High Performance Computing Environment and Applications in CAS

Xuebin CHI, Haili Xiao  ({chi, haili}@sccas.cn)

Computer Network Information Center (CNIC)
Center of Scientific Computing Application & Research (Coscar)
Chinese Academy of Sciences (CAS)

ISGC2017, March 9, Taipei
Outline

1. Supercomputing in CNIC (SCCAS)
2. HPC Infrastructures
3. HPC Applications
Part I

Supercomputing in CNIC
Supercomputing (SCCAS)

We have HPC Dept. In CNIC, CAS in Beijing, 40+ staffs

Missions

• Operation and maintenance of the Supercomputing Environment of CAS (China ScGrid)
• Development of visualization, HPC application software
• Provide HPC service

Our roles in the national HPC infrastructure of China

• Operation and Management Center of CNGrid (announced in 2005)
• Management of Supercomputing Innovation Alliance
Supercomputing Innovation Alliance

- Established at September 25, 2013, approved by MOST
- Initiated by supercomputing centers, application institutes, etc.
- Industry/university/research/application cooperation (产学研用)
- The Secretariat is located in CNIC
Supercomputing Innovation Alliance
- Membership (50+)
Intel Parallel Computing Center

• The only Intel Parallel Computing Center (IPCC) in Mainland China
• Intel & CNIC, started in Apr. 2015
• Focusing on applications using MIC
Parallel algorithm & software development

• Numerical linear algebra
• Computational fluid dynamics
• Fast multipole method
• Computational chemistry, physics, biology
• Fft, fmm, hpseps, ccf, sc_tangram
Part II

HPC Infrastructures
The Supercomputer History in SCCAS (1996-2016)

1996: SGI Power Challenge XL, 6.4G FLOPS
1998: Hitachi SR2201, 1.06 G FLOPS
2003: Lenovo DeepComp 8800, 5.3 T FLOPS
2008: Lenovo DeepContra/2000, 150 T FLOPS
2014: Era, 300 T FLOPS
2015: Era, 2 PetaFLOPS +

6.4G FLOPS → 5 T FLOPS → 150 T FLOPS → 2.3 P FLOPS
New Petascale Supercomputer - Era

- ERA - 元 (Yuan)
  - CAS HPC from T to P - new period
  - Peak performance - 2.36 Petaflops
  - The 6th generation supercomputer in SCCAS

- Installation
  - Site: Huairou Branch Center of CNIC
  - Two stages
    - Stage 1: announced on June 19, 2014
    - Stage 2: announced on April, 2016
Huairou Branch Center of CNIC

Facility Area: 19000 m², Power: 9 MW (extensible)
Software on Era

- Compiler, Math Libs, OpenMP, MPI
- HPC Software automatic installation tool - Clussoft
- Matlab, MolCAS, Q-Chem, Amber, CHARMM, Gaussian

**Clussoft 部分组件**

- **clussoft-base** — 编译器、函数库、并行库等
- **clussoft-benchmark** — 常用benchmark工具（CPU、内存、网络、I/O等）
- **clussoft-md** — 常用开源分子动力学、Monte Carlo等应用软件
- **clussoft-qchem** — 常用开源计算化学、物理、材料类应用软件
- **clussoft-bio** — 生命科学领域常用开源应用软件
- **clussoft-gpu** — GPGPU开发环境及应用软件
- **clussoft-mets** — 气象海洋类应用软件
Supercomputing Environment of CAS – China ScGrid

• Three-tier grid
  - 1 head center
  - 9 regional centers
  - 18 institution centers, 11 GPU centers

• Applications - 120
  - Computational Chemistry, Physics, Material science, Life science, CFD, Industrial computing

• Status (by Dec 2016)
  - #User > 900
  - #Job > 700 000
  - Walltime ≈ 150M cpu·hrs
**CNGrid environment**

- **15 sites**
  - SCCAS (Beijing, main node, operation center)
  - SSC (Shanghai, main node)
  - NSCTJ (Tianjin)
  - NSCSZ (Shenzhen)
  - NSCJN (Jinan)
  - NSCCS (Changsha)
  - THU (Beijing)
  - IAPCM (Beijing)
  - USTC (Hefei)
  - XJTU (Xi’an)
  - SIAT (Shenzhen)
  - HKU (Hong Kong)
  - SDU (Jinan)
  - HUST (Wuhan)
  - GSCC (Lanzhou)
CNGrid HPC

- Tianhe-1A
  - #1 TOP 500, 2010
  - 4701 TFlop/s, 186,368 cores
  - Tianjin

- Sunway Blue Light
  - #14 TOP 500, 2011
  - ShenWei processor
  - 1070.2 TFlop/s, 137,200 cores
  - Jinan

