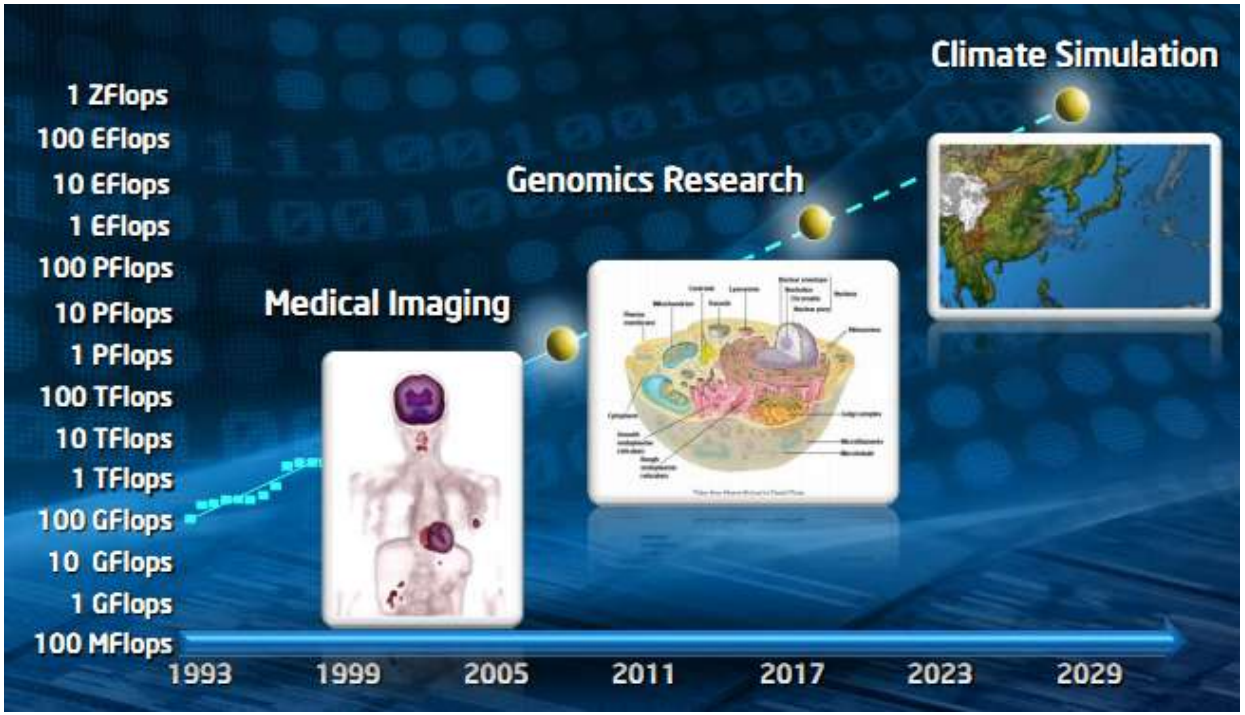
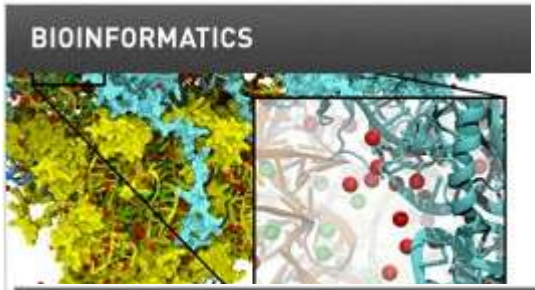
HPC Application Innovation Strategy

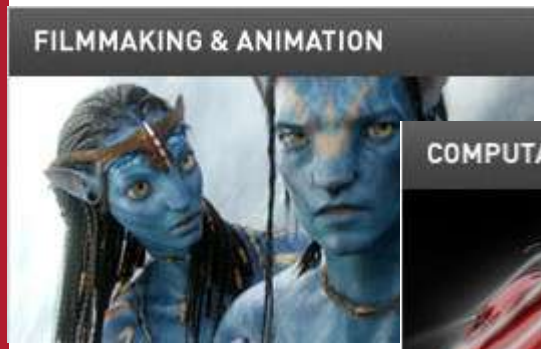
Leijun Hu

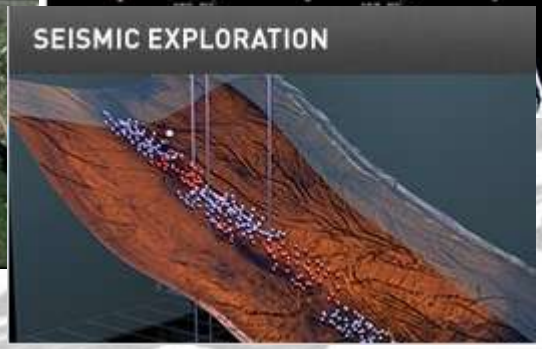
CTO, Inspur Information

From Petascale to Exascale







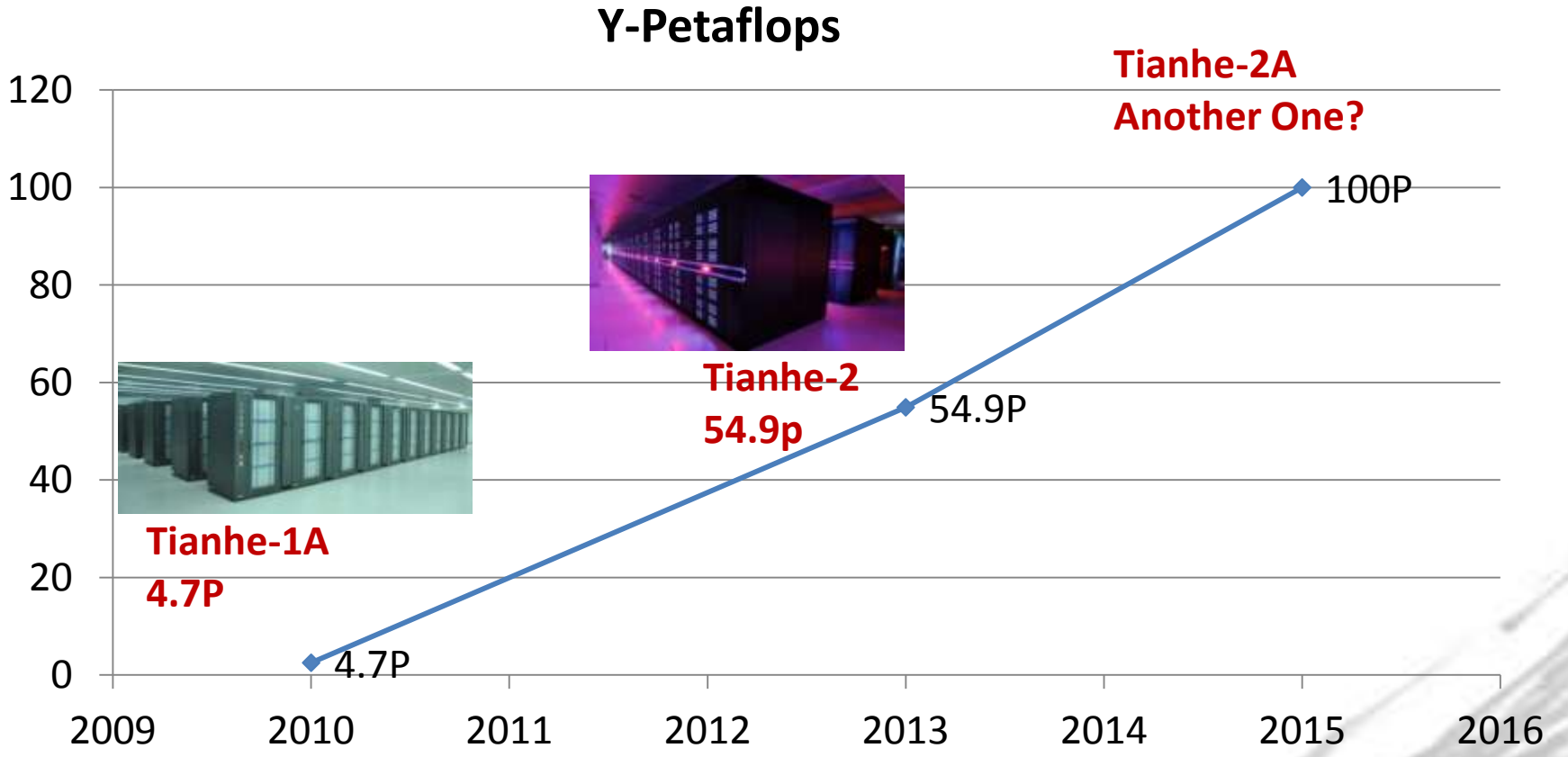


Tianhe-2

- 2013 33.86 Petaflops (Rmax)
  54.9 Petaflops (Rpeak)
- 2015 target 100 Petaflops (Rpeak)

