Open Supercomputer

PACS·CS

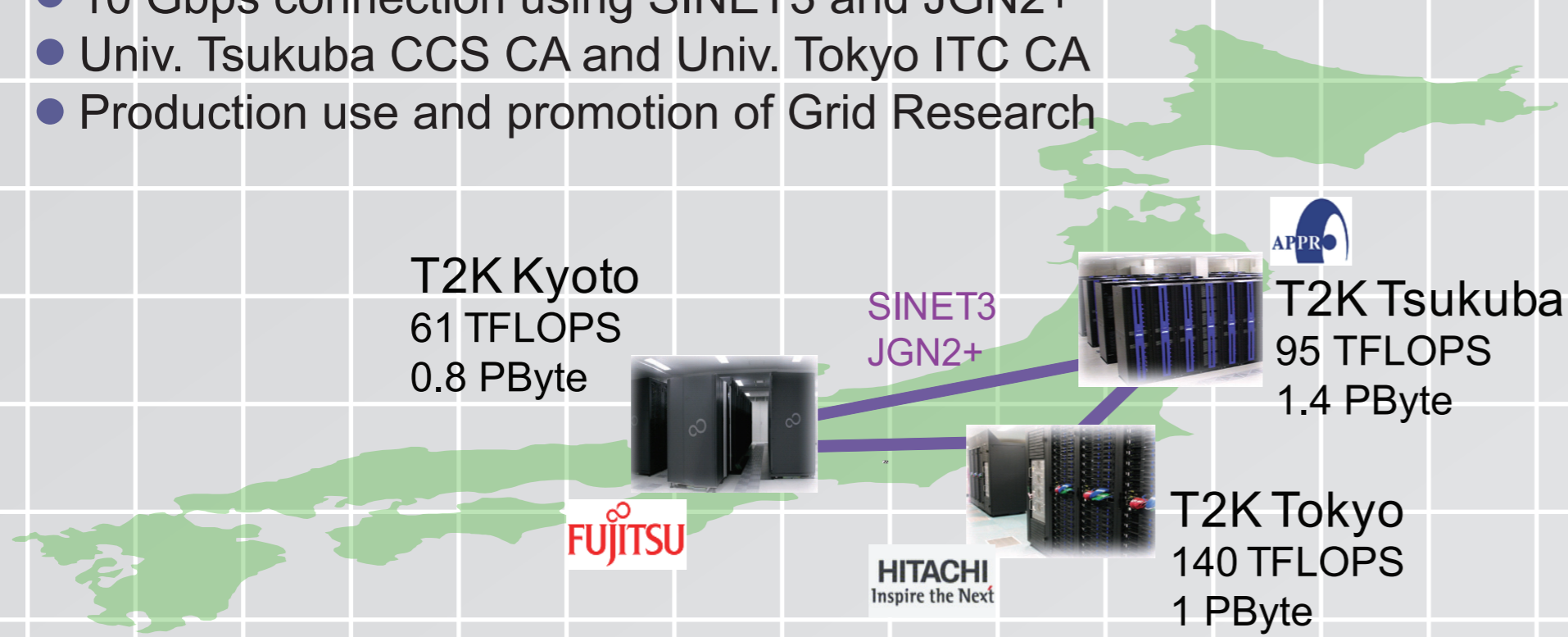# T2K Open Supercomputer & PACS-CS

T2K Open Supercomputer System: T2K Grid

Overview

The T2K Grid is a network and software infrastructure designed to facilitate collaboration between research and production towards seamless use of three T2K Open Supercomputers. Features include

- 10 Gbps connection using SINET3 and JGN2+
- Univ. Tsukuba CCS CA and Univ. Tokyo ITC CA
- Production use and promotion of Grid Research

T2K Kyoto
61 TFLOPS
0.8 PByte

SINET3
JGN2+

T2K Tsukuba
95 TFLOPS
1.4 PByte

FUJITSU

HITACHI
Inspire the Next

T2K Tokyo
140 TFLOPS
1 PByte

Grid Software

Standard Grid services are provided by the T2K Open Supercomputers.

- GSISSH
- GridFTP
- Virtual Organization Membership Service
- NAREGI-CA
- Gfarm Grid File System

the globus toolkit
www.globustoolkit.org

NAREGI

GRID DATA FARM

CS research is promoted for more advance Grid software including workflow tool, metascheduler and science portal.

Certificate Authority

The T2K Certificate Authority is operated by the Center for Computational Sciences, University of Tsukuba and the Information Technology Center, University of Tokyo.

