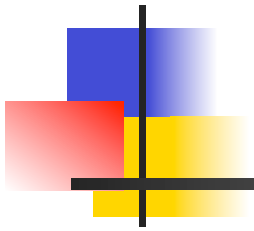


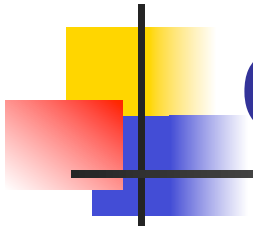
HPC and e-Infrastructure development in China

Depei Qian

Sino-German Joint Software Institute (JSI)

Beihang University

March. 24, 2011

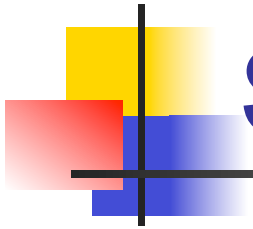


Outline

- Overview of 863's efforts on HPC and Grid
- HPC development
- Building up CNGrid
- HPC and Grid applications

China's High-tech Program

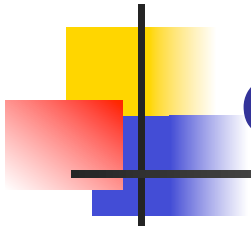
- The National High-tech R&D Program (863 Program)
  - proposed by 4 senior Chinese Scientists and approved by former leader Mr. Deng Xiaoping in March 1986
  - One of the most important national science and technology R&D programs in China
- Now a regular national R&D program planed in 5-year terms, the one just finished is the 11th five-year plan



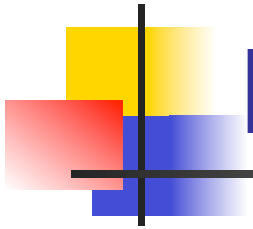
Shifts of research focuses

- 1987: Intelligent computers following the 5th generation computer program in Japan
- 1990: from intelligent computers to high performance parallel computers
  - Intelligent Computer R&D Center established
  - Dawning Computer was established in 1995
- 1999: from individual HPC system to national HPC environment
- 2006: from high performance computers to high productivity computers

Overview of 863 key projects on HPC and Grid

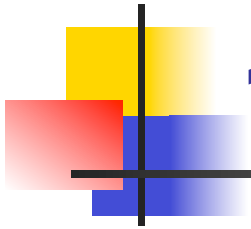


- "High performance computer and core software"
  - 4-year project, May 2002 to Dec. 2005
  - 100 million Yuan funding from the MOST
  - More than 2X associated funding from local government, application organizations, and industry
  - Outcomes: China National Grid (CNGrid)
- "High productivity Computer and Grid Service Environment"
  - Period: 2006-2010
  - 940 million Yuan from the MOST and more than 1B Yuan matching money from other sources



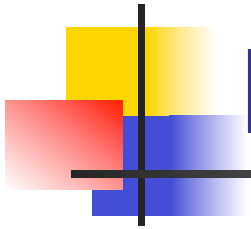
Major R&D activities

- Developing PFlops computers
- Building up a grid service environment--CNGrid
- Developing Grid and HPC applications in selected areas



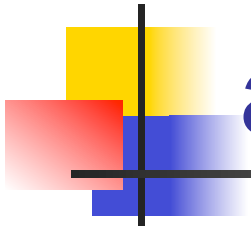
1. HPC development

- **Major issues in developing PetaFlops computer systems**
  - **High productivity**
  - **Support to a wide range of applications**
  - **Cost-effective in terms of development and operational costs**
  - **Low power consumption**



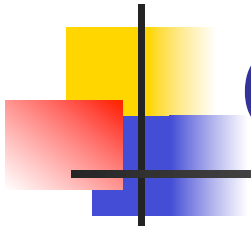
High productivity

- High performance
  - Pursuing not only peak performance, but also sustainable performance
- Time to solution
  - Reduce complexity in developing applications, shorten development cycle
  - Allow programmer to concentrate on problem instead of details of the computer
- Good program portability
  - Achieve high efficiency of ported programs
  - Automatic parallelization of programs
- Robustness
  - Improve reliability and stability
  - Tolerant to hardware and software failure
  - Self-recovery from failure



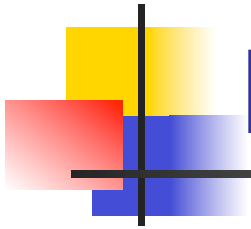
Support wide range of applications

- HPC systems are usually installed at general purpose computing centers and available to large population of users
  - Support wide range of applications and different computational models
  - Efficient for both general purpose and special purpose computing tasks
  - Support both fine and coarse parallelism



Cost effectiveness

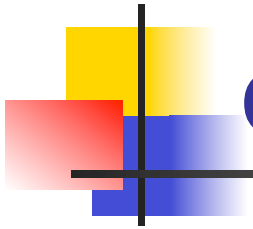
- Highly cost-effective in both developing and operation stages
- Lower the development cost
  - Use novel architectural design but as many commodity components as possible
- Lower the operation cost
  - Lower system operation and maintenance cost
    - Energy cost is the major limiting factor of running large systems
  - Easy to program
    - Reduce the application development cost
  - Better resource usage
    - Equal to prolonging of system lifetime



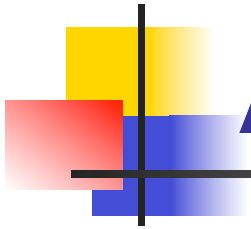
Low power consumption

- System power consumption
  - The dominant factor preventing implementation of high-end computers
  - Impossible to increase performance by expanding system scale infinitely
  - Energy cost is a heavy burden to operation of high-end systems
  - 2MW/PF limitation
- Power consumption in air conditioning
  - Cooling efficiency—water cooling

How to address those challenges?

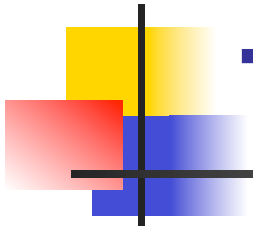


- Architectural support
- Technology innovation
  - Device
  - Component
  - system
- Hardware and software coordination



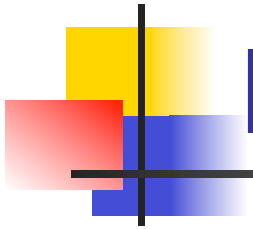
Architectural support

- Using the most appropriate architecture to achieve the goal
- Considering the performance and power consumption requirement
  - Hybrid architecture
    - General purpose + high density computing (cell or GPU) + accelerators (FPGA-based)
  - HPP architecture
    - Global address space
    - Multi-level of parallelism
- Programmability is a major issues of using hybrid architecture



Technology innovation

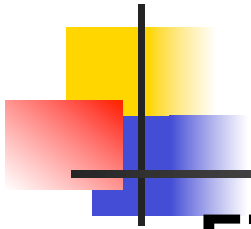
- Innovation at different levels
  - Device
  - Component
  - system
- New processors
- New interconnect technologies
- Low power devices
- Novel memory technologies



HW/SW coordination

- Using combination of hardware and software technologies to address the technical challenges
  - Better utilization of the hardware
  - More cost-effective
  - More flexible

Two phase development



- First phase: developing two 100TFlops machines
  - Dawning 5000A for SSC
  - Lenovo DeepComp 7000 for SC of CAS
- Second phase: three 1000Tflops machines
  - Tianhe IA: NUDT/Inspur/Tianjin Supercomputing Center
  - Dawning 6000: ICT/Dawning/South China Supercomputing Center (Shenzhen)
  - Sunway: Jiangnan/Shandong Supercomputing Center

Dawning 5000A

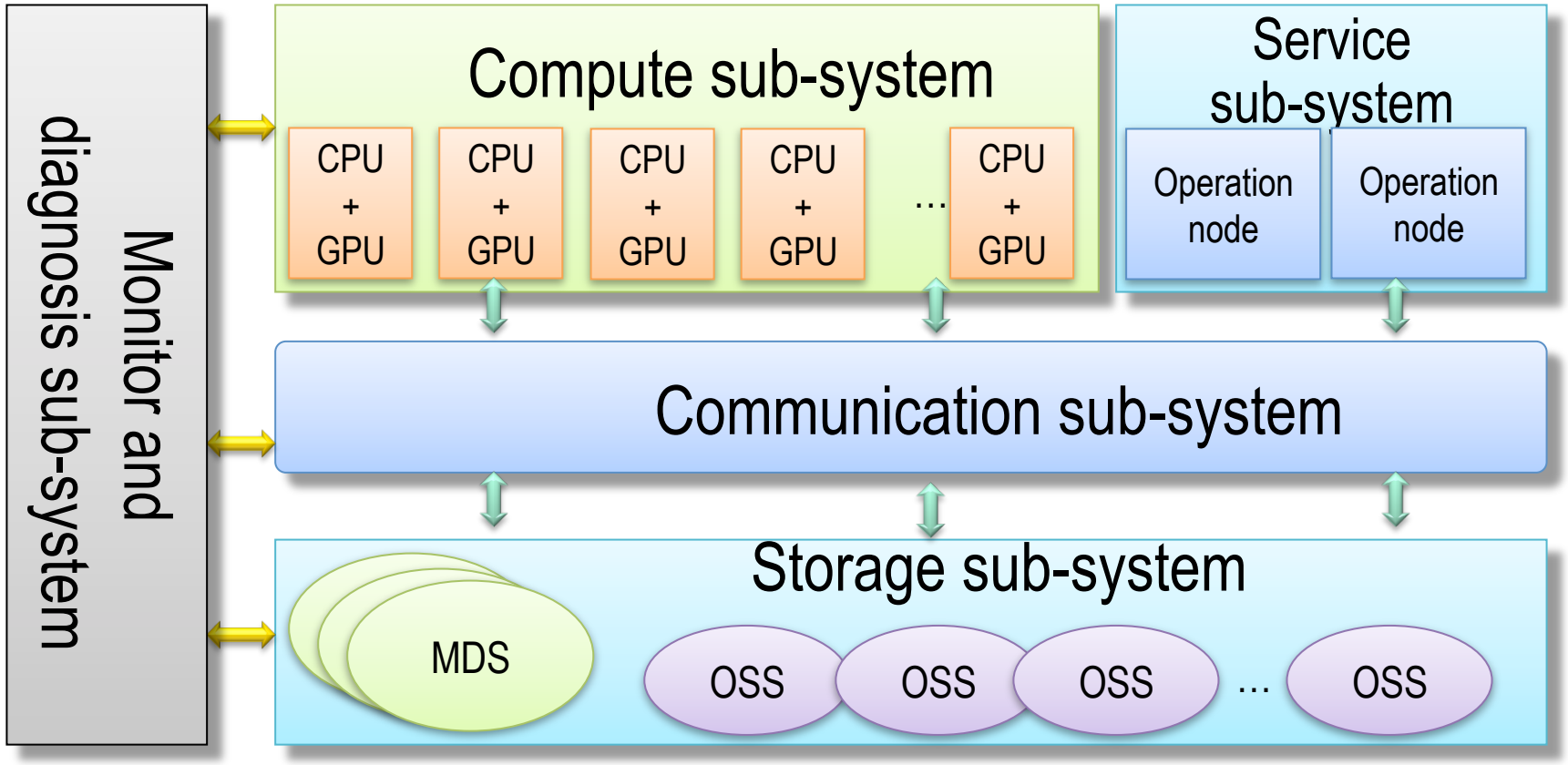
- Constellation based on AMD multicore processors
- Low power CPU and high density blade design
- High performance InfiniBand switch
- 233.472TF peak performance, 180.6TF Linpack performance
- #10 in TOP500 (Nov. 2008), the fastest machine outside USA

