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, Taibei

Outline

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

Part I

Supercomputing in CNIC

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Supercomputing Innovation Alliance - Membership (50+)

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Parallel algorithm & software development

- Numerical linear algebra
- Computational fluid dynamics
- Fast multipole method
- Computational chemistry, physics, biology
- Fft, fmm, hpseps, ccfd, sc_tangram

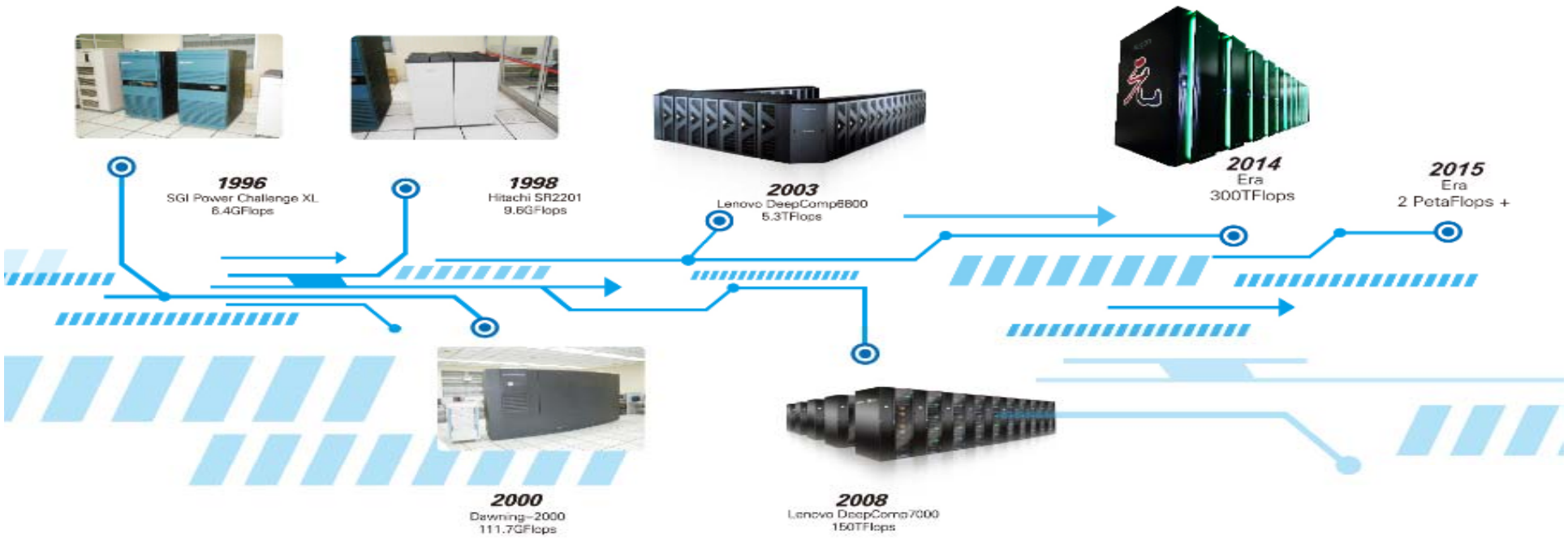
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

Part II

HPC Infrastructures

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

The Supercomputer History in SCCAS (1996-2016)



6.4GFlops → 5TFlops → 150TFlops → 2.3PFlops

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

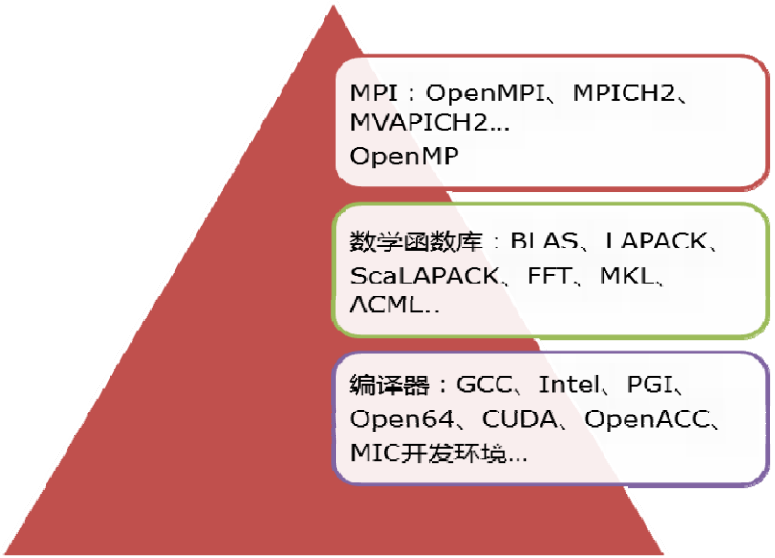
Huairou Branch Center of CNIC

Facility Area: 19000 m^2, Power: 9 MW(extensible)

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

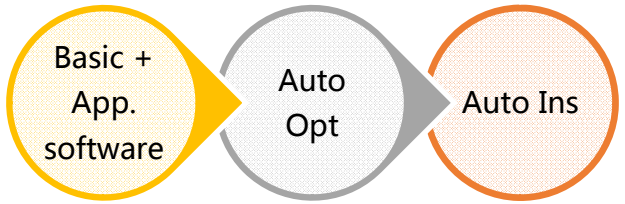
Software on Era

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



Clussoft 部分组件

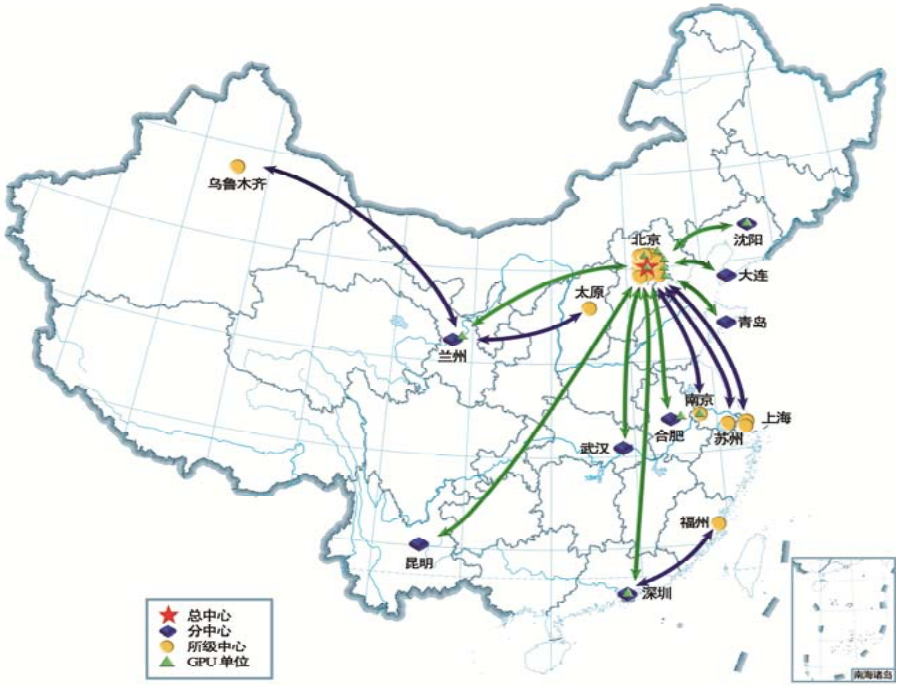
- clussoft-base——编译器、函数库、并行库等
- clussoft-benchmark——常用benchmark工具（CPU、内存、网络、I/O等）
- clussoft-md——常用开源分子动力学、Monte Carlo等应用软件
- clussoft-qchem——常用开源计算化学、物理、材料类应用软件
- clussoft-bio——生命科学领域常用开源应用软件
- clussoft-gpu——GPGPU开发环境及应用软件
- clussoft-mets——气象海洋类应用软件



中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

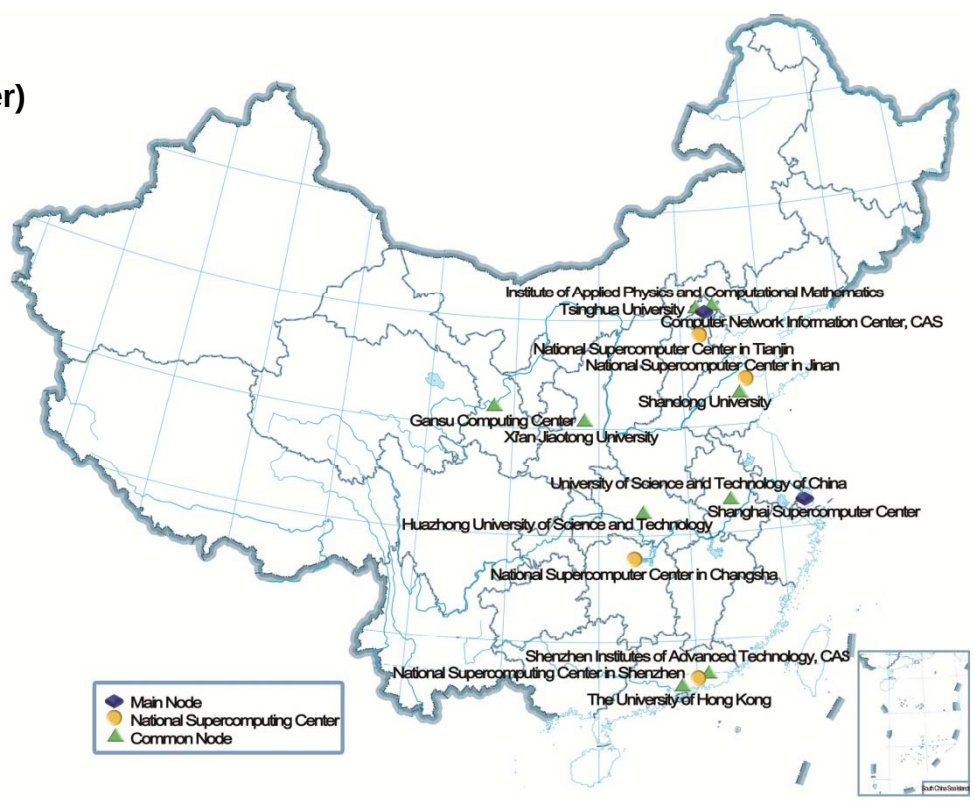


中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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)**