- A user certificate can be issued in just a few clicks for every account

University of Tsukuba CCS CA
http://www.ccs.tsukuba.ac.jp/ca/

Gfarm Grid File System
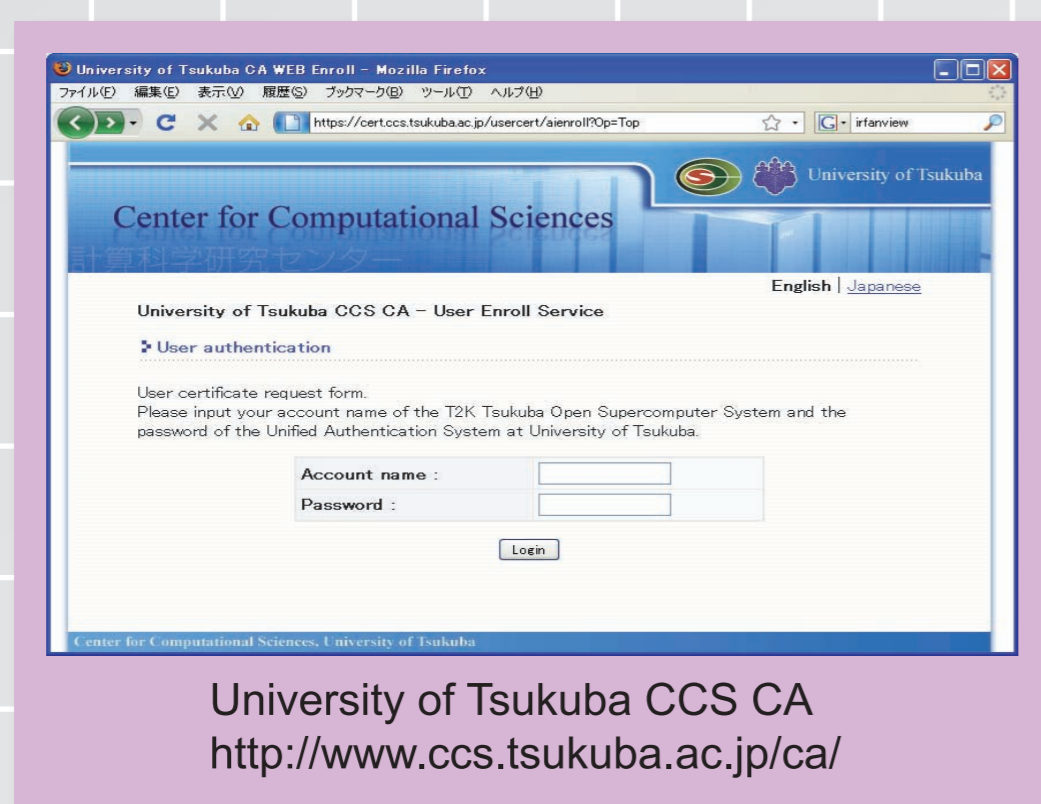
http://sf.net/projects/gfarm/

The Gfarm Grid File System is a global shared file system that can be mounted by T2K Open Supercomputers. Users can access the same shared file system at each site, and do not need to copy files and maintain consistency among them any more.