Lenovo DeepComp 7000

- Hybrid cluster architecture using Intel processors
  - SMP+cluster
- Two sets of interconnect
  - InInfiniBand
  - Gb Ethernet
- SAN connection between I/O nodes and disk array
- 157TF peak performance
- 106.5 TF Linpack performance (cluster)
- #19 in TOP500 (Nov. 2008)

Tianhe 1A

- **Hybrid system architecture**
  - **Computing sub-system**
  - **Service sub-system**
  - **Communication networks**
  - **Storage sub-system**
  - **Monitoring and diagnosis sub-system**

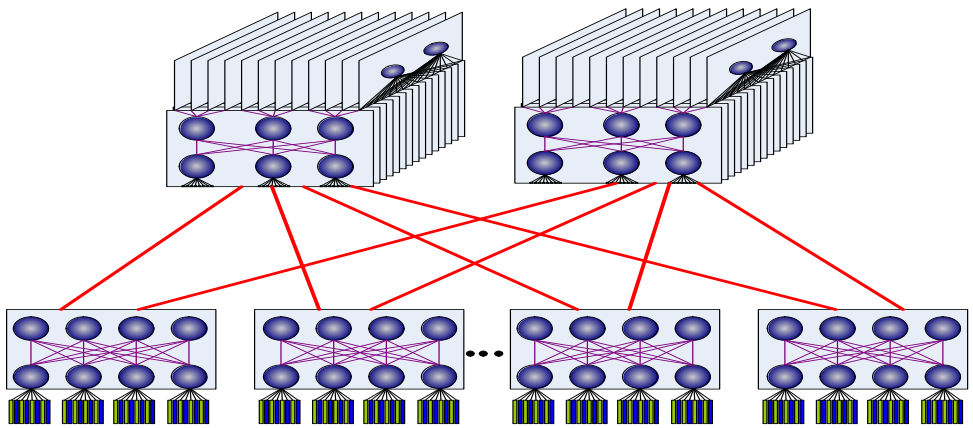


Tianhe 1A (cont')

- Computing Node
  - 2 6-core Intel processor
    - Xeon X5670 (Westmere)
    - 2.93GHz
  - 1 NVIDIA Fermi GPU
  - 32GB memory
- 7,168 computing nodes
- Perfromance
  - 4.7PF peak
  - 2.566PF Linpack
  - 54.6% efficiency
- #1 in TOP500 (2010.11)

Tianhe 1A (cont')

- Interconnect
  - Bi-directional bandwidth
    - 160Gbps, 2X IB QDR
  - Topology: fat tree



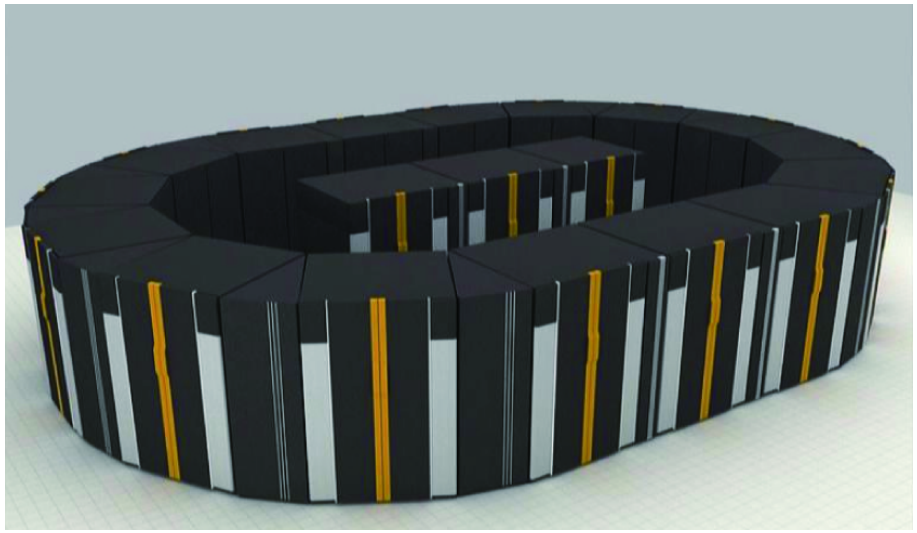
- Storage
  - 2PB
- Power consumption
  - 4MW
- Footprint
  - 700M^2
- Cooling
  - Water-cooling+fan

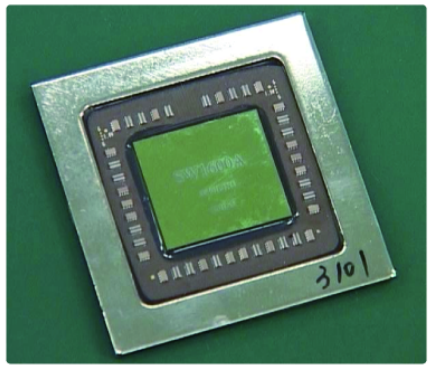
Dawning 6000

- Hybrid system
  - Service unit (Nebula)
    - 9600 Intel 6-core Westmere processor
    - 4800 nVidia Fermi GPGPU
    - 3PF peak performance
    - 1.27 Linpack performance
    - 2.6 MW
  - Computing unit
    - Domestic processor
- #2 (2010.6) and #3 (2010.11) in TOP500

Sunway （Blue Ray）

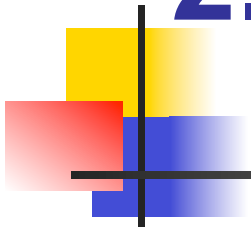
- Completed by the end of 2010
  - 1.1 PF peak, 738TF Linpack
  - Very compact system
    - 128TF/Rack
  - Implemented with domestic 16-core processors
  - Exploring possible architectures and key technologies for 10P-scale computers





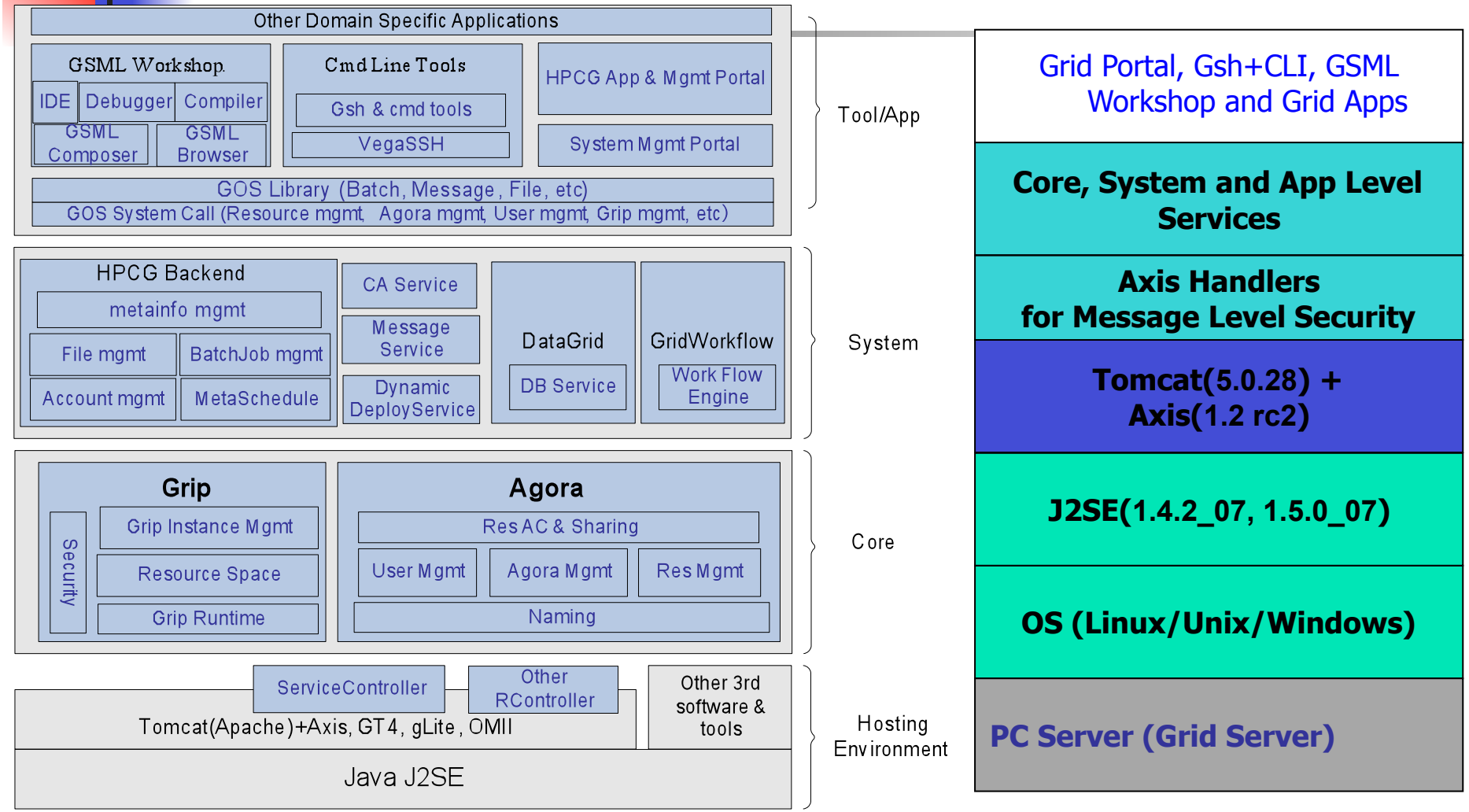
国产 “申威”16核CPU

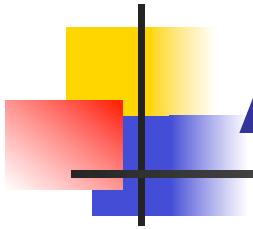
2. Grid software development



- Goal
  - Developing system level software for supporting grid environment operation and grid applications
  - Pursuing technological innovation
  - Emphasizing maturity and robustness of the software

