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

HPC Infrastructure and Applications in Chinese Academy of Sciences

Xuebin CHI, Haili XIAO, Rongqiang CAO, Yining ZHAO

(chi@sccas.cn)

Computer Network Information Center (CNIC)
Chinese Academy of Sciences (CAS)

Jan. 28, 2016, Geneva, Swiss

Outline

1. Supercomputing center in CNIC (SCCAS)
2. HPC infrastructures in China
3. Applications in CAS
4. Collaborations

Part I

Supercomputing Center in CNIC

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

Supercomputing Center (SCCAS)

Subsidiary branch of CNIC, CAS in Beijing, 100+ staffs

Missions

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

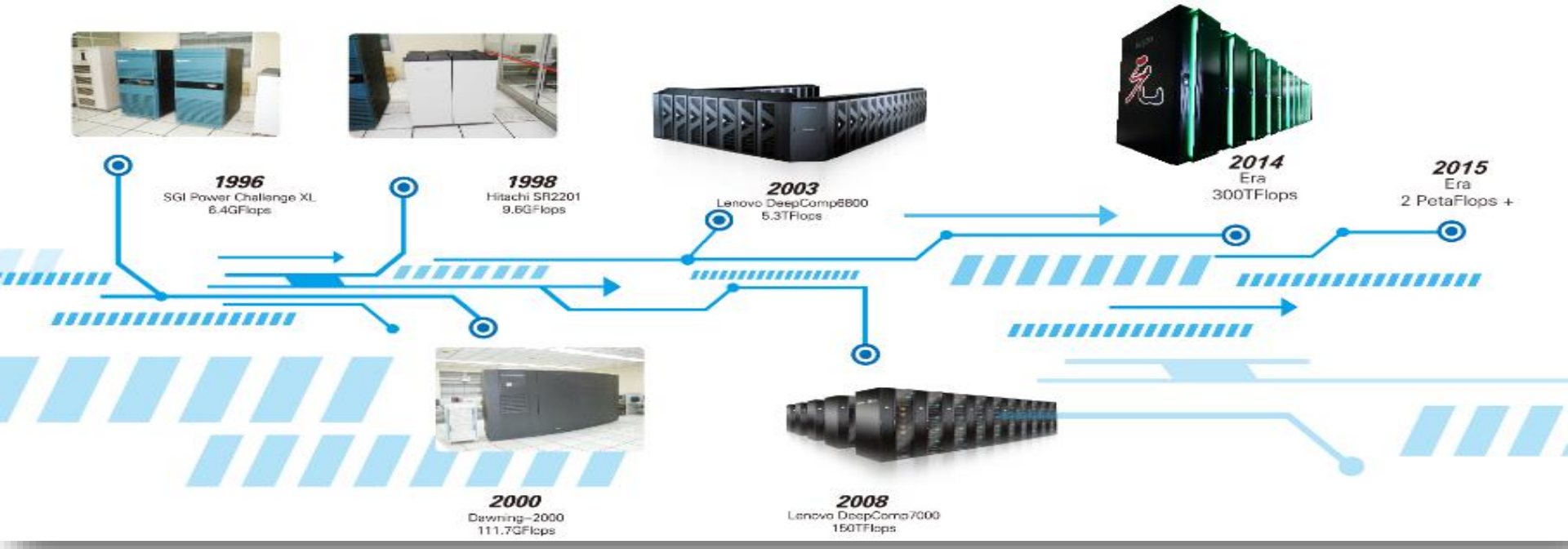
Our roles in the national HPC infrastructure of China

- Operation and Management Center of CNGrid (announced in 2005)
- The northern major node of CNGrid
- 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

The Supercomputer History in SCCAS (1996-2015)



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: will be announced on March, 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

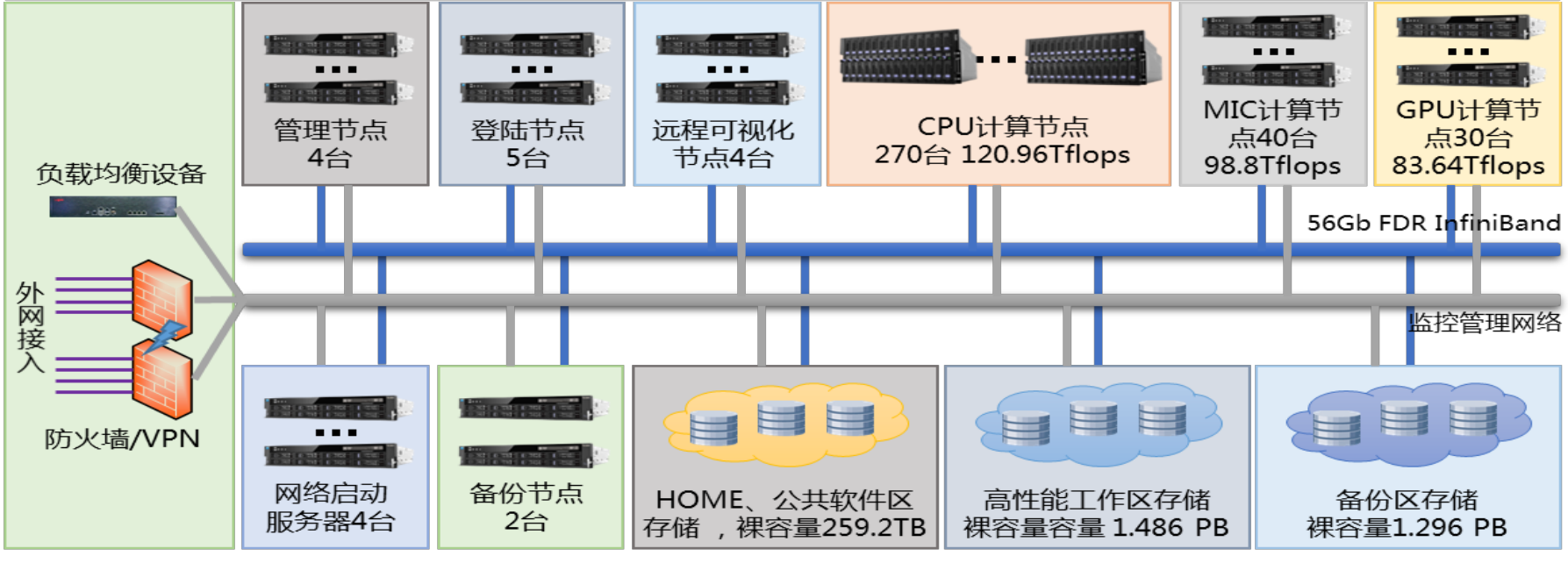
Huairou Branch Center of CNIC

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

Hardware @ Stage 1

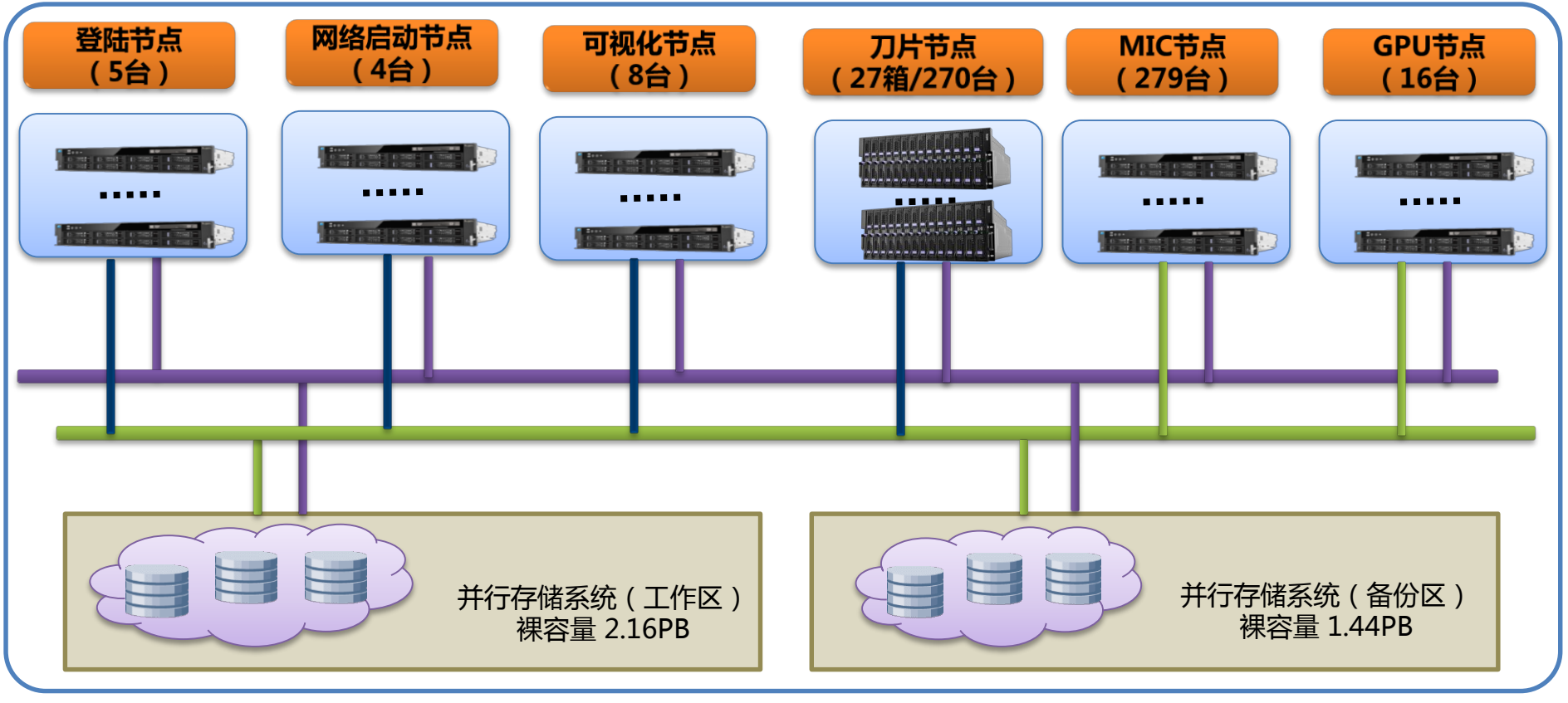
- 303.4 Teraflops (CPU + GPU/Intel Xeon Phi)
- Storage capacity - 3.041 PB
- Integrated bandwidth - 64.5GB/s
- 56Gbps FDR InfiniBand Interconnection
- New Gridview 3.0 Cluster Management System
- Highly efficient horizontal air-flow water cooling system