- No.1 @Top500 June, 2013
- Co-Developed by NUDT and Inspur

100P in 2015



According to China 863 High tech Plan,
2 sets of 100 Petaflops supercomputers in China in 2015.

Challenge for China HPC

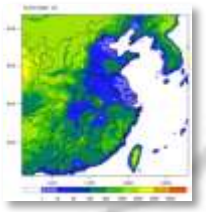
Tianhe1A, Tianhe2

HPC System Building

HPC Applications

HPC Talent Cultivation

ASC Student Supercomputer Challenge



- 12th five-year plan as Chinese government
- 100M RMB on parallel computing software R&D
- 200M RMB on software platform development

Tianhe-2, the arena for ASC14 Student Supercomputer Challenge Final Contest

82 teams from 5 continents registered for ASC14
16 teams are qualified for the ASC14 finalists:

Top16 finalists will

1 built there own cluster under 3KW and run 5 applications

2 Optimize one application on Tianhe-2 supercomputer

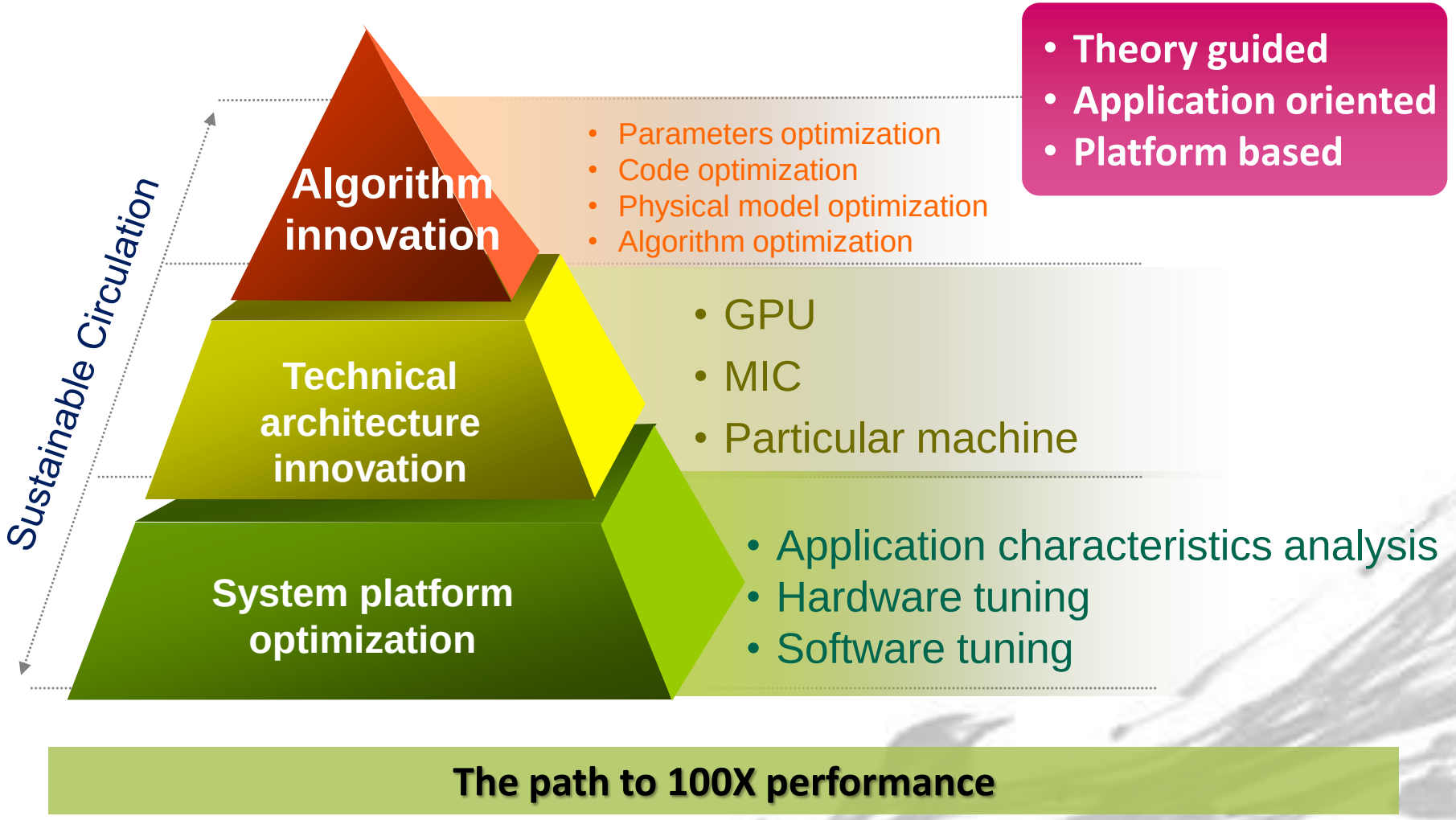
Welcome to join ASC 14 Final:

Sun Yat-Sen University (Guangzhou) on April 21-25,2014

Website：http://www.asc-events.org

inspur 浪潮

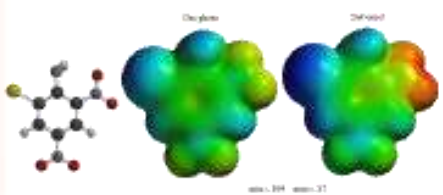
HPC application optimization strategy

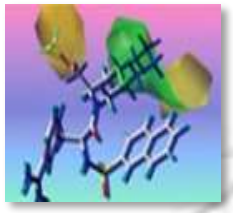
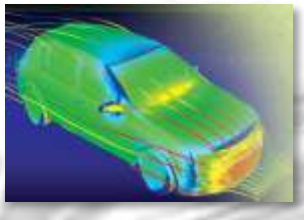


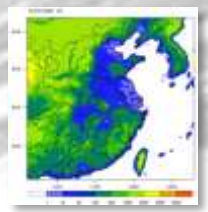
The path to 100X performance

HPC application requirement and challenge

- **Different professional discipline areas**
  - More specific and detailed on research division.
  - More technical and featured on different subjects.
- **Legacy software and code irreplaceable**
  - Legacy software and code is very important, can not be abandoned
- **How a large scale computing is running on commercial HPC system.**
  - Mathematic and physical modelling, parallel computing are more accurate and detail-oriented, accompanied with productive scaling issues, which require improving utilization of resources on HPC platform, includes: CPU, memory, network and IO etc.
  - To tune system configuration appropriately, based on the analysis for application bottleneck； To supply an evidence for parallel software development and coding update.





