Scaling applications from six application domains on Tianhe-2

Yunfei Du (杜云飞)
National Supercomputing Center in Guangzhou
yunfei.du@nscc-gz.cn
Outline

- Overview and Highlights of Tianhe-2
- Six Application Domains
- Conclusions
Overview

- **Tianhe-2 is a New Milestone of Chinese Supercomputer**
  - Manufacture: NUDT
  - Joint fund by the Gov of Guangdong province & Guangzhou city
  - Site: National Supercomputing Center in Guangzhou
Overview

**Hardware**

- Compute System
- Service System
- Storage System
- Interconnect System
- Monitor & Diagnose System

Express-2 proprietary interconnection network

IB Storage network

Operation Nodes

Compute Nodes Array

SAN & NAS

I/O Storage Nodes

Monitor
Overview

<table>
<thead>
<tr>
<th>Perf</th>
<th>54.9PFlops / 33.86PFlops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>16000</td>
</tr>
<tr>
<td>Mem</td>
<td>1.4PB</td>
</tr>
<tr>
<td>Racks</td>
<td>(125+8+13+24=170) (720m²)</td>
</tr>
<tr>
<td>Power</td>
<td>17.8 MW (1.9GFlops/W)</td>
</tr>
<tr>
<td>Cool</td>
<td>Close-coupled chilled water cooling</td>
</tr>
</tbody>
</table>

**TH-Net**

**Perf** 54.9PFlops / 33.86PFlops

**Nodes** 16000

**Mem** 1.4PB

**Racks** \(125+8+13+24=170\) (720m²)

**Power** 17.8 MW (1.9GFlops/W)

**Cool** Close-coupled chilled water cooling

**Hybrid Hierarchy shared storage System**

**TH-2** (125 x Rack)

**Rack** (8 x Frame)

**Frame** (8 x board)

**Compute board**

**ION**

**CPM**

**PHI** #45000

**IVB** #32600

**FT-1500** #4096

**APM**

**Compute board**

**Frame** (8 x board)

**TH-Net**

**Global Storage**

**ID Storage Network**
1. Architecture

- Multipurpose-Heterogeneous Architecture
  - Based on MPP, using MIC accelerating computing
  - Using customized high speed interconnect network, supporting high productive Science & Tech computing
  - Based on I/O accelerating module using SSD, storage network, enhance the S&T computing, supporting big data and high throughput service
2. Compute node

**Neo-heterogeneous Compute Node**
- CPU + MIC
- Similar ISA, Different components
- Compatibility, Flexibility, Usability

**Accelerator**

<table>
<thead>
<tr>
<th>GPU</th>
<th>Vs</th>
<th>MIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data parallel</td>
<td>Multi-thread, SIMD</td>
<td></td>
</tr>
<tr>
<td>Cuda, OpenCL</td>
<td>MSI, OpenMP, OpenCL</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>Offload, Native, Symmetric, Shared</td>
<td></td>
</tr>
</tbody>
</table>
3. Interconnect Network

- **TH-Express2** — scalar to 100PF system
  - Self-design ASIC: NIC & NRC
  - Optic and electronic hybrid network
  - Fat tree topology, balanced extension to 100PFlops
3. Interconnect Network

- **Customized communication Lib and MPI**
  - reduced connection-less user level communication protocol using virtual port;
  - collective offload mechanisms based on triggered descriptor queue operations;
  - communication protocol based on multiple channel data transfer according to application communication mode;
  - multiple level zero-copy transfer technology based on memory registration management;

- **Get over 96% efficiency of hardware level**

- **Decrease 75% latency of collective comm**

- **High BW & Low Latency**
4. Storage & File system

- **Increasing I/O requirements**
  - Large scale Pre/Post data sets
  - Visualization and Analysis

- **I/O Bottleneck**
  - Scalability
  - Efficiency
  - Performance
  - Economic and durability

- **Hybrid Hierarchy parallel storage architecture**
4. Storage & File system

- Unified namespace for hybrid and hierarchy storage
- Affinity-based asynchronous collective-IO
- Application guided data layout and movement
- Various API, Posix, MPIIO, HDF5, Netcdf …

- Unified Filesystem: H²FS
- Single I/O BW > 2GB/s
- Aggregative I/O BW > 500GB/s
5. Software Env for HPC

- **Intra-node**
  - OpenMP, Pthread

- **Inter-node**
  - TH-MPI, GA

- **Multi-Domain Framework**
  - Domain-Specific tools
  - Task allocation based topology
  - Load Balance & comm optimized
  - Fault tolerance
6. Software Env for Big Data

- **Cloud computing platform**
  - Data processing & Information Service
- **micMR Framework for Map/Reduce over MIC**
- **Support Hadoop & Data mining**
- **Multiple data management models based on H2FS**
Summary for Tianhe-2

