Introduction of PRAGMA routine-basis experiments

Yoshio Tanaka (yoshio.tanaka@aist.go.jp)
PRAGMA
Grid Technology Research Center, AIST, Japan

Grid Technology Research Center AIST

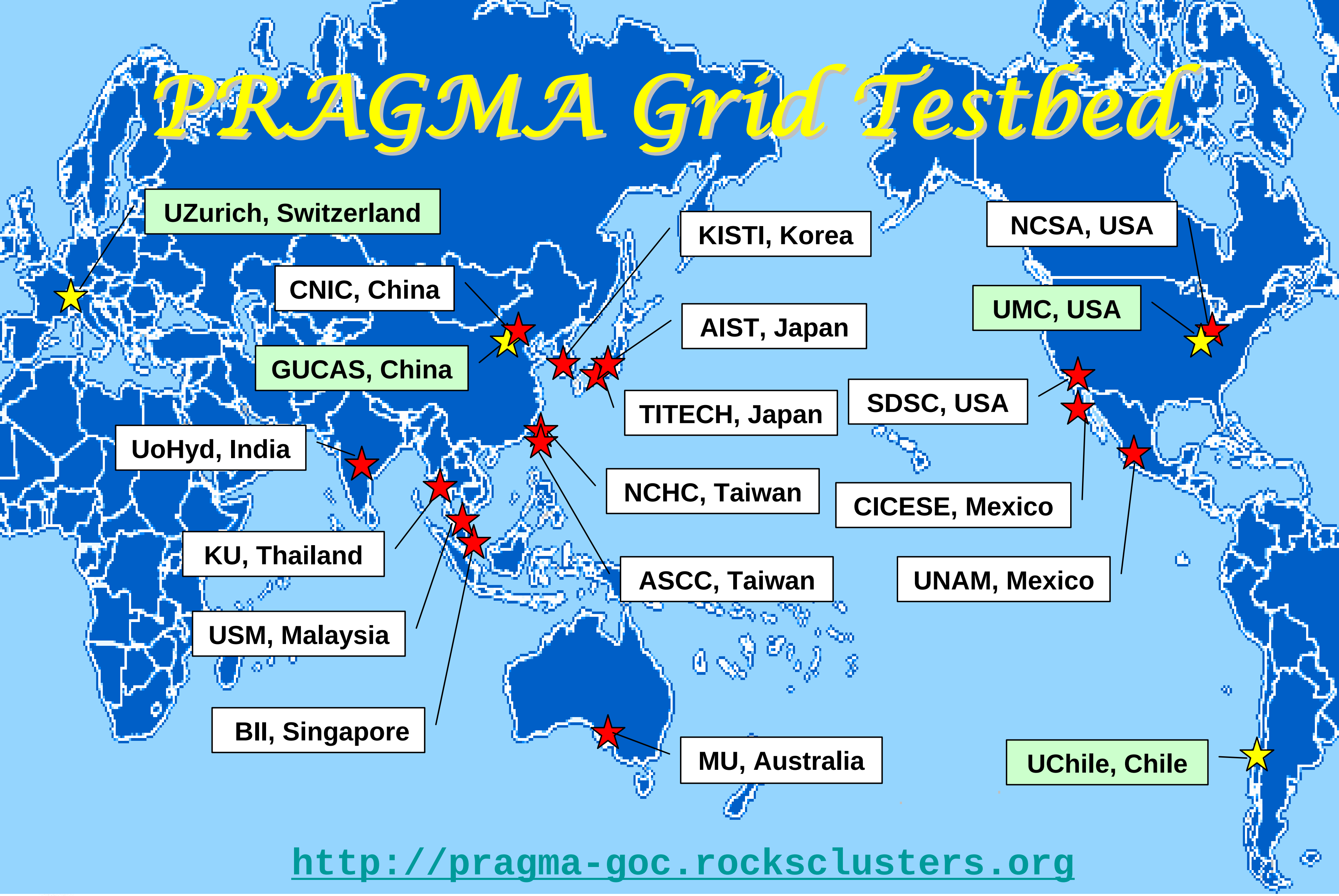
Grid Communities in Asia Pacific – at a glance –