中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

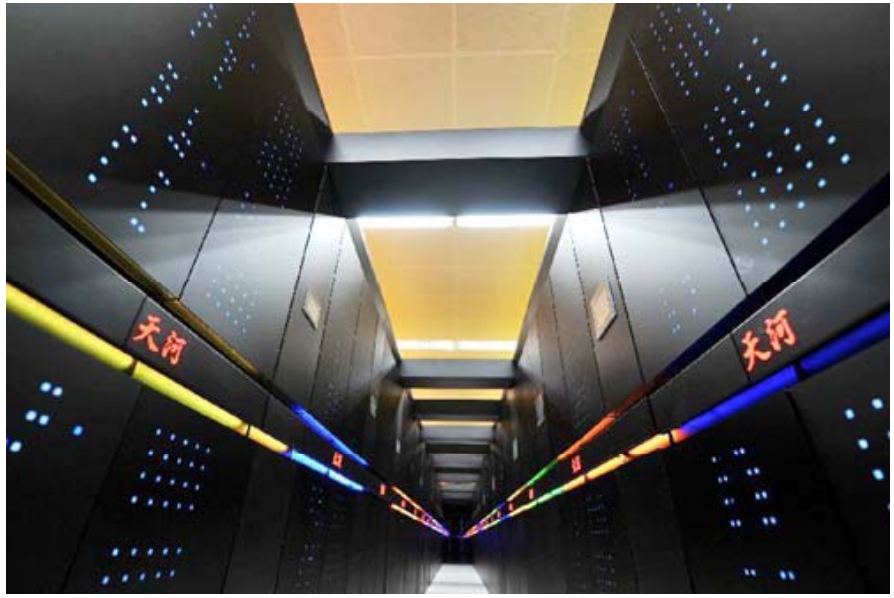
- Nebulae
- #2 TOP 500, 2010
- 2984.3 TFlops/s, 120,640 cores
- Shenzhen

- 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

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

CNGrid HPC



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

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

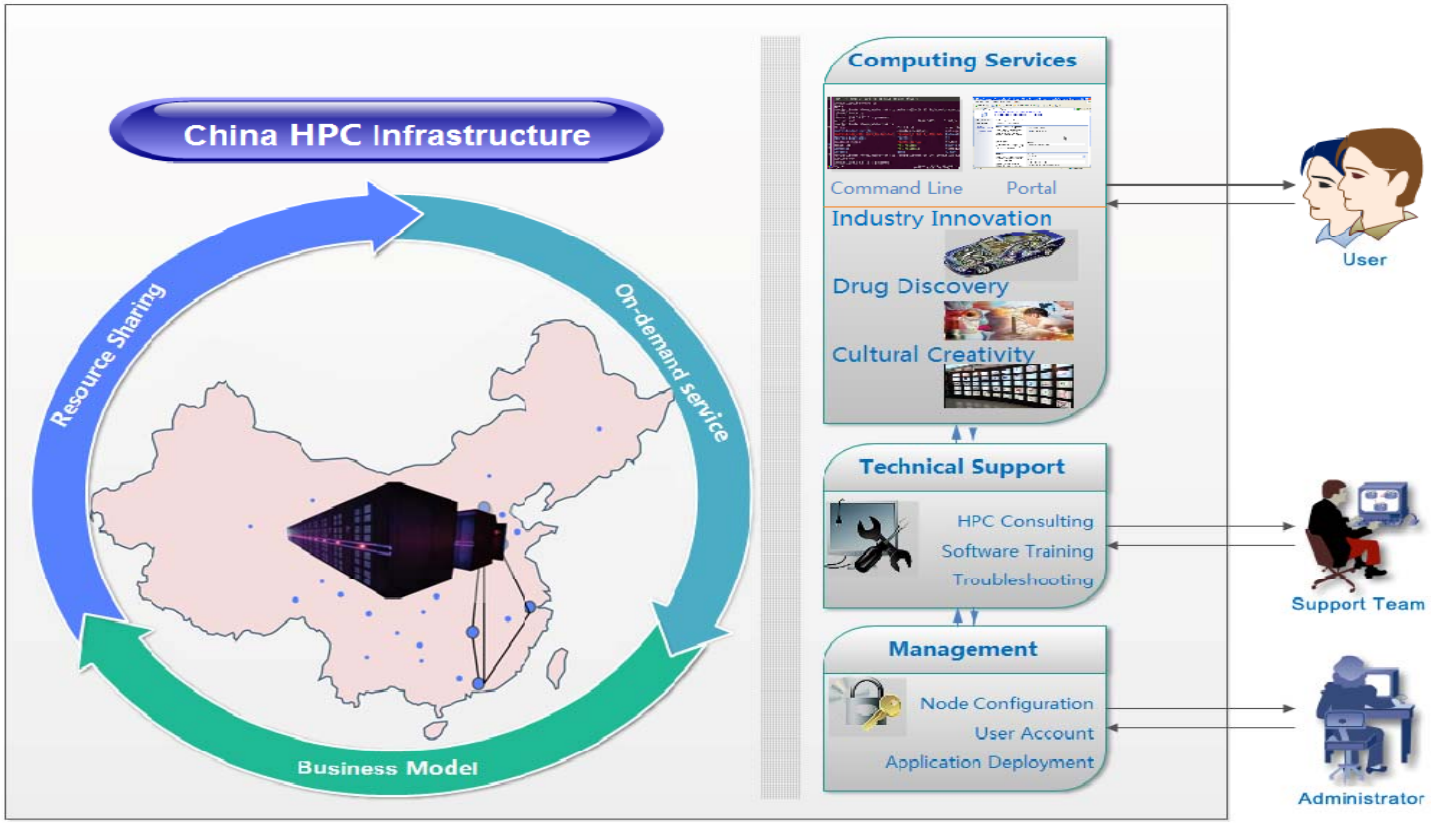
CNGrid HPC

- Sunway TaihuLight
- #1 TOP 500, 2016
- Sunway processor: SW26010
- 125.436PFlops, 10,649,600 cores
- 15,371 kW
- Wuxi

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

CNGrid new model

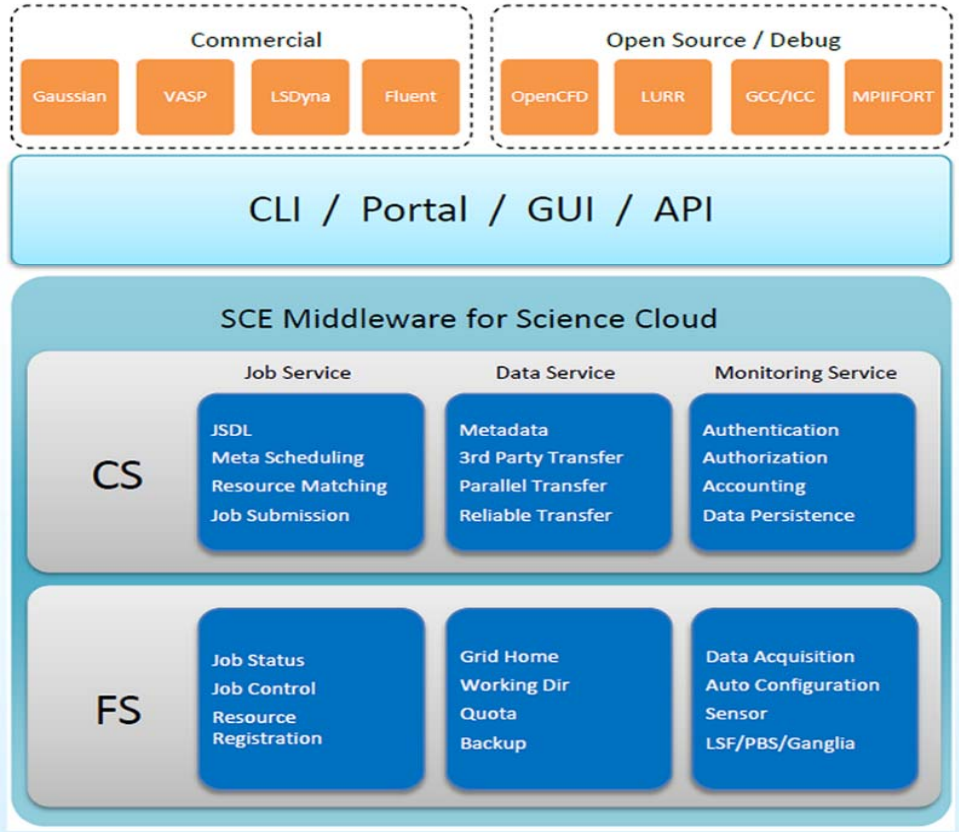


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

SCE - Middleware for HPC Cloud

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

International Patent
（PCT/CN2011/071640）



Part III

HPC Applications

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Phase Field Simulations on the Sunway TaihuLight

compact Exponential Time Differencing (cETD) scheme

+

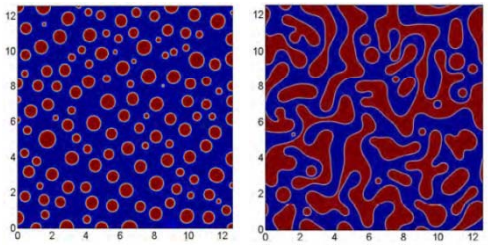
localized integration subdomain coupling