One system

- Scientific Computation
  - MPP, CPU+Xeon phi+High speed interconnect

- Big Data
  - IO Acceration Nodes + SSD

- High security
  - Domestic FT-1500 CPU
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Material Science and Engineering

- There is foundation: quantum mechanics, classical dynamics, statistical mechanics.
**Purposes**

- **Modeling for materials, parameters**
- **Compared with experiments, guiding research**
- **Multiscale platforms for invent and design the new materials**
- **Optimizing the development time for new materials, the standard workflow for developing new materials**

**Modeling/ Simulation** ➔ **Verification** ➔ **Production**
## Software for Material

<table>
<thead>
<tr>
<th>Name</th>
<th>Google Scholar Cites</th>
<th>Cite/Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAMD</td>
<td>6043(2005)</td>
<td>604.3</td>
</tr>
<tr>
<td>LAMMPS</td>
<td>7922(1995)</td>
<td>396.1</td>
</tr>
<tr>
<td>DL-POLY</td>
<td>297(2006)</td>
<td>33</td>
</tr>
<tr>
<td>AMBER</td>
<td>3442(2005) 1206(2003)</td>
<td>344.2 100.5</td>
</tr>
<tr>
<td>CHARMM</td>
<td>3381(1983) 2346(2009)</td>
<td>105.7 391</td>
</tr>
</tbody>
</table>
LAMMPS

- solid-state materials (metals, semiconductors)
- soft matter (biomolecules, polymers), coarse-grained or mesoscopic systems
- Multiscale (atomic, meso, or continuum scale)
- Scalability is good
- Prepost and postpro remain to be improved
- Lammps-MIC

49152 cores
GROMACS

- biochemical molecules like proteins, lipids and nucleic acids
- simulate the Newtonian equations of motion for systems with hundreds to millions of particles
- Many postprocessing tools

![Graph showing GROMACS cluster performance]
CASE 1: NEMO5

- A multi-scale simulation tool for NanoElectronics Modeling
  - Basic science in ultra-scaled physics oriented devices such as single atom transistors
  - Deployment of apps in nanoHUB that are powered by NEMO5 and are being used by over 12,000 users

- Optimization
  - Four levels of MPI parallelism
    - space, energy, momentum, voltage
  - We use the RGF solver
    - The time-consuming operations: zgemm and zgesv
    - They takes nearly all the execution time
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Purposes

- The structure simulation and function modelling for biological macromolecule
- Drug design and screening
- Protein sequence analysis
- Gene Sequence analysis and alignment
- Analysis and modeling for gene regulatory networks
- The bigdata analysis for medical health
## Open Source Software

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAMD</td>
<td>2.9, 2.10-CVS, 2.10-MIC</td>
<td>Molecular dynamics for simulation of biomolecular systems</td>
</tr>
<tr>
<td>BLAST</td>
<td>2.2.30</td>
<td>similarity between biological sequences</td>
</tr>
<tr>
<td>Gromacs</td>
<td>4.5.3, 4.6.3, 5.0.1, 5.0.4, 5.0.4-MIC</td>
<td>simulate the Newtonian equations of motion for particles</td>
</tr>
<tr>
<td>Modeller</td>
<td>9.14</td>
<td>homology or comparative modeling of protein</td>
</tr>
<tr>
<td>BWA</td>
<td>0.5.10-MIC, 0.7.10, 0.7.12</td>
<td>Sequences alignment</td>
</tr>
<tr>
<td>Nwchem</td>
<td>6.5.0</td>
<td>kinetics and dynamics of chemical transformations</td>
</tr>
<tr>
<td>Lammmps</td>
<td>14-Feb14, 1Aug13, 25Sep14, 9Oct14, 9Oct14-MIC</td>
<td>kinetics and dynamics of chemical transformations</td>
</tr>
</tbody>
</table>
# Open Source Software

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>samtools</td>
<td>1.2</td>
<td>Sequences alignment</td>
</tr>
<tr>
<td>AutoDock</td>
<td>4.2.6</td>
<td>Molecular docking</td>
</tr>
<tr>
<td>AutoDuck_Vina</td>
<td>1.1.2</td>
<td>Molecular docking</td>
</tr>
<tr>
<td>GATK</td>
<td>3.3.0</td>
<td>analyze high-throughput sequencing data</td>
</tr>
<tr>
<td>picard</td>
<td>1.129</td>
<td>Gene sequences alignment</td>
</tr>
<tr>
<td>plink</td>
<td>1.07, 1.9</td>
<td>Whole genome data analysis toolset</td>
</tr>
<tr>
<td>vcf tools</td>
<td>0.1.12b</td>
<td>Cnv analysis tools</td>
</tr>
</tbody>
</table>
CASE 1: Million-core scale gene regulatory networks

- The first ever genome-scale approach for construction of gene regulatory networks of Arabidopsis thaliana using Bayesian network structure learning
- Whole genome regulatory networks help us understand gene regulatory mechanisms on a genome-wide scale
- Large-scale CPU/MIC heterogeneous parallel computing
  - 8192 Nodes, 1.6 millions of cores
  - Measured performance: 4.81 Pflops
  - 220 millions of elements 3 hours
  - Nearly linear weak scaling, 83% strong scaling parallel efficiency
CASE 2: Large-scale SNP Detection

- SNP is a single nucleotide variation in DNA, SNP detection is a fundamental and essential step in whole genome analysis.