- **ApGrid: Asia Pacific Partnership for Grid Computing**
  - Open Community as a focal point
    - more than 40 member institutions from 15 economics
  - Kick-off meeting: July 2000, 1st workshop: Sep. 2001
- **PRAGMA: Pacific Rim Applications and Grid Middleware Assembly**
  - NSF funded project led by UCSD/SDSC
    - 19 member institutions
  - Establish sustained collaborations and advance the use of the grid technologies
  - 1st workshop: Mar. 2002, 10th workshop: next month!
- **APAN (Asia Pacific Advanced Network) Grid Committee**
  - Bridging APAN application communities and Grid communities outside of APAN
  - Grid WG was launched in 2002, re-organized as a committee in 2005
- **APGrid PMA: Asia Pacific Grid Policy Management Authority**
  - General Policy Management Authority in the Asia Pacific Region
    - 16 member CAs
  - A founding member of the IGTF (International Grid Trust Federation)
  - Officially started in June 2004
- **APEC/TEL APGrid**
  - Building social framework
  - Semi-annual workshops
- **APAN (Asia Pacific Advanced Network) Middleware WG**
  - Share experiences on middleware.
  - Recent topics include ID management and National Middleware Efforts.
  - Approved in January 2006.

PRAGMA routine-basis experiments

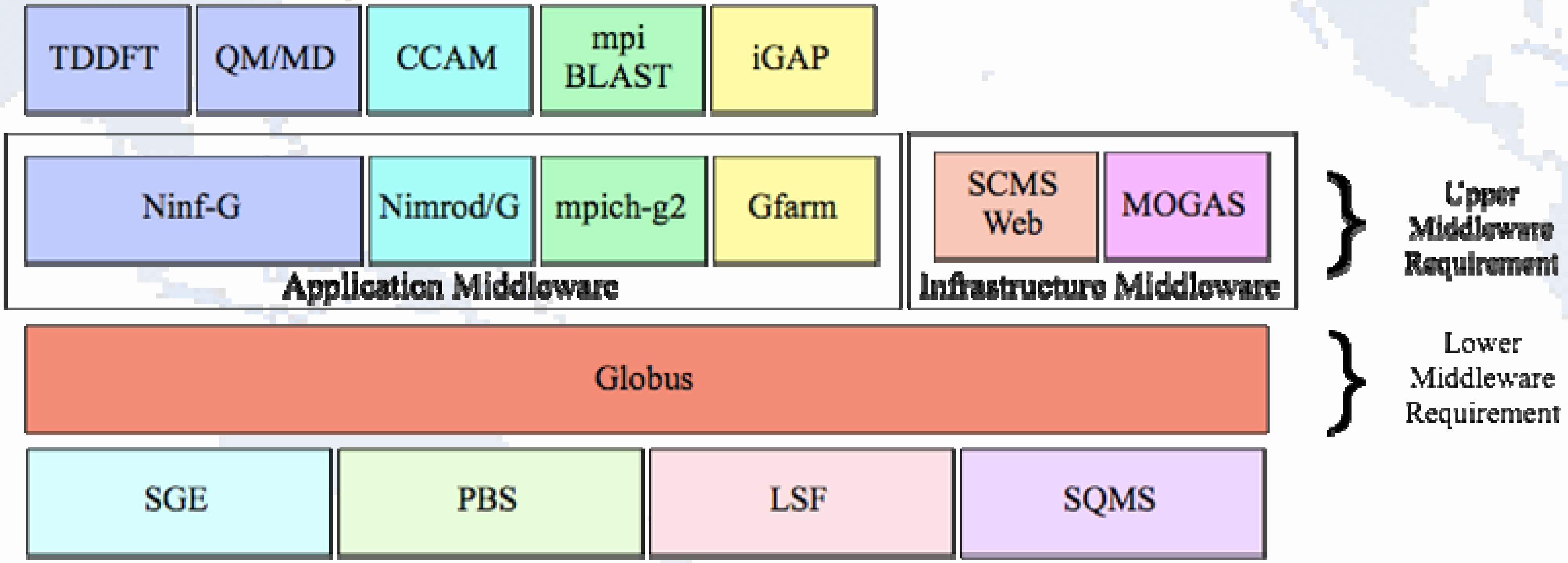
Most slides are by courtesy of Mason Katz and Cindy Zheng (SDSC/PRAGMA)