- Current capacity is 1.1 PByte

T2K Tsukuba
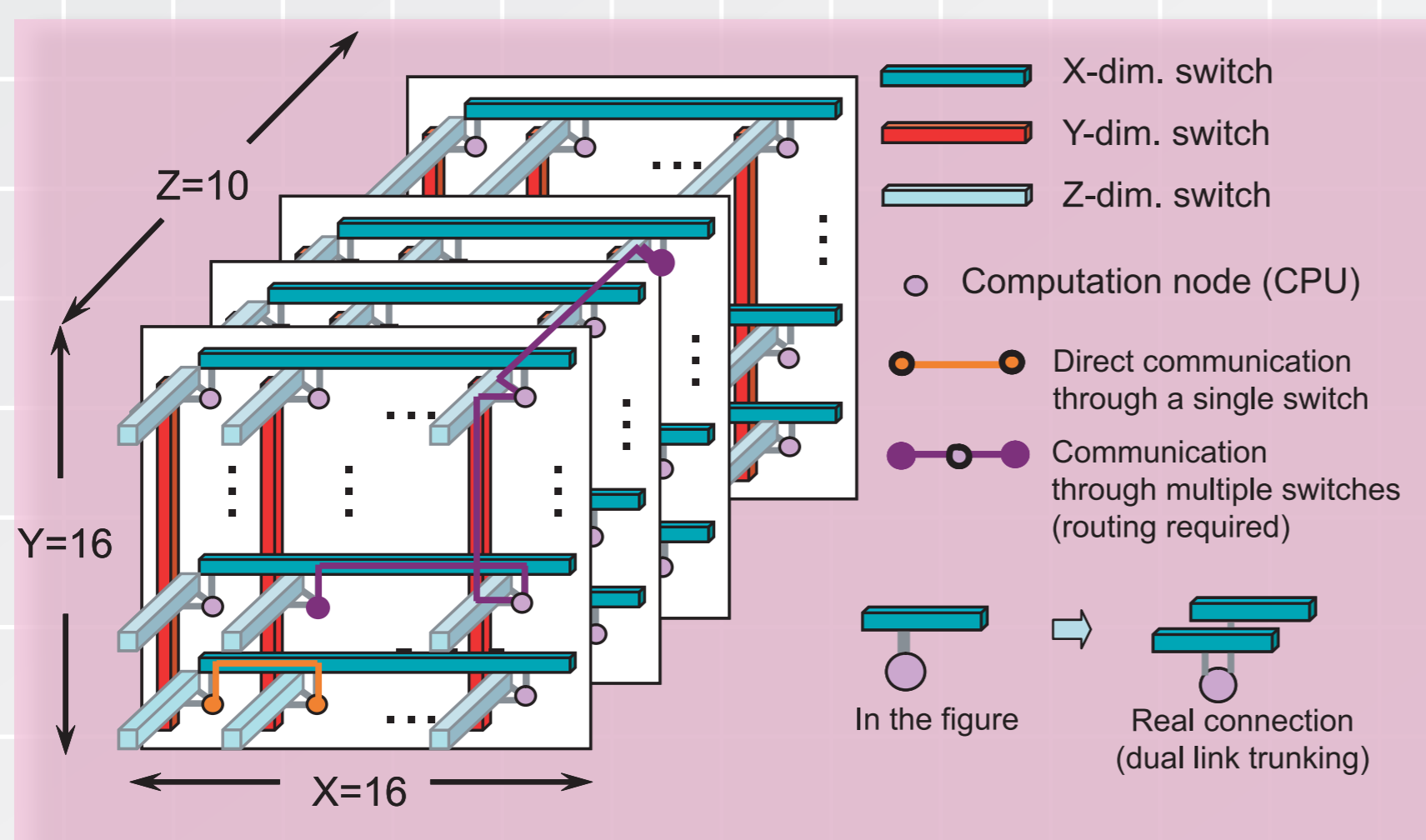800 TByte

T2K Kyoto
8 GByte

T2K Tokyo
300 TByte
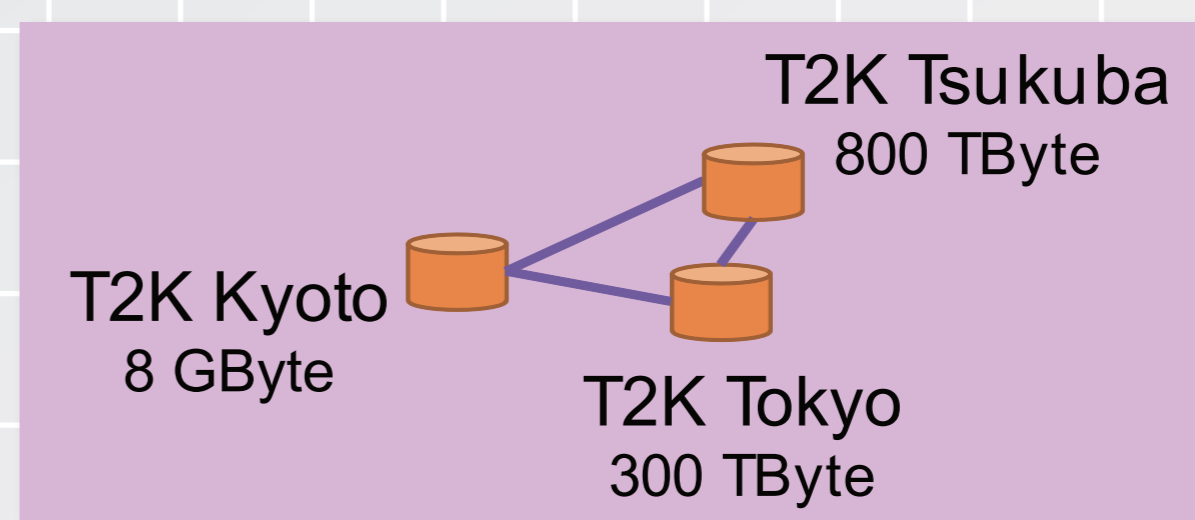
Development of PACS-CS

Large-scale simulations in computational science require a high performance system with a balanced floating point capability and network performance. PACS-CS is a massively parallel system designed to achieve such a balance using cost-effective commodity components.

number of nodes	2560 (16 x 16 x 10)
Peak performance	14.3 TFLOPS
Node configuration	Single CPU / node
CPU	Intel LV Xeon EM64T 2.8 GHz 1 MB L2 cache
Memory	DDR2/400 2 GB/node (5.12 TB/system)
Network	3-dimensional Hyper-Crossbar Network
Link bandwidth	750 MB/s (3-D simultaneous trans.)
Local HDD	160 GB/node (RAID-1, usable space)
Total system size	59 racks
Power consumption	550 kW
Operating system	Linux (Fedore Core 3) + SCore middleware
MPI	YAMPI and MPICH

Z=10

Y=16

X=16

- X-dim. switch
- Y-dim. switch
- Z-dim. switch
- Computation node (CPU)
- Direct communication through a single switch
- Communication through multiple switches (routing required)

In the figure

Real connection (dual link trunking)

Each computation node is equipped with three (X-, Y- and Z-dimension) paired on-board GbE NICs (6 ports in total) for data communication. Nodes on a single line of a dimension are connected by an L2 GbE switch. For 3-D nearest neighboring communication, the node can communicate with surrounding nodes simultaneously with aggregated 750 MB/s of theoretical peak bandwidth.

A dedicated network layer PM/Ethernet-HXB provides network trunking with paired links and high-speed routing in 3 dimensions. This network is suitable for direct physical mapping of problems with spatial domain decomposition.