- Nebulae
  - #2 TOP 500, 2010
  - 2984.3 TFlops/s, 120,640 cores
  - Shenzhen

- Dawning 5000A
  - #11 TOP 500, 2008
  - 233.5 Tflop/s, 30,720 cores
  - Shanghai

- DeepComp 7000
  - #19 TOP 500, 2008
  - 146.0 TFlop/s, 12,216 cores
  - Beijing
CNGrid HPC

- Tianhe-2
- #1 TOP 500, 2013-2015
- 54,902.4 TFlop/s, 3,120,000 cores
- 17,808.00 kW
- Guangzhou
CNGrid HPC

- Sunway TaihuLight
- #1 TOP 500, 2016
- Sunway processor: SW26010
- 125.436 PFlops, 10,649,600 cores
- 15,371 kW
- Wuxi
CNGrid new model
SCE - Middleware for HPC Cloud

- Developed by SCCAS
- SCE
  - Scientific computing
  - Lightweight
  - Stable
- Diversity
  - CLI
  - Portal
  - GUI
  - API

*International Patent (PCT/CN2011/071640)*
Part III

HPC Applications
Phase Field Simulations on the Sunway TaihuLight

compact Exponential Time Differencing (cETD) scheme

+ localized integration subdomain coupling

Scalable compact Localized Exponential Time Differencing (ScLETD) scheme

+ Efficient implementation on leading class supercomputer

Extreme-scale CH simulations

Final List - ACM Gordon Bell Prize 2016!
10M-Core Scalable Fully-Implicit Solver for Nonhydrostatic Atmospheric Dynamics

- **2016 ACM Gordon Bell Prize** (Institute of Software, CAS)
  - Algorithm innovations on many-core friendly fully implicit solver
  - Scaled to **10.6M cores** on Sunway TaihuLight (cf: **1.6M cores** in GB’15)
  - Sustained **7.95 DP-PF** (cf: **0.69 DP-PF** in GB’15)
  - At 488m-res with **772B DOFs** (cf: **602B DOFs** in GB’15)
  - Achieved **89.5X** speedup to explicit solver ~ Exa-scale perf in explicit

![Graph showing resolution against total number of cores](image-url)
Meteorological Simulations

- **CMA Unified Atmospheric Chemistry Environment CUACE/Dust**
  - Analyzing the atmosphere changes in China
  - Our work is parallel, coupling with MM5, and performance optimization, 3dvar data assimilation software package
  - CUACE-dust has been used for real time forecasting in CMA.(72 hours, 2times every day)
  - Surface dust outbreak, dust transport, cohesion, sedimentation, cleanup procedure
  - Good accuracy and has greatly reduced losses from atmospheric hazards
Meteorological Simulations

- Development of Regional Integrated Environment Modeling System (RIEMS)

- [BC] SEA SALT optical thickness of BC

- [SO2] SO4 optical thickness of sulfate
CAS earth system model

- We participate in the development of CAS earth system model
- Run CMIP6 experiments on “era”
ATLAS (A Toroidal LHC ApparatuS) is one of the seven particle detector experiments constructed at the Large Hadron Collider (LHC), a particle accelerator at CERN (the European Organization for Nuclear Research) in Switzerland. The experiment is designed to take advantage of the unprecedented energy available at the LHC and observe phenomena that involve highly massive particles which were not observable using earlier lower-energy accelerators. It was one of the two LHC experiments involved in the discovery of a particle consistent with the Higgs boson in July 2012.
CNGrid & ATLAS

- CNGrid support ATLAS experiment
  - SCEAPI works as a bridge between ARC-CE middleware and CNGrid resources
  - ATLAS simulation jobs run on Chinese HPCs including TianHe-1A and ERA
CCFD- parallel CFD software

Aerodynamic Computation
Multi-Body Separation
Aeroelastic Flutter
Scale up to over 10,000 cores

**CCFD-MGMB**
- Multi-block structured grids
- Implicit time stepping (via pseudo-time iteration)
- Multi-grid acceleration

**CCFD-MBS**
- Chimera grids
- Parallel grid assembling

**CCFD-AE**
- Grid deformation
- Couple with structural analysis software

---

TianHe II test
DLR-F6 model

---

**Graph:**
- **Speedup** vs **Cores**
  - **Ideal**
  - **CCFD-MGMB**

MIC-accelerated DPD

Flowchart of DPD simulation

Progress in IPCC

- Implement our DPD code on a single MIC, and achieve more than 5 times speedup than a single CPU core.
SC_Tangram: a Charm++-Based Parallel Framework for Cosmological Simulations

- Hide complex parallel technologies.
- Provide a platform for composing components together into a complex application.