↓

Scalable compact Localized Exponential Time Differencing (ScLETD) scheme

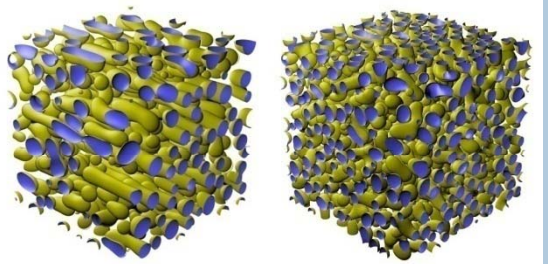
+

Efficient implementation on leading class supercomputer



↓

Extreme-scale CH simulations



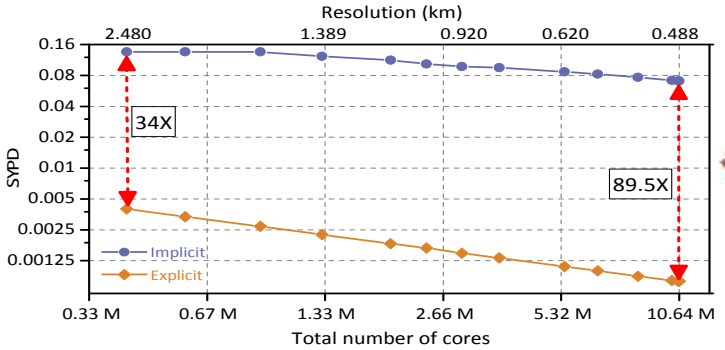
Final List - ACM Gordon Bell Prize 2016!

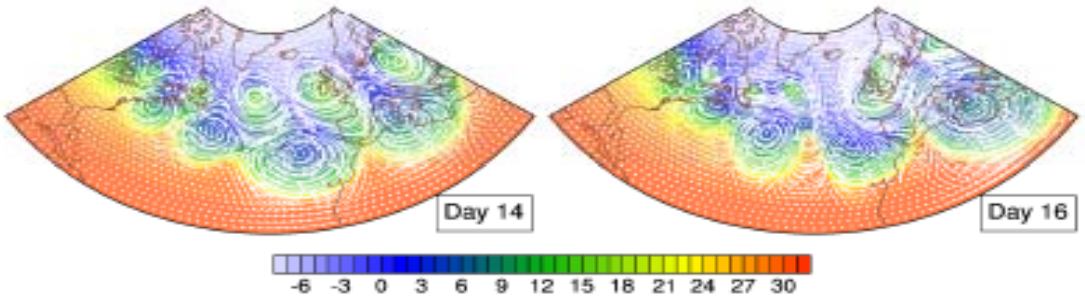
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

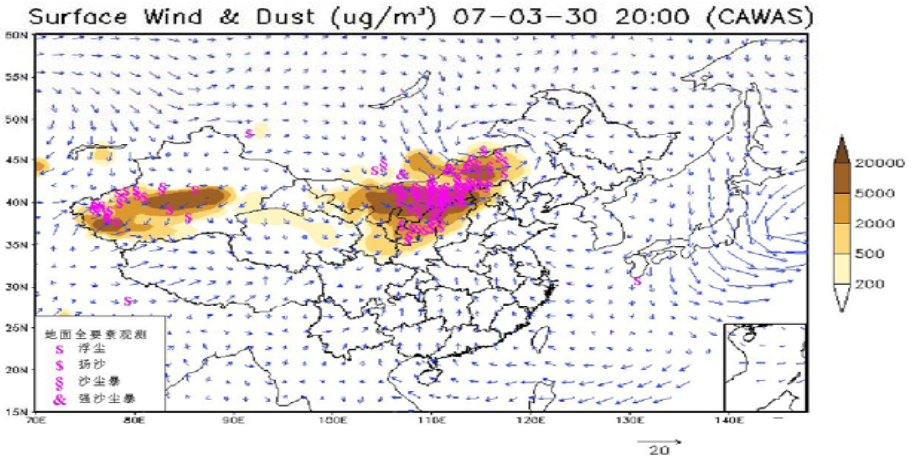


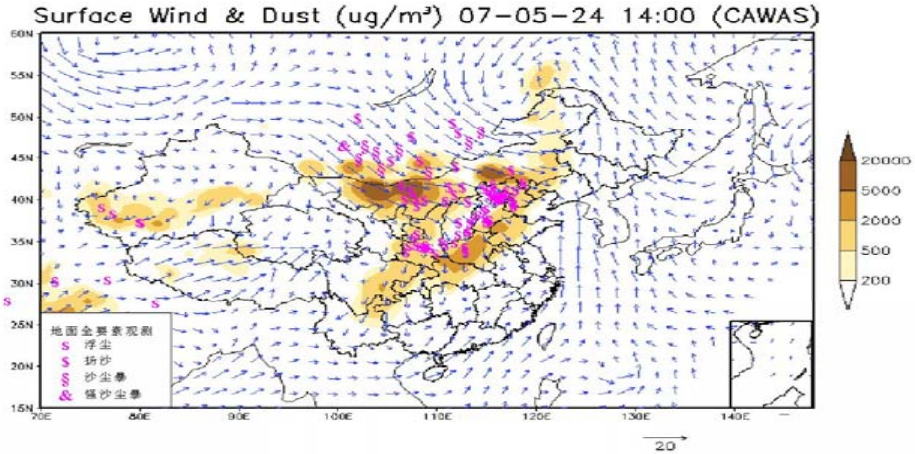


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

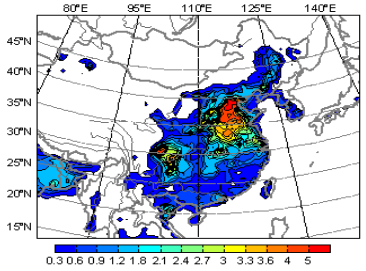




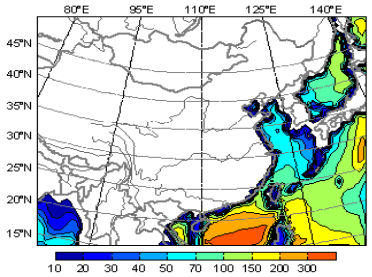
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Meteorological Simulations

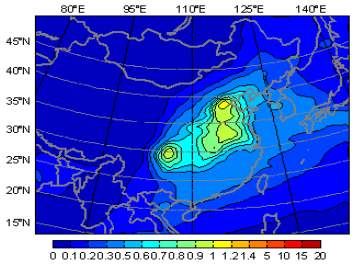
- **Development of Regional Integrated Environment Modeling System-RIEMS**



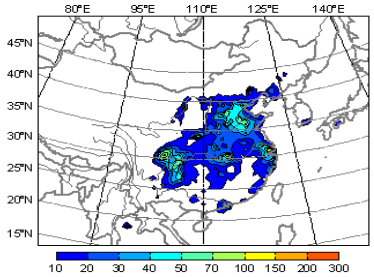
BC



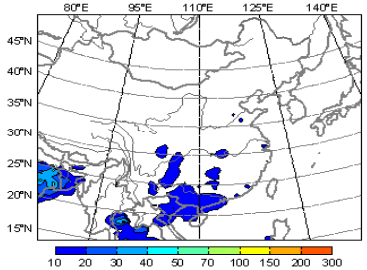
SEA SALT



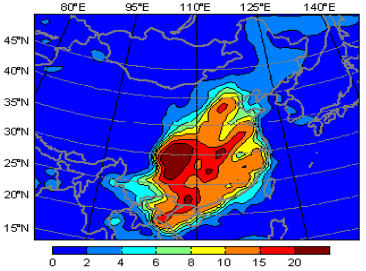
optical thickness of BC



SO2



SO4



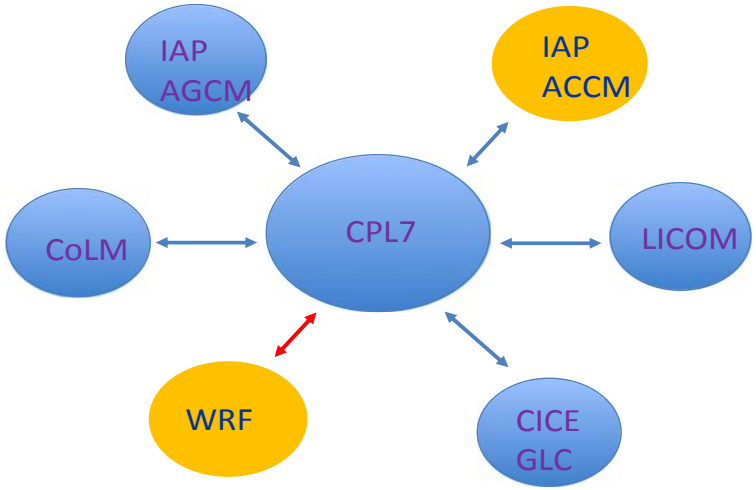
optical thickness of sulfate

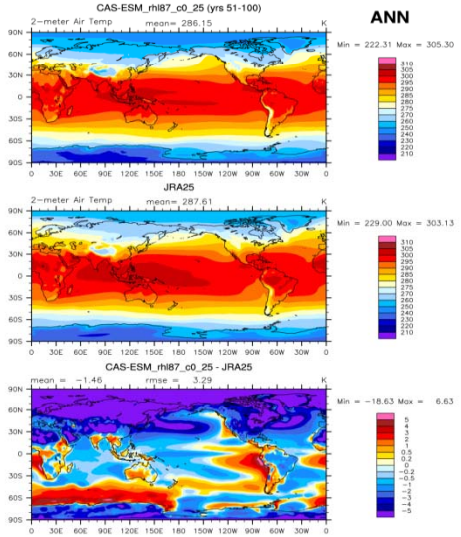
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