System platform optimization: Application Character Analysis

Preparation Specific computing platform, typical application (Application), suitable computing examples (workload)

Application Tuning

Application features analysis; Adjust application index; Adjust parallel methods; Adjust application load

Feature Collection

Collect the performance data of CPU/ Memory/ Netowrk/ IO in system levels, application levels and micro-architecture levels

Platform Tuning

Platform feature analysis; change hardware configuration; adjust hardware index, adjust middle ware

Results Provide actual-measured data and analysis support, best system architecture solution in designing industry

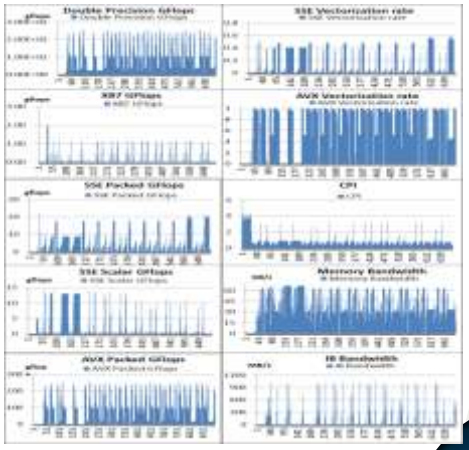
Application-based main performance features guide hardware and software optimization

Numerical Analysis Method

Application In NHM+IBA	Max(Memory usage and Bandwidth per core)	Scalability range (process number)	MPI% in 8process, p2p,collective	DISK IO	Vector and float	CPI	Others
VASP	1GB, 3.7GB/s	<16, 512>	p2p:<3,12> co: 0	Burst Write (>xxGB) in final output	30% Double float	1.2	Huge cache miss Cache size sensitive
Gaussian 03	2GB, 0.7GB/s	<32,64>	Linda thread	<0.01,2>MB/s/process	0% ~ 30% Double float	0.6 ~0.8	Different module has differ character
WIEN2K	Lapw1: 2.7GB/s Lapw2: 0.4GB/s	32	script para Mpi para	<0.5,1>MB/s/process for xxGBs	83% Double float	0.5	2 modules has different character
Material Studio	2.3 GB/s	64		2kB/s/process	83% Double float	0.62	This is for CASTEP
Amber 10	0.2GB/s	<64,256>	p2p:1.4 co: 7.2	2.3KB/s/process	15% Double float	0.73	
GROMACS 4.0	0.3GB/s	64	P2p: 6.7 Co: 5.1	4.7KB/s/process	54% single float	0.7	Enable double and decrease 40% perf
CPMD	3GB/s	128	P2p:0 Co:6	1.5KB/s/process	25% Double float	1.0	
Blast	1GB, 0.5GB/s	Scale well depend on workload	little	huge	integer	0.7	
Espresso	1.3GB/s	16	P2p:0 Co:15	0.5MB/s/process	64% Double float	0.5	
CHARMM	0.5GB, 0.6GB/s	64	P2p:1.1 Co:5.4	1.5KB/s/process	3% Double float	0.9	
DACAPO	0.5GB/s	16	P2p:0.2 Co:24			0.9	

