High Performance Computing in a GRID Environment

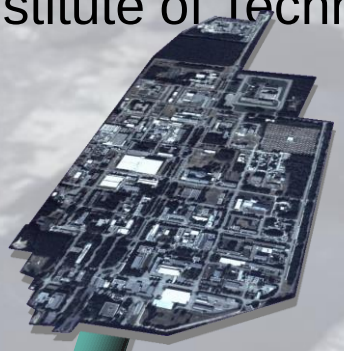
Institut for Scientific Computing

Frank Schmitz and Olaf Schneider

Agenda

- The merger of University and Forschungszentrum Karlsruhe
- What is Grid computing
- Example for a data grid
- Architecture
- Focus of Grid projects
- Grid history at Forschungszentrum Karlsruhe
- GridKa (www.gridka.de)
- INT.EU.GRID (www.interactive-grid.eu)
- CampusGrid (www.campusgrid.de)
- D-Grid (www.d-grid.de)
- Conclusion

Karlsruhe Institute of Technology (KIT)



KIT…

University

11	fakulties
120	institutes
4,000	employees
18,500	students
250 mio €	budget

Forschungszentrum

10	programs (research areas)
21	large institutes
4,000	employees
310 mio €	budget

science

education

structures for the administration

institutes

services

infrastructure

What is Grid Computing

- Foster's three point checklist: A Grid is a system that
  - coordinates resources that are not subject to centralized control
  - using standard, open, general-purpose protocols and interfaces
  - to deliver nontrivial qualities of service

Grid Computing is not Cluster Computing!

analogy with the electric power grid

Analogy World Wide Web

The World Wide Web provides *seamless access* to information stored in different geographical locations and different formats.

The Grid provides *seamless access* to any kind of computing power and data storage capacity distributed over the globe.