PRAGMA Grid Testbed



http://pragma-goc.rocksclusters.org

Application vs. Infrastructure Middleware



PRAGMA Grid resources

http://pragma-goc.rocksclusters.org/pragma-doc/resources.html

	Institution Name	Country /Region	Host Name	# nodes	# CPUs	Memory (GB)	User Disk Space (GB)	Job manager	CPU Model	CPU Speed (MHz)	System Type	OS Type	OS Release
	AIST	Japan	ume	33	66	34.460	209	SGE	i686	1396	Cluster	Linux	2.4.20-30.8.legacysmp
	ASCC	Taiwan	pragma001	4	4	2.936	196	PBS	i686	2390	Cluster	Linux	2.4.21-15.0.4.Elsmp
	BII	Singapore	marlin	5	5	2.505	27	SGE	i686	1836	Cluster	Linux	2.4.21-15.EL
	CICESE	Mexico	solaris	8	8	2.048	11	SGE	Sparc	500	Cluster	Solaris	5.9
	CNIC	China	pragma	8	32	64.000	48	LJRS	ia64	1300	Cluster	Linux	2.4.21-3.5qsnet
	GUCAS	China	igeon	16	32	36.000	1000	SGE	x86_64	3000	Cluster	Linux	
	KISTI	Korea	jupiter	17	17	15.103	481	PBS	i686	1694	Cluster	Linux	2.4.20-28.7
	KU	Thailand	amatal	15	15	8.020	138	SQMS	i686	1009	Cluster	Linux	2.4.20-31.9
	MU	Australia	mahar	50	50	50.828	57	PBS	i686	2993	Cluster	Linux	2.4.22-xfs
	NCHC	Taiwan	ase	9	18	9.288	154	PBS	i686	1666	Cluster	Linux	2.4.26-686-smp
	NCSA	USA	tgc	13	52	26.169	1685	PBS	i686	2400	Cluster	Linux	2.4.21-15.ELsmp
	SDSC	USA	rocks-52	16	62	32.493	275	SGE	i686	2388	Cluster	Linux	2.4.21-20.Elsmp
	SDSC	USA	rocks-47	4	4	8.140	24	SGE	ia64	900	Cluster	Linux	2.4.21-20.EL
	Titech	Japan	gsic-presto	9	9	7.566	118	PBS	i686	1195	Cluster	Linux	2.4.26
	UChile	Chile	syntagma	16	32	34.000	2000	SGE	ia64	1600	Cluster	Linux	2.6.9-11smp
	UMC	USA	obsidian	16	32	32.000	500	SGE	i386	2800	Cluster	Linux	2.4.20-8smp
	UNAM	Mexico	malicia	6	6	3.510	78	PBS	i686	1894	Cluster	Linux	2.4.19-ipvs
	UoHyd	India	amber	9	9	4.287	110	PBS	i686	2394	Cluster	Linux	2.4.18-14
	USM	Malaysia	aurora	17	34	17.700	70	SGE	i686	1396	Cluster	Linux	2.4.28-husseinl
	USM	Malaysia	hawk	17	17	8.700	70	SGE	i686	2800	Cluster	Linux	2.4.20-8smp
	UZurich	Switzerland	ocikbsbg	8	16	16.840	28	SGE	i686	2800	Cluster	Linux	2.4.21-20.ELsmp
	UZurich	Switzerland	ocikbsbg	8	16	16.840	28	SGE	i686	2800	Cluster	Linux	2.4.21-20.ELsmp
Total	19	13	22	367	662	981.080	7307						

Features of PRGMA Grid

- Grass-root approach
  - No single source of funding for testbed development
  - Each site contributes its resources (computers, networks, human resources, etc.)
- Operated/maintained by administrators of each site.
  - Most site admins are not dedicated for the operation.
- Small-scale clusters (several 10s CPUs) are geographically distributed in the Asia Pacific Region.
- Networking is there (APAN/TransPAC), but performance (throughput and latency) is not enough.
- Aggregated #cpus is more than 600 and still increasing.
- Really an international Grid across national boundary.
- Give middleware developers, application developers and users many valuable insights through experiments on this real Grid infrastructure.

Why Routine-basis Experiments?

- Resources group Missions and goals
  - Improve interoperability of Grid middleware
  - Improve usability and productivity of global grid
- PRAGMA from March, 2002 to May, 2004
  - Computation resources

    10 countries/regions, 26 institutions, 27 clusters, 889 CPUs
  - Technologies (Ninf-G, Nimrod, SCE, Gfarm, etc.)
  - Collaboration projects (Gamess, EOL, etc.)
  - Grid is still hard to use, especially global grid
- How to make a global grid easy to use?
  - More organized testbed operation
  - Full-scale and integrated testing/research
  - Long daily application runs
  - Find problems, develop/research/test solutions

