Development and Collaboration of NSCC-TJ

Meng Xiang-Fei

Leader of HPC Application from NSCC-TJ

mengxf@nscc-tj.gov.cn
Supercomputer center in China

HPC Developing Plan in China

Overview of NSCC-TJ

Application of NSCC-TJ

Research and Development of NSCC-TJ

Future Plan of NSCC-TJ
Supercomputer center in China
HPC Developing Plan in China
Overview of NSCC-TJ
Application of NSCC-TJ
Research and Development of NSCC-TJ
Future Plan of NSCC-TJ
Supercomputer Center in China

National Supercomputer center

NSCC-TJ, NSCC-SH, NSCC-SD, NSCC-CS, ...

Local Supercomputer center

SSC, ...

Domain computer center

CAS, CMA(China Metrological Administration), …
HPC Developing Plan in China

NSFC

- Basic algorithms and computable modeling for high performance scientific computing
- Network based research environment
- Many-core parallel programming

863 Program (High Science and Tech.)

- High productivity computer and Grid service environment, HPC Software R&D
- Multi-core/many-core programming support

973 Program (Basic Research and Application)

- Parallel algorithms for large scale scientific computing
- Virtual computing environment
863 key projects on HPC and Grid: 2002-2010

Phase I: “HPC computer & core software”
- 4-year project, May 2002 to Dec. 2005
- 100 million Yuan funding from the MOST
- More than 2X associated funding from local government, application organizations, and industry
- Major outcomes: China National Grid (CNGrid)

Phase II: “High productivity Computer and Grid Service Environment”
- Period: 2006-2010 (extended to 2012), TH-1A ...
- 940 million Yuan from the MOST and more than 1B Yuan matching money from other sources
863 key projects on HPC: 2011-2015

**Supercomputer**
- Heterogeneous architecture
- Programmability is really an issue
- 100PF by the end of 2015

**Software**
- Fusion simulation
- Simulation for aircraft design
- New medic discovery
- Digital media
- Structural mechanics for large machinery
- Electro-magnetic environment simulation

**Programming framework**
- Convenient parallel program development
- Compiler optimization
Overview of NSCC-TJ
Overview of NSCC-TJ

Sponsored by

- Chinese government: MOST, MOF…….
- Local government: Tianjin Binhai New Area

Public information infrastructure

- To accelerate the economy, science innovation and industry of China
- To provide high performance computing service to whole China and even to all over the world

Open platform for science research and education
# TH-1A system Configuration

<table>
<thead>
<tr>
<th>Items</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processors</td>
<td>14336 Intel CPUs + 7168 nVIDIA GPUs + 2048FT CPUs</td>
</tr>
<tr>
<td>Memory</td>
<td>262 TB</td>
</tr>
<tr>
<td>Interconnect</td>
<td>Proprietary high-speed interconnecting network</td>
</tr>
<tr>
<td>Storage</td>
<td>4PB (Lustre File System)</td>
</tr>
<tr>
<td>Cabinets</td>
<td>120 Compute / service Cabinets</td>
</tr>
<tr>
<td></td>
<td>14 Storage Cabinets</td>
</tr>
<tr>
<td></td>
<td>6 Communication Cabinets</td>
</tr>
</tbody>
</table>
Roadmap of Supercomputer in China

<table>
<thead>
<tr>
<th>HPC TOP500 排名</th>
<th>501</th>
<th>20</th>
<th>40</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>年份</td>
<td>1997</td>
<td>2000</td>
<td>2007</td>
<td>2010</td>
</tr>
</tbody>
</table>
Overview of NSCC-TJ

Compute sub-system

- CPU + GPU
- CPU + GPU
- CPU + GPU
- CPU + GPU
- CPU + GPU

Service sub-system

- Operation node
- Operation node

Communication sub-system

Storage sub-system

- MDS
- OSS
- OSS
- OSS
- OSS
- OSS
Overview of NSCC-TJ

Compiler system
- Heterogeneous programming frame
- OpenMP
- MPI
- C/C++
- Fortran77
- Compiler optimization

Programming environment
- Parallel Debugger
- Performance analysis tool
- Science computing library

Visualization system
- Analysis of large scale data
- Parallel drawing and render
- Interactive control module