Hardware @ Stage 2

- 1.988 Pflops (542.4 + 1446.34)
- Storage - 3.6 PB
- 100Gbps EDR InfiniBand

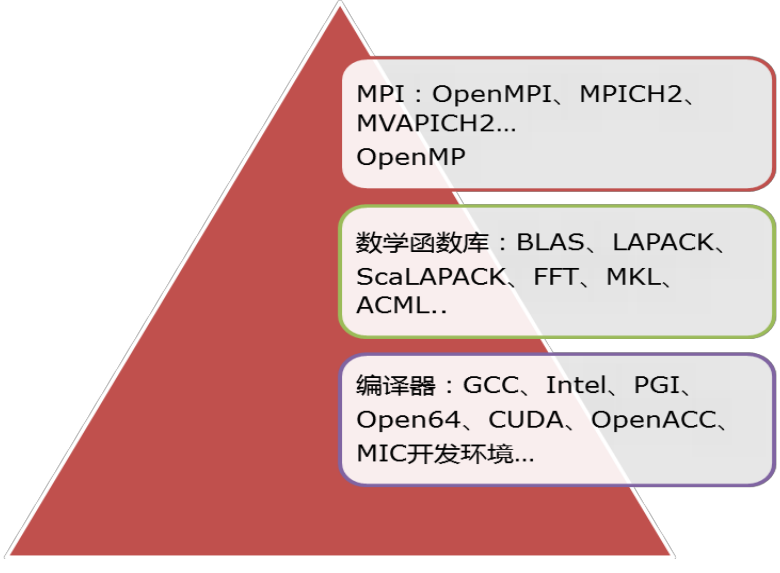
InfiniBand网
10Gb管理网
1Gb管理网



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

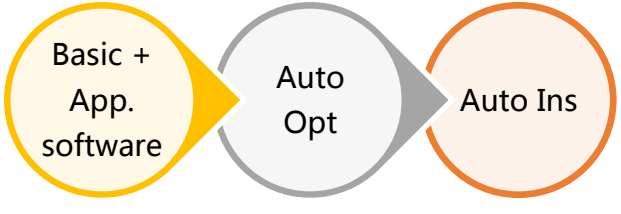
Software

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



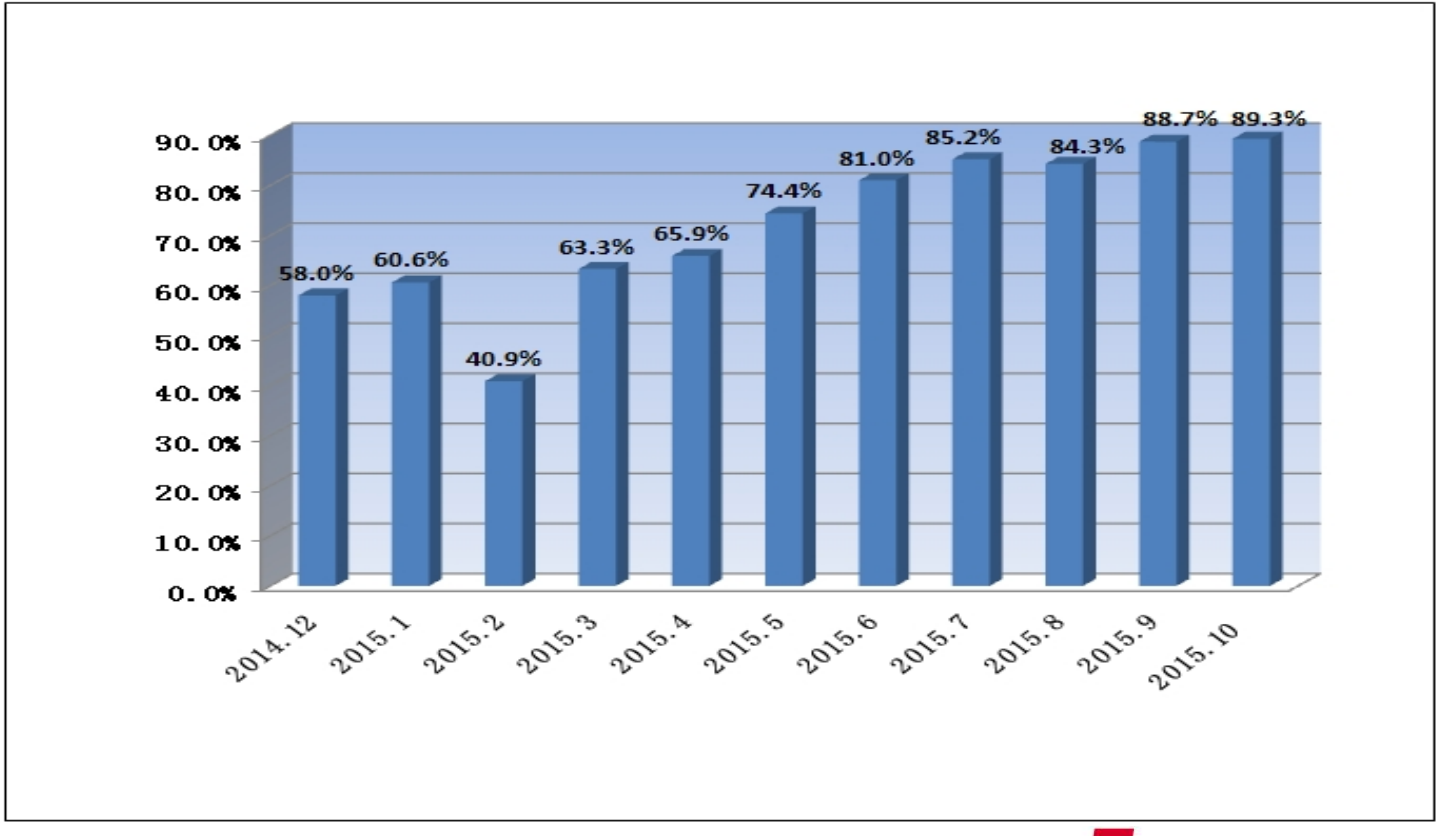
Clussoft 部分组件

- 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

Usage of ERA



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

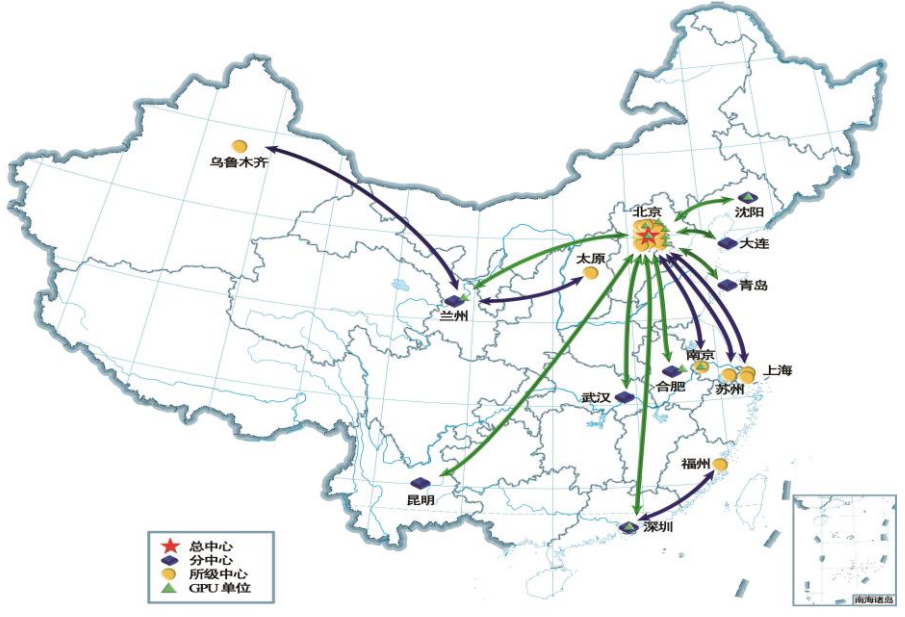
Part II

HPC Infrastructures in China

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

Supercomputing Environment of CAS – China ScGrid

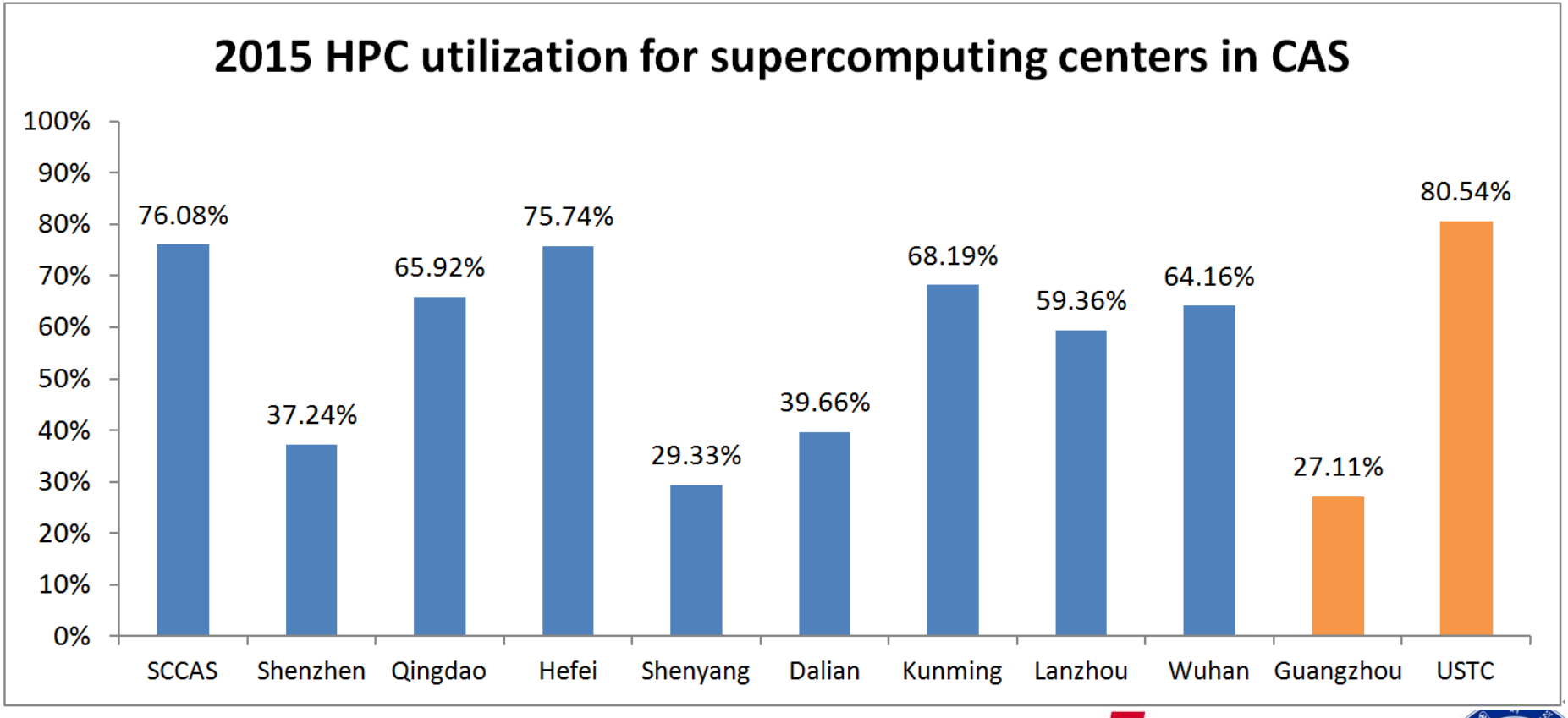
- Three-tier grid
  - 1 head center
  - 9 regional centers (NEW: USTC + Guangzhou)
  - 18 institution centers, 11 GPU centers
- Applications - 120
  - Computational Chemistry, Physics, Material science, Life science, CFD, Industrial computing
- Status (by 2015)
  - User 500
  - #Job > 550 000
  - Walltime >100 000 000 CPU·Hours