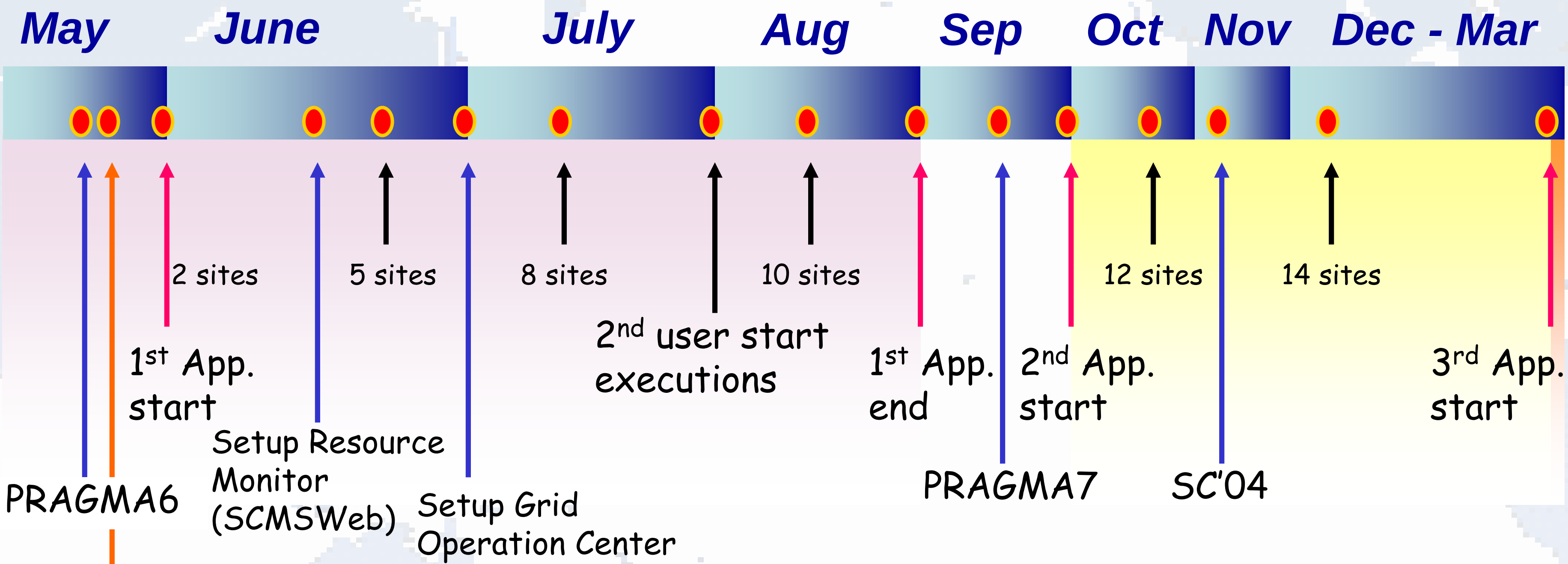
Routine-basis Experiments

- Initiated in May 2004 PRAGMA6 workshop
- Testbed
  - Voluntary contribution (8 -> 17)
  - Computational resources first
  - Production grid is the goal
- Applications
  - TDDFT, mpiBlast-g2, Savannah, QM/MD
  - iGAP over Gfarm
  - Ocean science, Geoscience (proposed)
- Learn requirements/issues
- Research/implement solutions
- Improve application/middleware/infrastructure integrations
- Collaboration, coordination, consensus

Rough steps of the experiment

1. Players:
   - A conductor
   - An application driver
   - Site administrators
2. Select an application and an application driver
3. The application driver prepares a web page that describes software requirements (prerequisite software, architecture, public/private IP addresses, disk usage, etc.) of the application. Then, the application driver informs the conductor that the web page is ready.
4. The conductor ask site administrators to (1) create an account for the driver (including adding an entry to grid-mapfile, and CA certificate/policy file), and (2) install necessary software listed on the web site.
5. Each site admin let the conductor and the application driver know when she/he has done account creation and software installation.
6. The application driver login and test the new site. If she/he finds any problems, she/he will directly contact to the site admin.
7. The application driver will start main (long-run) experiment when she/he decides the environment has been ready.

Progress at a Glance



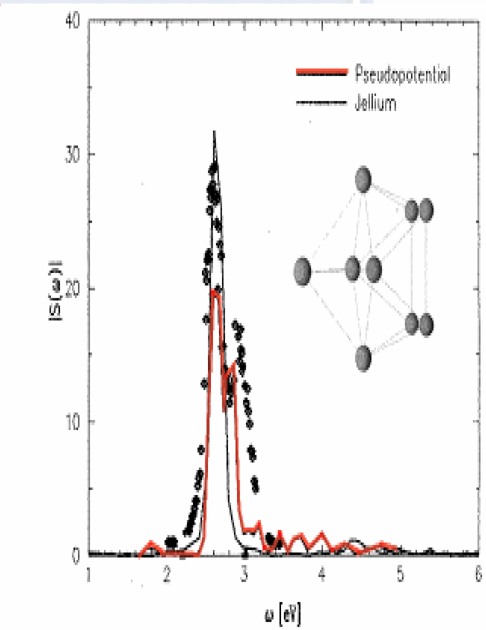
On-going works

1. Site admins install required software
2. Site admins create users accounts (CA, DN, SSH, firewall)
3. Users test access
4. Users deploy application codes
5. Users perform simple tests at local sites
6. Users perform simple tests between 2 sites