CNGrid GOS Architecture



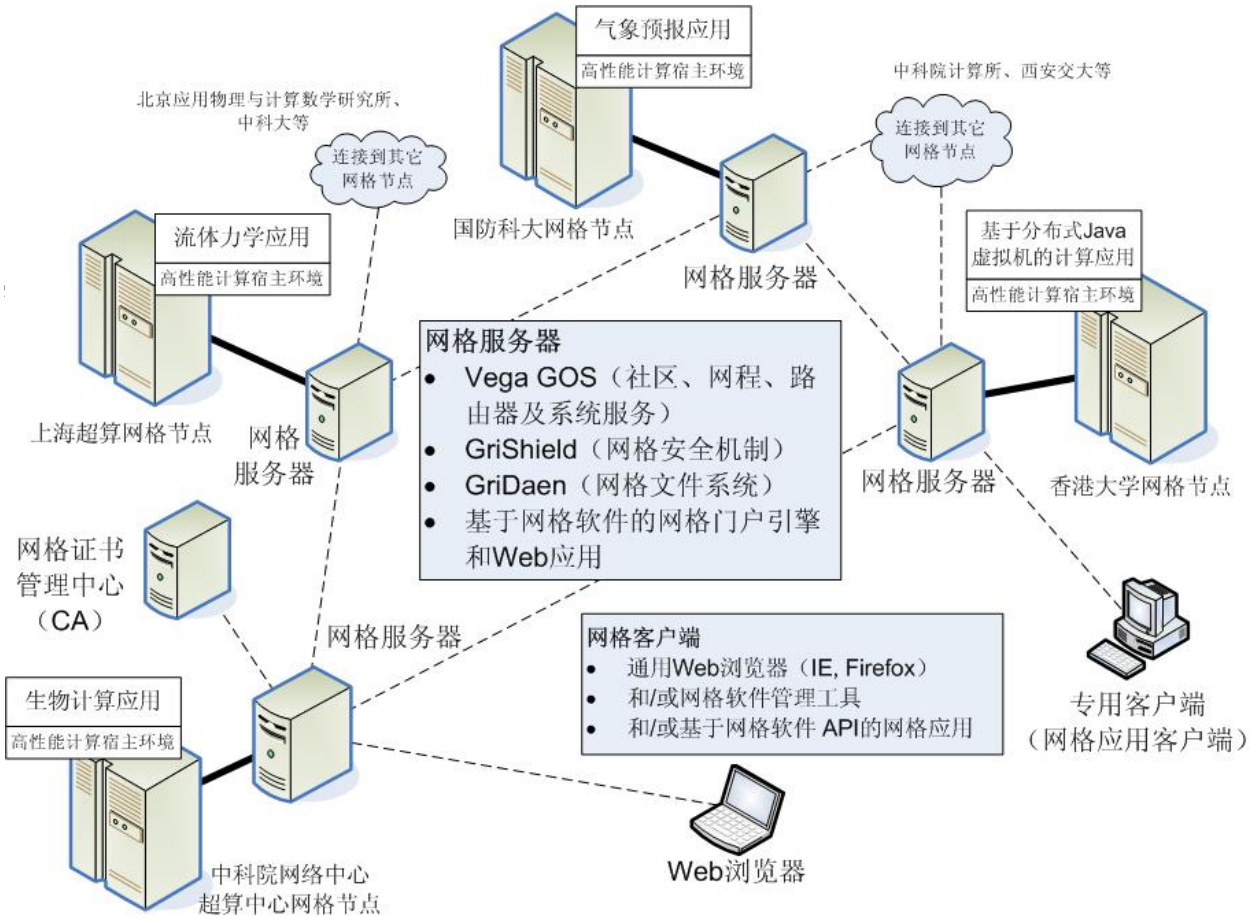


Abstractions

- Grid community: Agora
  - persistent information storage and organization
- Grid process: Grip
  - runtime control

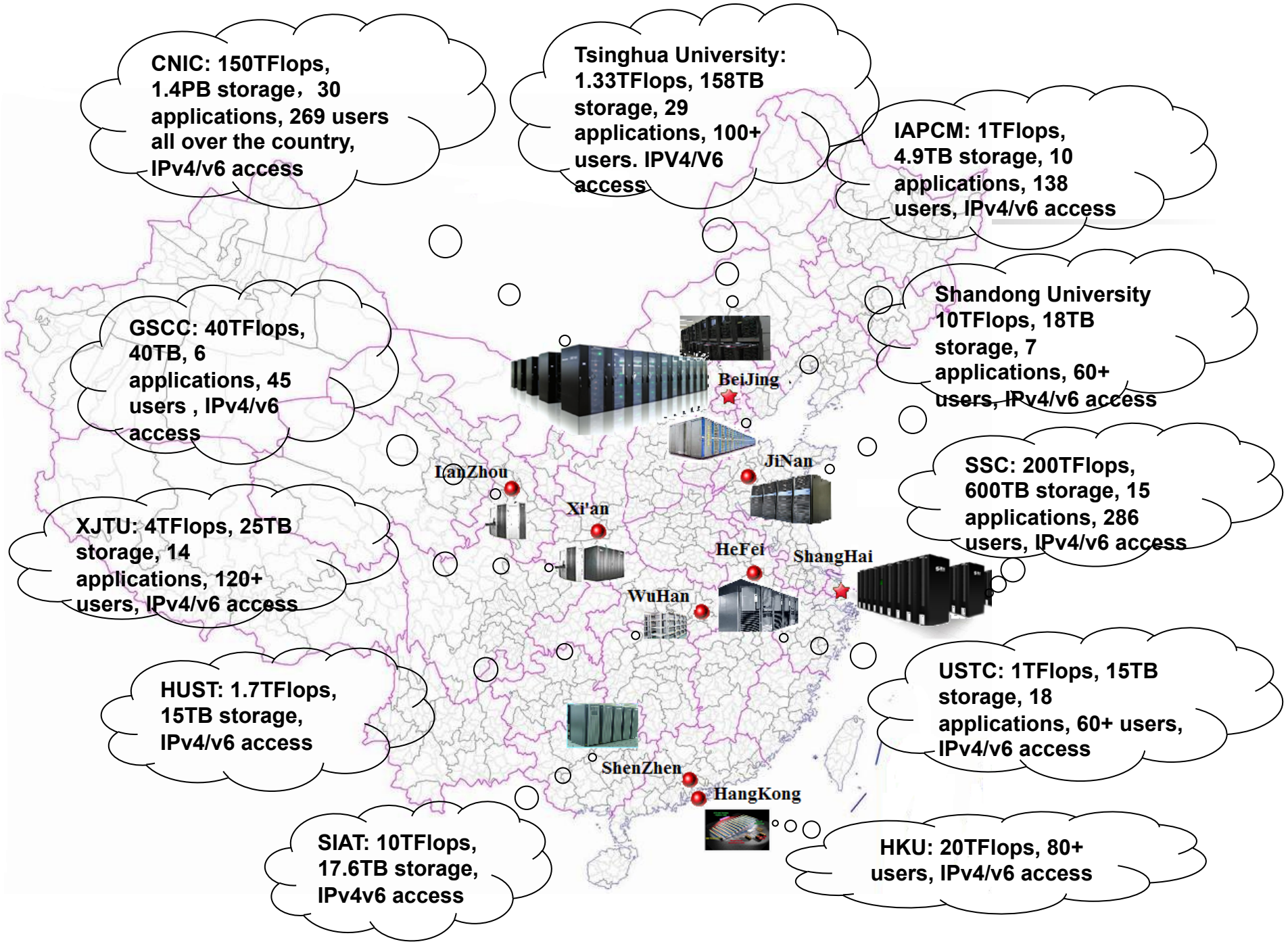
CNGrid GOS deployment

- CNGrid GOS deployed on 11 sites and some application Grids
- Support heterogeneous HPCs: Galaxy, Dawning, DeepComp
- Support multiple platforms
  Unix, Linux, Windows
- Using public network connection, enable only HTTP port
- Flexible client
  - Web browser
  - Special client
  - GSML client