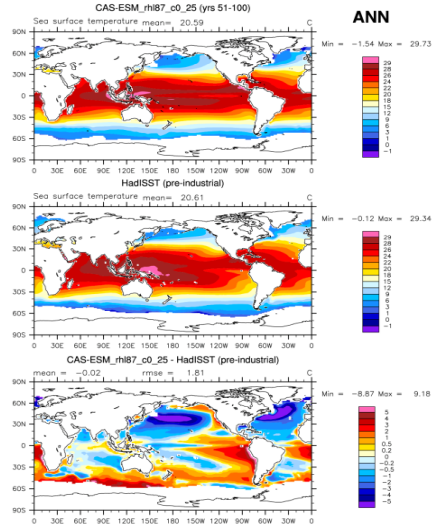
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

CAS earth system model

- We participate in the development of CAS earth system model
- Run CMIP6 experiments on “era”



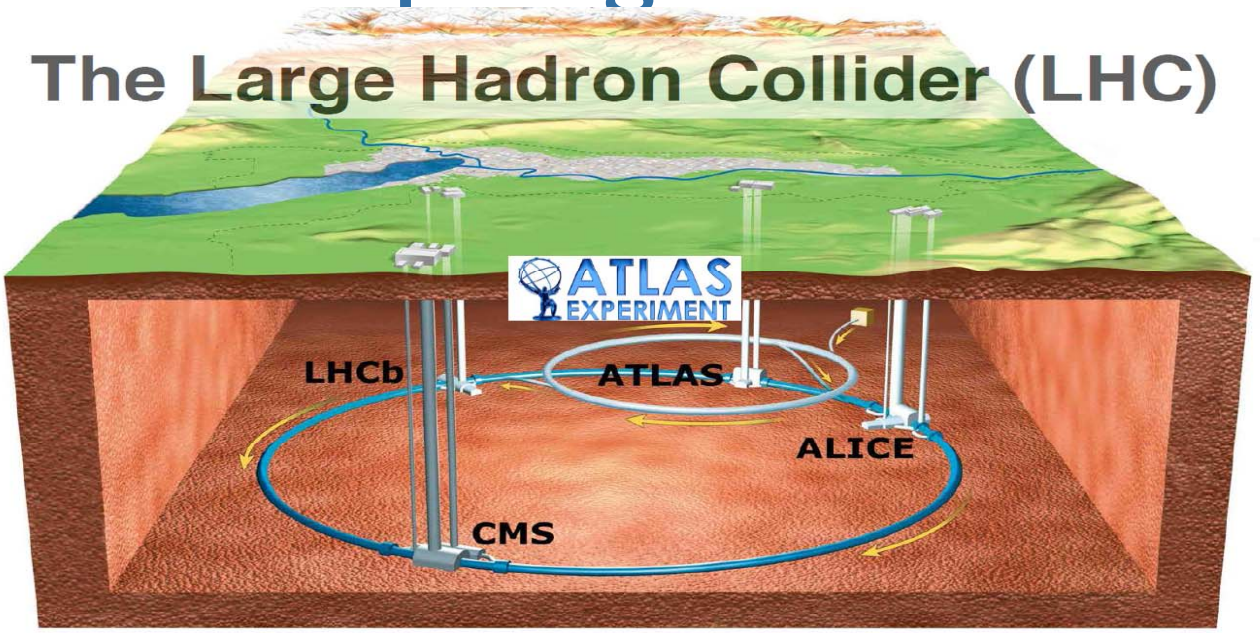




中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

ATLAS Monte Carlo Simulation Computing Platform

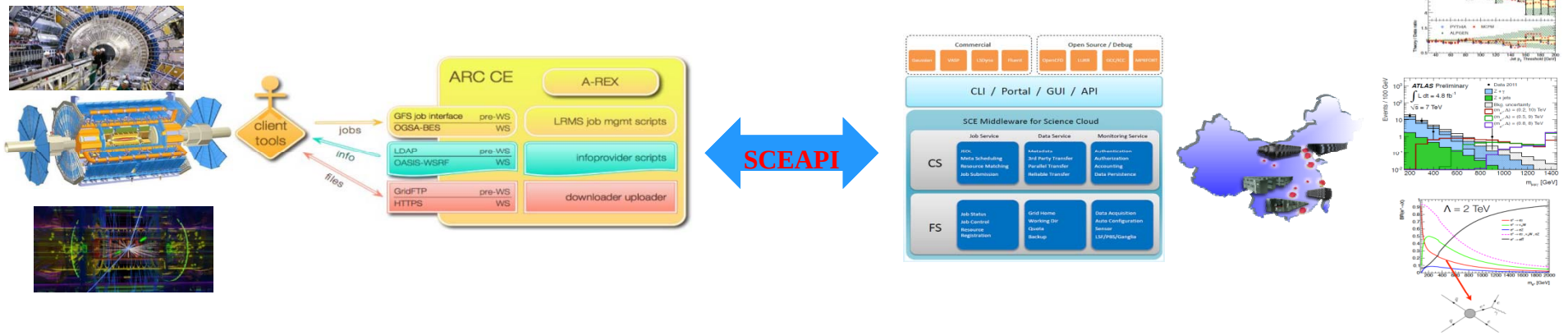


ATLAS (A Toroidal LHC ApparatuS) is one of the seven particle detector experiments constructed at the Large Hadron Collider (LHC), a particle accelerator at CERN (the European Organization for Nuclear Research) in Switzerland. The experiment is designed to take advantage of the unprecedented energy available 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.

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

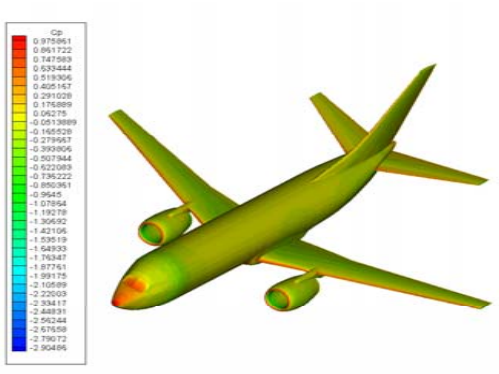
CNGrid & ATLAS

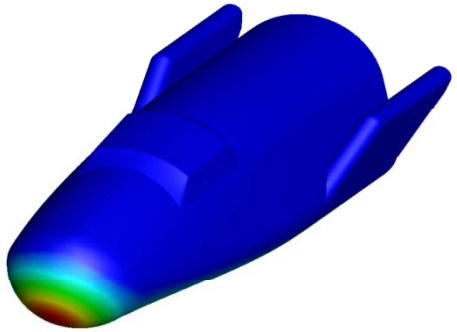
- CNGrid support ATLAS experiment
  - SCEAPI works as a bridge between ARC-CE middleware and CNGrid resources
  - ATLAS simulation jobs run on Chinese HPCs including TianHe-1A and ERA