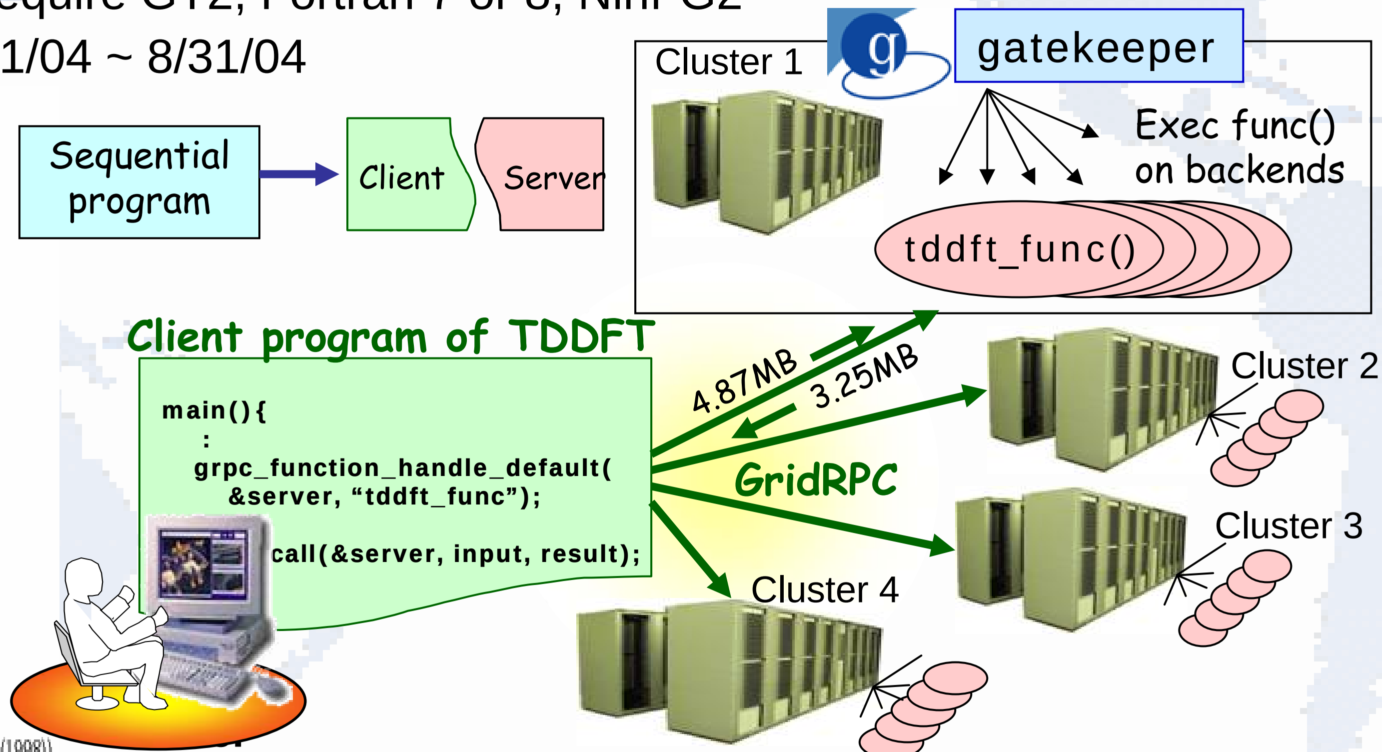
Join in the main executions (long runs) after all's done

1st application
Time-Dependent Density Functional Theory (TDDFT)

- Computational quantum chemistry application
- Driver: Yusuke Tanimura (AIST, Japan)
- Require GT2, Fortran 7 or 8, Ninf-G2
- 6/1/04 ~ 8/31/04



(F. Calvayrac, P.-G. Reinhard, and E. Suraud, J. Phys. B 31, 1367 (1998))



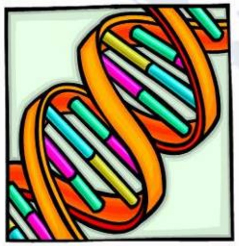
http://pragma-goc.rocksclusters.org/tddft/default.html