Inspur T-Eye: Application Character analyzer

The speedometer for scientists:
Easy use、fast、simple、visible



Software development

Software optimization

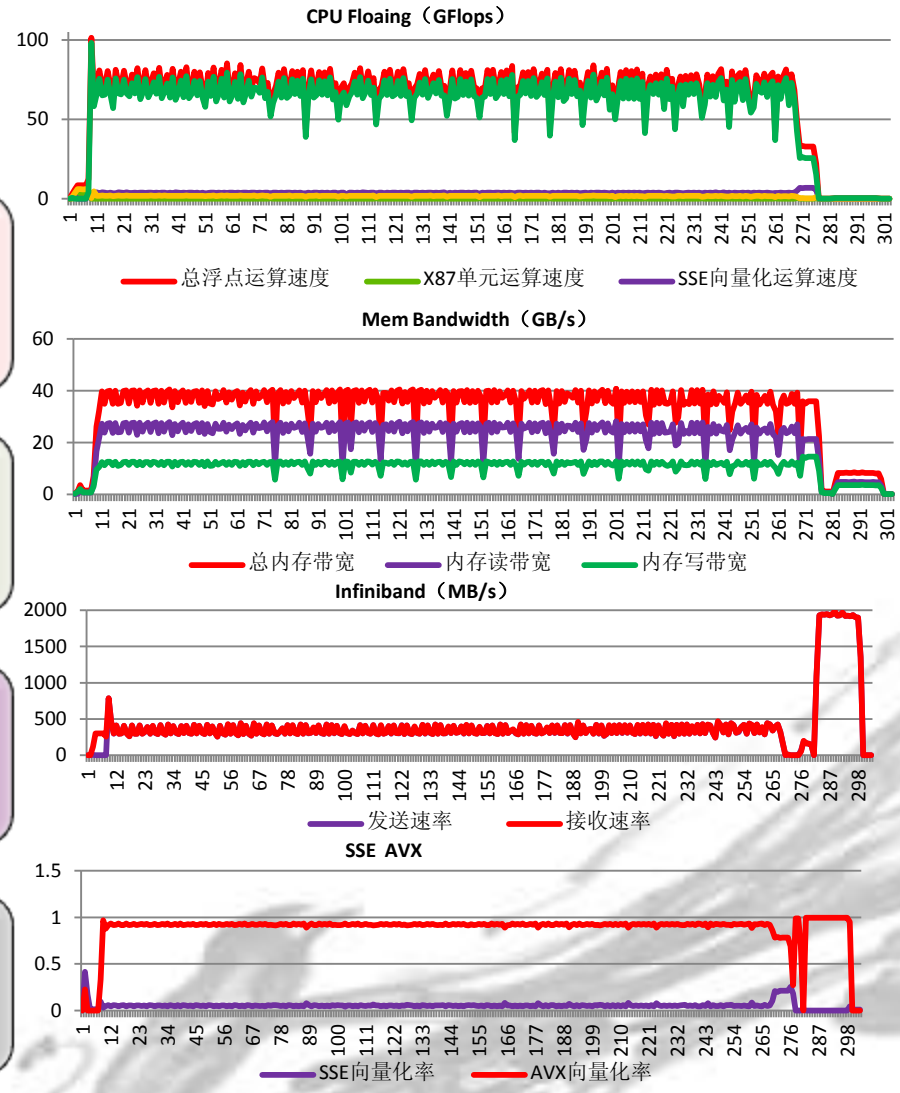
Performance evaluation

Cluster performance evaluation

Inspur T-Eye: Application Character analyzer

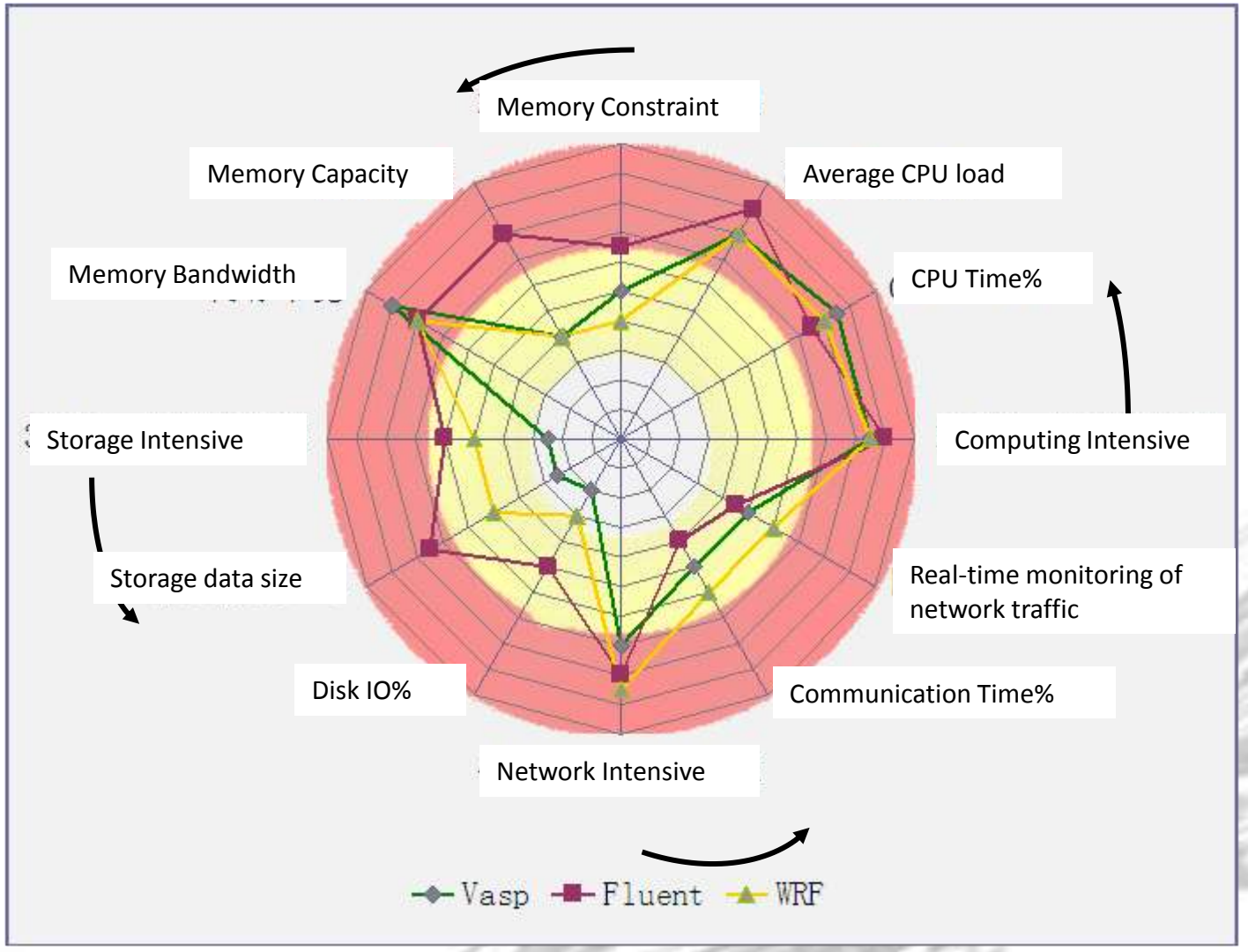
40+ micro arch & system indicators

CPU	• Usr%, sys%, idle%, iowait% • X87 GFLOPS, SP/DP SSE scalar/packed GFLOPS, SP/DP AVX scalar/packed GFLOPS • SP/DP SSE VEC, SP/DP AVX VEC • CPI
Mem	• used, cached, buffered • Mem Bandwidth
Interconnect	• Gigabit, Infiniband • TCP/IP, UDP, RDMA, IPoIB • GE Rec/Send IB Rec/Send • Packet Numer
Filesystem	•Local: Read\Write, Data block size •NFS: NFS Client Read、Write