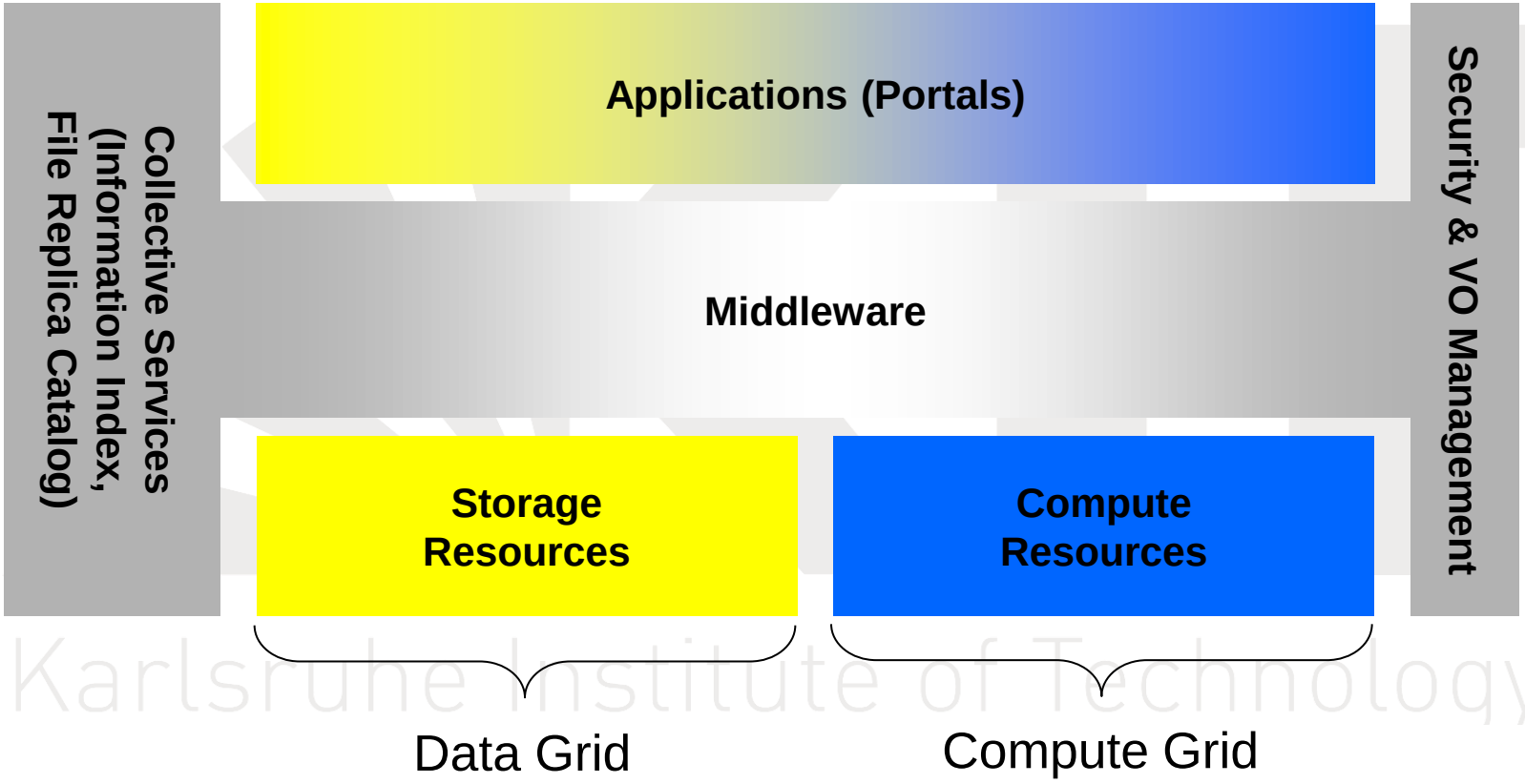
→ Word Wide Grid

A motivating example (Data Grid)

- CERN is building the Large Hadron Collider (LHC) the most powerful instrument ever built to investigate elementary particles physics
- Data Challenge:
  - One Megabyte of data digitised for each collision
  - 1010 collisions recorded each year = 10 Petabytes/year of data !!!
  - LHC data correspond to about **20 million CDs each year!**
- Simulation, reconstruction, analysis: LHC data handling requires a computing power equivalent to ~100,000 of today's fastest PC processors!
- (106 mega; 109 giga; 1012 tera; 1015 peta)

→ LHC Computing Grid (LCG)

Grid Architecture



Focus of Grid projects

- Software Development
  - applications
  - high level services
  - portals and clients
  - middleware components
  - build of middleware distributions

- Infrastructure
  - build up dedicated resources for a Grid
  - integration of legacy systems into existing Grids
  - deploy additional software on an existing Grid
  - operation of a Grid

the globus toolkit®
www.globustoolkit.org

eGee
Enabling Grids for E-sciencE

LCG

CROSSGRID

UNICORE

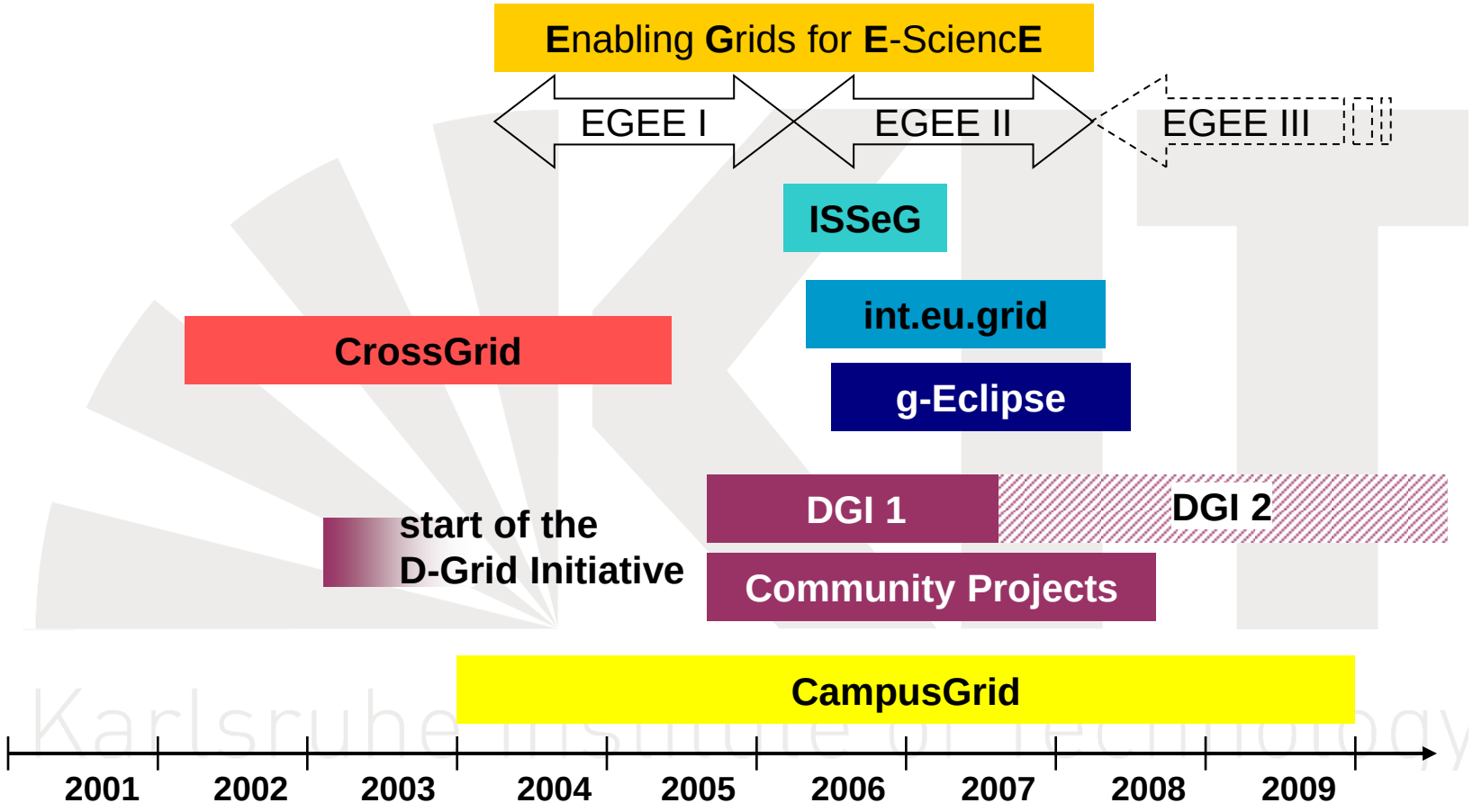
Distributed European Infrastructure for Supercomputing Applications