Routine Use Applications

2nd Application - mpiBLAST

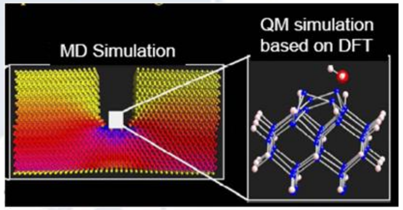
A DNA and Protein sequence/database alignment tool

- Drivers (ASCC, Taiwan)
  - Hurng-Chun Lee
  - Chi-Wei Wong
- Started 9/20/04
- SC04 demo
- New requirements
- New mpiBLAST-G2
- Automate installation/setup/testing



http://pragma-goc.rocksclusters.org/biogrid/default.html

3rd Application – QM/MD



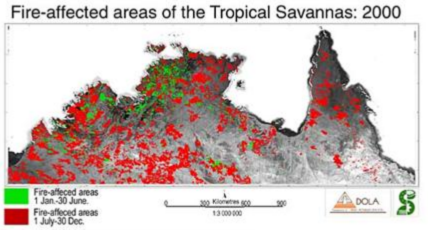
- Quantum mechanical calculation within a molecular dynamics simulation
- Grid-enabled by Hiroshi Takemeya, AIST, Japan

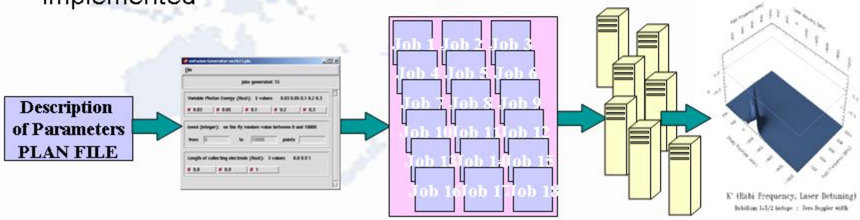
- Ninf-G based
- Driver:
  Hiroshi Takemeya
- Started 4/19/05
- MPI
- Multiple applications
- Fault tolerance
  Auto migration

4th Application – Savannah Case Study

Study of Savannah fire impact on northern Australian climate

- Climate simulation model
- Based on Nimrod-G
- 1.5 month CPU * 90 experiments
- Driver: Colin Enticott (Monash University, Australia)
- Requirements have been implemented