Parallel Operating System
- Communication system
- Security isolation
- Resource management system
- Parallel file system

Kernel
China Supercomputing development Strategy: Cooperating Development and Innovation
Support High Tech. and Industry – Joint Lab, Union, Innovation center
Overview of NSCC-TJ

Support Basic Science Research–Sub-center at University
Supercomputer center in China

HPC Developing Plan in China

Overview of NSCC-TJ

Application of NSCC-TJ

Research and Development of NSCC-TJ

Future Plan of NSCC-TJ
User Number of Research Group and Company more than 410
## Application of NSCC-TJ

### Supported Projects by NSCC-TJ

<table>
<thead>
<tr>
<th>NSFC</th>
<th>863, 973 Program</th>
<th>Other Key Projects Funded by Central Government</th>
<th>International or Local Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;400 item</td>
<td>&gt;45 item</td>
<td>&gt;20 item</td>
<td>&gt;10 item</td>
</tr>
</tbody>
</table>

### Profile of user number

- Basic science research (Physics, Chemical, Astronomy, etc) - 38%
- Bio-medical research - 20%
- New material, new energy research - 10%
- Computing fluid dynamics - 8%
- Engineering design, simulation and analysis - 8%
- Environment science - 7%
- Weather and climate forecasting - 6%
- Petroleum exploration - 5%
- Animation - 2%
- Other - 1%
Magnetic confinement fusion research

- International ITER Project: China, EU, USA, Japan, Korea, Russian and India;
- Application Scale: GTC program running on up to 50,000 cores;
- Proprietary Programming (CPU+GPU Version).
Petroleum seismic data processing

- Program: single/double-way wave Prestack depth migration (RTM), proprietary
- 85860 cores (whole TH-1A)
- 2600 Km², 2.2TB data; 10000Km², 100TB data
- New programming: (CPU+GPU) version has good scalability and reaches 4x speedup based-on TH-1A
- IO based on memory, one IO routine (rearrange data and sum) reaches 3X speedup on TH-1A (reduce form four day to one day on 2000 nodes)

Surface: 2600 Km² depth: 5Km
Large-scale Simulation of the Global SWEs

- CPU: Using 82,944 cores; Parallel efficiency: 60%
- CPU+GPU: Towards Peta-scale global SWE simulation, 809 TFlops in double precision in 3750 nodes
- Real topography of the Earth, zonal flow
  - Day 15, 10,240*10,240*6 mesh (1km res)

<table>
<thead>
<tr>
<th></th>
<th>CPU-only (1-core)</th>
<th>CPU-only (12-core)</th>
<th>CPU-GPU</th>
<th>CPU-GPU tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.5 Tflops</td>
<td>126 Tflops</td>
<td>658 Tflops</td>
<td>809 Tflops</td>
<td></td>
</tr>
</tbody>
</table>

32.8% of peak!
Global Climate Change

- User: Institute of Atmospheric Physics, Chinese Academy of Sciences.
- LICOM pattern: 2D MPI & OpenMP programming.
- Up to 10,000 Cores, complete an numerical simulation based-on 50 years statistics.
- Take 600 hours, 9x speedup performance.
- Capable of completing the high-resolution Ocean Circulation Pattern numerical simulation with massive data.

SSHA standard deviation: simulation (cm)
Genomics research

- User: BGI (Academe), Etc. Chinese BGI cooperates with Denmark to set up the Academe in Denmark.
- Human Genetics, Animals and Plants, Microorganism
- Cooperation with Bill-Gates Fund: The thousand year plate of UN, human foodstuff and health
- Genic data base: 1PB
New medicine design

- Institute of Material Medical, Shanghai, China
- Simulation on TH-1A with close coordination of the experiment results
  - Confirmation of a new drug interaction sites;
  - Directly through the drug design, without any chemical modification, obtained lead compounds for drug treatment of epilepsy which has a good activity ability in animal

Mouse model of epilepsy

Injection of new drugs
PKUFFT Application in CFD

- PKUFFT published at ACM ICS 2010 is the fastest algorithm for GPU clusters
- PARRAY published at ACM PPoPP 2012 is generalized to many application areas as the advanced technology of for GPU cluster
- Resilience Technology has already supported the communication of “TH-1A”, including the GPU clusters
- The scale of the Directly Simulating Turbulent Flows, $14333^3$ 3D, surpassed the $4096^3$ 3D data, which could lead the research of aerospace, shipbuilding, climate simulation and so on.