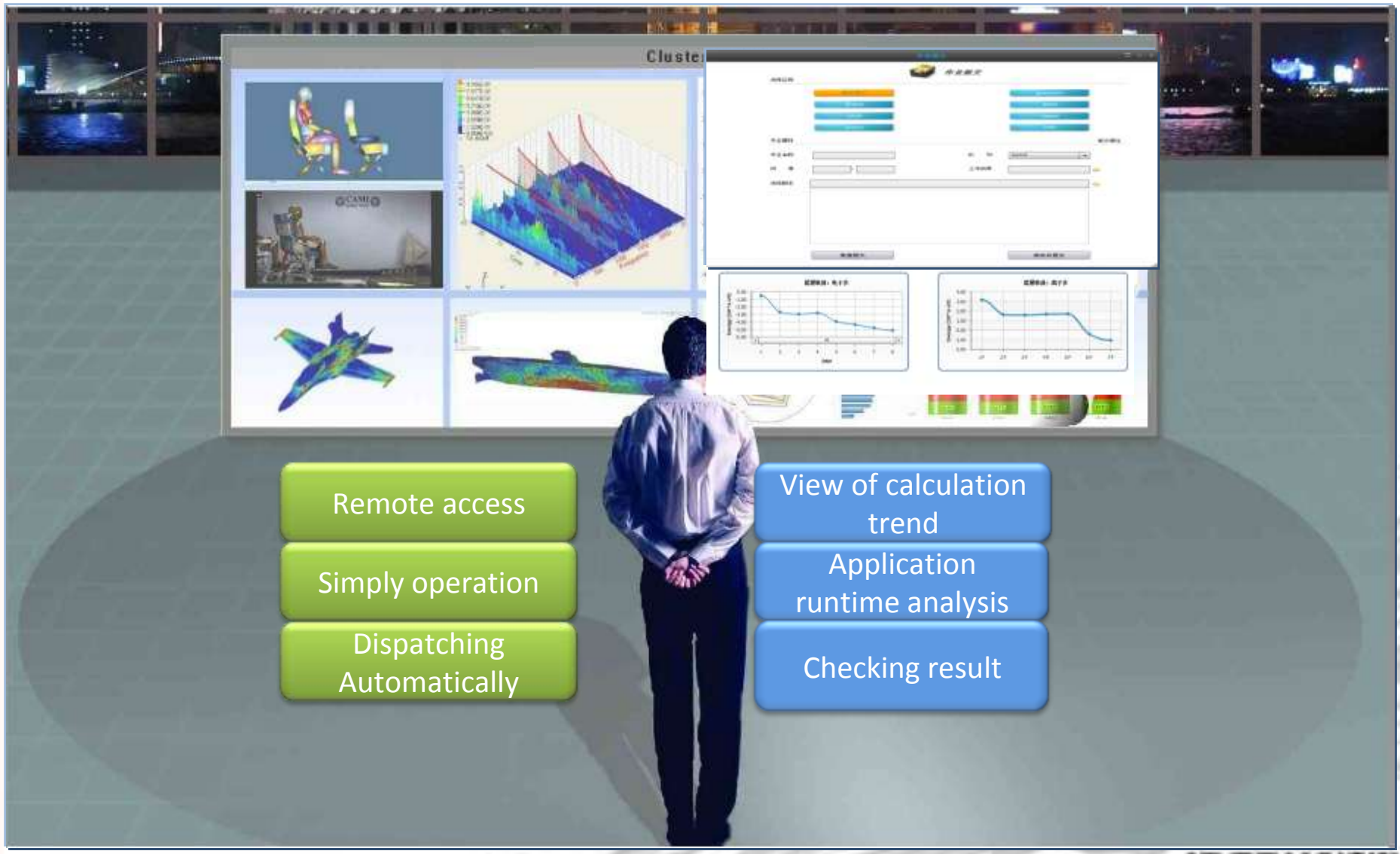


HPC Applications Radar chart

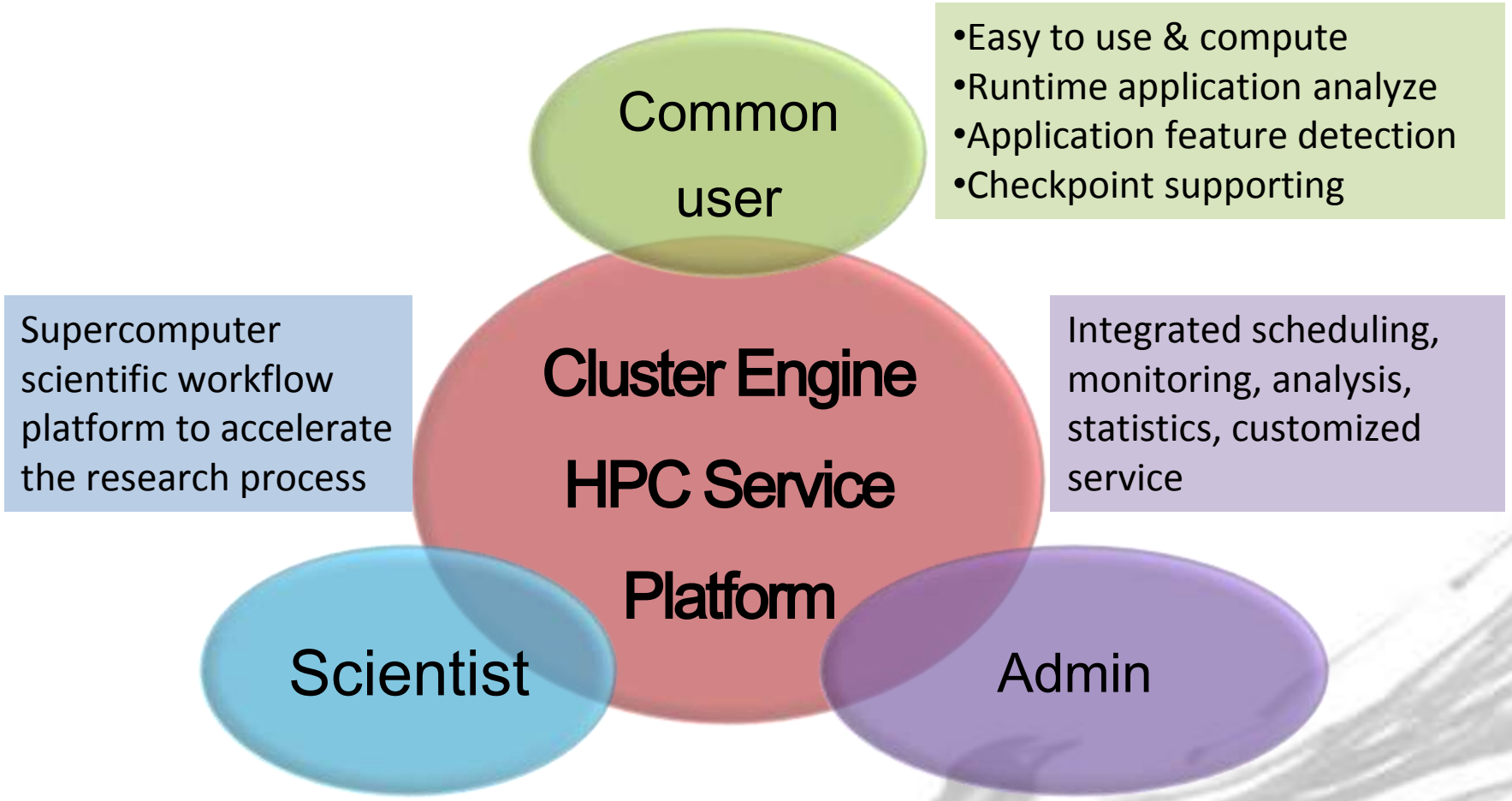
Analysis chart of industry application in life science, computing chemistry, CAE, numerical meteorology



Cluster Engine – HPC service platform

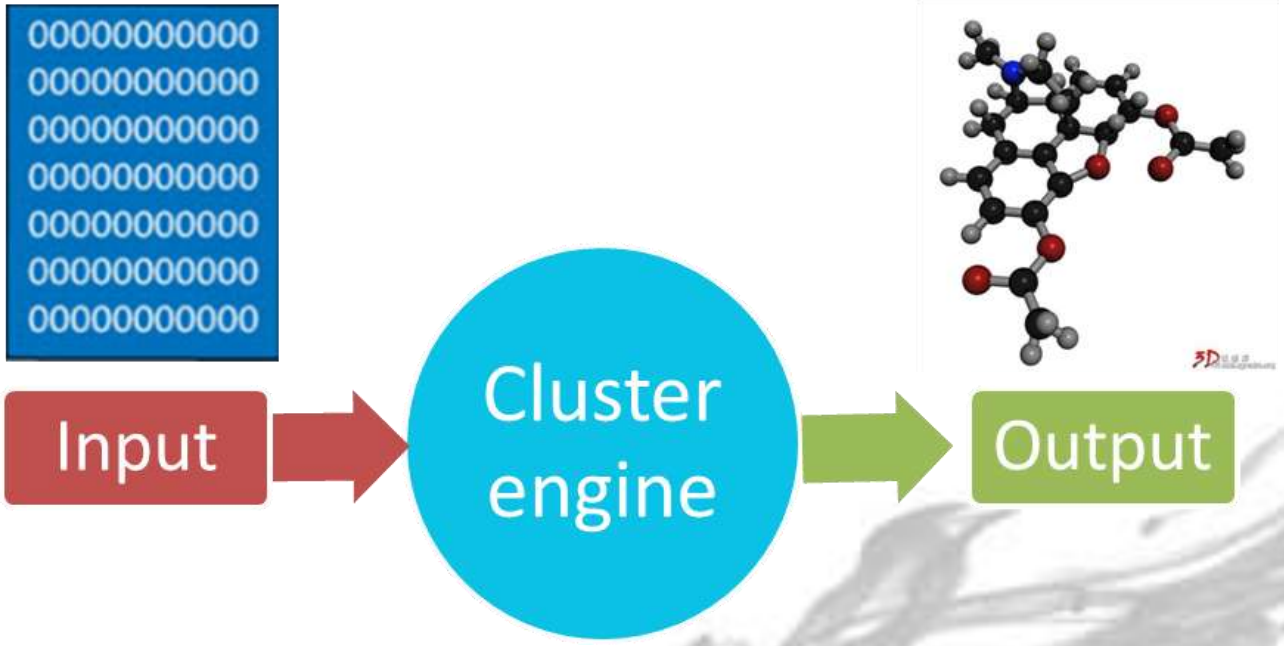


Cluster Engine – HPC service platform



Scientists' requirements for HPC services

- Scientists：
  1. Special users
  2. Not HPC professional users， not familiar with HPC work procedure
- Workflow ：



Cluster Engine service: HPC workflow

Cluster User | Network | Cluster Engine

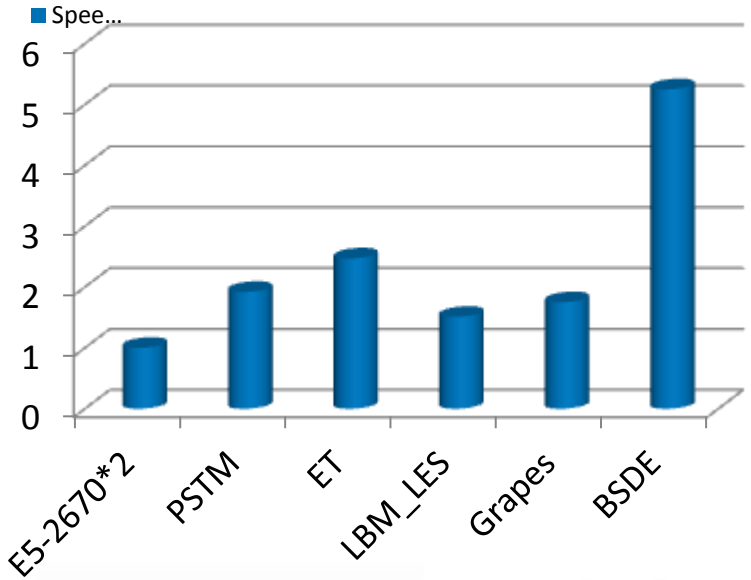
Cluster Engine

1. Fluent 12core
2. ATOM 12core
3. VASP 24core

vasp 6 | vasp 6 | vasp 6 | vasp 6 | c06

Example generated → Job submission → Queue & schedule → Job operation → Job completed