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

Overview of System Usage

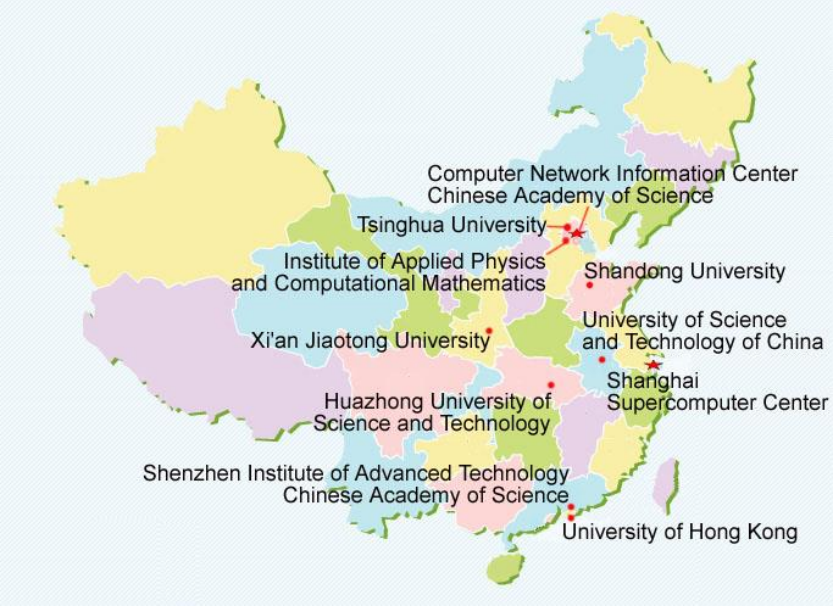


中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

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

CNGrid environment

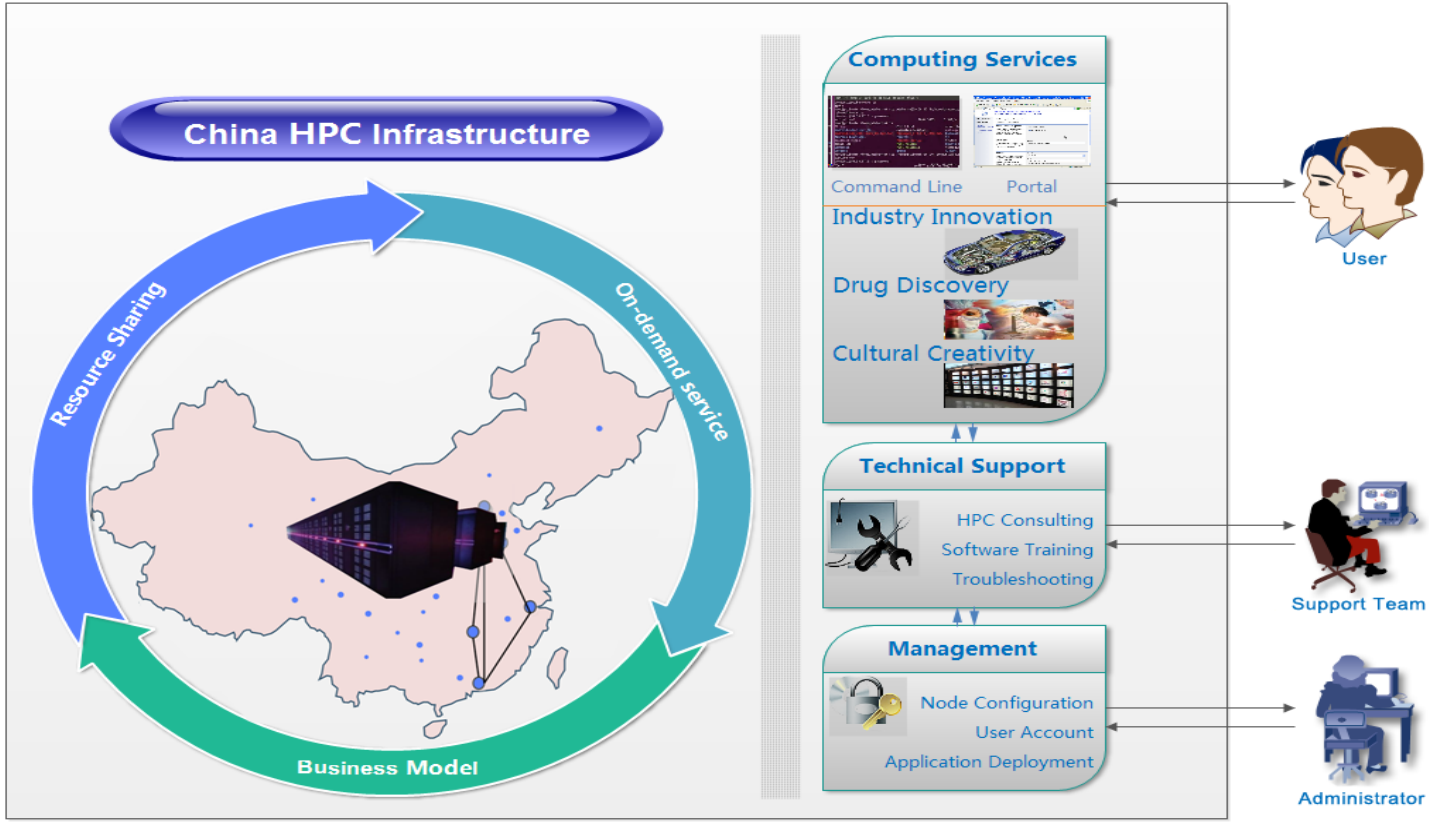
- **14 sites**
  - **SCCAS (Beijing, major site)**
  - **SSC (Shanghai, major site)**
  - **NSCTJ (Tianjin)**
  - **NSCSZ (Shenzhen)**
  - **NSCJN (Jinan)**
  - **Tsinghua University (Beijing)**
  - **IAPCM (Beijing)**
  - **USTC (Hefei)**
  - **XJTU (Xi'an)**
  - **SIAT (Shenzhen)**
  - **HKU (Hong 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 new model

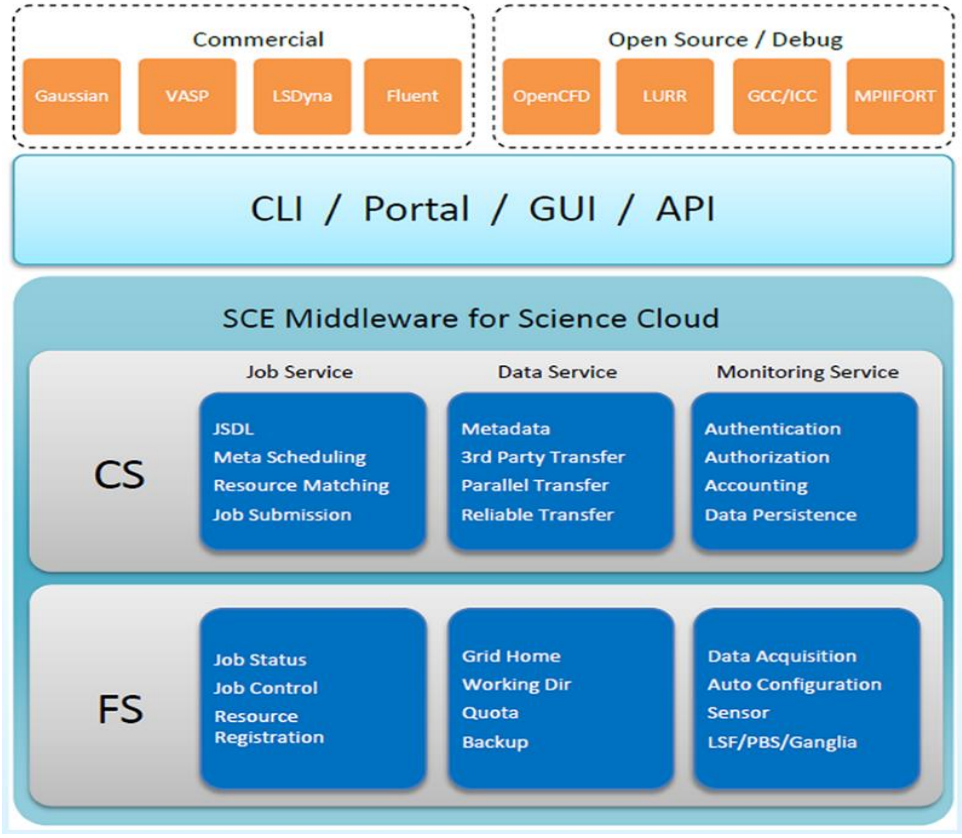


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

SCE - Middleware for Science Cloud

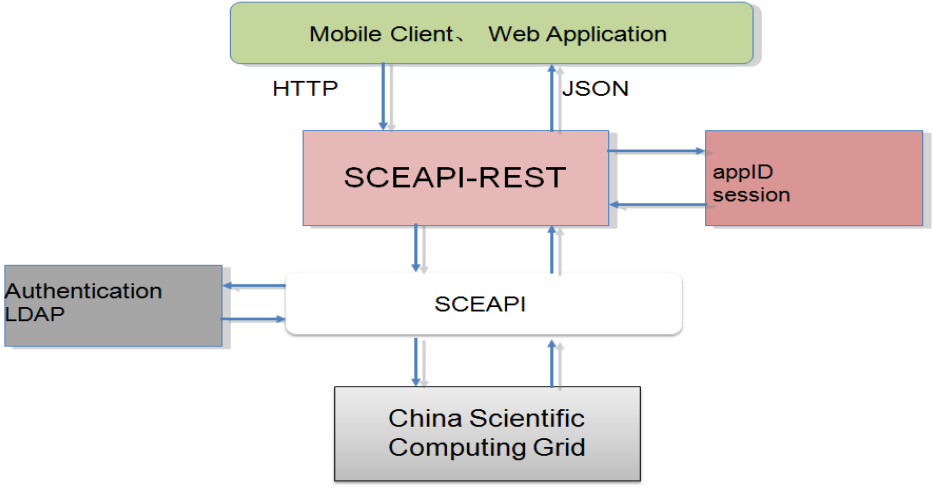
- Developed by SCCAS
- SCE
  - Scientific computing
  - Lightweight
  - Stable
- Diveristy
  - CLI
  - Portal
  - GUI
  - API