<table>
<thead>
<tr>
<th>TH-1A Node</th>
<th>Turbulent Flow</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2048</td>
<td>$4096^3$</td>
<td>The largest scale of turbulent flows at that time</td>
</tr>
<tr>
<td>4096</td>
<td>$8192^3$</td>
<td></td>
</tr>
<tr>
<td>7168 (including GPU)</td>
<td>$14336^3$</td>
<td>The largest scale, flow parameter approximately equal the real turbulent flows</td>
</tr>
</tbody>
</table>
Heat recovery coke oven

- chemical engineering institute, Tianjin University
- 3D numerical simulation based-on TH-1A,
- Large-scale, complex procedure of simulation, non-steady state

3D numerical simulation view
Multilevel heterogeneous programming model and software

- **Hardware structure**
  - Node-node, symmetry form
  - Node inside, hybrid from

- **Programing model**
  - Node-node, MPI
  - Node inside, share memery
    - Pure CPU-thread
    - CPU-thread(attempering GPU)
Hiberarchy Model of Supercomputing Application Environment

Hiberarchy parallel initialization

- Host nodes: arrangement initialization tasks
- Sub-nodes: finished part of initialization

Hiberarchy parallel computing

- Computing resource dynamic allocation
- Running node fault tolerance

Hiberarchy I/O management

- Local memory: temporary I/O
- Excursion strategy: high-load, no-block
Magnetic confinement fusion simulation

- Design Heterogeneous Programming model
  - Memory access
    - Texture prefetch for 2D spatial locality
    - Reorganizing arrays to increase texture cache hit
  - CPU-GPU data transfer
    - Porting temporary arrays to GPU
    - Using register for arrays only used in one kernel
  - Parallel algorithms for CUDA
    - Hierarchical scan solution to stream compaction problem
    - Concurrent execution

This work has been selected as top 10 application software On Titan and accepted to presentation by ISC13.
Research and development of NSCC-TJ

Fusion Tech. of Big Data and HPC, CC

- Cloud Computing
- Big Data Research and Application
  - Data Collection
  - Data Service
  - Supercomputer and HPC
  - Hierarchy High Performance Storage
Big-data Technology R&D

- Collection and service:
  - Model of big data collection and service based on cloud computing;

- Analysis and process:
  1) Develop HPC software of big data analysis and process
  2) Model the performance of executing HPC application on cloud platform
  3) Research high cost-performance model of data analysis and process
  4) Research a statistic model that is scalable for distributed big-data set

- Storage:
  - Building a Scalable Storage based on hardware technology and CC

Big-data Application

- Genome, geophysics, Climate, Intelligent City, …;
- Global systems science
Research and development of NSCC-TJ

High speed interconnecting network

3rd-level storage
(Memory based Storage: 1P)

2nd-level storage
(Lustre: 10P)

1st-level storage
(500P)
Strategy to extend High-end Information Tech. Application (HITA) in China and even the world

- User and Collaborator: University, Academe and Industry
- NSCC-TJ: HITA service and technical R&D

- Provide HPC, CC, Big-data service for users
- Research and develop HPC, CC, Big-data technology and construct application platform
- Training: improve application ability of users and educate more high-end IT talents
- Collaboration: strengthen international cooperation and lead the HITA development
Area of International Collaboration

- Focus on the important science research area, based on Tianhe-1A and other HPC resource to strength the application cooperation, such as Life science, Energy Development, Climate Change, Physics and so on;

- HPC technology R&D:
  1) Design and optimize programming model for heterogenic supercomputers;
  2) Research the challenges of large-scale scientific computing: I/O wall, Fault-tolerant, Energy efficiency, etc
  3) Model the performance of executing HPC application on cloud platform

- Big data research and application: Fusion technology of Big data and HPC, CC
Meng Xiang-Fei
Leader of HPC Application from NSCC-TJ
mengxf@nscc-tj.gov.cn