagnose, and debug the entire system in a real-time and secure way.

(6) The infrastructure system uses a high-density assemblage technique, in which node cases can be plugged in a double-side backplane. A liquid cooling air-conditioning mechanism is adopted to dissipate heat loads from air tight racks. There are 140 cabinets in Tianhe-1 system, including 112 computing cabinets, 8 service cabinets, 6 communication cabinets, and 14 I/O cabinets. The entire system weighs about 160 tons and occupies $700m^2$. The temperature of the running environment is 10℃~35℃, and the humidity is 10%~90%.

(7) The operating system of Tianhe-1 is 64-bit Kylin Linux, which is oriented to high-performance parallel computing optimization, and supports power management

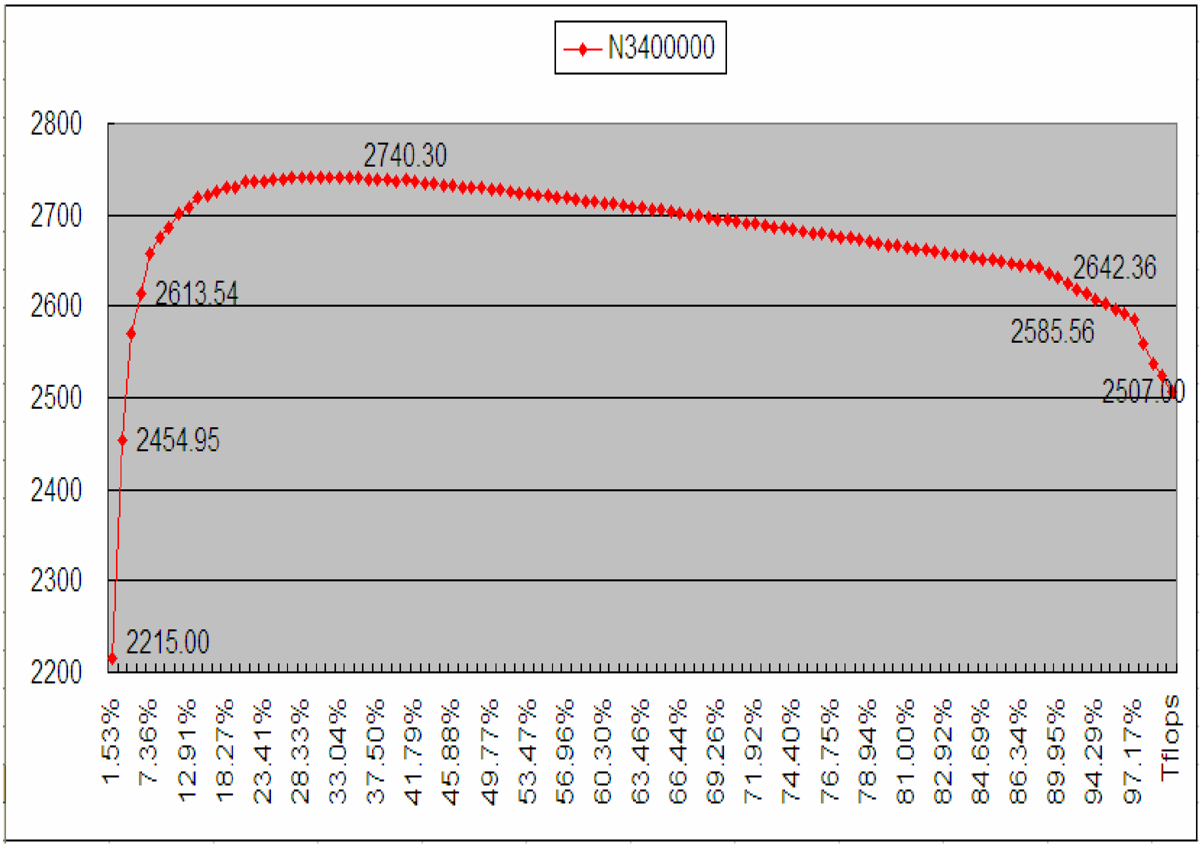
and high-performance virtual computing zone. The operating system broadly supports the third-party application software.

(8) The compiling system of Tianhe-1 supports the C, C++, Fortran77/90/95, and Java languages; it also supports OpenMP and MPI parallel programming, as well as the heterogeneous cooperative programming framework, which achieves high performance by using CPUs and GPUs cooperatively.