International Patent
（PCT/CN2011/071640）



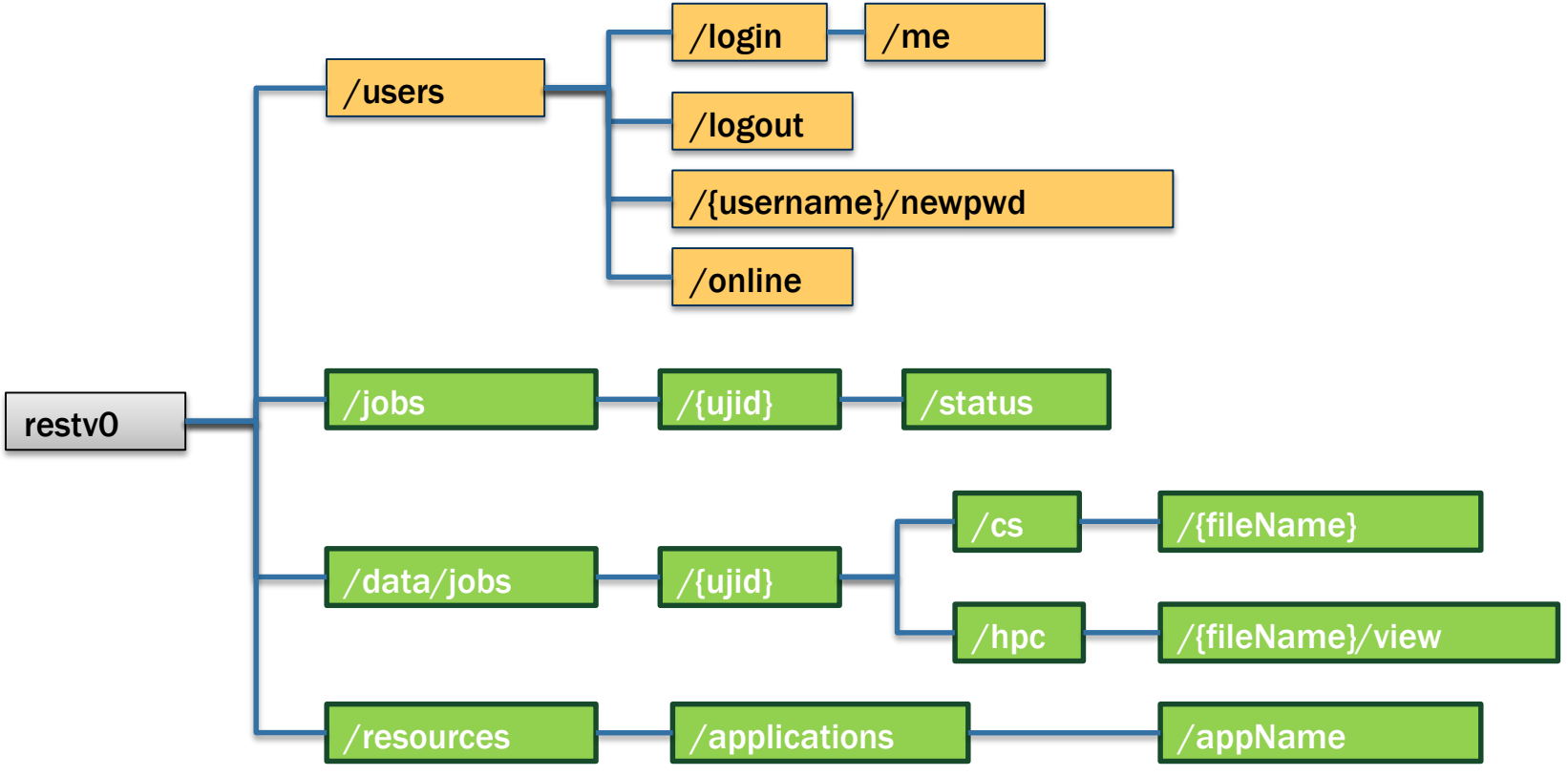
SCEAPI - HPC Cloud API based on SCE

- RESTful API
  - Lightweight Web Service
  - OS independent
    - Windows, Linux
    - iOS, Android
  - Language independent
    - Java, C/C++
    - PHP, Python, Ruby
    - ...
  - Support App. Community
  - Support mobile APPs



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

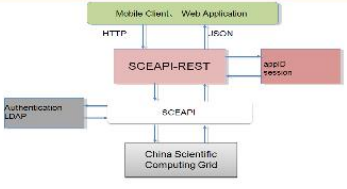
SCEAPI - HPC Cloud API



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

SCEAPI - HPC Cloud API

Software as a service



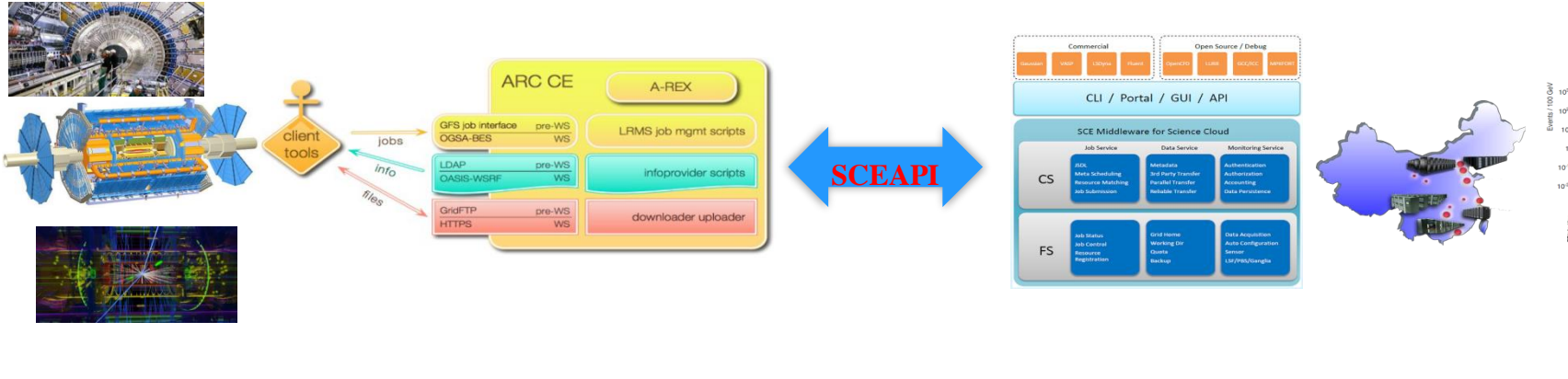
Platform as a service

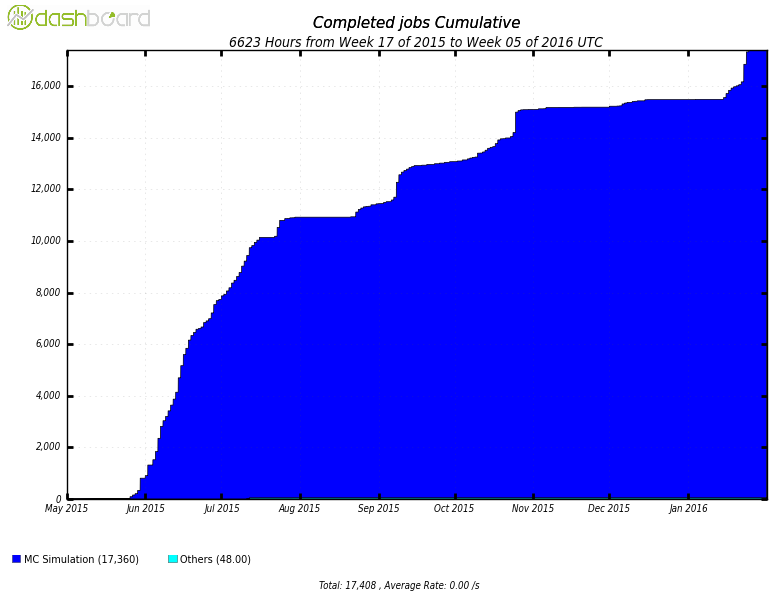
Infrastructure as a service

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

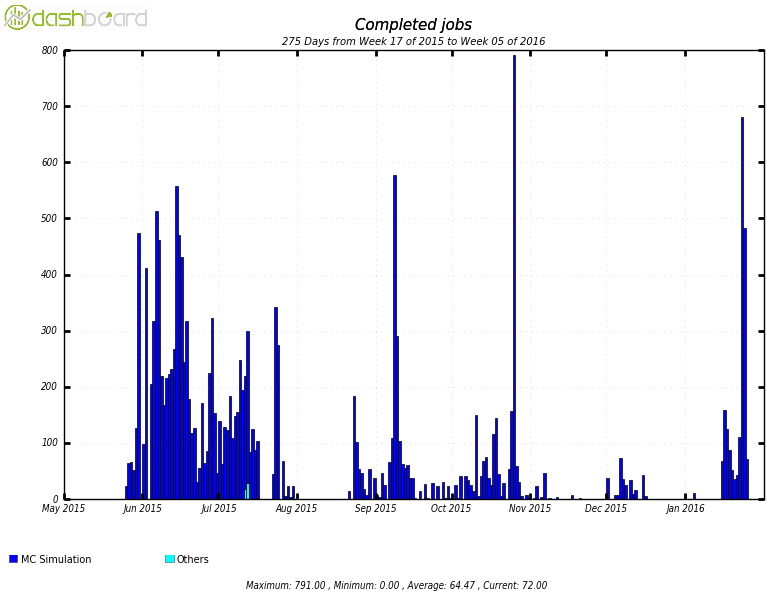
CNGrid & ATLAS

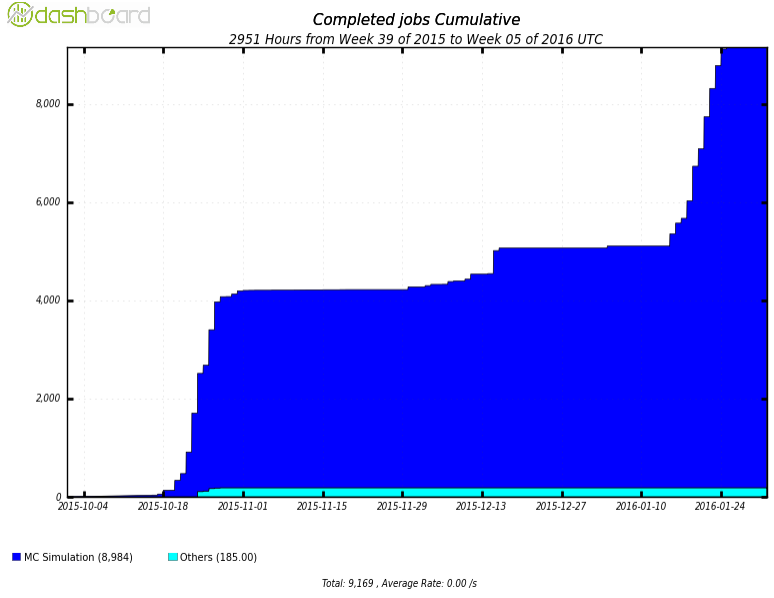
- CNGrid support ATLAS experiment
  - SCEAPI works as a bridge between ARC-CE middleware and CNGrid resources
  - ATLAS simulation jobs run on Chinese HPCs including TianHe-1A and ERA



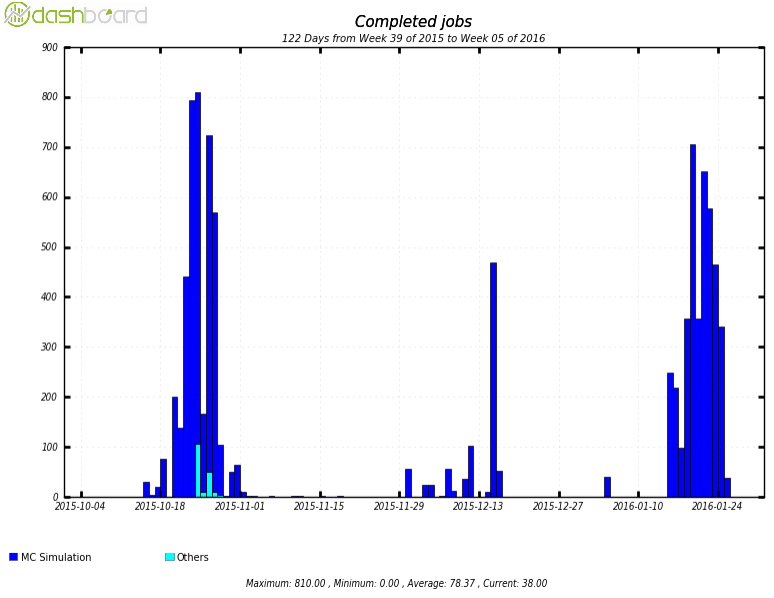


Era
sccas





Tianhe-1A
Tianjin

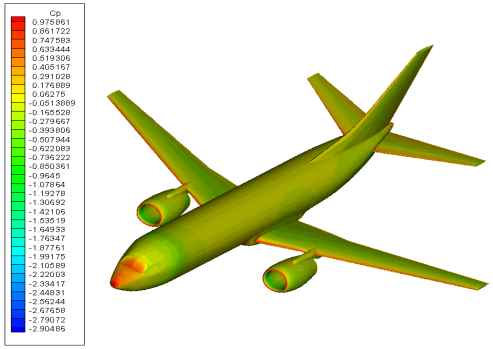


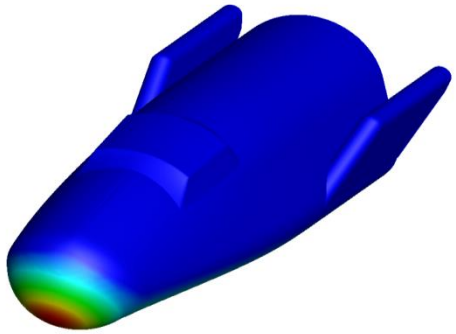
Part III

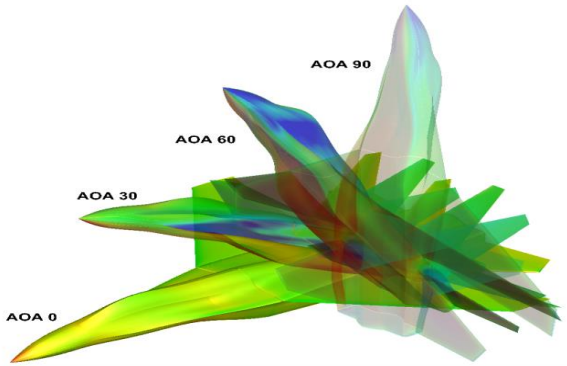
HPC science and engineering applications in CAS