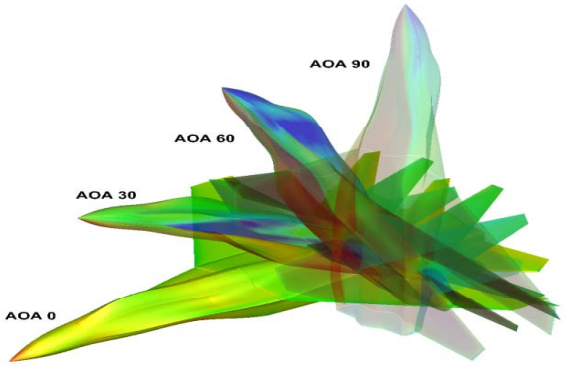


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

CCFD- parallel CFD software

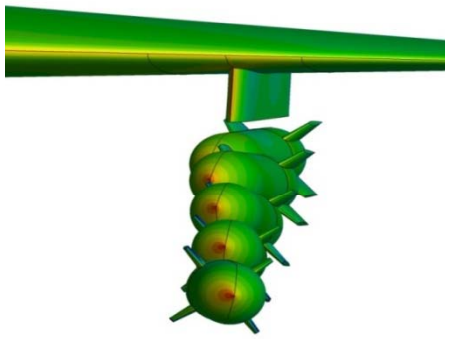


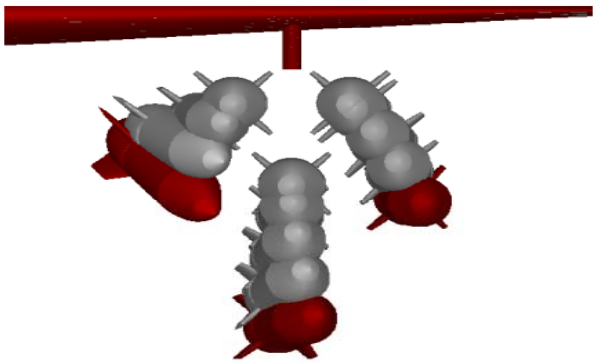




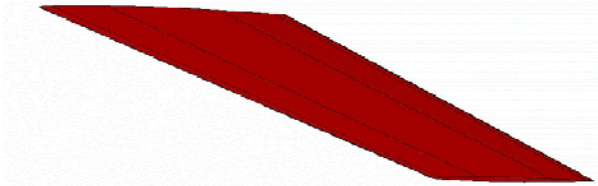
Aerodynamic Computation

Multi-Body Separation





Aeroelastic Flutter

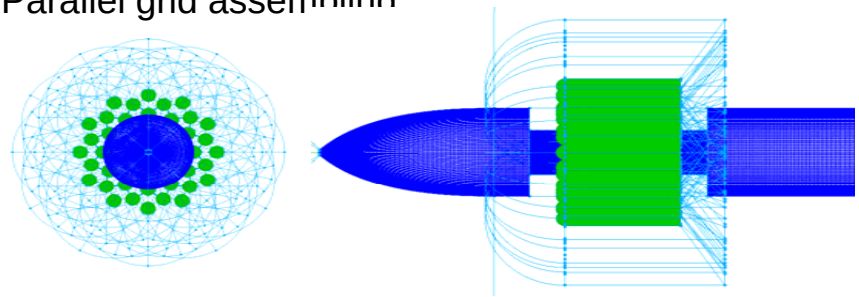


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

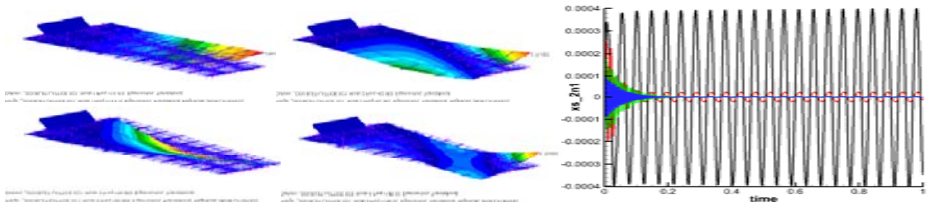
Scale up to over 10,000 cores

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

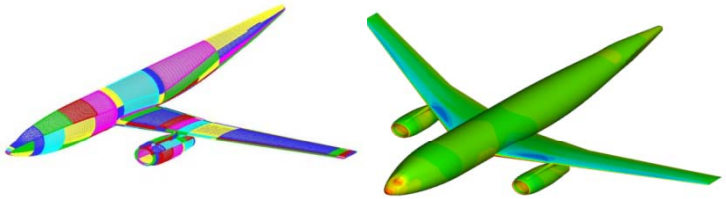
CCFD-MBS
- Chimera grids
- Parallel grid assembling

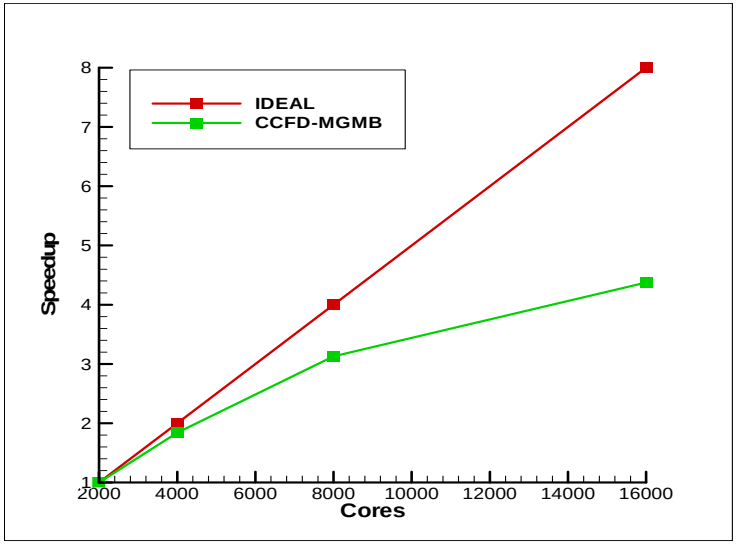


CCFD-AE
- Grid deformation
- Couple with structural analysis software



TianHe II test
DLR-F6 model

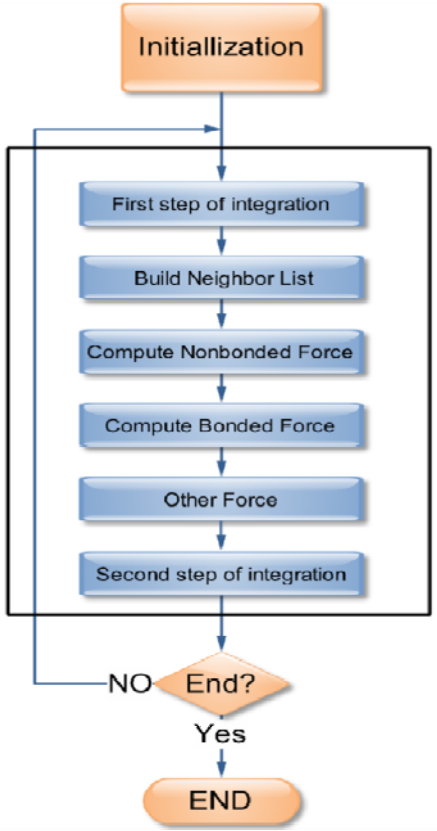




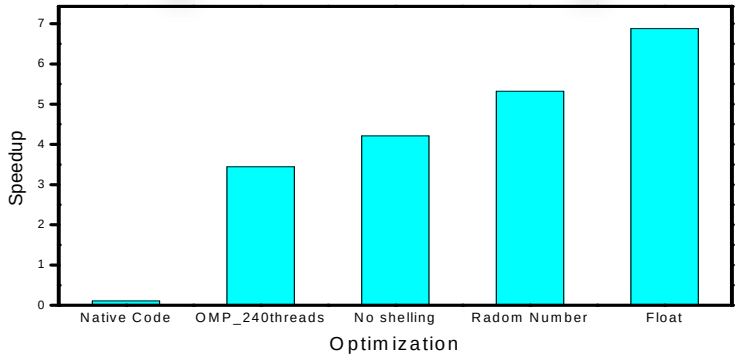
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

MIC-accelerated DPD

◆**Flowchart of DPD simulation**



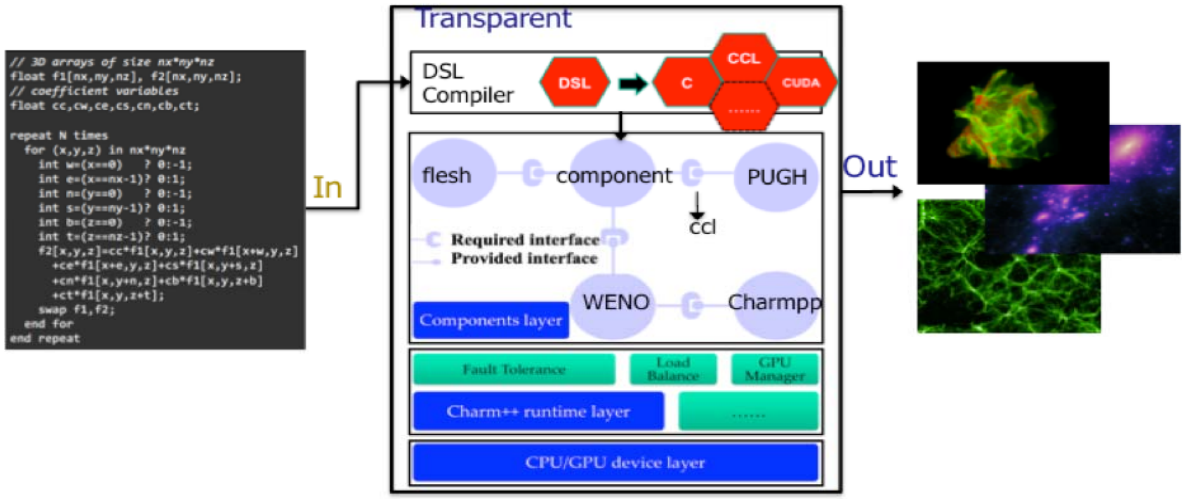
◆ **Progress in IPCC**



- 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.**



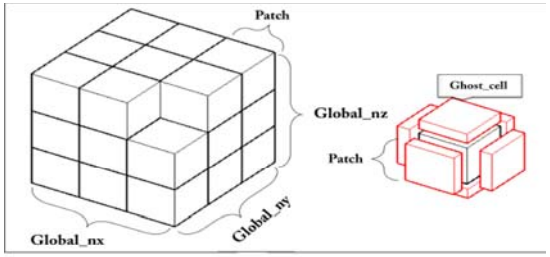
1. Modularity
   - componentization
   - **collaboration**
2. Runtime Adaptivity
   - Fault Tolerance
   - Load Balance
3. Domain Specification
   - Cosmological hydrodynamics
   - N-body

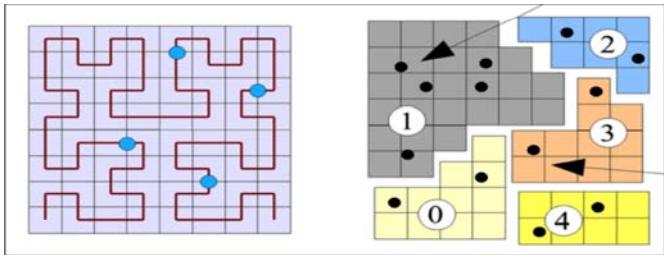
FIGURE 1. MULTI-LAYERS DESIGN OF SC_TANGRAM