Tianhe-1 system relies on independent innovations, facilitates high practicality and ease of use, and features characteristics such as high originality, high performance, high energy efficiency, high security, and high usability.

(1) High originality: Most components of Tianhe-1 system were developed with self-intellectual property, including the three chips of large-scale integrated circuits (FeiTeng-1000 CPU, high-radix network routing chip NRC, and high-speed network interface chip NIC), four types of nodes (the computing node, the service node, the I/O management node, and the I/O storage node), two networks (the communication network, and the maintenance and diagnostic network), and 15 types of different PCB boards. Most software, including the operating system, the compiling system, the parallel programming environment, and the scientific virtualization system, were also developed with self-intellectual property.

(2) High performance: The theoretical peak performance of the system is 4700TFlops. According to the authoritative testing of Chinese TOP 100 organization, the result of the LINPACK benchmark is 2507TFlops, which was ranked the first in the world on the TOP 500 list released in November 2010.



(3) High power efficiency: The power efficiency of Tianhe-1 is 620.54MFlops/W, which was ranked the fourth highest in the world according to the Green 500 list released in 2010. It was preceded only by three IBM BlueGene (PowerXCell 8i) systems which were co-ranked the first.

(4) High security: Kylin operating system has passed the B2 security verification. It provides a separated working environment for users with different requirements.

(5) High usability: Tianhe-1 system provides a unified view for the global resources management, user-friendly system management interface, an integrated development environment for parallel programming, and a virtualized network computing environment.

In the development of Tianhe-1 system, many breakthroughs were achieved and the overall technology has reached the international advanced level. The major contributions are as follows:

(1) The heterogeneous cooperation architecture. An efficient architecture which integrates GPUs with CPUs for high-performance parallel computing was exploited for the first time in the world. It advances the technologies of resource management collaboration and mixed programming under the heterogeneous architecture, fully exploits the computational power of CPUs and GPUs, and thus meets the various demands of a diversity of applications.

(2) The 64-bit multi-core multi-thread CPU with self-intellectual property. Tianhe-1 advances the design technologies of both the multi-core multi-thread architecture and PSoC (Parallel System on Chip). The 8-core and 64-thread

FeiTeng-1000 CPU, which integrates the DDR3 SDRAM controller, the PCIe 2.0 I/O interface and the CPU direct link interface, was developed by NUDT. Its main frequency is 1GHz, and the peak performance is 8GFlops. The result of SPECint_rate2000 is 37.7, and that of SPECfp_rate2000 is 41.6.

(3) The high-speed communications based on high-radix routing. With the breakthroughs in the architecture of NoC (Network on Chip), we have developed efficient communication protocols, a high-radix tile NoC, and a high-density inter-chip network. It achieves the bi-directional bandwidth of 160Gbps, and the single backboard switching density of 61.44Tbps, which are 2 times and 2.37 times of the off-the-shelf commercial switch IB QDR, respectively.

(4) The multi-level parallel compiling optimization. We have designed a multi-core and multi-thread scheduling mechanism for optimizing resource utilization, an algorithm for multi-level parallel dynamic load balancing, and an algorithm for the inter-procedure analysis of the entire program. The compiler supports JASMIN programming framework. It is easy to use and very efficient for programming and running applications.

(5) The high-performance virtualized zone. Tianhe-1 has made breakthroughs in the techniques of user container, load balancing and virtualized network terminal, and has innovatively implemented the functions of security isolation and customized user environment, which improves the security and usability of Tianhe-1 system effectively.

(6) The mechanism of power consumption control with integrated hardware and software. An integrated framework of power consumption management was designed. Several methods were adopted, including monitoring the system self-feedback cooling adjustment, the CPU frequency and voltage adjustment, and the adaptive power consumption state transition of nodes to reduce the system power consumption effectively.

Tianhe-1 system has been installed to provide services at NSCC-TJ, and has been successfully applied to many domains, such as oil exploration, high-end equipment development, bio-medical research, animation design, exploitation of new energy sources, research of new materials, engineering design and simulation analysis, weather forecast, remote data processing, and financial risk analysis. Tianhe-1 will make significant contributions to the national scientific progress, the national strategic technical innovation of new industries, the development of national economy, and the training of innovative hi-tech workforce.

Introduction of National Supercomputer Center in Tianjin

NSCC-TJ, November 2010

National Supercomputer Center in Tianjin (NSCC-TJ), which was authorized by the Ministry of Science and Technology (MOST) of China in May 2009, was constructed collaboratively by Tianjin Binhai New Area and National University of Defense Technology (NUDT).