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

CCFD- parallel CFD software

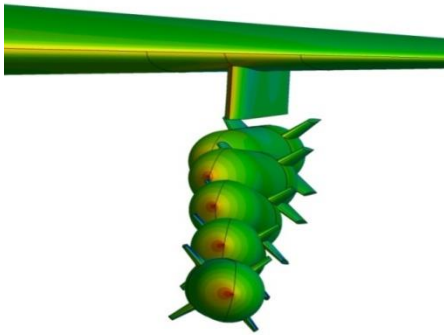


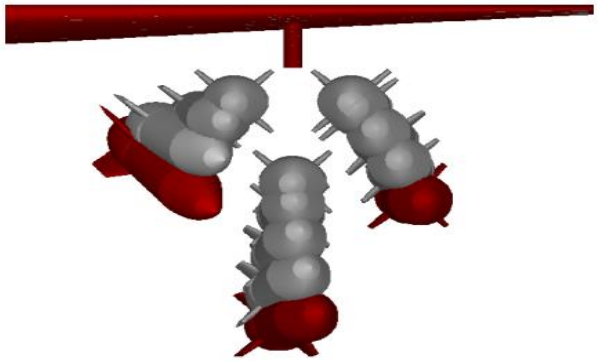




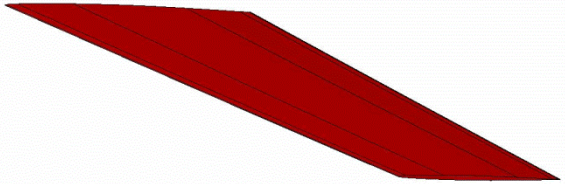
Aerodynamic Computation

Multi-Body Separation





Aeroelastic Flutter

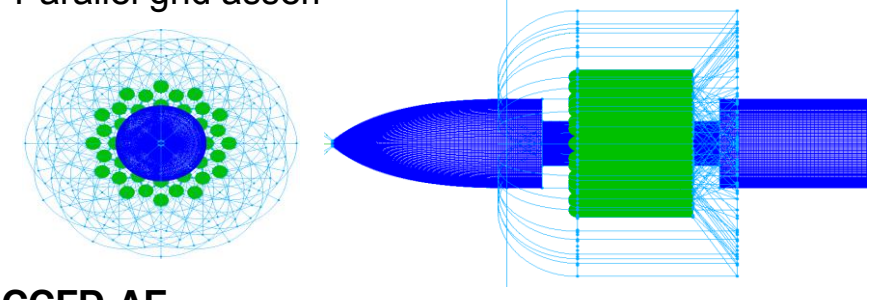


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

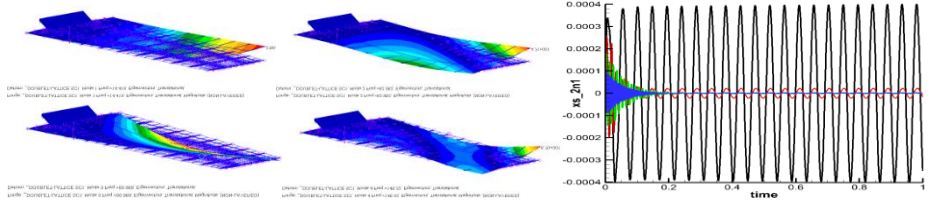
Scale up to over 10,000 cores

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

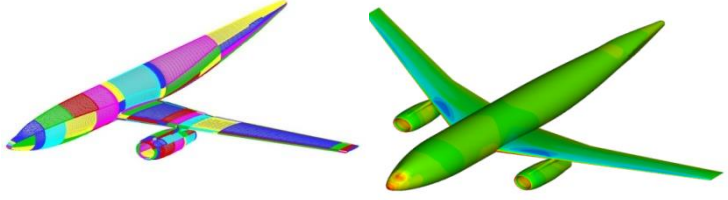
CCFD-MBS
- Chimera grids
- Parallel grid assembling