Heterogeneous Application Development



MIC = (1.5x～5.25x) 16 cores CPU

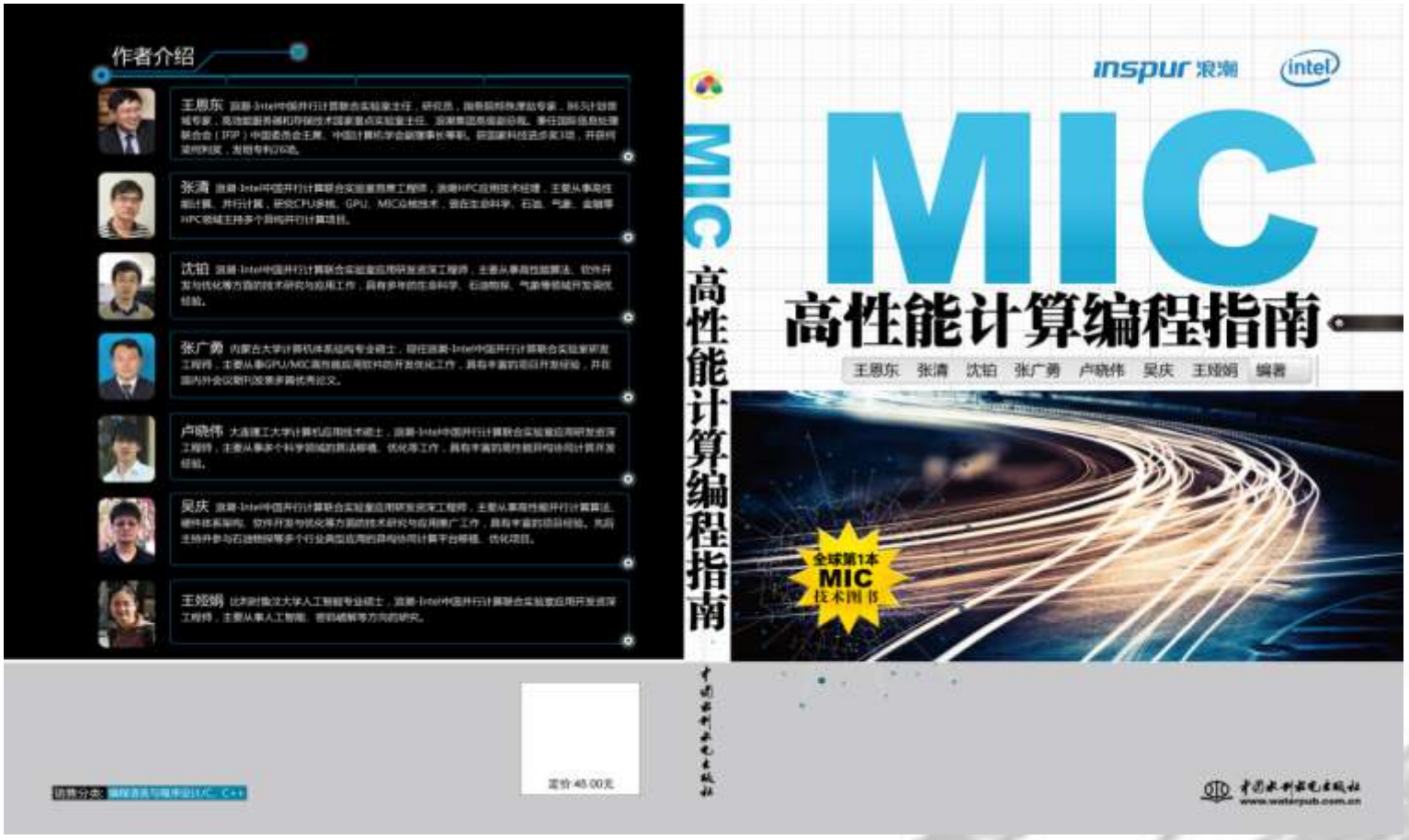
Intel-Inspur Parallel Computing Joint-Lab

- Face to Exascale computing
- CPU multi-core computing research
- MIC many-core computing research

Nvidia-Inspur Cloud Supercomputing Center

- GPU supercomputing application
- Scientific Computing application
- Big Data application
- Machine Learning application

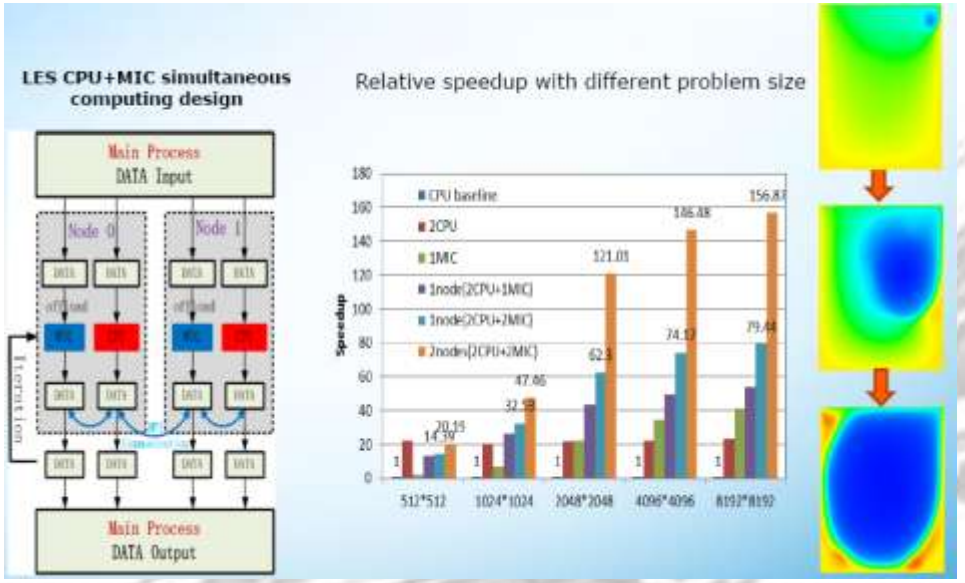
1st MIC Programming Book in Dec. 2012



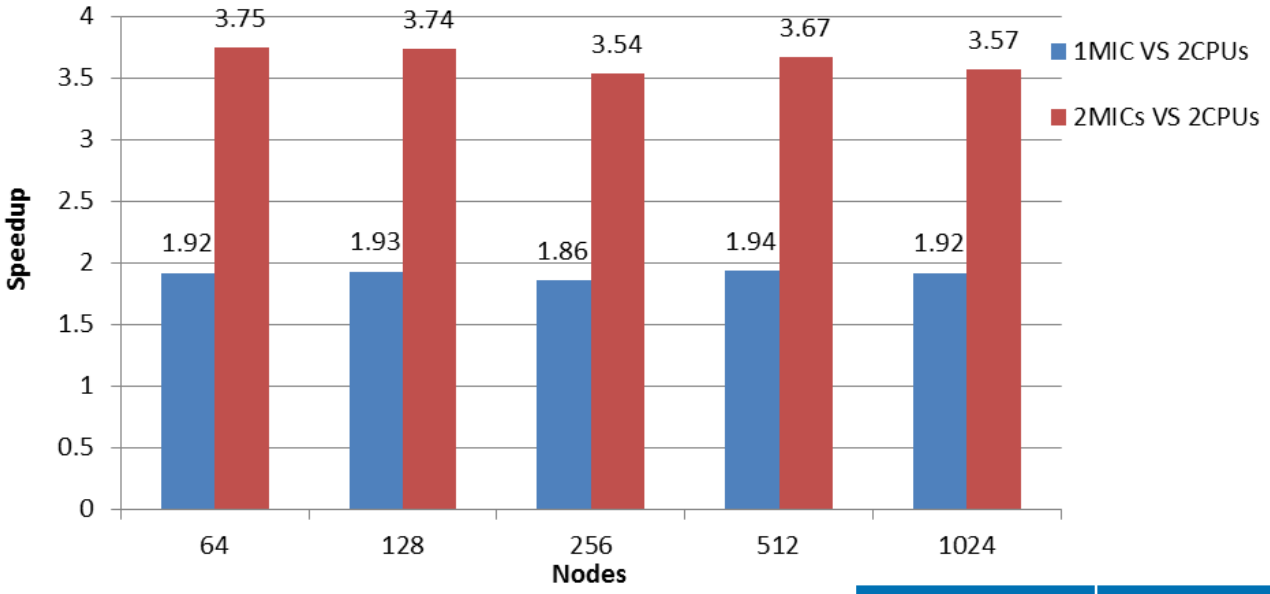
English version will be published by Springer in 2014