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

SC_Tangram: a Charm++-Based Parallel Framework for Cosmological Simulations

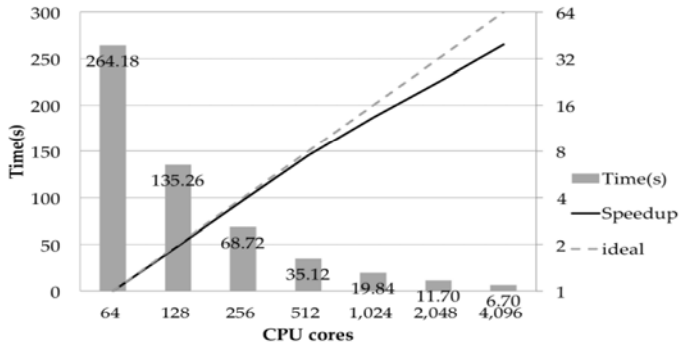
- **Domain Specific Data Types**



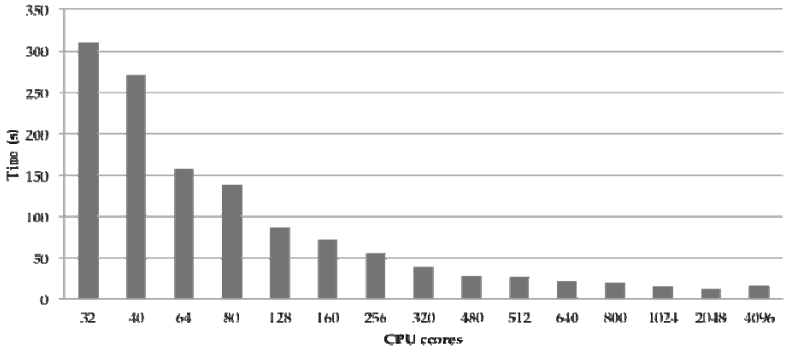


DATA TYPES "GF" FOR UNIFORMED MESH AND "GP" FOR PARTICLES IN MESH ON CPU-CLUSTERS

- **Applications**



EXECUTION TIME OF TEN STEPS ON THE SCALE OF 1024^3 AND STRONG SCALING

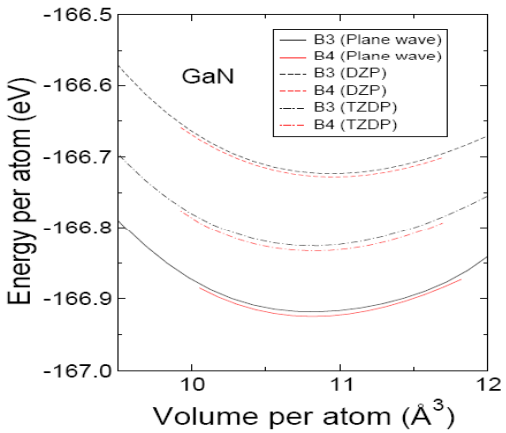


EXECUTION TIME OF ONE STEP ON THE SCALE OF 134217728 PARTICLES ON THE MESH

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

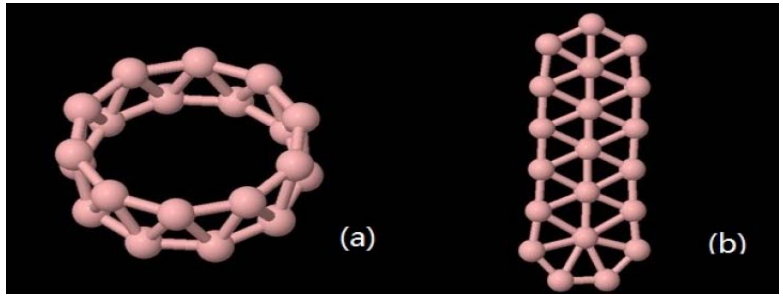
Application of HPSEPS in first principle calculation software MESIA

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**



	Plane wave	DZP	SIESTA
double ring (a)	0.00 eV	0.00 eV	0.00 eV
candidate1 (b)	2.74 eV	2.52 eV	-0.13 eV

The energy sequence calculated using SIESTA is incorrect, MESIA gave the correct one.

We developed MPI-GPU eigensolver for dense eigenproblem

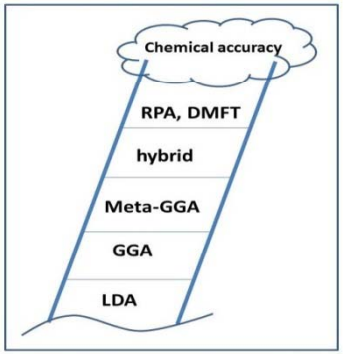
Computational Throughput: 16GPUs= 512CPUs

中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

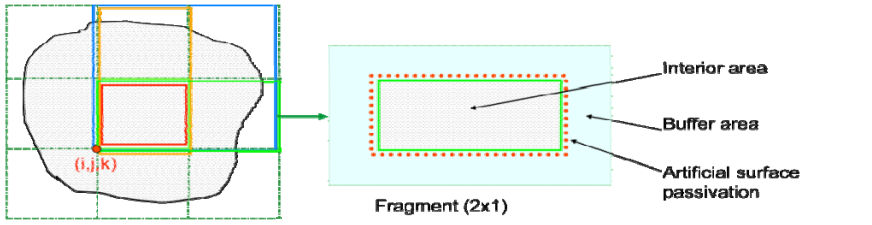
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

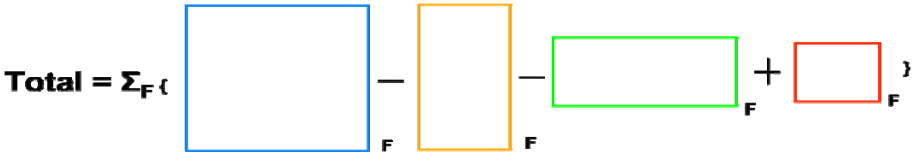
Large scale three dimensional fragment method on GPU

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.





Boundary effects are (nearly) cancelled out between the fragments

$$System = \sum_{i,j,k} \{F_{222} + F_{211} + F_{121} + F_{112} - F_{221} - F_{212} - F_{122} - F_{111}\}$$

LS3DF: Linear Scaling Three Dimensional Fragment Method

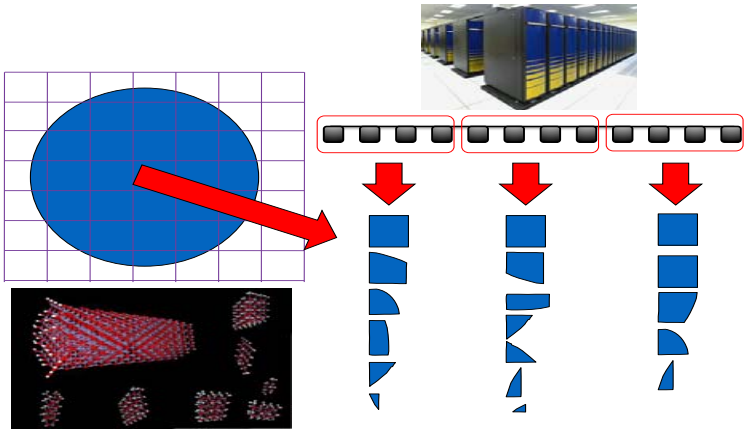
Collaborate with Lin-Wang Wang, LBNL

This project is supported by INCITE program and CSC

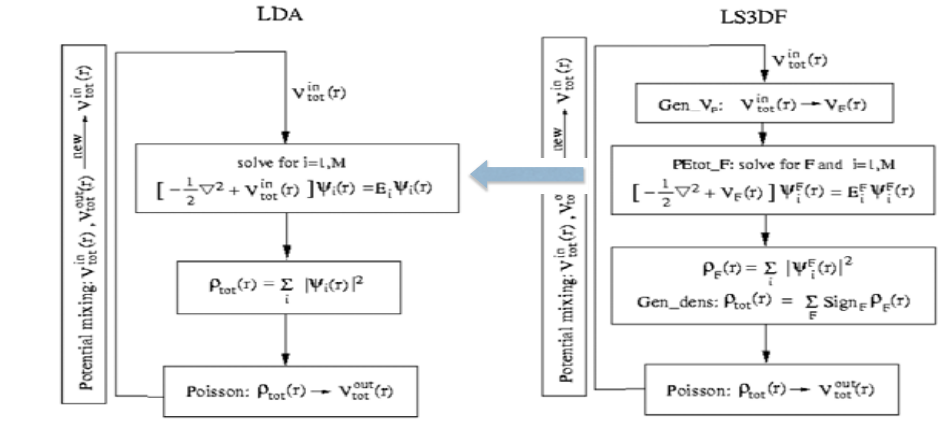
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

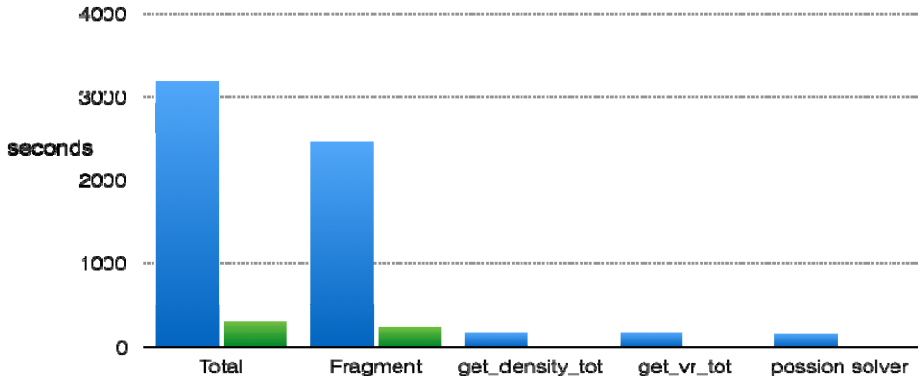
Large scale three dimensional fragment method on GPU



LS3DF data distribution



LS3DF algorithm compared with LDA algorithm



LS3DF-GPU speedup compared with CPU code

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.