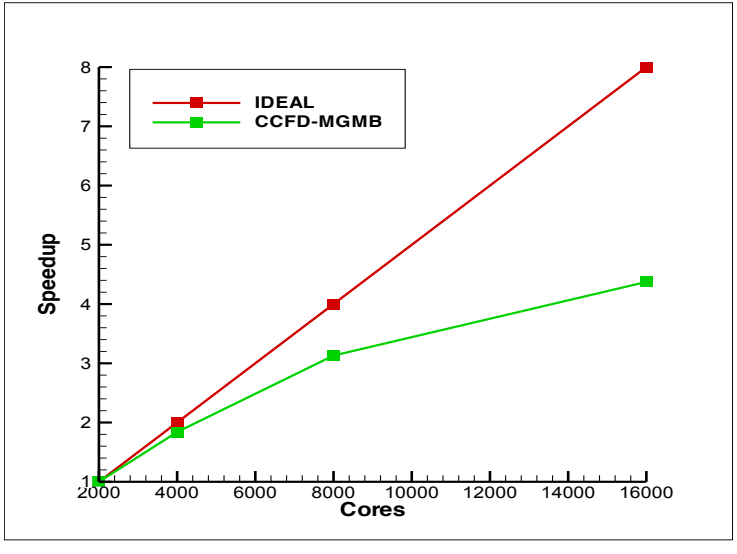


CCFD-AE
- Grid deformation
- Couple with structural analysis software



TianHe II test
DLR-F6 model

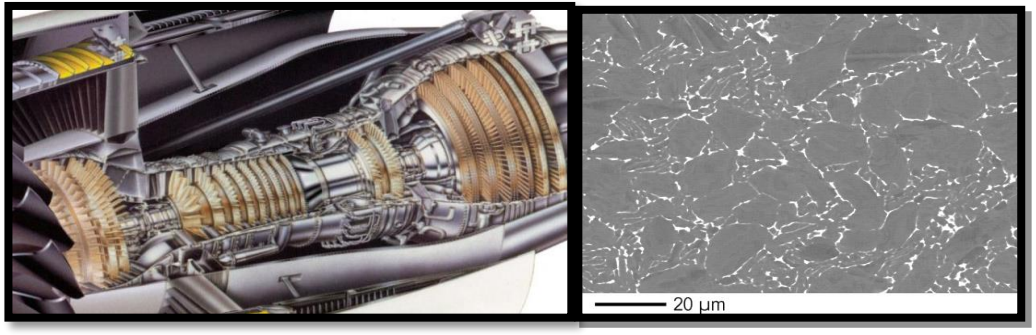


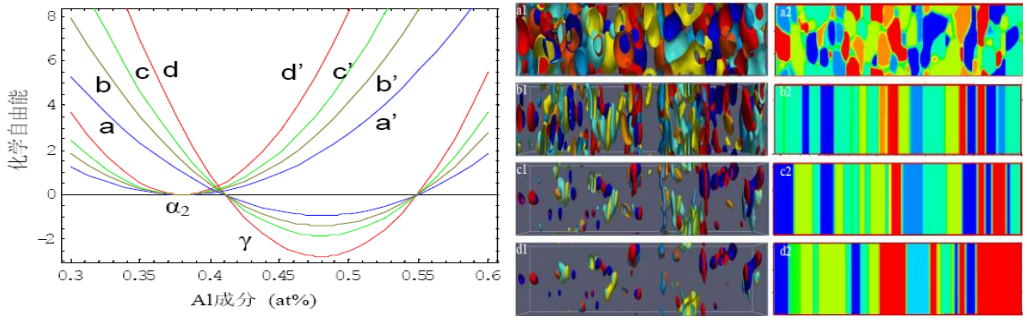


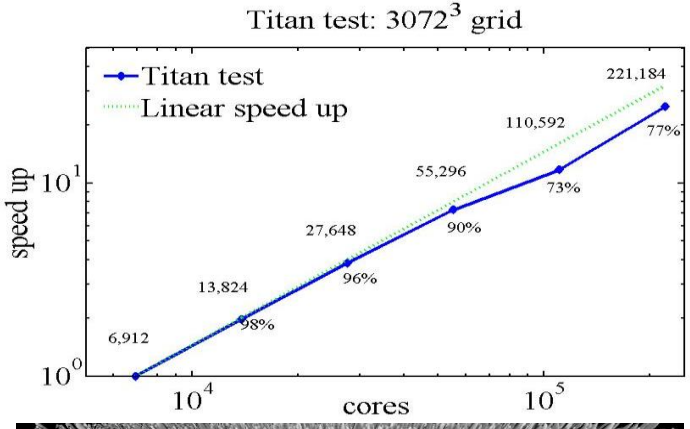
Phase Field Simulation

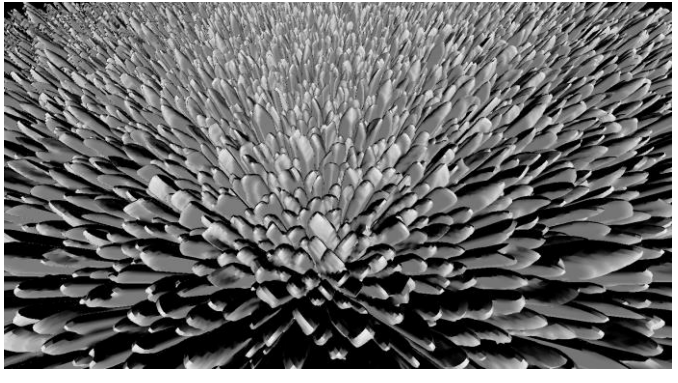
Micro structure evolution in Ti-alloy

- Collaborated with Institute of Metal Research, CAS
- Scale up to 220,000 cores







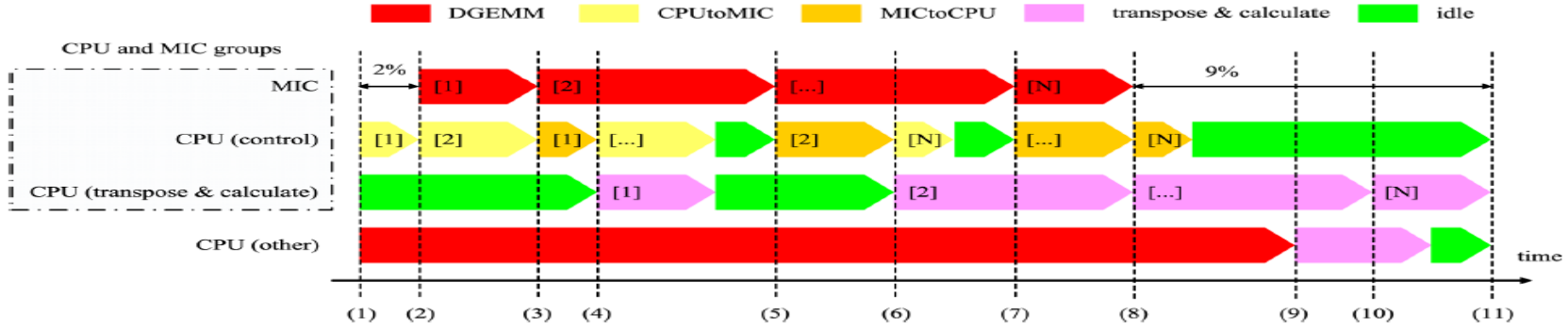


中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

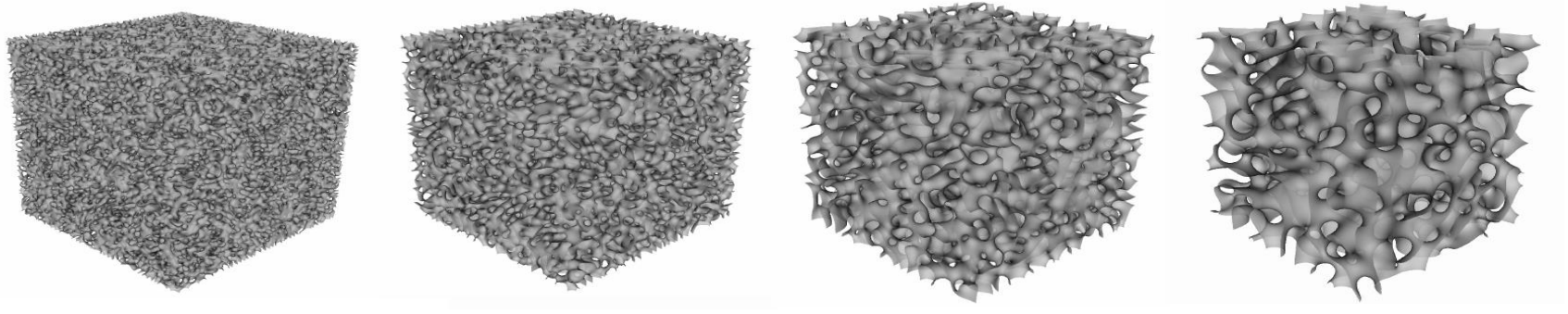
Phase Field Simulation with cETD

compact Exponential Time Differencing(cETD)

- Explicit large time step
- Stable & accurate
- High performance on CPU+MIC
- Supported by the Intel Parallel Programming Center(IPCC) program



Over 1,300GFlops(DP) on 2CPU+2MIC, 52% peak.

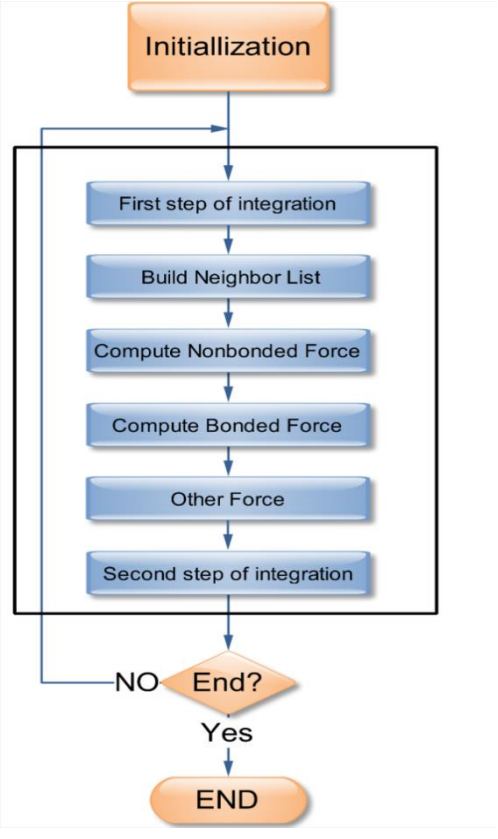


Phase separation on 1024^3 grids, 3hrs with 2CPU+2MIC.

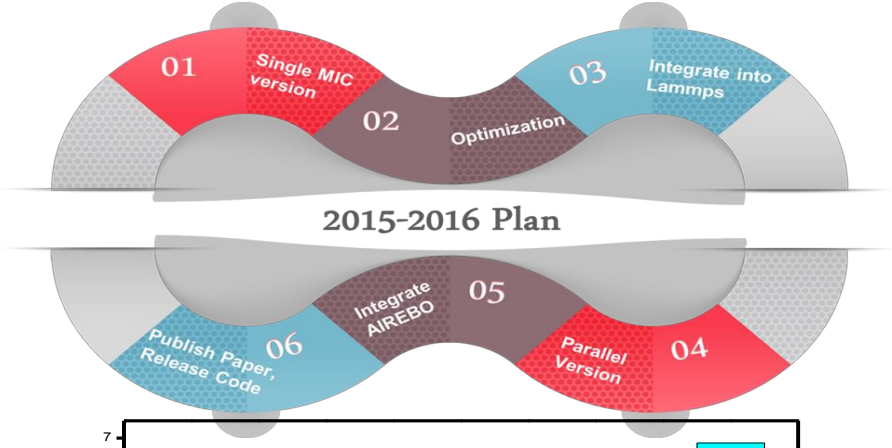
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

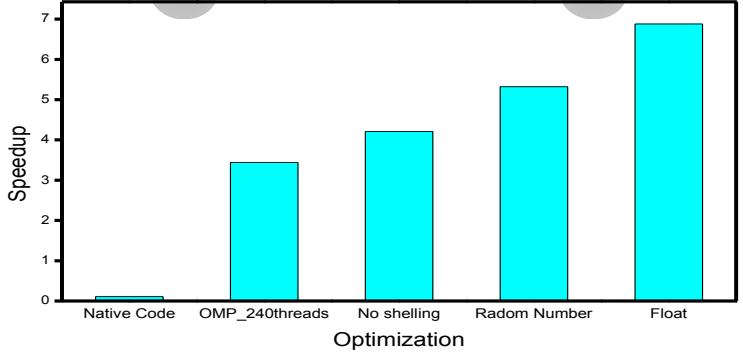
MIC-accelerated DPD

◆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.

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

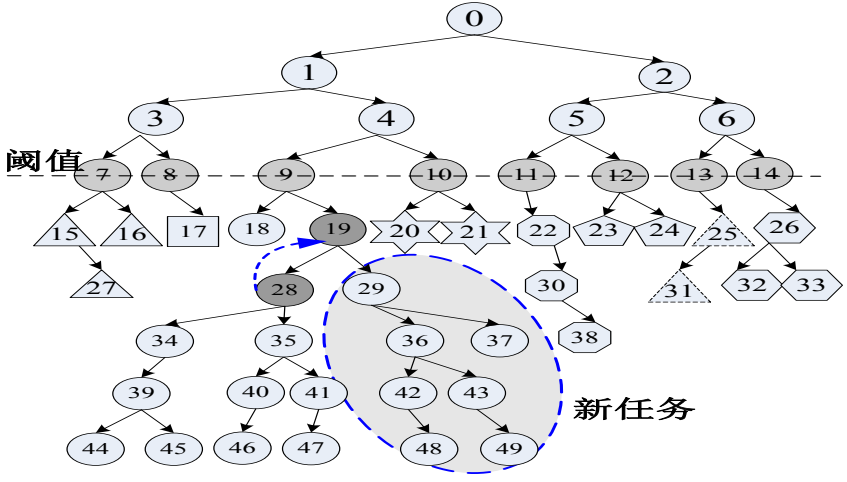
Computer Network Information Center
Chinese Academy of Sciences

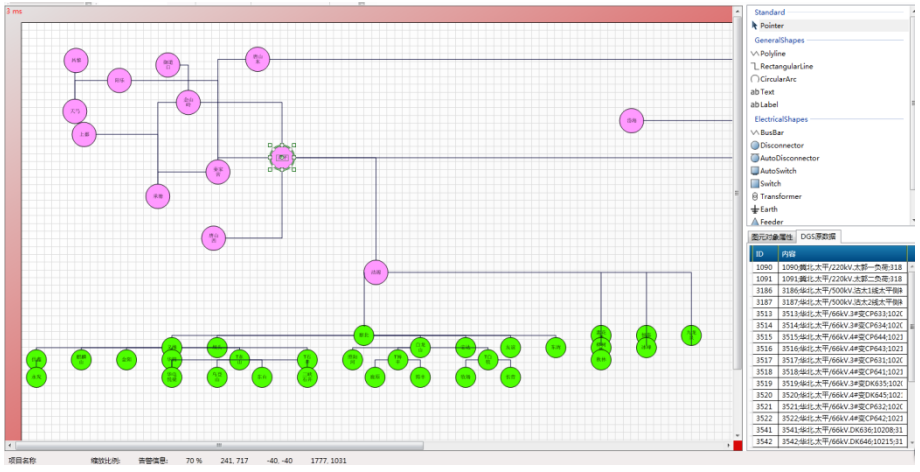
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

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

New Energy Power Generation

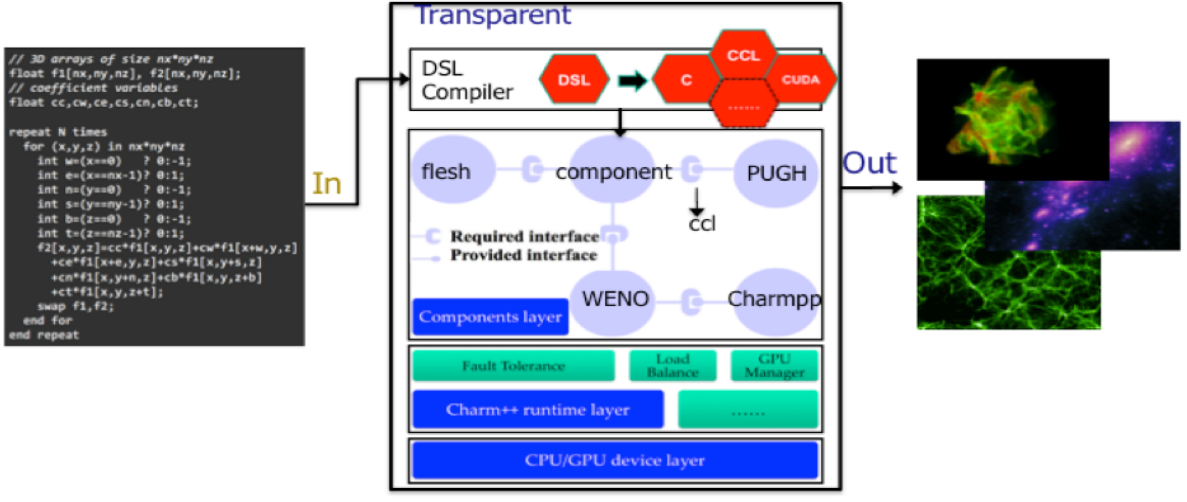
- CMIP（Chinese Mixed Integer Programming Solver）
- Mid long term wind power forecast
- Power network topology visualization





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.**



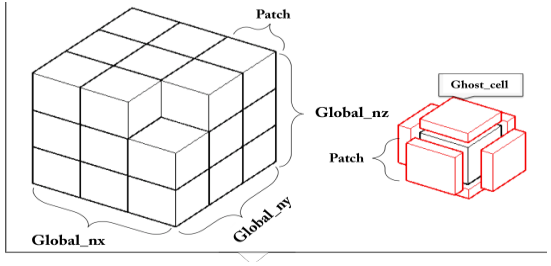
1. Modularity
   - componentization
   - **collaboration**
2. Runtime Adaptivity
   - Fault Tolerance
   - Load Balance
3. Domain Specification
   - Cosmological hydrodynamics
   - N-body