1. Modularity
   - componentization
   - collaboration

2. Runtime Adaptivity
   - Fault Tolerance
   - Load Balance

3. Domain Specification
   - Cosmological hydrodynamics
   - N-body

**Figure 1. Multi-Layers Design of SC_Tangram**
SC_Tangram: a Charm++-Based Parallel Framework for Cosmological Simulations

- Domain Specific Data Types
  - DATA TYPES “GF” FOR UNIFORMED MESH AND “GP” FOR PARTICLES IN MESH ON CPU-CLUSTERS

- Applications

**Execution Time of Ten Steps on the Scale of 1024^3 and Strong Scaling**

**Execution Time of One Step on the Scale of 134217728 Particles on the Mesh**
Application of HPSEPS in first principle calculation software MESIA

HPSEPS, a parallel eigenproblem solver developed by SC, CAS is adopted in a multi-scale first principle calculation software MESIA, developed by the Key Lab. Of Quantum Information, CAS.

◆ MESIA produced the correct energy sequences for B3 (Zinc Blend) and B4 (Wurtzite)

• B20 cluster

We developed MPI-GPU eigensolver for dense eigenproblem

Computational Throughput: 16GPUs = 512CPUs
Large scale three dimensional fragment method on GPU

1. Accuracy
   (climb Jacob's ladder)
2. Temporal scale (from fs to seconds)
   (new algorithms, like accelerated MD)
3. Size scale (mesoscale problems)
   (divide & Conquer methods)

Titan GPU: 88% of total computing power
But NO plane wave code on GPU.

LS3DF: Linear Scaling Three Dimensional Fragment Method

Collaborate with Lin-Wang Wang, LBNL

This project is supported by INCITE program and CSC
Large scale three dimensional fragment method on GPU

On Titan Supercomputer:
3877 atom Si system, 1500 computing nodes (total 24000 CPU cores) compared with 1500 GPU cards, LS3DF_GPU has a speedup of 10.5x.
Fast Parallel Direct Solver for Large linear system

**HSS algorithm**

- *hierarchically semiseparable matrix*, Chandrasekaran, Gu, Xia, et al
- **three steps**: HSS compression, ULV factorization & **ULV solver**
- **Complexity**: $O(kN^2)$ for step 1, $O(kN)$ for step 2&3; **storage**: $O(kN)$, $k$: block rank
- Recursive Low-rank compression by tree ($c1$&$c2$: children of node $j$)

$$D_j = \begin{pmatrix} D_{c1} & U_{c1} B_{c1} V_{c2}^T \\ U_{c2} B_{c2} V_{c2}^T & D_{c2} \end{pmatrix}, \quad U_j = \begin{pmatrix} U_{c1} R_{c1} \\ U_{c2} R_{c2} \end{pmatrix}, \quad V_j = \begin{pmatrix} V_{c1} W_{c1} \\ V_{c2} W_{c2} \end{pmatrix}$$

*HSS matrix structure & HSS tree*
Result: HSS solver vs ScaLapack

Dense linear system:

\[
\begin{bmatrix}
E_{inc}(u_n)
\end{bmatrix} = \begin{bmatrix}
Z_{mnl}
\end{bmatrix} \begin{bmatrix}
\mathbf{J}_n
\end{bmatrix},
\]

where \( H \) is cylinder Hankel function, \( \text{inc} \) means incident field

\( N: 32768, \) HSS tree level: 8, block rank: 32

Total runtime, compared with pzgesv of ScaLapack (using MKL)
New Energy Power Generation

- New energy simulation system
  - new energy time series modeling
  - time series power generation simulation
  - stochastic power generation simulation
- save at least ¥10 billion every year
- Increase new energy at least 1 billion kwh
  = saving the coal nearly 400,000 tons
  reducing carbon dioxide emissions by 800,000 tons
New Energy Power Generation

- CMIP (Chinese Mixed Integer Programming Solver)
- Mid long term wind power forecast
- Power network topology visualization
Visualization Platform for Large Data

- **GPVis**
  - Parallel and GPU acceleration
  - Support Scalars and Vectors
  - Plenty of visualization method
    - Isosurface, Isoline, Colormap
    - Streamline, pathline, LIC
    - Parallel coordinate, figures
  - Flexible user interface
  - Two usage modes
    - C/S
    - Web
Visual Analysis Examples with GPVis

- Visual analysis of ENZO (El Nino-Southern Oscillation) ensemble simulation to verify formation of ENZO
  - Flexible slice operation
  - Isoline on Isosurface
  - Pathlines with efficient sends setting
  - Simulation data, measured wind field, terrain
Visualization of Large Scale Ocean Flow
Our plan

• Last 5 year plan, we have completed many applications using more than 10 thousands cores, this 5 year plan, we will continue to support applications in scientific research areas

• Build cloud for science and technology
  – 50+ pflops, 100 Gb network

• Establish systems for intellectual CAS

• Modify the current Grid environment

• Enroll exascale project
Thank you!

chi@sccas.cn