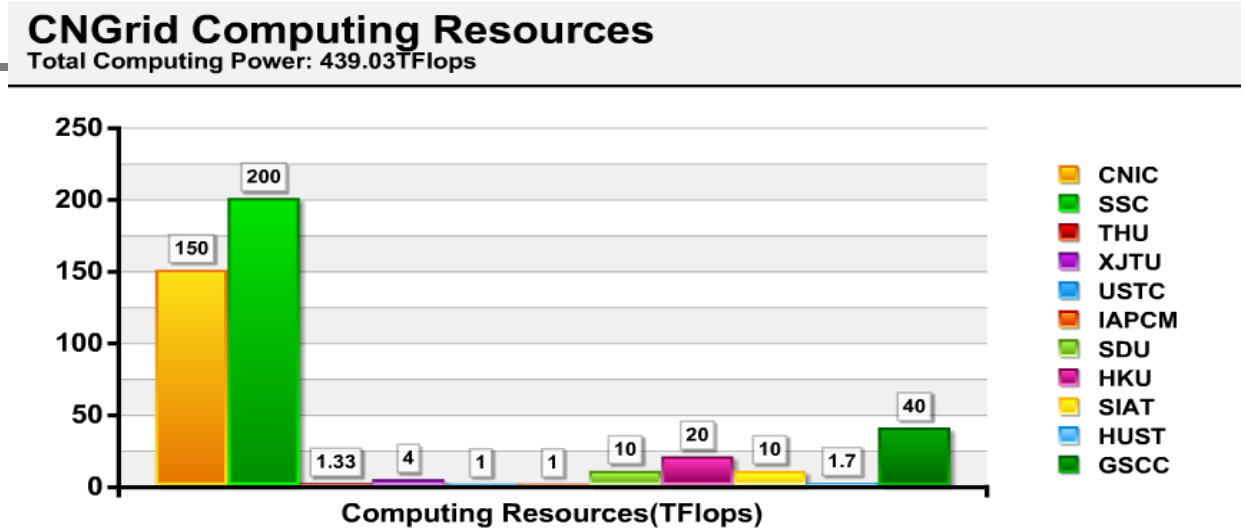
3. CNGrid development

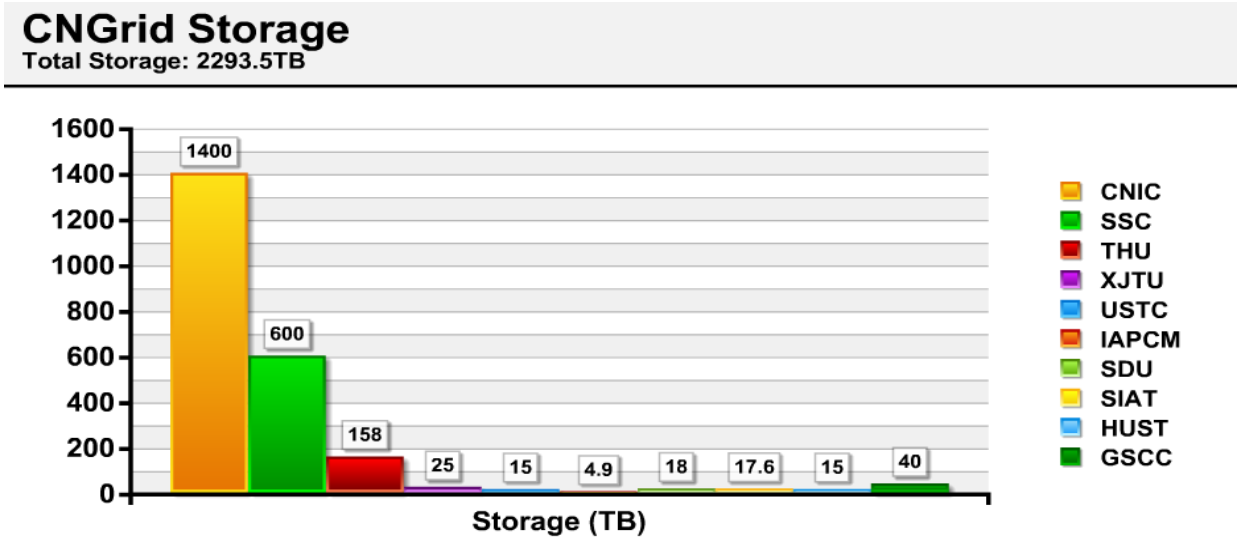
- 11 sites
  - CNIC, CAS (Beijing, major site)
  - Shanghai Supercomputer Center (Shanghai, major site)
  - Tsinghua University (Beijing)
  - Institute of Applied Physics and Computational Mathematics (Beijing)
  - University of Science and Technology of China (Hefei, Anhui)
  - Xi'an Jiaotong University (Xi'an, Shaanxi)
  - Shenzhen Institute of Advanced Technology (Shenzhen, Guangdong)
  - Hong Kong University (Hong Kong)
  - Shandong University (Jinan, Shandong)
  - Huazhong University of Science and Technology (Wuhan, Hubei)
  - Gansu Provincial Computing Center
- The CNGrid Operation Center (based on CNIC, CAS)


CNIC: 150TFlops, 1.4PB storage，30 applications, 269 users all over the country, IPv4/v6 access
Tsinghua University: 1.33TFlops, 158TB storage, 29 applications, 100+ users. IPV4/V6 access
IAPCM: 1TFlops, 4.9TB storage, 10 applications, 138 users, IPv4/v6 access
GSCC: 40TFlops, 40TB, 6 applications, 45 users , IPv4/v6 access
Shandong University 10TFlops, 18TB storage, 7 applications, 60+ users, IPv4/v6 access
BeiJing
JiNan
LanZhou
Xi'an
SSC: 200TFlops, 600TB storage, 15 applications, 286 users, IPv4/v6 access
XJTU: 4TFlops, 25TB storage, 14 applications, 120+ users, IPv4/v6 access
HeFei
ShangHai
WuHan
USTC: 1TFlops, 15TB storage, 18 applications, 60+ users, IPv4/v6 access
HUST: 1.7TFlops, 15TB storage, IPv4/v6 access
ShenZhen
HangKong
SIAT: 10TFlops, 17.6TB storage, IPv4v6 access
HKU: 20TFlops, 80+ users, IPv4/v6 access

CNGrid: resources

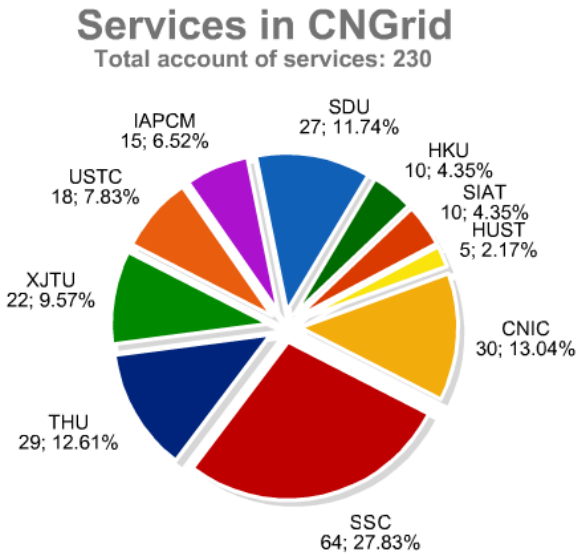
- 11 sites
- >450TFlops
- 2900TB storage
- Three PF-scale sites will be integrated into CNGrid soon

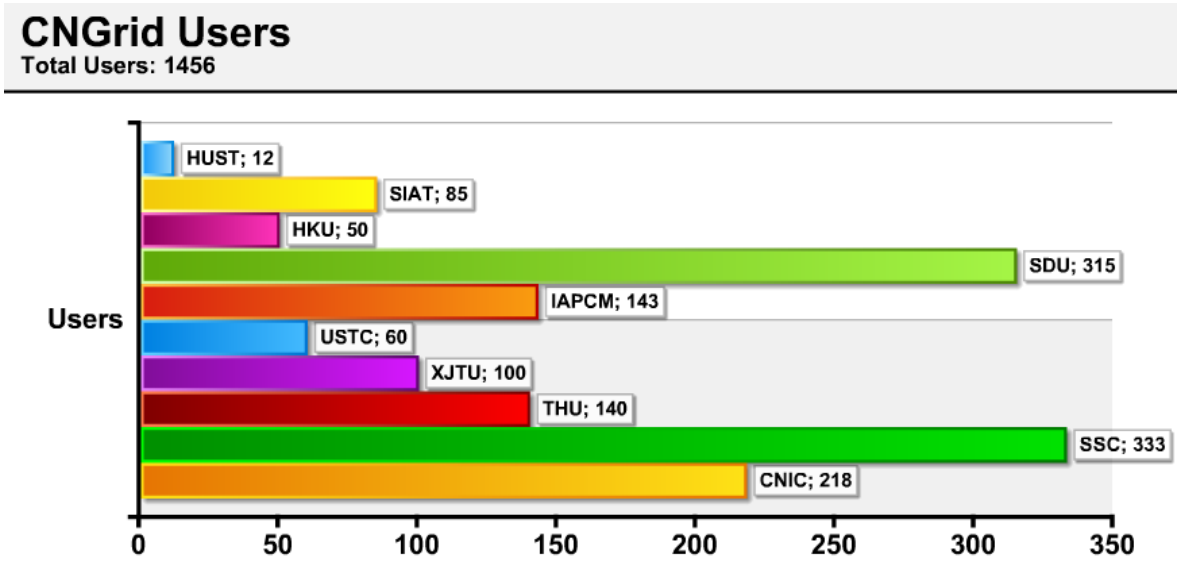




CNGrid：services and users

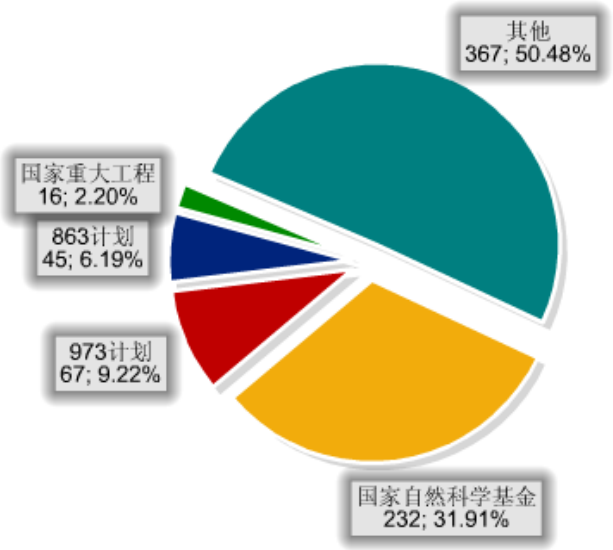
- 230 services
- >1400 users
  - China commercial Aircraft Corp
  - Bao Steel
  - automobile
  - institutes of CAS
  - universities
  - ……

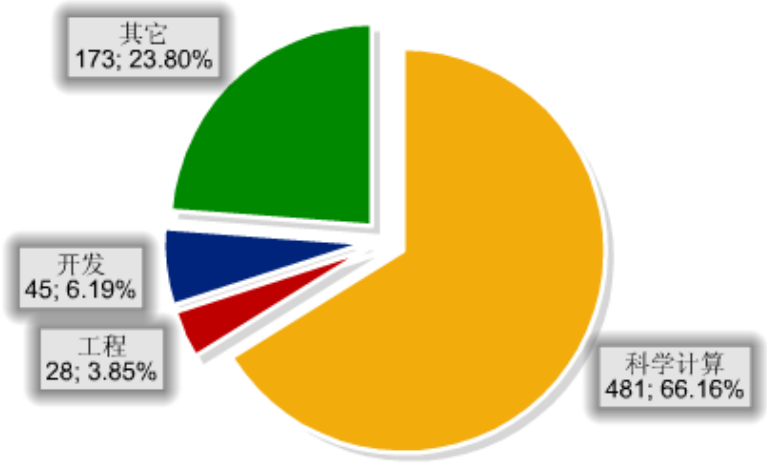




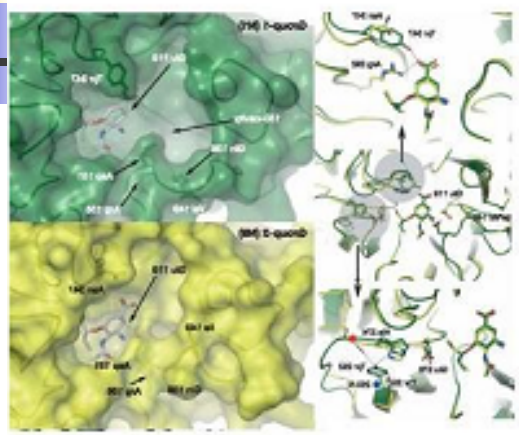
CNGrid：applications

- Supporting >700 projects
  - 973, 863, NSFC, CAS Innovative, and Engineering projects