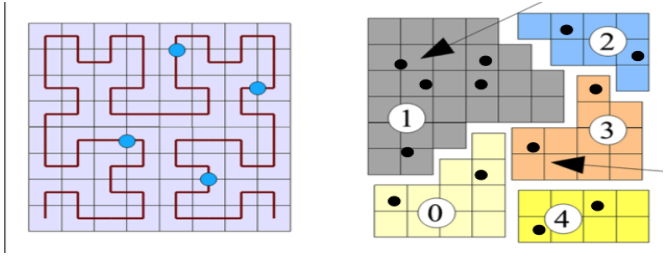
FIGURE 1. MULTI-LAYERS DESIGN OF SC_TANGRAM

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

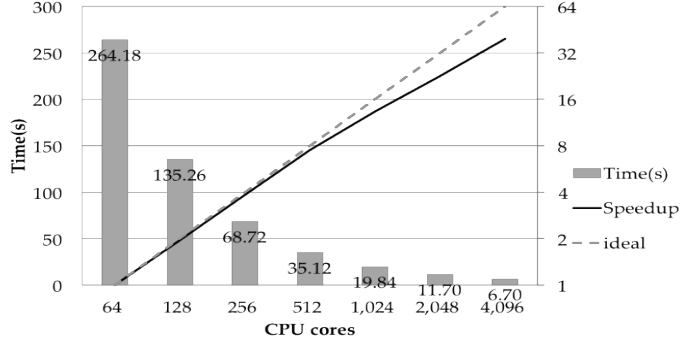
- **Domain Specific Data Types**



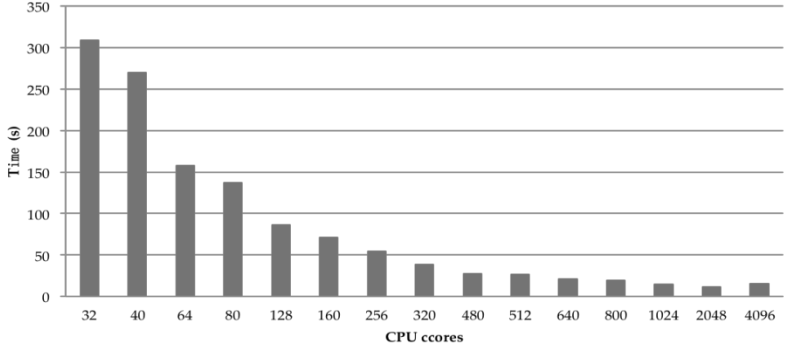


DATA TYPES "GF" FOR UNIFORMED MESH AND "GP" FOR PARTICLES IN MESH ON CPU-CLUSTERS

- **Applications**



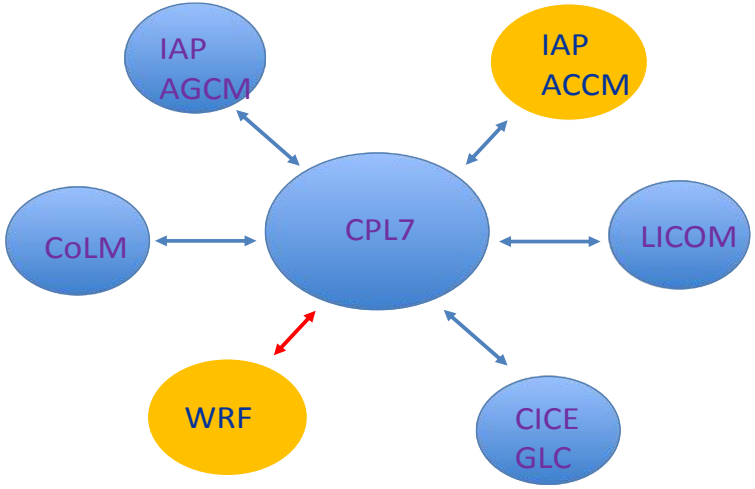
EXECUTION TIME OF TEN STEPS ON THE SCALE OF 1024^3 AND STRONG SCALING

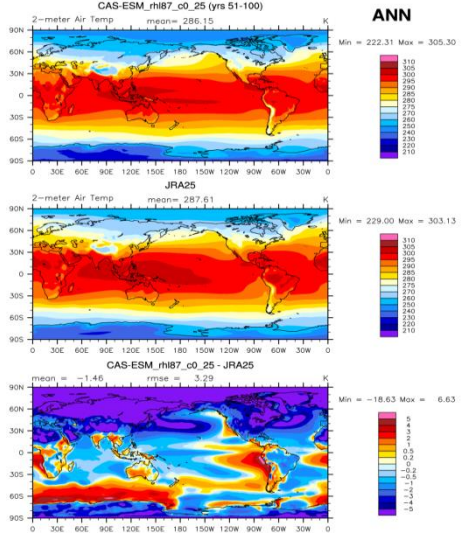


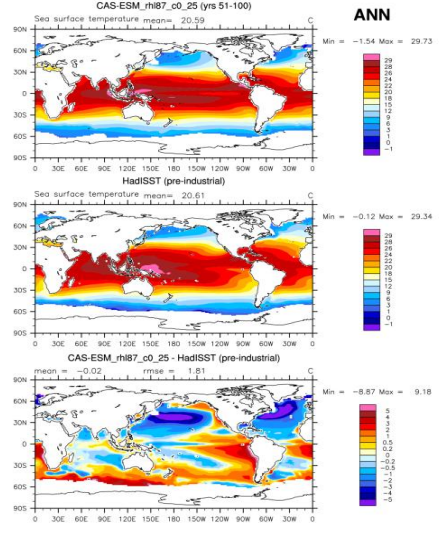
EXECUTION TIME OF ONE STEP ON THE SCALE OF 134217728 PARTICLES ON THE MESH

CAS earth system model

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



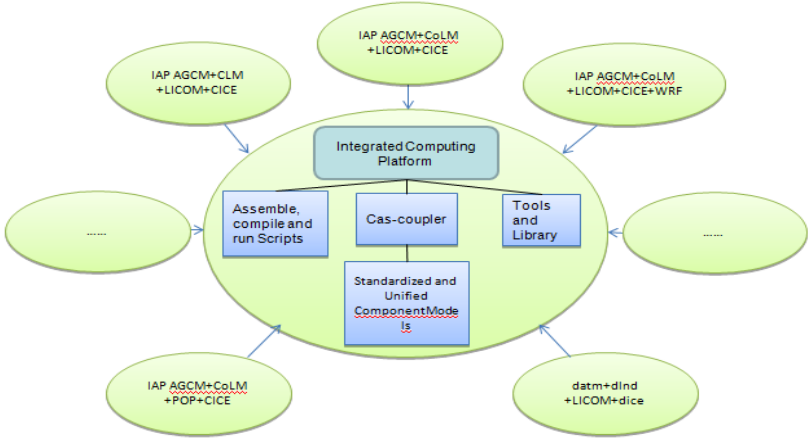




中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

High performance integrated computing platform

- Assemble, compile and run Scripts
- Cas-coupler
  - 2-D coupler(based on MCT)
  - 3-D coupler
  - Coupler creator
- Standard and unified component models
  - Parallel and optimization algorithms for IAP AGCM,CoLM,LICOM,RIEMS,...
  - Standard and unified rules and interface
- Tools and Library
  - High performance communication library
  - Parallel I/O library
  - Performance debugging tools



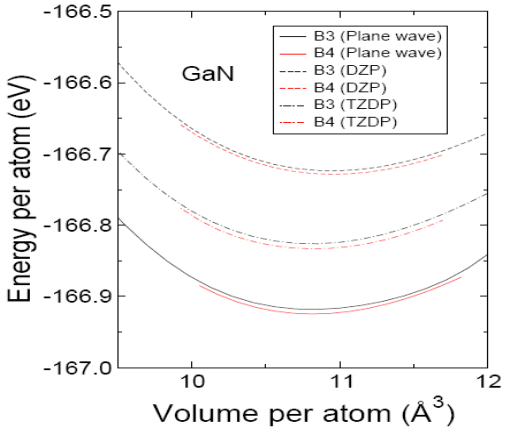
Prototype system of ESM simulation facilities

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

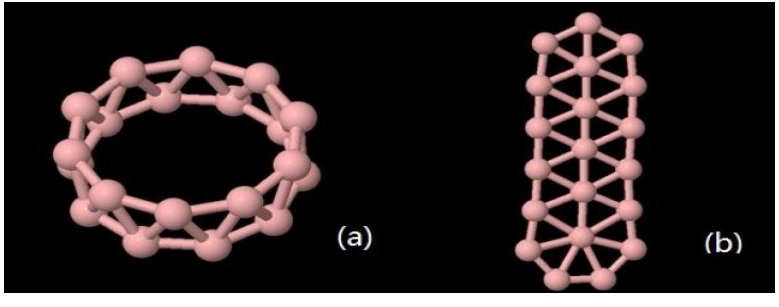
Application of HPSEPS in first principle calculation software MESIA

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.

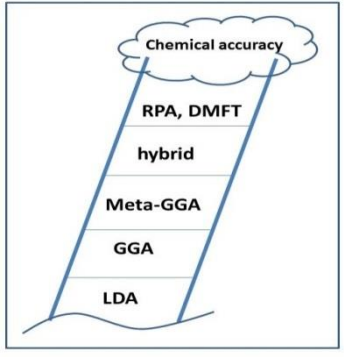
We developed MPI-GPU eigensolver for dense eigenproblem
Computational Throughput: 16GPUs= 512CPUs

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

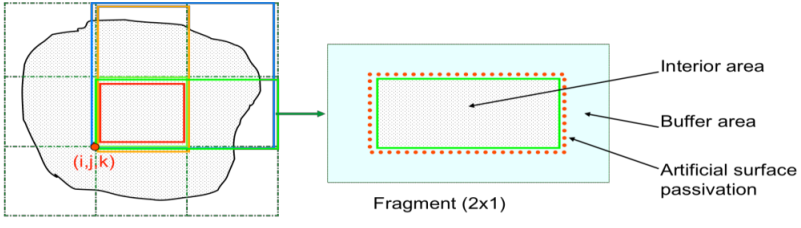
中国科学院超级计算中心
Supercomputing Center of Chinese Academy of Sciences

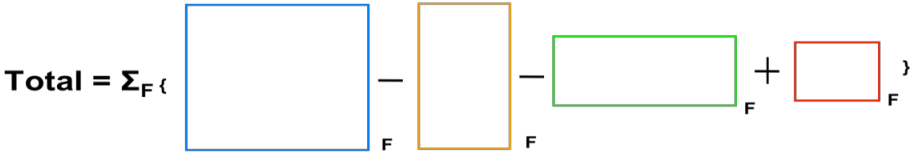
Large scale three dimensional fragment method on GPU

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

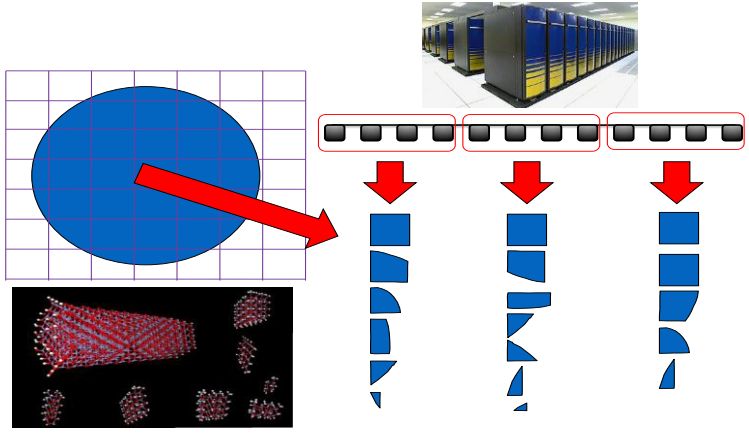
Collaborate with Lin-Wang Wang, LBNL

This project is supported by INCITE program and CSC

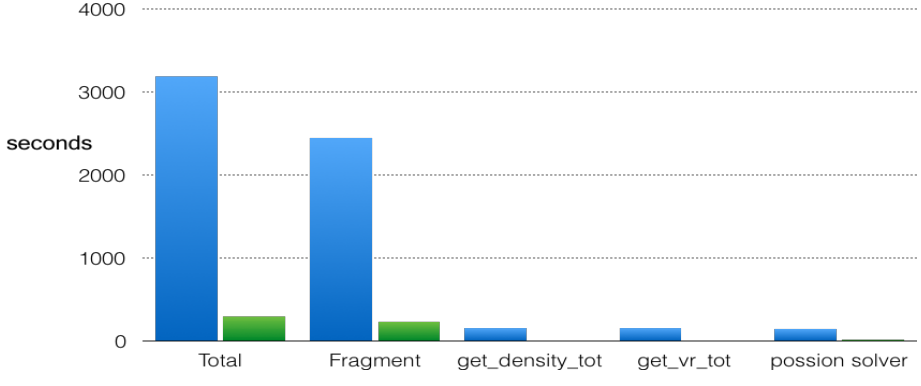
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