Tianhe-2 application(1)： LBM_LES

- LBM_LES background:
  - Lattice Boltzmann Method can simulate Large Eddy Simulation， this method is the key algorithm of LES
  - Application case：Inspur collaborated and developed LES（ Large Eddy Simulation ）algorithm with NPC on MIC platform.
  - The only MIC demo in IDF12
  - MIC cluster demo of CFD application in SC12
  - Accomplished test on Tianhe-2 in this year



LBM_LES on Tianhe-2

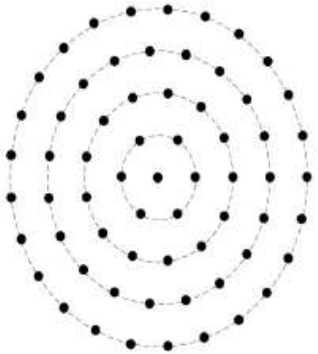


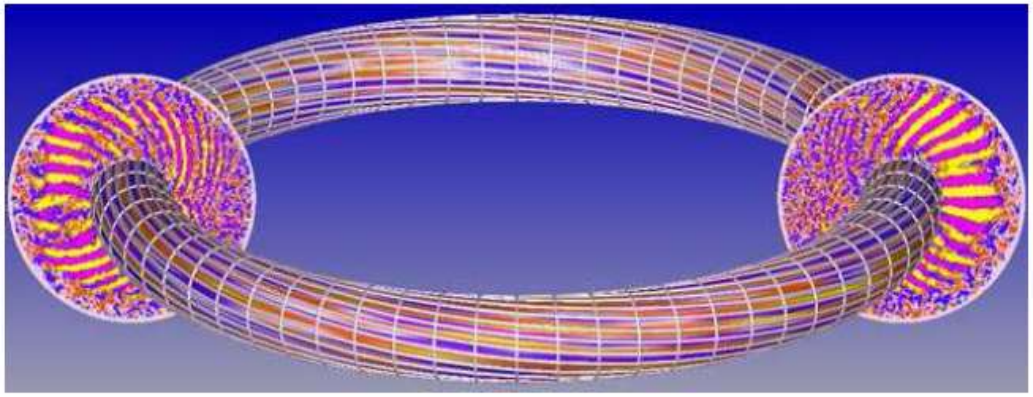
- Grid size dealt with reached Billion-grade
- Performance of 2MIC VS 2CPU:3.6 times

nodes	Grid size		
	2CPU	1MIC	2MICs
64	4.29E+09	4.29E+09	8.59E+09
128	8.59E+09	8.59E+09	1.72E+10
256	1.72E+10	1.72E+10	3.44E+10
512	3.44E+10	3.44E+10	6.87E+10
1024	6.87E+10	6.87E+10	1.37E+11

TianHe-2 Application(2): GTC

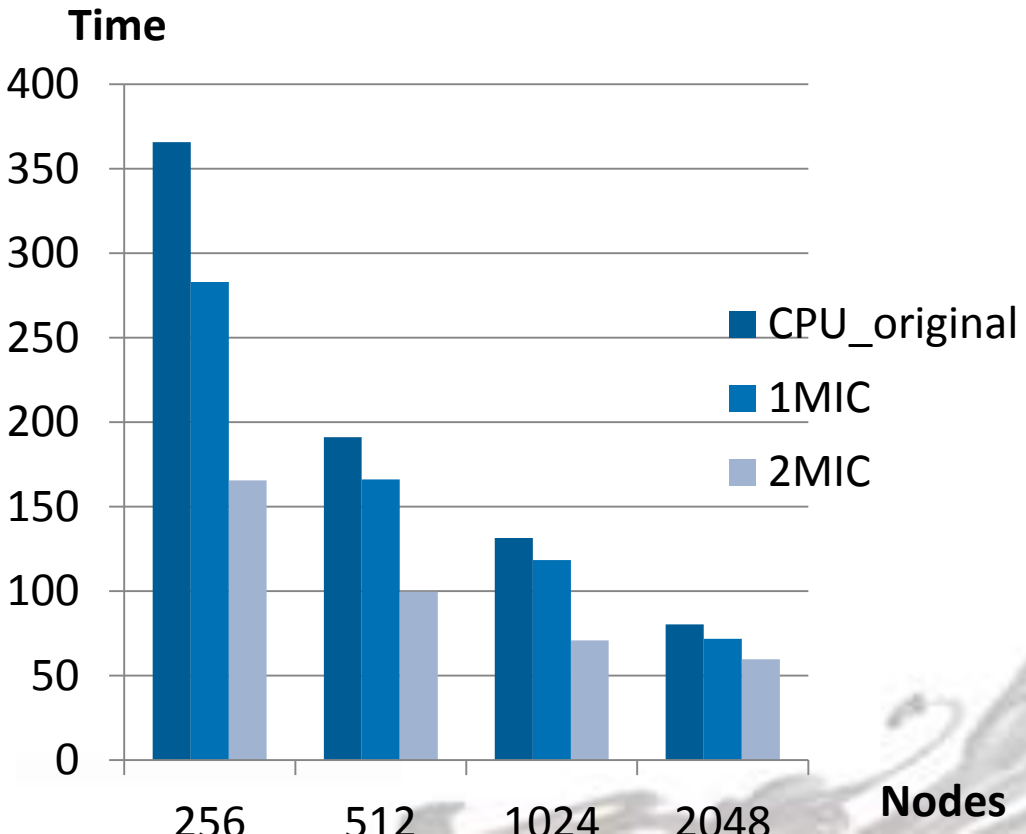
- GTC background:
  - Gyrokinetic Toroidal Code
  - large-scale magnetic confined fusion numerical simulation software，Cyclotron toroidal plasma code
  - Simulation of GTC is Magnetic confinement fusion problems.
  - Inspur collaborated and developed GTC algorithm which is as one of 100p applications with NUDT, National Supercomputing Center in Tianjin and Peking University on MIC platform. It is the first MIC version