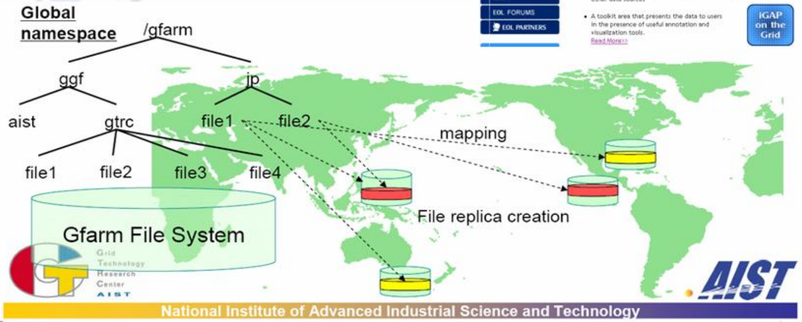




http://pragma-goc.rocksclusters.org/savannah/default.html

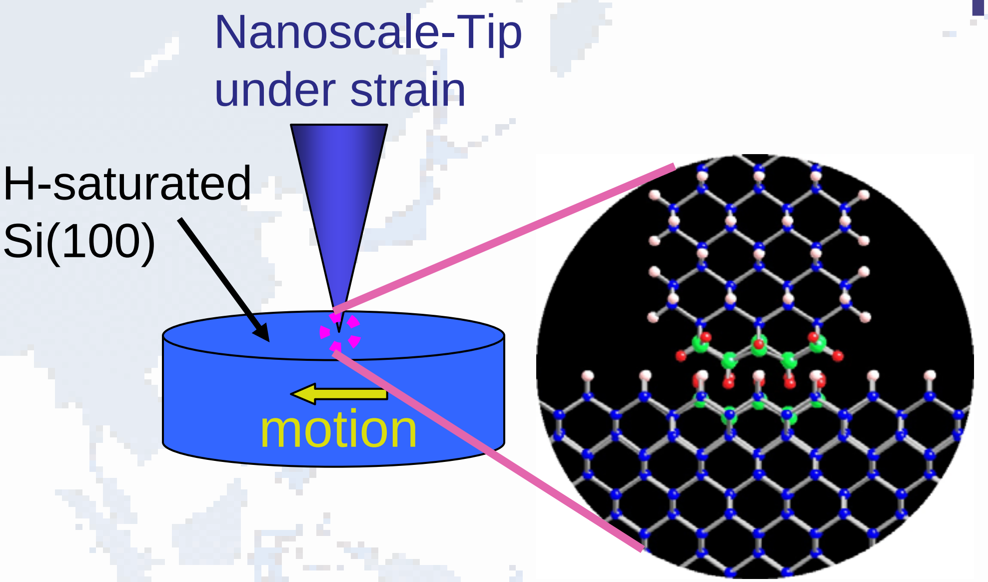
5th Application – iGAP/Gfarm

- iGAP and EOL (SDSC, USA)
  - Genome annotation pipeline
- Gfarm – Grid file system (AIST, Japan)
- Demo in SC04 (SDSC, AIST, BII)
- Preparation start in testbed March 2005
- Drivers: Wilfred Li and Osamu Tatebe

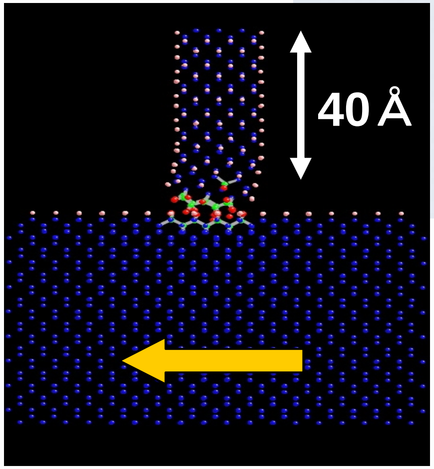


AIST
National Institute of Advanced Industrial Science and Technology

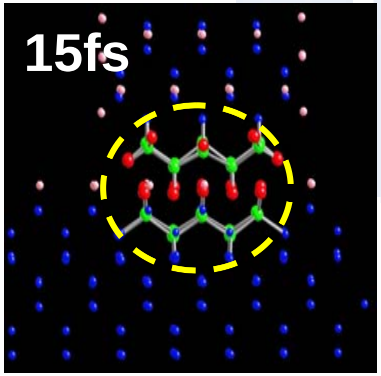
QM/MD simulation of Atomic-scale Stick-Slip Phenomenon



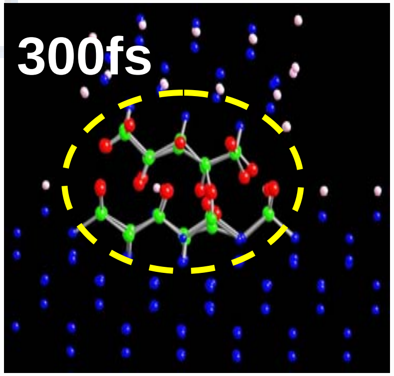
Initial Status



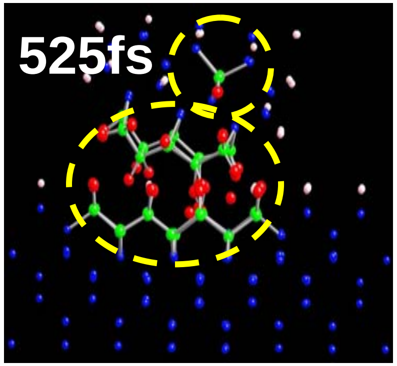
After 520 fs (image)