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

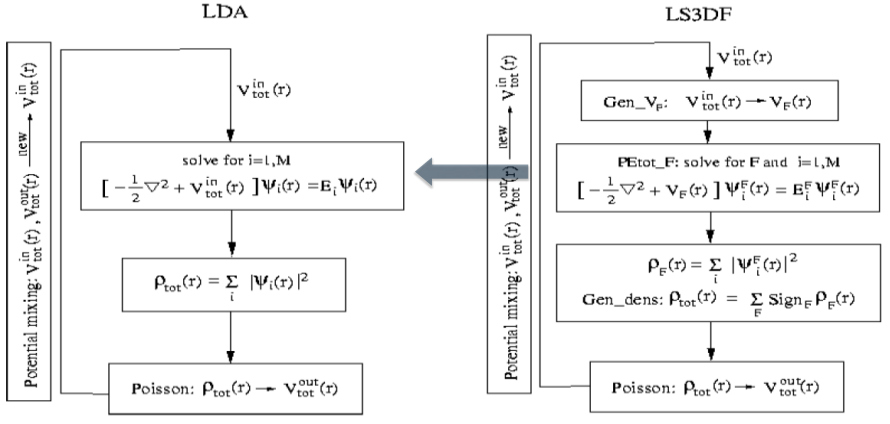
Large scale three dimensional fragment method on GPU



LS3DF data distribution



LS3DF-GPU speedup compared with CPU code



LS3DF algorithm compared with LDA algorithm

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.

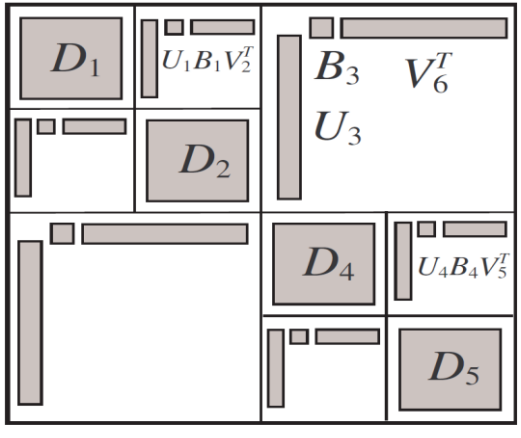
中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

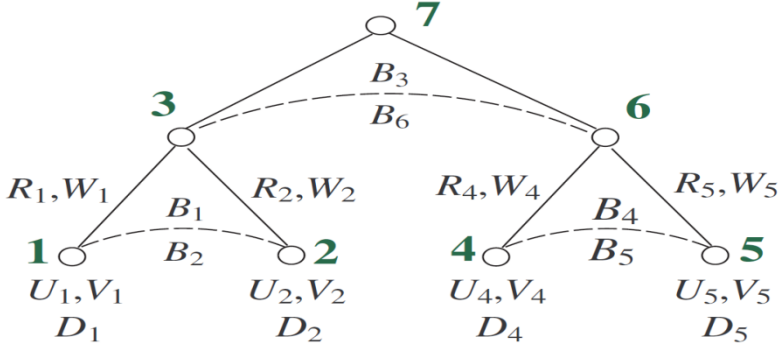
Fast Parallel Direct Solver for Large linear system

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

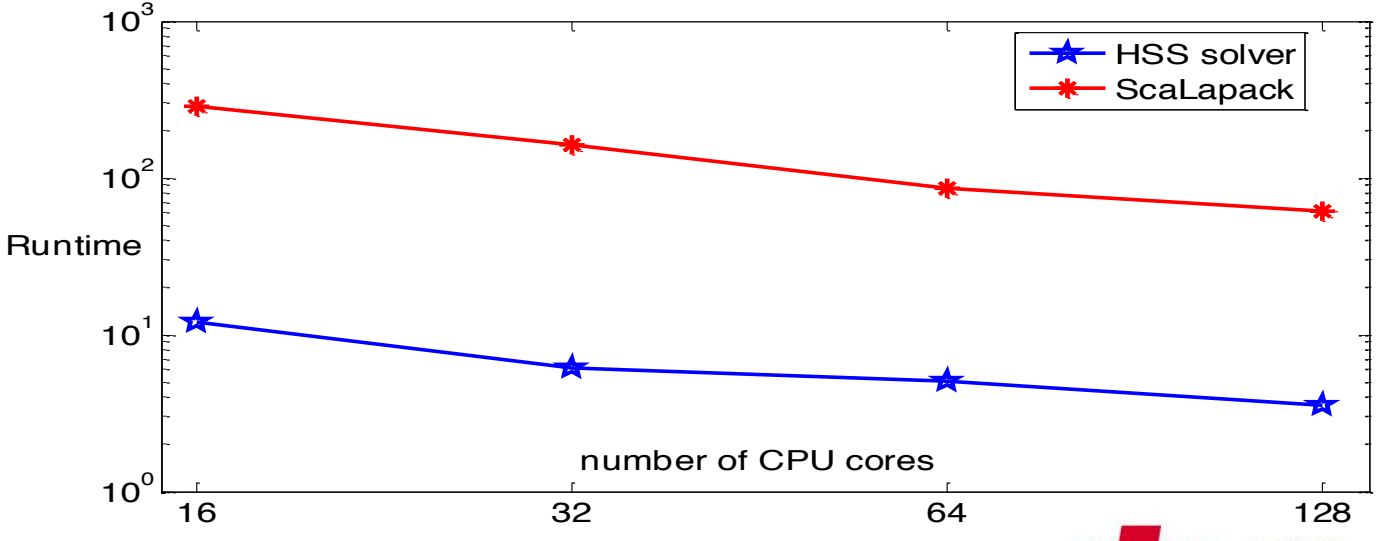
Result: HSS solver vs ScaLapack

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)

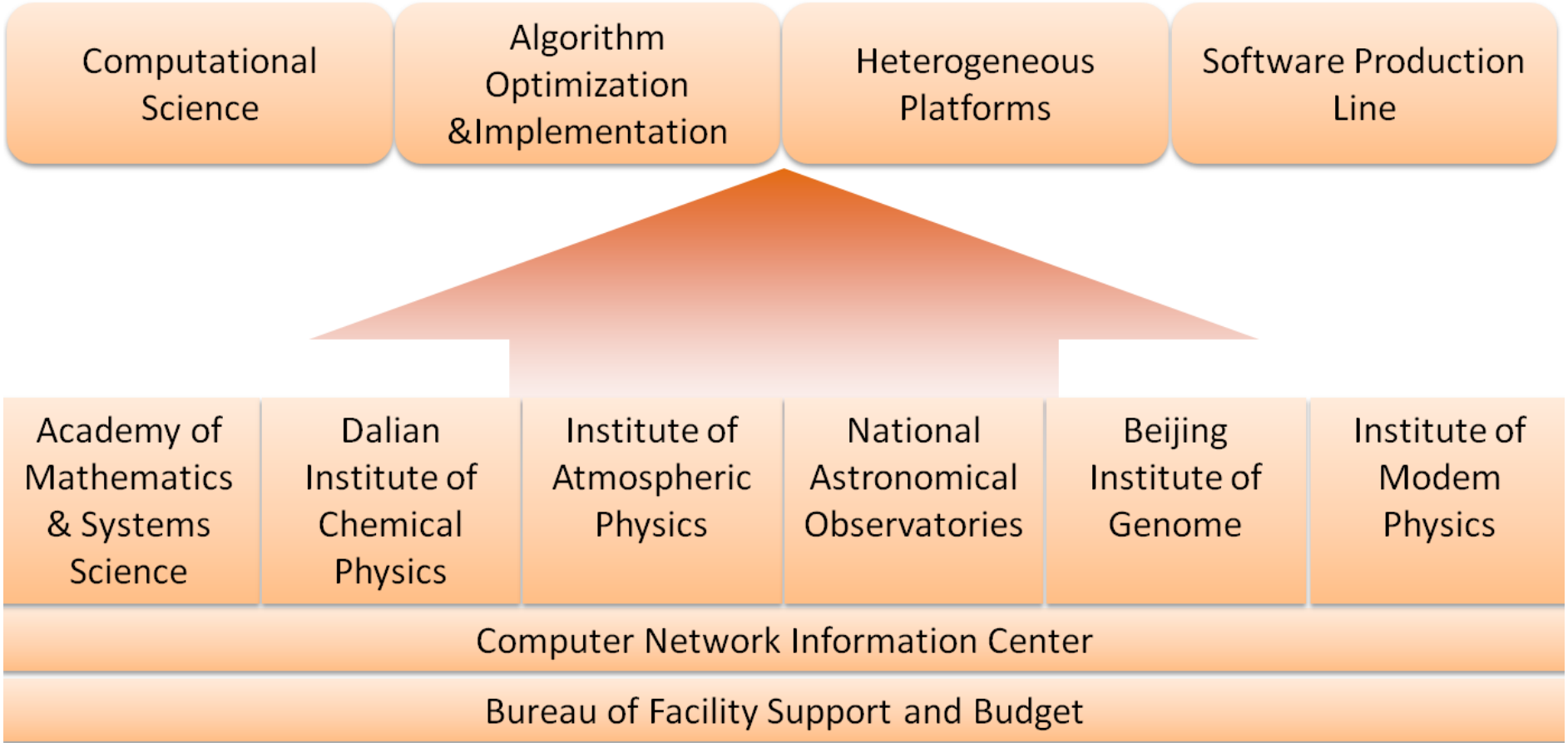


Part IV

Collaboration inside and outside China

CAS Computational Science Application Research Center

Approved by CAS and launched in 2014



Supercomputing Innovation Alliance

- The Alliance, approved by the Ministry of Science and Technology, is the Industry-University-Research-Application cooperation organization. The Alliance was established in September 25, 2013, initiated by the national or local Supercomputing Centers, high-performance computing application research institutes, and related enterprises of total 55 units.

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

Supercomputing Innovation Alliance - Organization Structure

Honorary chairman: Academician Jin Yilian

Chairman: Professor Qian Depei

Secretary-General: Research Professor Chi Xuebin

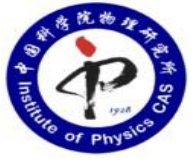
Vice Secretary-General: Research Professor Xie Xianghui

Relying institutions: CNIC, CAS

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

CHANGES 2016

- CHANGES (CHinese-AmericaN-German E-Science and cyberinfrastructure Workshop)
  - JSC-NCSA-CNIC collaborations
  - Since 2012
  - HPC/BigData/Vis.
  - CHANGES16
    - Fall of 2016, Juelich, Germany
    - Preparation meeting at ISC15

NCSA

JÜLICH
FORSCHUNGSZENTRUM

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

Thank you!

chi@sccas.cn