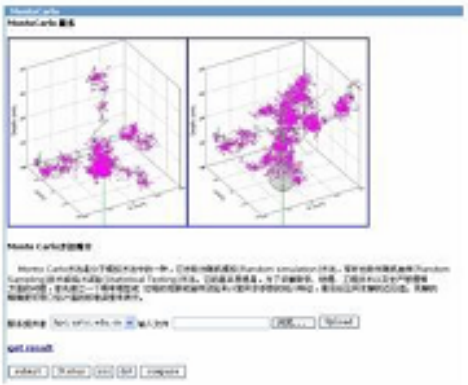




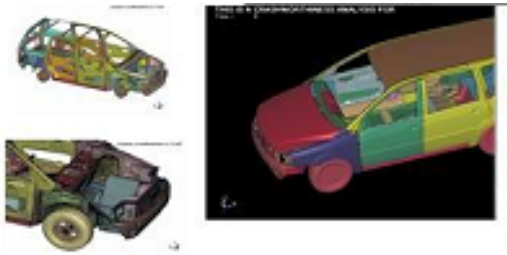
CNGrid applications



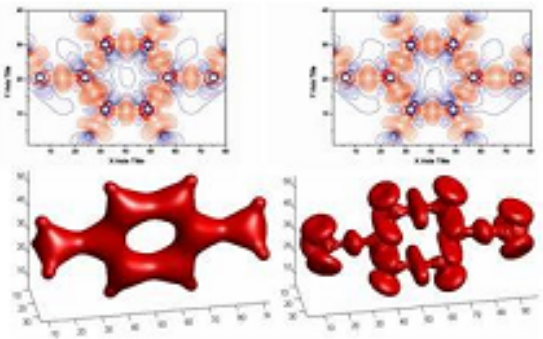
Bird flu drug screening



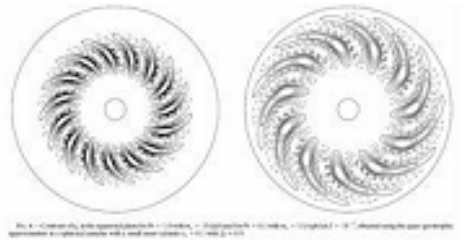
Monte Carlo simulation



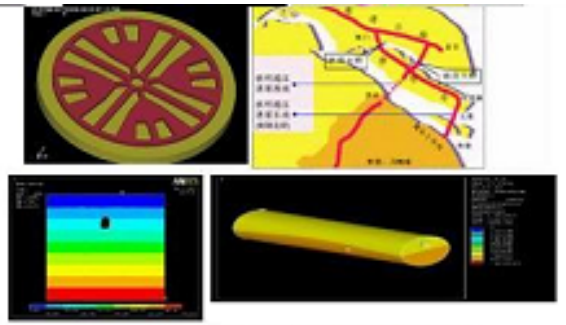
Car design safety analysis



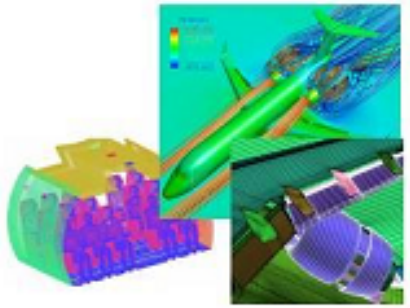
Computational physics



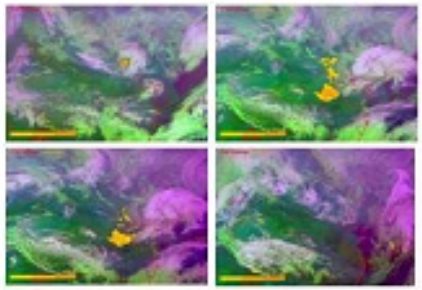
Magnetic hydrodynamics



Tunnel Construction simulation



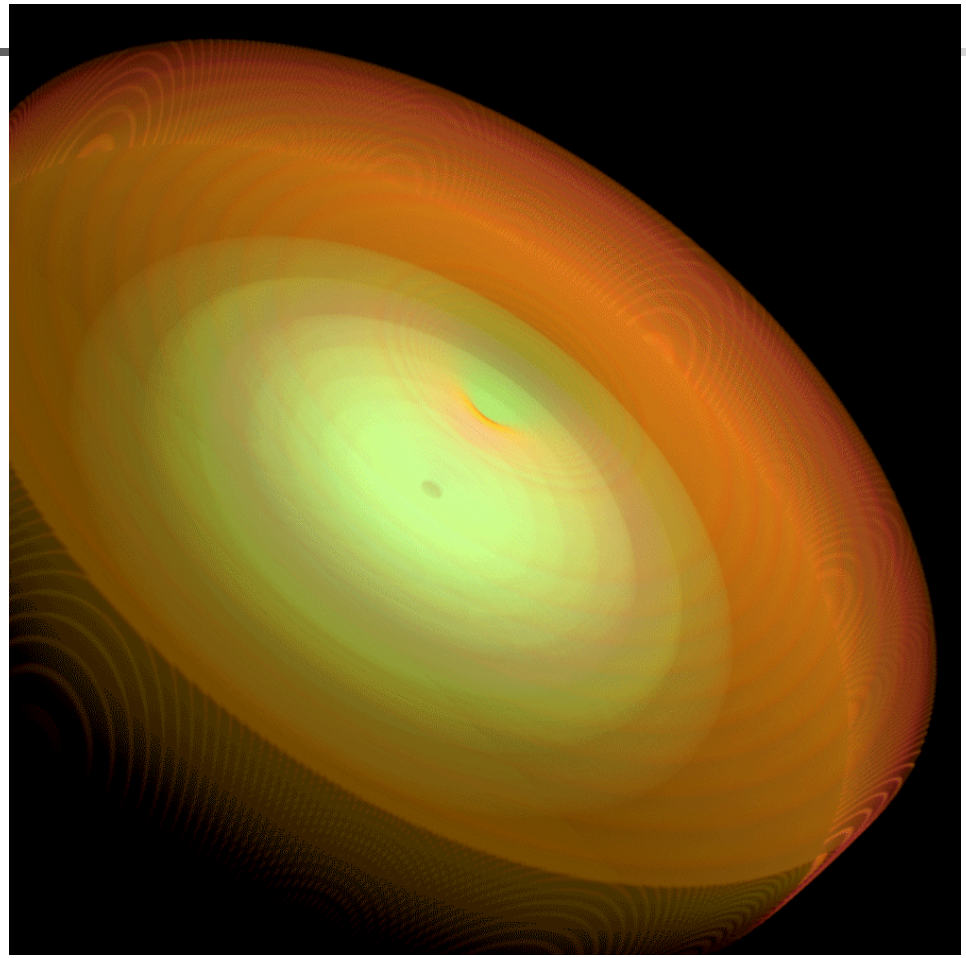
Aircraft design aerodynamics



Sand storm weather forecast

CNGrid Applications

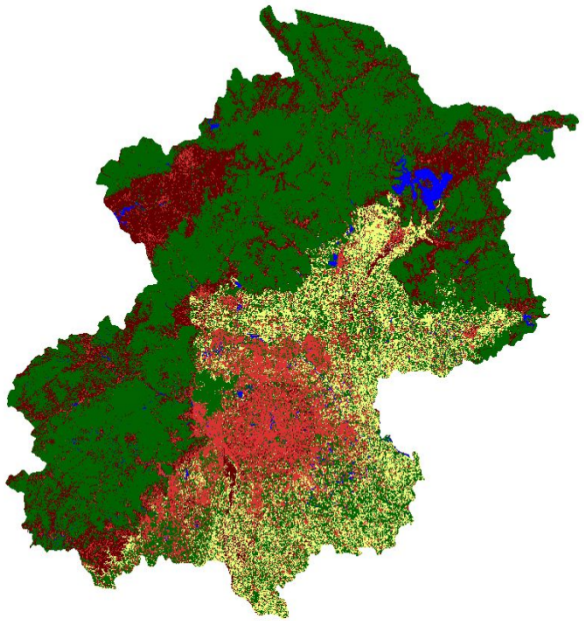
Galaxy wind simulation



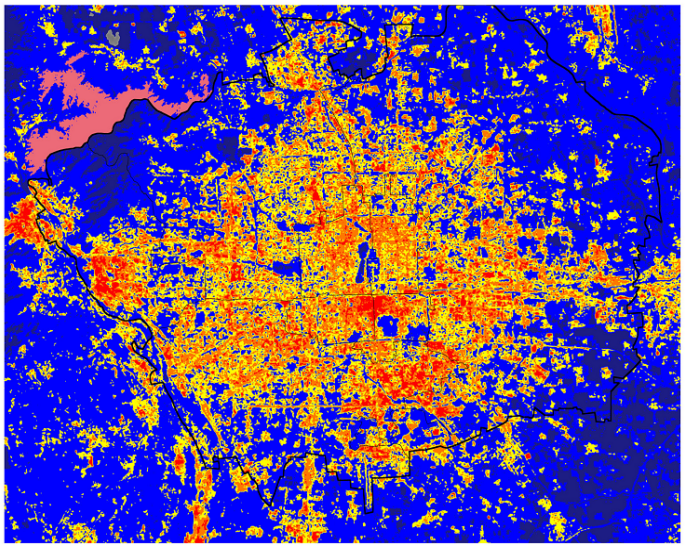
Using 8192 cores

Ecological Research

- Effect of Urban “heat island” to Beijing city planning
- Sand storm study: source and spead