- Optimization
  - Asynchronous data transform strategy between CPU and MIC
  - Vectorizing SNP Detection on MIC
  - Show nearly linear speedup with the number of MIC increasing

- SNP presents promising scalability on up to 8,192 nodes (196,608 CPU cores and 1,376,256 MIC cores). The parallel efficiency is more than 60.7%.
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Energy

- The analysis of vehicles, materials for energy generation, storage and utilization
- Developing energy materials
- Analyzing the procedure of generating energy
- The location for wind power plant and solar power plant
- Bigdata analysis, Whole Life Cycle Management of Assets for energy project
Softwares

- **Open source**
  - Code-Aster
  - Code-Saturne
  - GTC
  - fds
  - Elmer
  - Salome
  - OpenFOAM
  - CAST3M
  - libmesh
  - Palabos
  - OpenLB
  - Deal.II
  - CGNS

- **Commercial software**
  - ANSYS
  - EuroPlexus
CASE 1: Envision Corporation, wind power plant

- Using GreenwichCFD to simulate the location of fans
  - CFD cloud service based on HPC
- The optimization for located each fan
  - economical efficiency $> 5$
- Performance and scalability is better
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
CAE Application Platforms

- FEM for structural design and strength check
- Using LBM, FEM and FVM to simulate flowfield and Electromagnetic Field
- Temperature stress analysis, piezoresistive analysis and fluid-solid coupling
- Using Wittenburg, Schiehlen, Kane to simulate multibody dynamics
**CASE 1: Simulation of Aerodynamics for High speed train**

- **Background**
  1. To design the high-speed train;
  2. The best shape of high-speed train, to improve the performance of aerodynamics;
  3. The stability of train when crosswind

- **Problem**
  Size: 40 millions meshes

- **Comments**
  To scientific design for train
  Good scalability, stability and performance
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Earth Science and Environmental Engineering

- To Evolution simulation for geography, ocean and atmosphere
- Five categories
  - Geography
  - Weather
  - Ocean
  - Universe
  - Tools
Scales for the platform

Scale

University

Climate

Ocean

Geo
CASE 1: Cosmology

TH2 simulating the whole universe with dark matter and neutrino-Cosmological Large scale structure

One of the most concerns of the international astrophysics in recent years in the world.

the world’s largest N-body simulation in all of scientific research field: largest in number of particles not volume.

Major Run

- 13824 nodes, 331,776 cores
- Running 54 hours
- 3Trillion particles
- Volume : $(1.2 \ Gpc/h)^3$
- Running 13.7 billion year age of Universe
Case 2: Earthquake simulation: SeisSol -- TUM

- Simulation of the 1992 Landers earthquake
- Modeling frequencies up to 10Hz relevant to civil engineering
- Large-scale CPU/MIC heterogeneous parallel computing
- Computing scale and efficiency
- 8192 heterogeneous nodes, 1.6 millions of cores
- Measured performance: 8.6 PFLOPS, 2014 Gordon Bell Final List
- System utilization > 31%
- Parallel efficiency (strong): 60%
- Parallel efficiency (weak): 98%
Six Application Domains

- Material Science and Engineering
- Life Sciences and Personalised medicine
- Energy and related Technologies
- Digital design and Manufacturing
- Earth Science and Environmental Engineering
- Smart City
Smart city

KylinCloud

E-gov.  Big Data  Server

Kylin OS

Paas
- Deployment
- Auto deploy
- Auto config
- Auto scala

Iaas
- services
- schedule
- High ava
- Computation
- Storage
- network

Security

2015/7/17 KylinCloud—面向行业用户的安全云解决方案
Large scale deployment

- Nearly ten thousand nodes
- Administration tools
- KylinCloud on Tianhe-2
CASE 1: Cloud for rendering

- Shorten the time for rendering to improve the efficiency for making animation

- Key technologies
  - Flexible deployment
  - Interconnect
  - The schedule policy for rendering
  - Optimizing the performance of IO

- Effects
  - The production system for several Chinese animations
Future development for NSCC-GZ

- production - study - research combination
- Multidisciplinary, computational science
- HPC, cloud and big data
  - Openstack+Docker+hadoop/spark
Thank You!