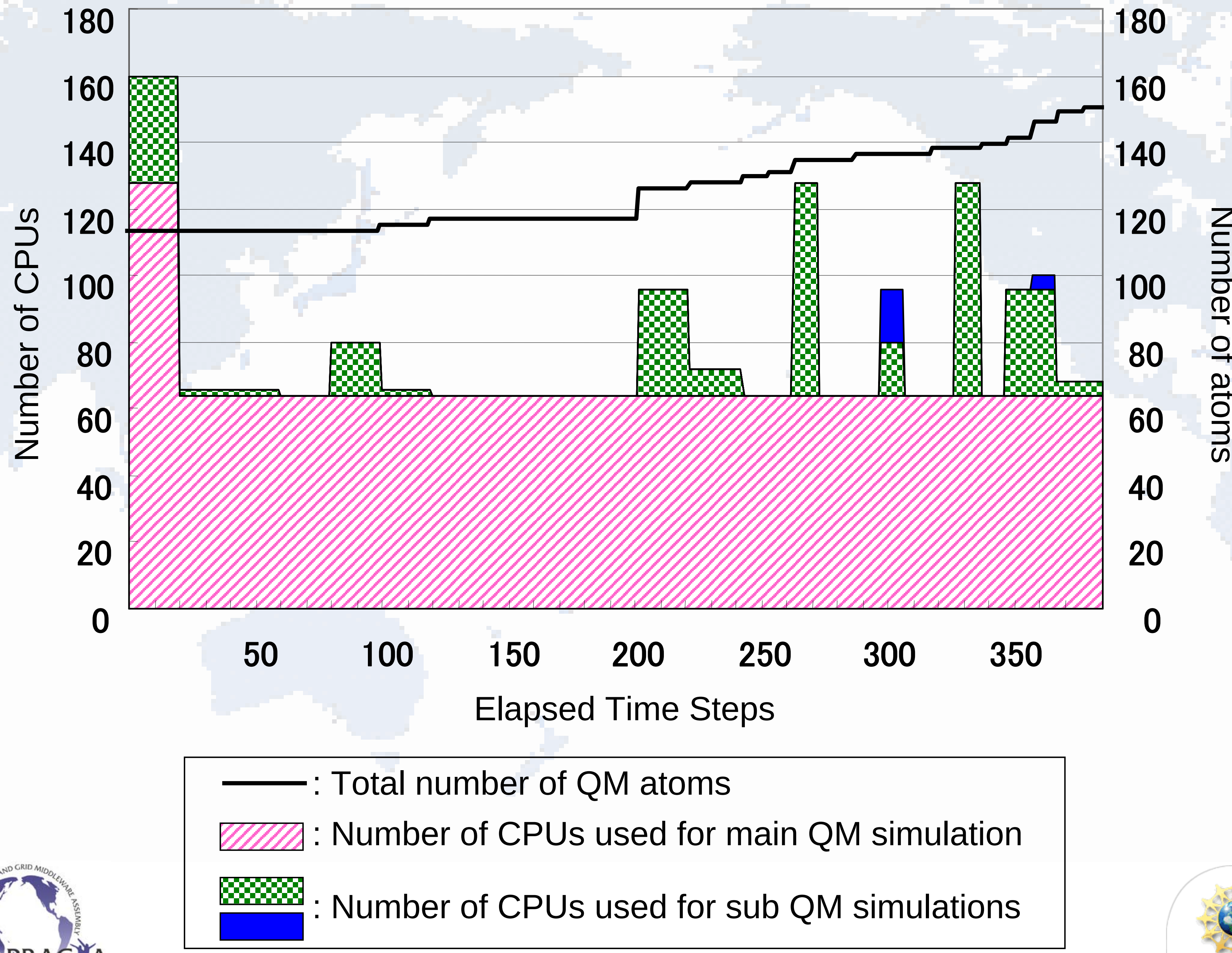
(1) Number of atoms in a QM region is small



(2) Number of atoms in a QM region has been increased



(3) One QM region has been splitted into two QM regions



180
160
140
120
100
80
60
40
20
0
Number of CPUs
Number of atoms
50
100
150
200
250
300
350
Elapsed Time Steps
: Total number of QM atoms
: Number of CPUs used for main QM simulation
: Number of CPUs used for sub QM simulations

Lessons Learned

http://pragma-goc.rocksclusters.org/

- Grid Operator's point of view
  - Preparing a web page is good to understand necessary operation.
    - But should be described in more detail as much as possible
  - Grid-enabled MPI is ill-suited to Grid
    - difficult to support co-allocation
    - private IP address nodes are not usable
    - (performance, fault tolerance, etc…)
- Middleware developer's point of view
  - Observed may kinds of faults (some of them were difficult to detect)
    - Improved capabilities for fault detection
      - e.g. heartbeat, timeout, etc.
- Application user (driver)'s point of view
  - Heterogeneity in various layers
    - Hardware/software configuration of clusters
      - Front node, compute nodes, compile nodes
      - public IP, private IP
      - File system
      - Configuration of batch system
      - …
    - Need to check the configuration when the driver accessed to the site for the first time
  - Not easy to trace jobs (check the status of jobs / queues)
  - Clusters were sometimes not clean (zombie processes were there).

Summary: Collaboration is the key

- Non-technical, most important
- Different funding sources
- How to get enough resources
- How to get people to act, together
  - how to motivate them to participate in
- Mutual interests, collective goals
- Cultivate collaborative spirit
- Key to PRAGMA's success

- Experiences on the routine-basis experiments helped experiments on multi-grid interoperation between PRAGMA and TeraGrid.
  - Details will be presented by Phil P. this afternoon ☺