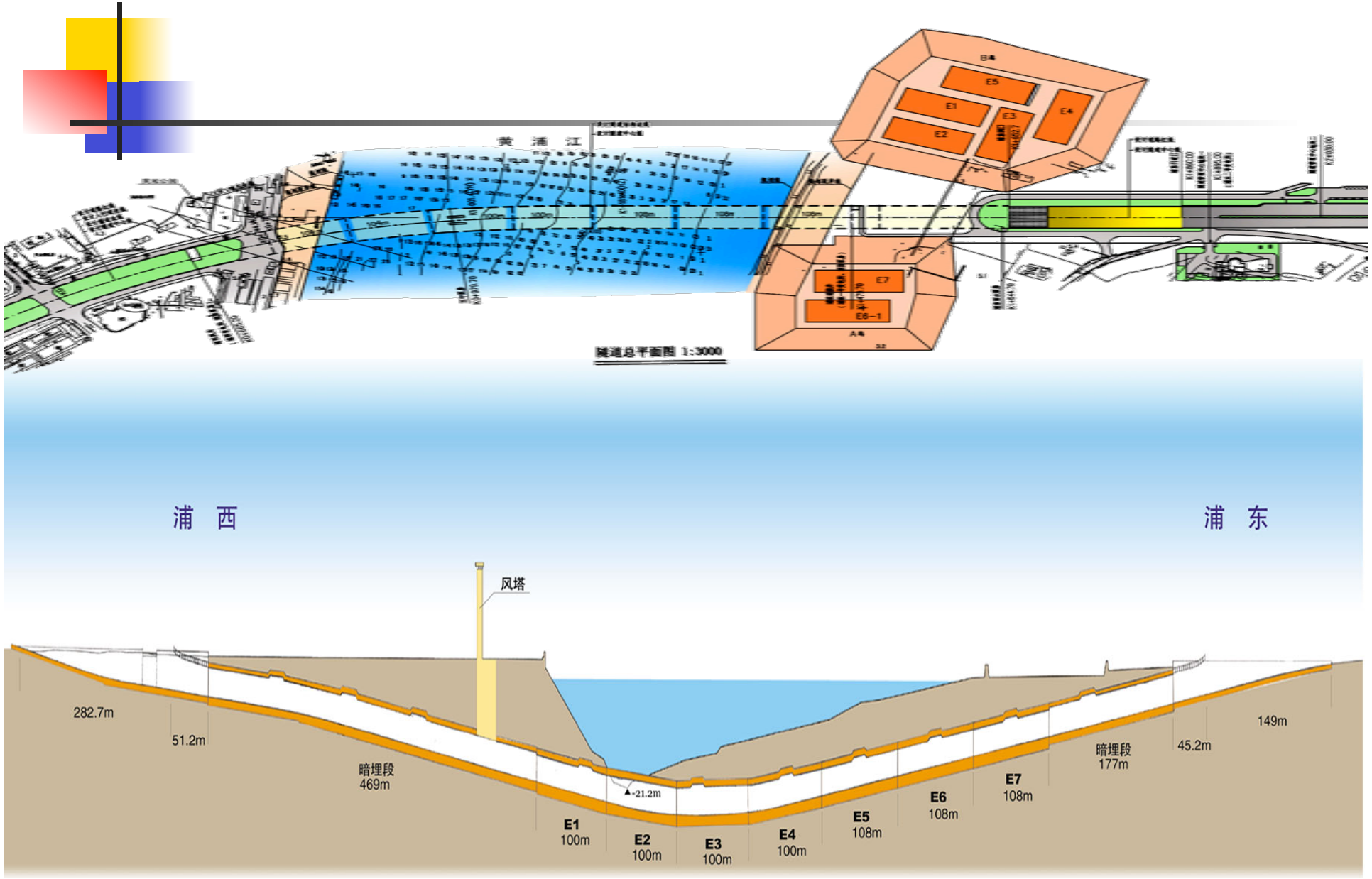


Beijing earth’s surface coverage

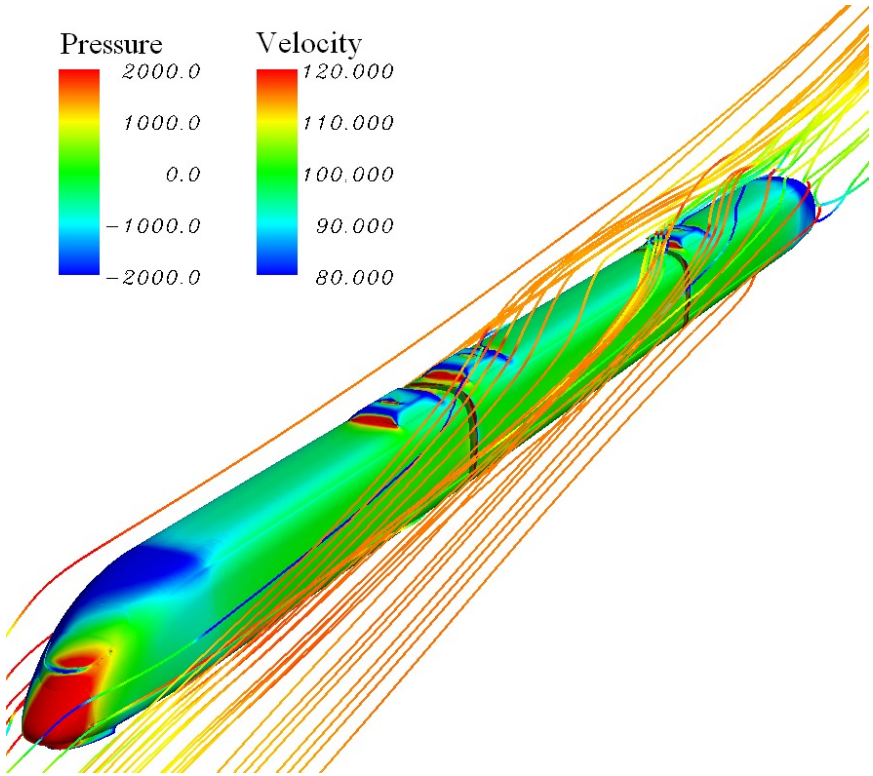


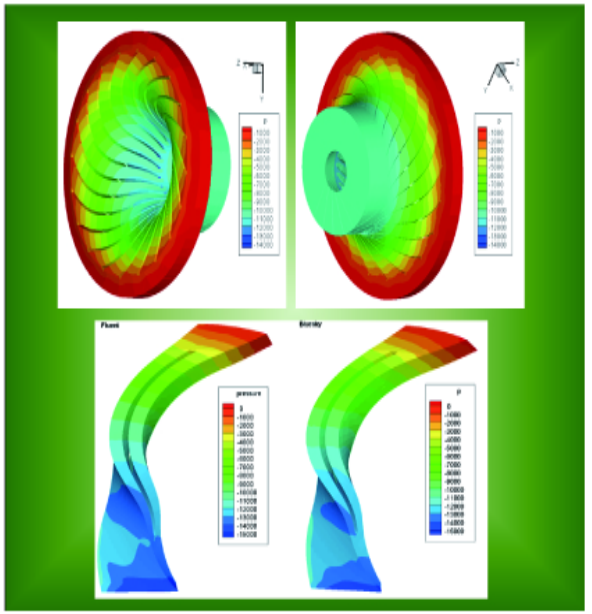
Beijing City Heat island distribution

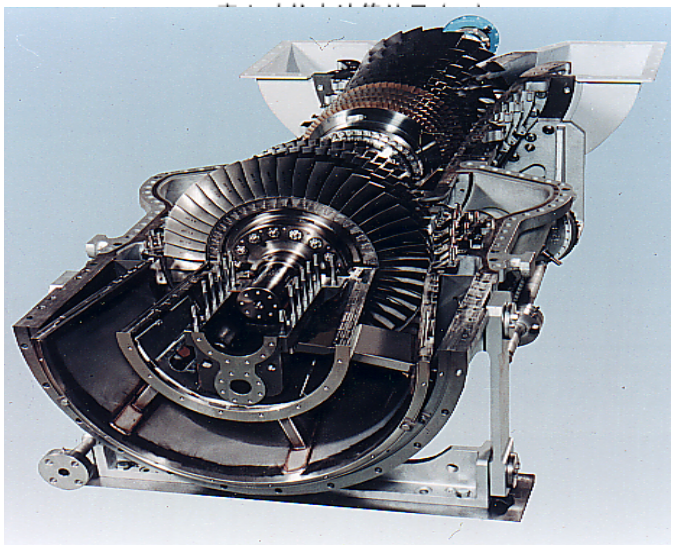
Engineering computing



Industrial applications

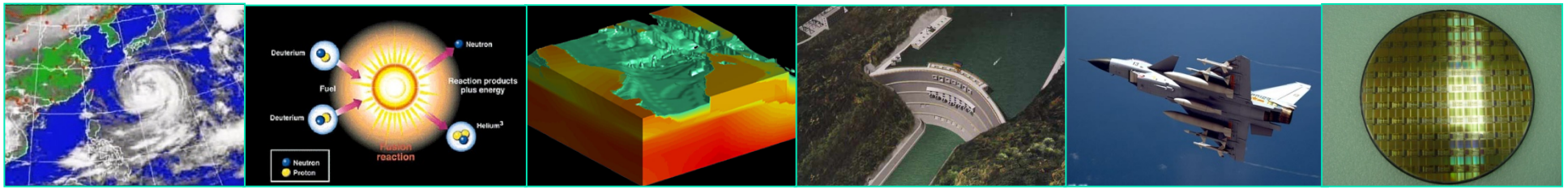






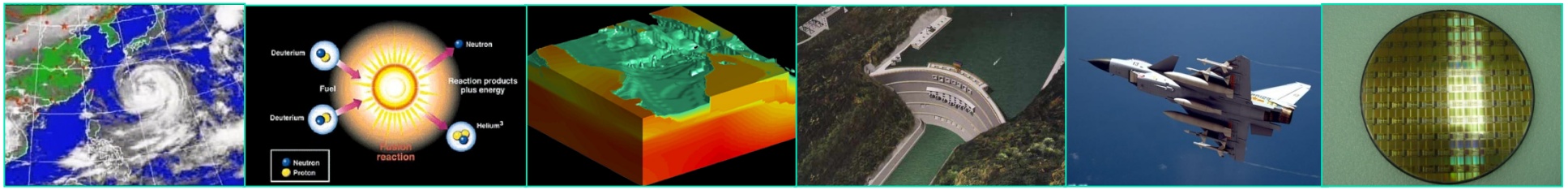
4: Grid and HPC applications

- Developing productive HPC and Grid applications
- Verification of the technologies
- Applications from selected areas
  - Resource and Environment
  - Research
  - Services
  - Manufacturing



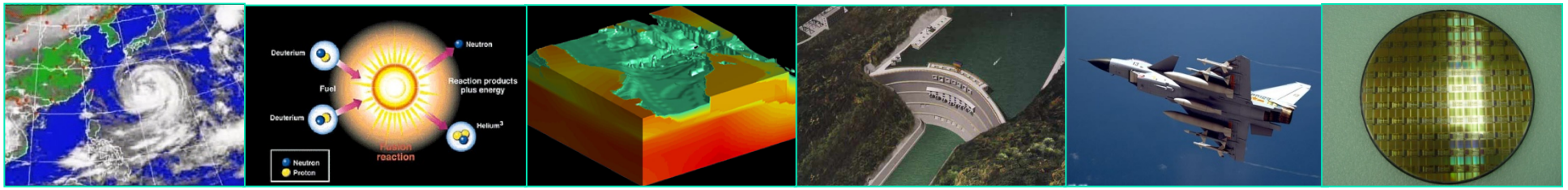
Grid applications

- Drug Discovery
- Weather forecasting
- Scientific data Grid and its application in research
- Water resource Information system
- Grid-enabled railway freight Information system
- Grid for Chinese medicine database applications
- HPC and Grid for Aerospace Industry (AviGrid)
- National forestry project planning, monitoring and evaluation