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Fast Parallel Direct Solver for Large linear system

HSS algorithm

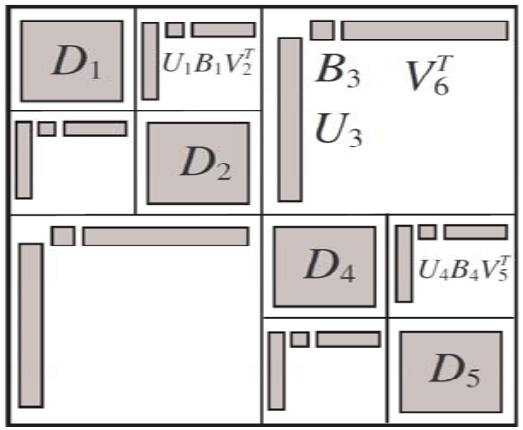
✓*hierarchically semiseparable matrix,* Chandrasekaran, Gu, Xia, et al

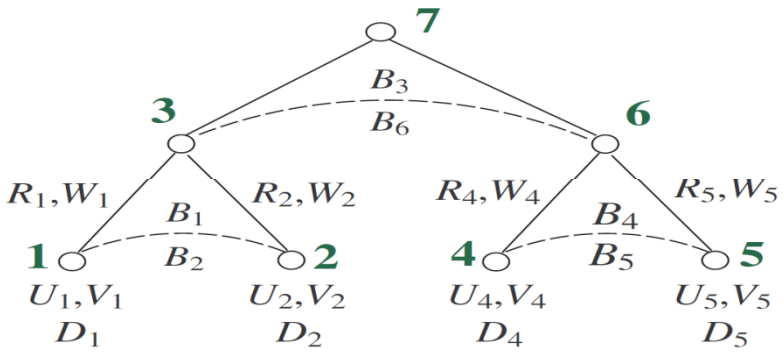
✓three steps: HSS compression, ULV factorization & *ULV solver*

✓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)

$$D_j = \begin{pmatrix} D_{c_1} & U_{c_1}B_{c_1}V_{c_2}^T \\ U_{c_2}B_{c_2}V_{c_2}^T & D_{c_2} \end{pmatrix}, \quad U_j = \begin{pmatrix} U_{c_1}R_{c_1} \\ U_{c_2}R_{c_2} \end{pmatrix}, \quad V_j = \begin{pmatrix} V_{c_1}W_{c_1} \\ V_{c_2}W_{c_2} \end{pmatrix}$$





HSS matrix structure & HSS tree

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

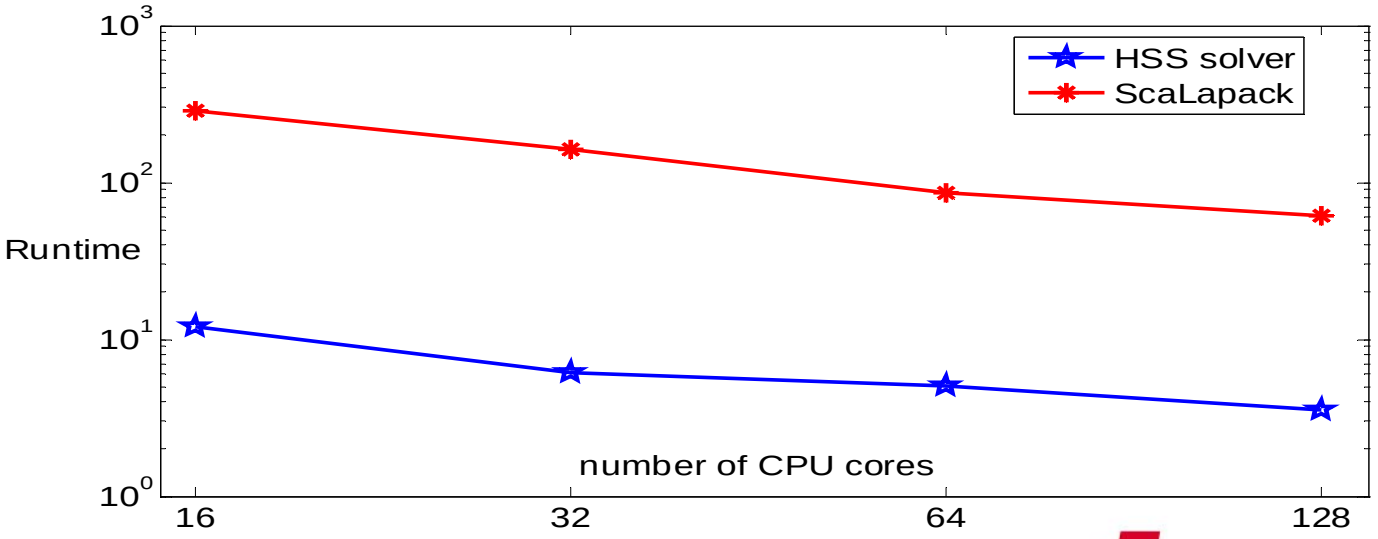
Result: HSS solver vs ScaLapack

Dense linear system: $[E_{inc}(u_m)] = [Z_{mn}][j_n], Z_{mn} = \frac{k\eta}{4} d_n H_0^{(2)}(kR_{mn})$

✓in which H is cylinder Hankel function, inc means incident field

✓N: 32768, HSS tree level: 8, block rank: 32

✓Total runtime, compared with pzgesv of ScaLapack (using MKL)



中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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

国家电网 STATE GRID
中国电力科学研究院 CHINA ELECTRIC POWER RESEARCH INSTITUTE

国家电网 STATE GRID
国网冀北电力有限公司 STATE GRID JIBEI ELECTRIC POWER COMPANY

Computer Network Information Center
Chinese Academy of Sciences

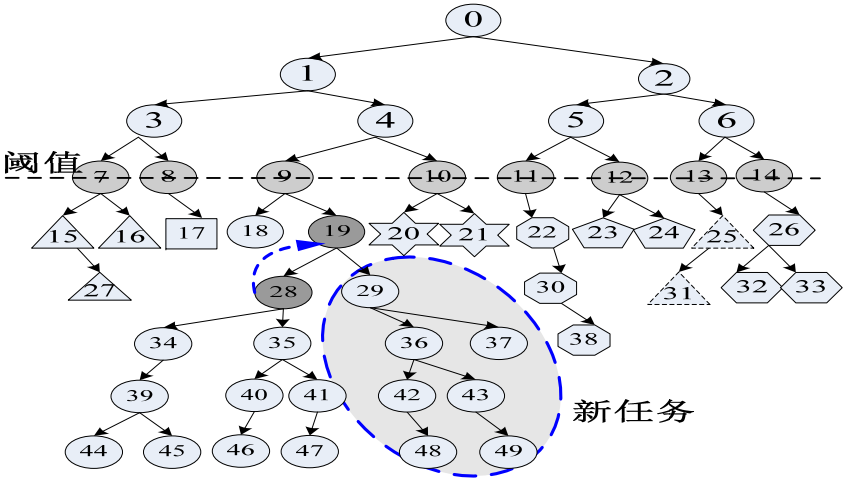
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

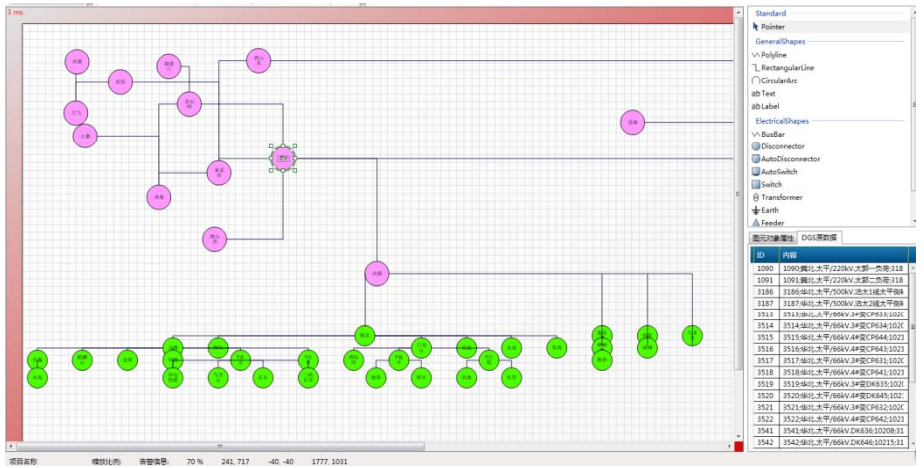
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

New Energy Power Generation

- CMIP（Chinese Mixed Integer Programming Solver）
- Mid long term wind power forecast
- Power network topology visualization