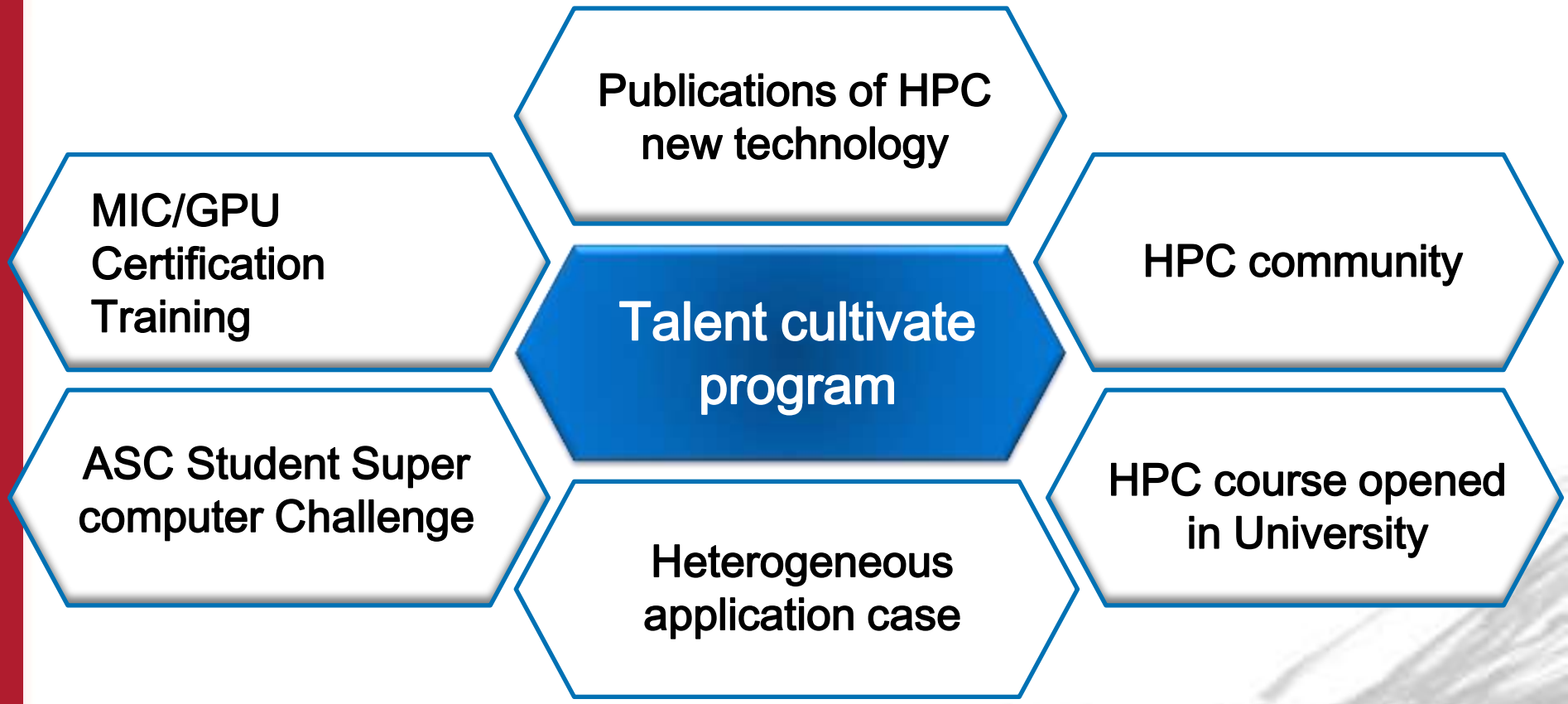
GTC Phase I Test on TianHe-2

- Scalability
  - Performance of 1MIC = 27-31 CPU cores
  - Performance of 2MIC VS 2CPU: 2.2X
  - 200K cores parallelism



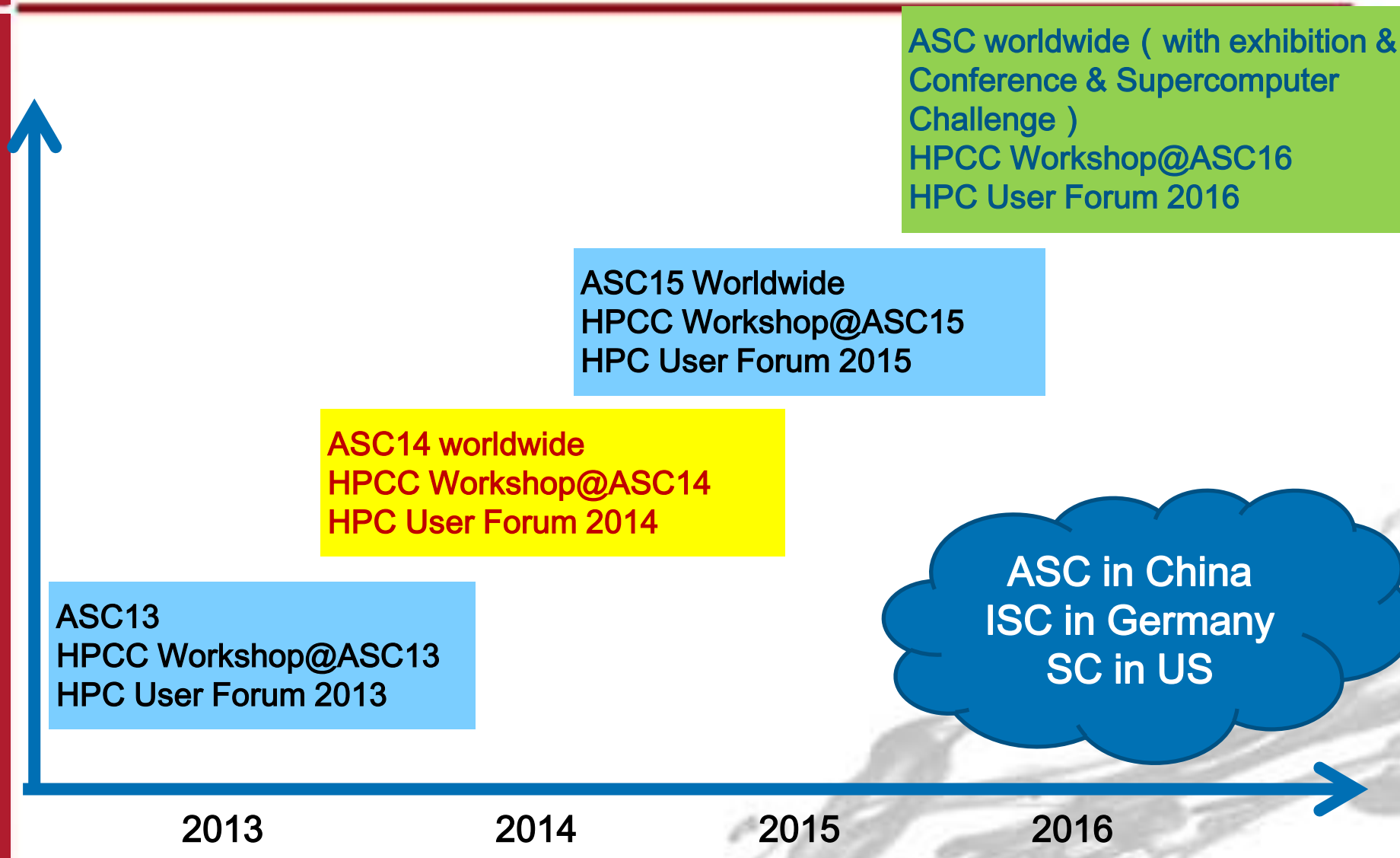
Next phase:
1. Whole system test on Tianhe-2
2.Scalability test with large cases
3.Three MIC cards test

HPC talent cultivate program



High level talent cultivation

ASC Roadmap


ASC worldwide (with exhibition & Conference & Supercomputer Challenge)
HPCC Workshop@ASC16
HPC User Forum 2016
ASC15 Worldwide
HPCC Workshop@ASC15
HPC User Forum 2015
ASC14 worldwide
HPCC Workshop@ASC14
HPC User Forum 2014
ASC13
HPCC Workshop@ASC13
HPC User Forum 2013
ASC in China
ISC in Germany
SC in US
2013
2014
2015
2016

2013 ASC Student Supercomputer Challenge

32 China Universities + 11 Worldwide Universities

國立清華大學
National Tsing Hua University

清華大學
Tsinghua University

國防科學技術大學

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

University of Mumbai

UNIST

同濟大學
TONGJI UNIVERSITY

南開大學

上海理工大學

中山大學
SUN YAT-SEN UNIVERSITY

西北工業大學

南京大學
NANJING UNIVERSITY

Previous Champions

2014 ASC Student Supercomputer Challenge

82 Universities from 5 continents

ASC14: The talents' amazing potential

3D Elastic Wave Equation Optimization on CPU+MIC

University	Original Runtime(Serial)	Optimized Runtime (4 Nodes)
Taiyuan University of Technology	9, 399s	21. 70s
Huazhong University of Science and Technology	9, 399s	39. 84s
Nanyang Technological University	9, 399s	49. 89s
Beihang University	9, 399s	62. 37s
Ural Federal University	9, 399s	44. 37s
ZheJiang university	9, 399s	72. 87s
Shanghai Jiao Tong University	9, 399s	83. 17s

How would the youth perform on Tianhe-2?

inspur 浪潮

谢谢！