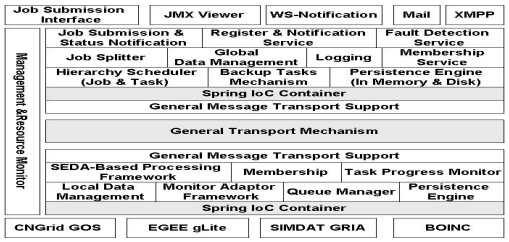


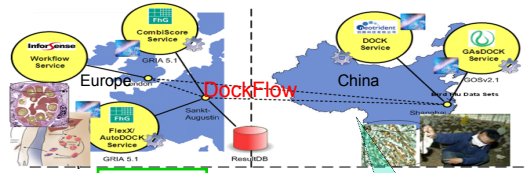
HPC applications

- Computational chemistry
- Computational Astronomy
- Parallel program for large fluid machinery design
- Fusion ignition simulation
- Parallel algorithms for bio- and pharmacy applications
- Parallel algorithms for weather forecasting based on GRAPES
- 10000+ core scale simulation for aircraft design
- Seismic imaging for oil exploration
- Parallel algorithm libraries for PetaFlops systems

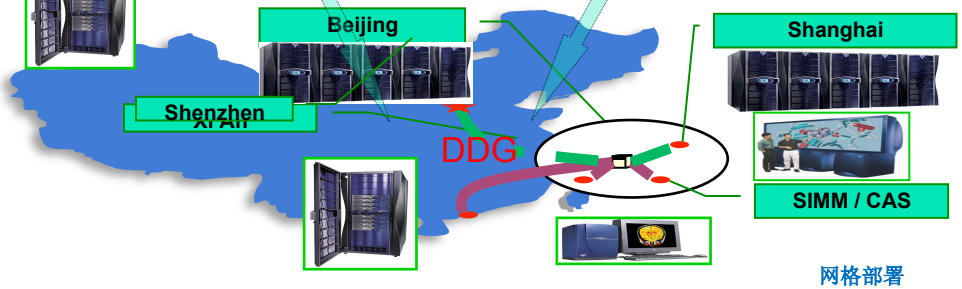


Drug Discovery Grid





国际合作

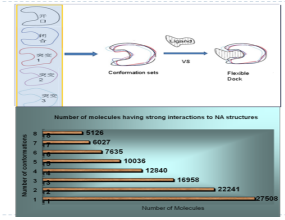


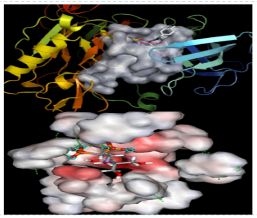
网格部署

体系结构

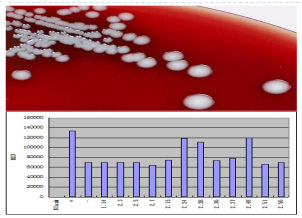
用户界面

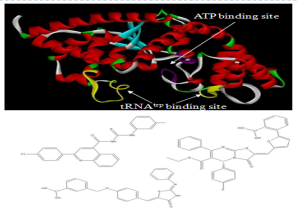
禽流感神经氨酸酶抑制剂设计



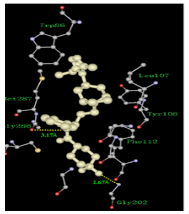


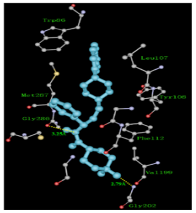
针对表皮葡萄球菌的抑制剂设计

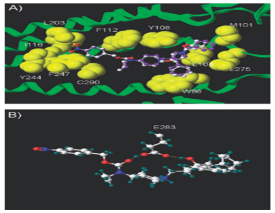




针对HIV重要靶点CCR5的抑制剂设计



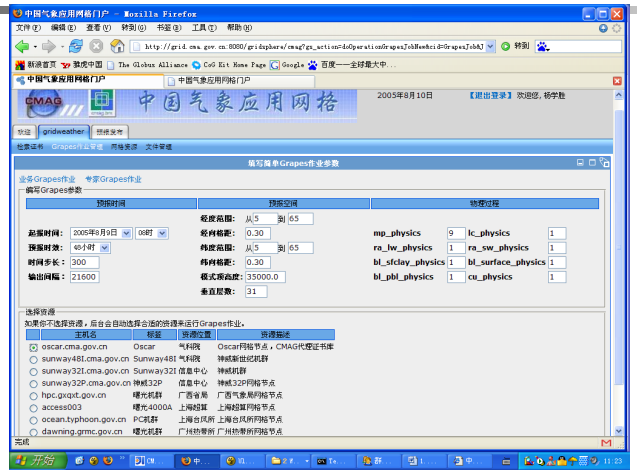


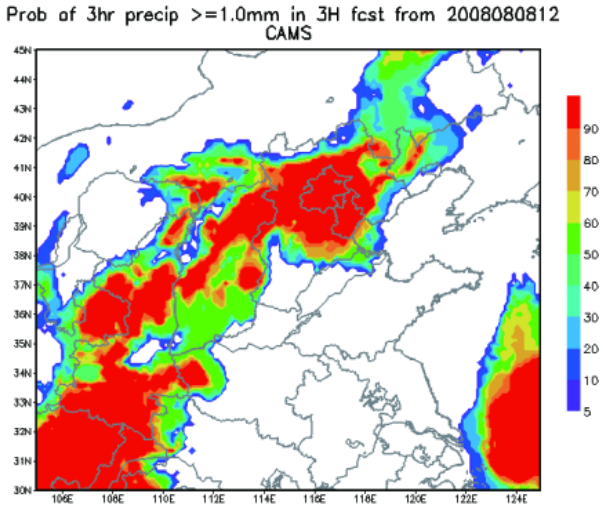


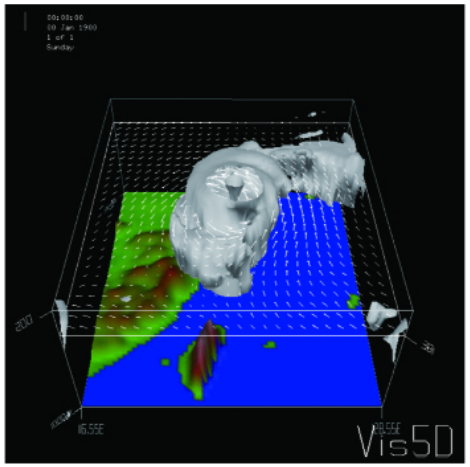
典型应用

China Meteorological Application Grid (CMAG)

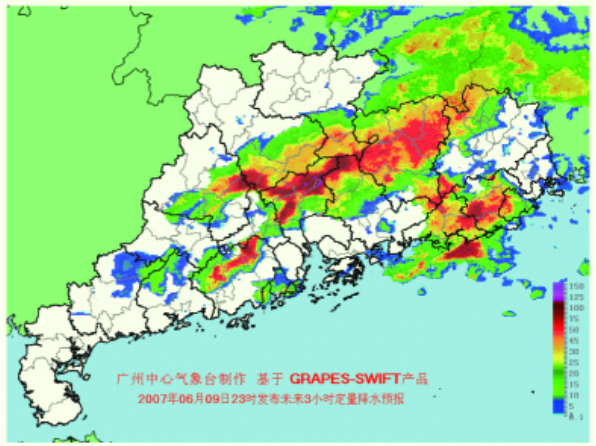
- A platform for collaborative research on new weather prediction model
- Providing new weather forecast services (time and location-specific) to less developed areas