Constructing NSCC-TJ collaboratively is a significant act on carrying out President Hu Jintao's idea of joint civil-military development, and is a successful practice of applying defense science and technology to the promotion of national strategic new industries, and to the development of national economy.

The construction of NSCC-TJ began at the end of June 2009. The Center is a public institution of Tianjin Binhai New Area superintended by the Scientific and Technological Committee of Binhai New Area. The Center has 40 staff members. NSCC-TJ is located in the No.5 Building of TEDA Service Outsourcing Industrial Park and has a building area of about 6,500m^2.

Tianhe-1 supercomputer system was built by NUDT, which is the host computer system of NSCC-TJ and has been supported by National High-Tech Program (the 863 Program) and Tianjin Binhai New Area. The phase I system was built successfully in September 2009, and the theoretical peak performance of this system is 1206TFlops, while the result of its LINPACK test is 563.1TFlops. It was the first petaflop supercomputer in China. On the biannual TOP 500 supercomputer list released in November 2009, it was ranked the fastest supercomputer in Asia, and the fifth fastest computer in the world. The completion of the phase I system indicated that it was a breakthrough for China to develop supercomputers from one hundred trillion flops per second to quadrillion flops per second. China has become the second country with the petaflop supercomputer R&D capability in the world besides the United States. The phase II system had been upgraded and enhanced in NSCC-TJ in August 2010. The theoretical peak performance of the enhanced system was 4700TFlops, and its LINPACK test reached 2507TFlops. The enhanced system partly adopted the domestic FeiTeng-1000 CPUs, and was ranked the first on the TOP 500 list released in November 2010. The system has made a significant progress from being the first in Asia to being the first in the world. This indicates that China's overall technology of supercomputer development with self-intellectual property has reached the international advanced level, which is a historic breakthrough.

The construction of NSCC-TJ was greatly supported by MOST, the municipal committee and government of Tianjin, Tianjin Binhai New Area, and NUDT. The development of Tianhe-1 supercomputer system was funded with ¥200 million by National High-Tech Program (the 863 Program) and ¥200 million by Binhai New Area. Tianjin Economic Development Area (TEDA) also provided ¥100 million for the infrastructure construction and the running preparation of the Center. With a long-term accumulation of high-performance computer development technology, NUDT organized an excellent research team, and successfully built Tianhe-1 in a relatively short period of time.

In order to promote the application of supercomputing, the Scientific and Technological Committee of Tianjin Binhai New Area has funded NSCC-TJ with ¥30 million from the special project of "The Construction of Innovative Platforms and Environments" in 2010. The Economy and Information Technology Commission of Binhai New Area has set up a special project of "Supercomputing Applications" in its first information program of 2010 and 2011. The Scientific and Technological Committee of Tianjin will support the supercomputing applications through the key projects of Tianjin Science & Technology Pillar Program. The Tianjin Development and Reform Commission and the Tianjin Economic and Information Technology Commission have listed the extension of supercomputing application in the 12th Five-Year Development Plan of Tianjin.

NSCC-TJ's application domains include oil exploration, high-end equipment development, bio-medical research, animation design, exploitation of new energy sources, research of new materials, engineering design and simulation analysis, weather forecast, remote data processing, and financial risk analysis. NSCC-TJ has cooperated with enterprises, as well as scientific institutions and colleges, and has been providing high-performance computing services for more than 20 customers, including Tianjin FAW XIALI Automobile Co., Ltd., Tianjin Motor Dies Company Limited, and Tianjin Samsung Electro-Mechanics Company Limited. NSCC-TJ is also cooperating with Chinese Academy of Sciences, National animation industrial park in the Sino-Singapore Tianjin Eco-city, Tianjin International Joint Academy of Biotechnology & Medicine, Tianjin Meteorological Bureau, Tianjin University, Tianjin University of Science & Technology, Nankai University, Peking University, and

Tsinghua University, etc.

NSCC-TJ will focus on the requirements of applications, and keep on improving the infrastructure of the center to further improve the ability to provide public services. In addition, it will be based on Tianjin and oriented to the entire country in order to perform construction and operation according to the idea of hi-tech public services, information industry development, and professional information workforce training. It will also provide powerful high-performance computing support for improving the national scientific innovation capability and promoting strategic industrial technology innovation.