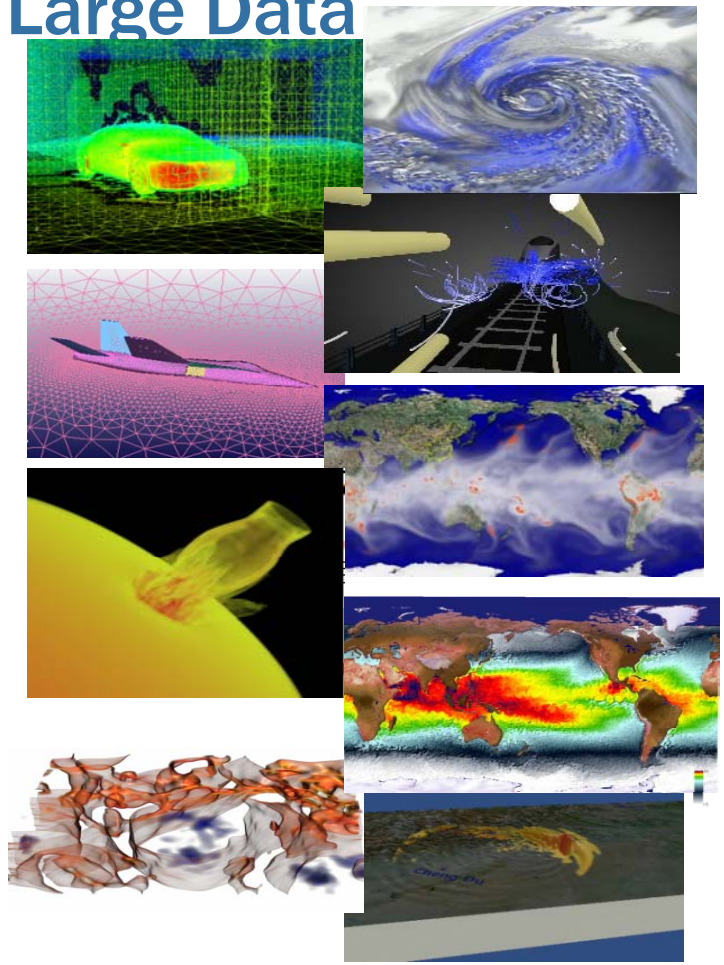


中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Visualization Platform for Large Data

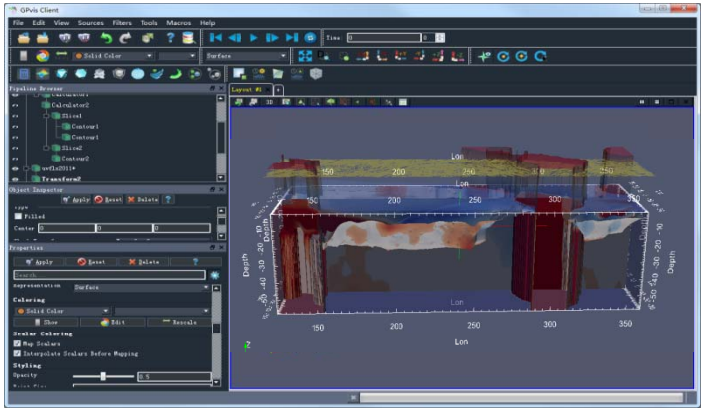
- GPVis
  - Parallel and GPU acceleration
  - Support Scalars and Vectors
  - Plenty of visualization method
    - Isosurface, Isoline, Colormap
    - Stremline, pathline, LIC
    - Parallel coordinte, figures
  - Flexible user interface
  - Two usage modes
    - C/S
    - Web

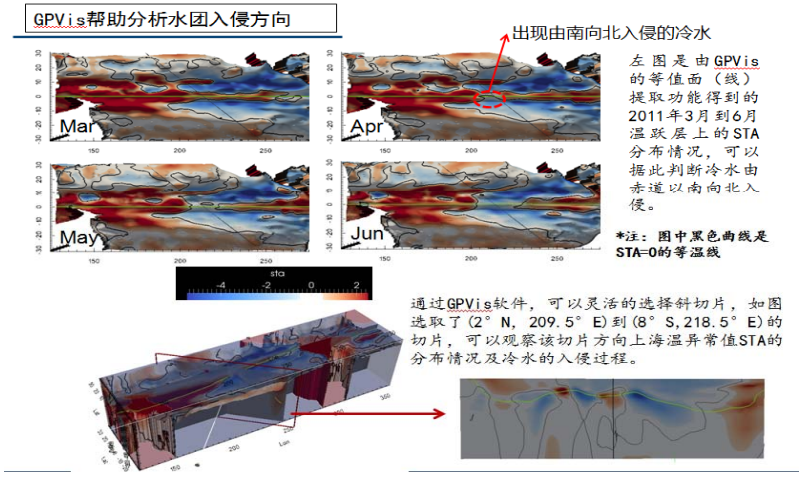


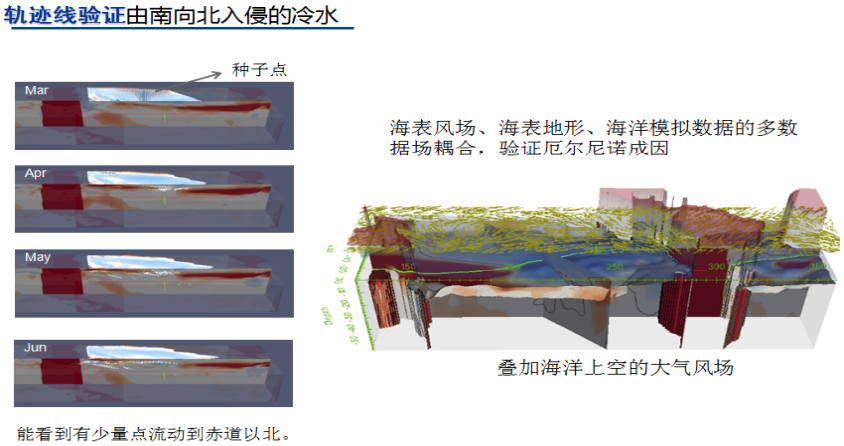
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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 Isosurfce
  - Pathlines with efficient sends setting
  - Simulation data , measured wind field , terrain



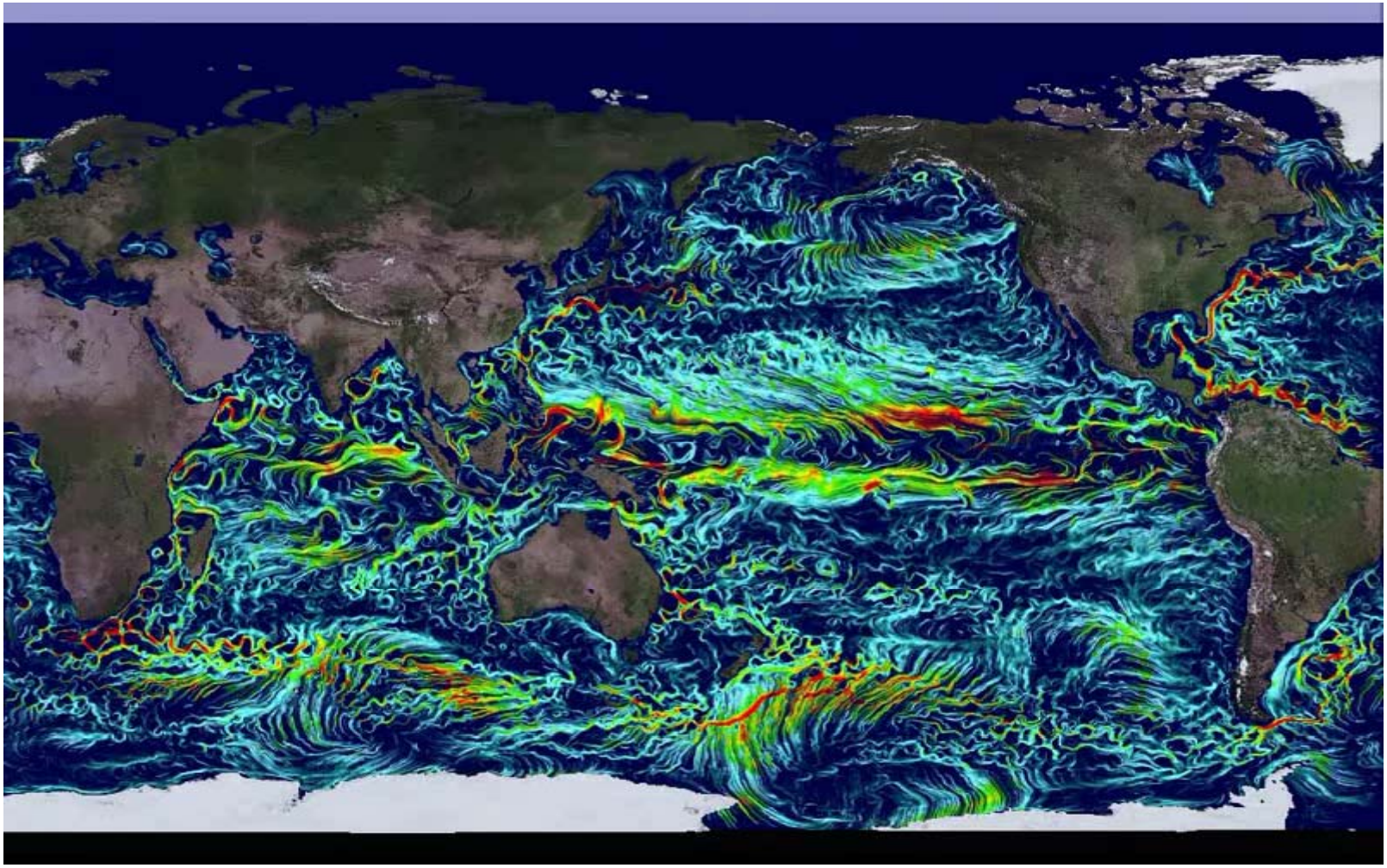




中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

Collaborator ：Atmospheric Physics, CAS

Visualization of Large Scale Ocean Flow



中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

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