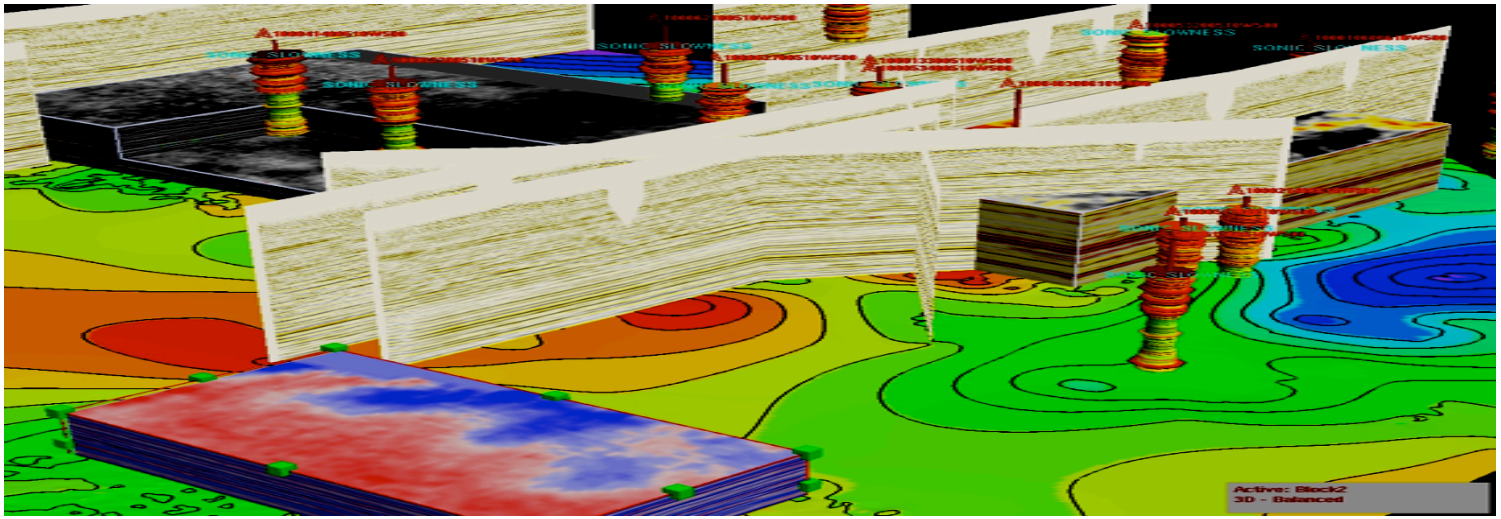


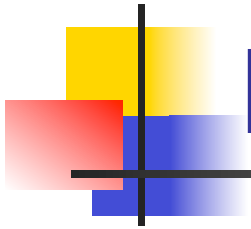
GRAPES计算的桑美台风结构



GRAPES 的降水预报

Seismic oil exploration image processing



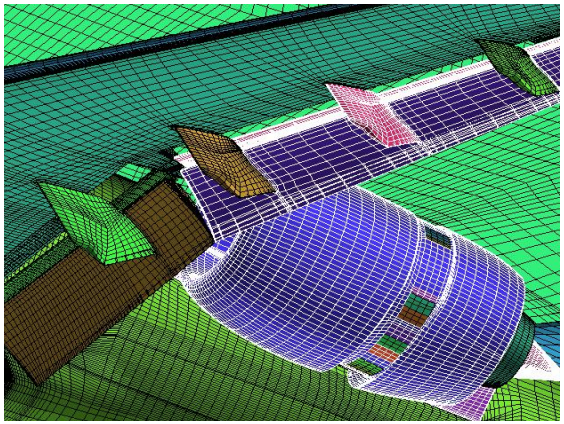


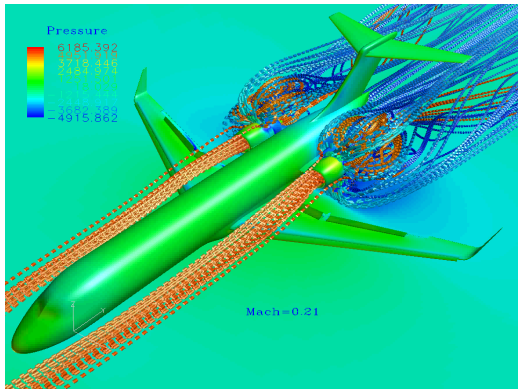
Domain application Grid

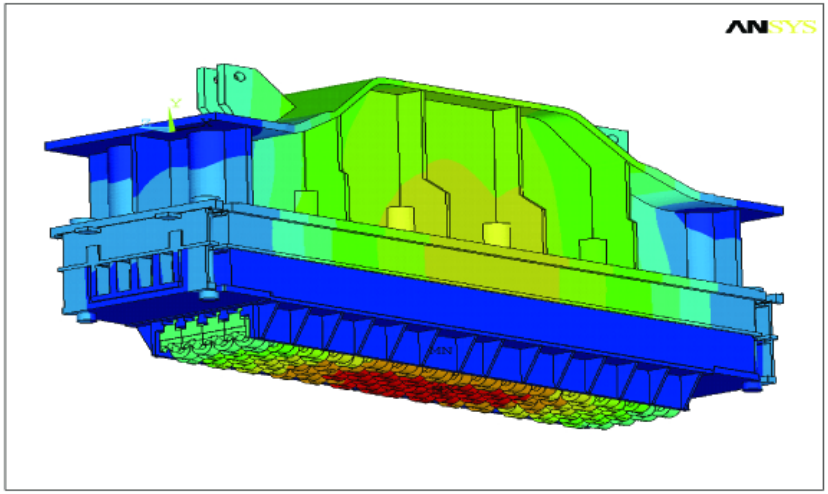
- Domain application Grids for
  - Simulation and optimization
    - automobile industry
    - aircraft design
    - steel industry
  - Scientific computing
    - Bio-information application
    - computational chemistry
- Introducing Cloud Computing concept
  - CNGrid—as IaaS and partially PaaS
  - Domain application Grids—as SaaS and partially PaaS

Domain application Grid: Simulation & Optimization

- Integrating software for product design and optimization, supporting simulation and optimization of industrial products
- Implementing resource scheduling, user management, accounting and service center, exploring new mode for sustainable development
- Used by Shanghai Automobile, Bao Steel, and aircraft industry in Shanghai

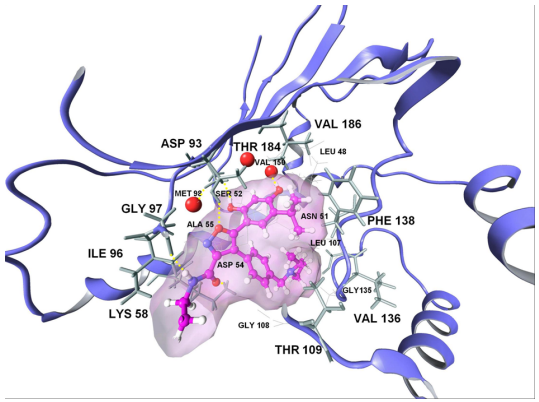


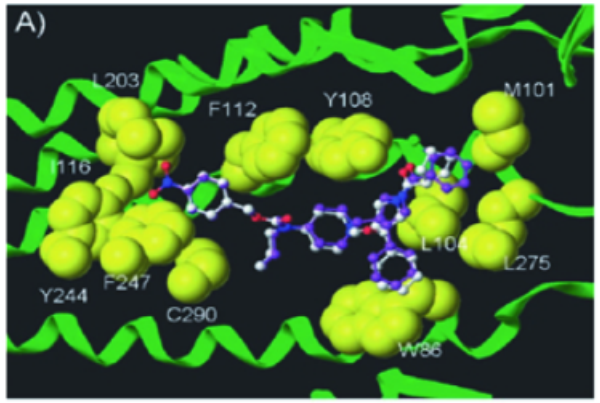


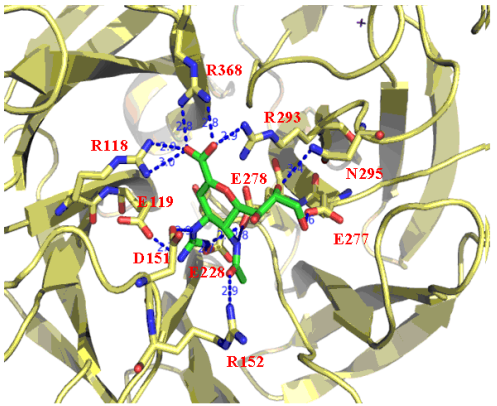


Domain application Grid: Bioinfo. & Comp. Chemistry

Gaussian	半经验计算和从头计算使用最广泛的量子化学软件
VASP	使用赝势和平面波基组，进行从头量子力学分子动力学计算的软件包
NAMD	并行度最好的大规模并行分子动力学模拟软件
Abinit	从头算计算软件
GAMESS(US)	计算速度最快的从头算量子化学软件
AMBER	最好的生物分子力场软件
Gromacs	计算速度最快的分子动力学模拟软件
ADF	专门作密度泛函计算的软件
NWChem	大规模并行量子化学软件
Molpro	国际上广泛使用的专业级高精度电子结构量化计算软件
Q-Chem	电子结构从头算软件，可以对分子的基态和激发态进行第一原理计算
TurboMol	可对激发态进行准确计算的量子化学软件
LAMMPS	大规模原子(分子)并行模拟器
WIEN2K	使用密度泛函理论计算晶体结构的量子力学软件
Stereodynamics-QCT	半经验轨线立体动力学计算软件

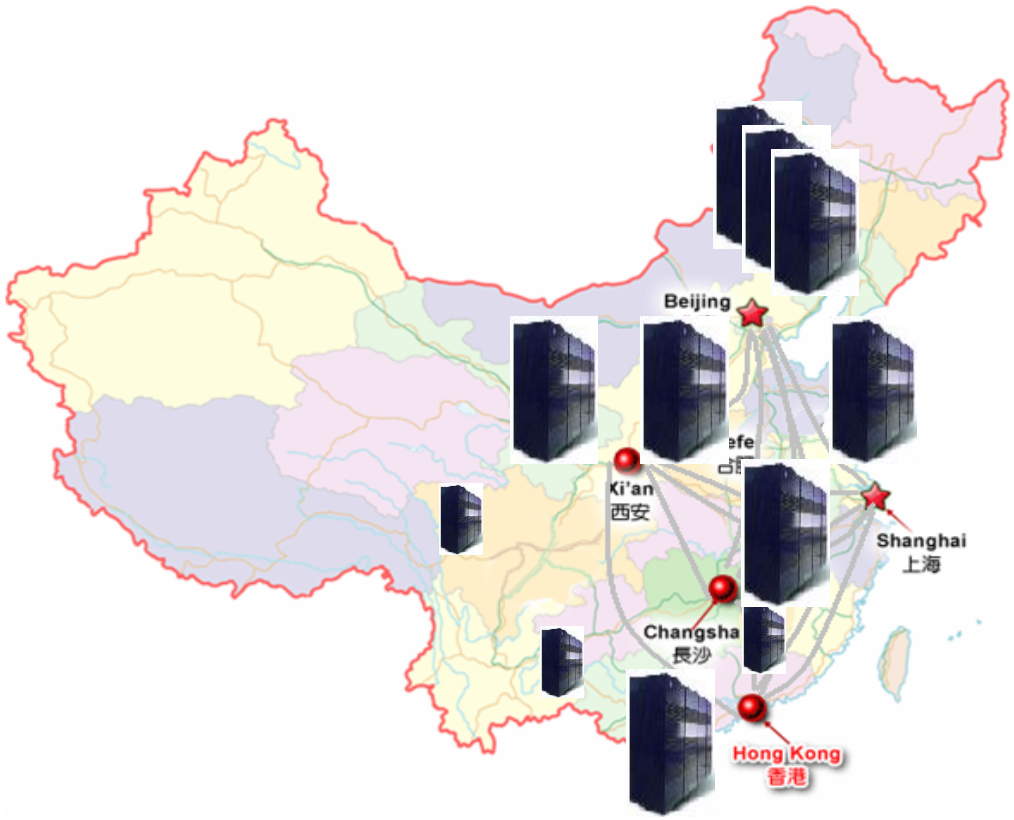




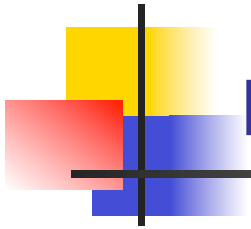


CNGrid (2006-2010)

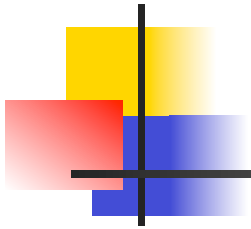
- HPC Systems
  - Two 100 Tflops
  - 3 PFlops
- Grid Software: CNGrid GOS
- CNGrid Environment
  - 14 sites
  - One OP Centers
  - Some domain app. Grids
- Applications
  - Research
  - Resource & Environment
  - Manufacturing
  - Services



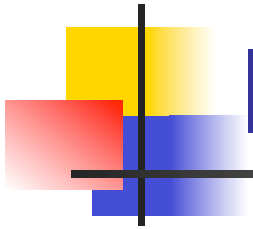
China's current status in the related fields



- Significant progress in HPC development and Grid service environment
- Still far behind in many aspects
  - kernel technologies
  - applications
  - multi-disciplinary research
  - talent people
- Sustainable development is crucial



- Need to put more attention to programming of CPU/GPU hybrid systems
- Need to find the way of sustainable development for CNGrid
- Need to further promote applications
- Need to educate/train more talents



Next 5-year plan

- China's 863 program has identified priority topics in both HPC and cloud computing
- A key project on cloud computing has been launched
  - "Key technologies and systems of cloud computing (1st phase)
    - Network operating systems
    - Network search engines
    - Network based language translation

Next 5-year plan (cont')

- A strategic study has been conducted on high productivity computers and application environment
- A proposal for new HPC key project has been submitted to the MOST
- Emphasizing balanced development in high productivity computers, application environment, and HPC applications
- We wish to continue our effort in this field

Thank you!