D-GRID

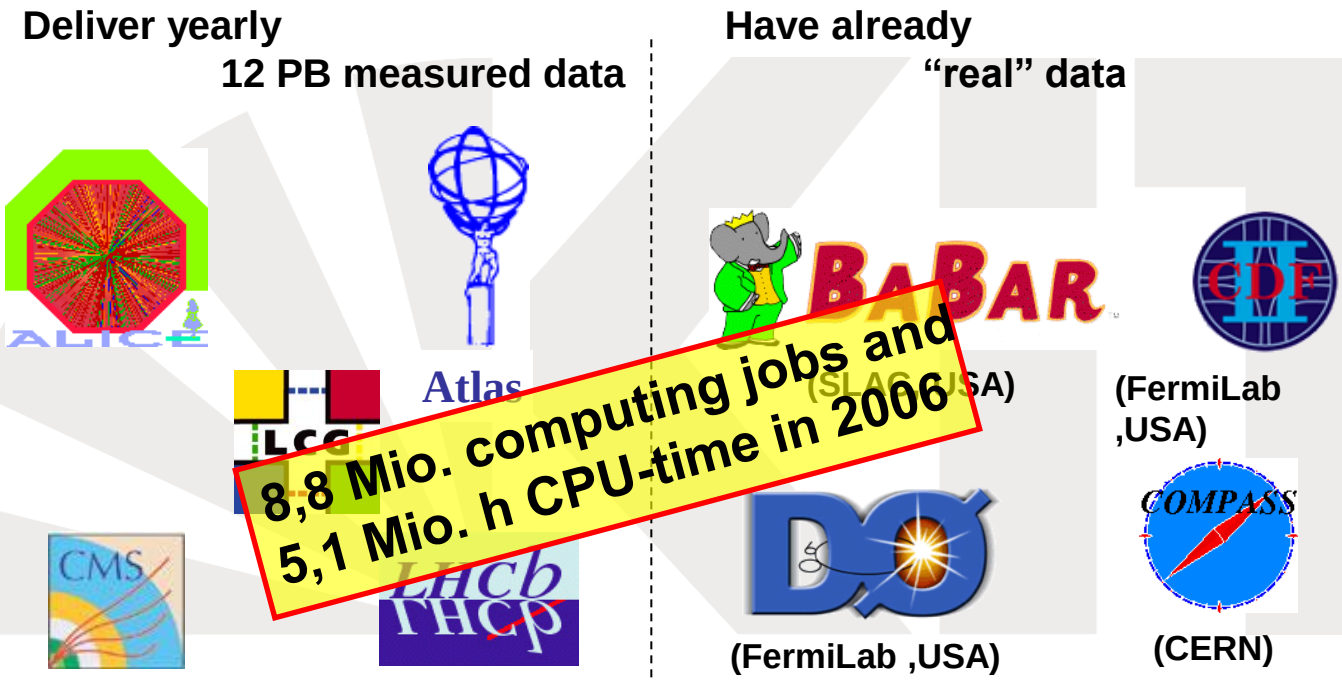
CampusGrid

History of Grid at FZK

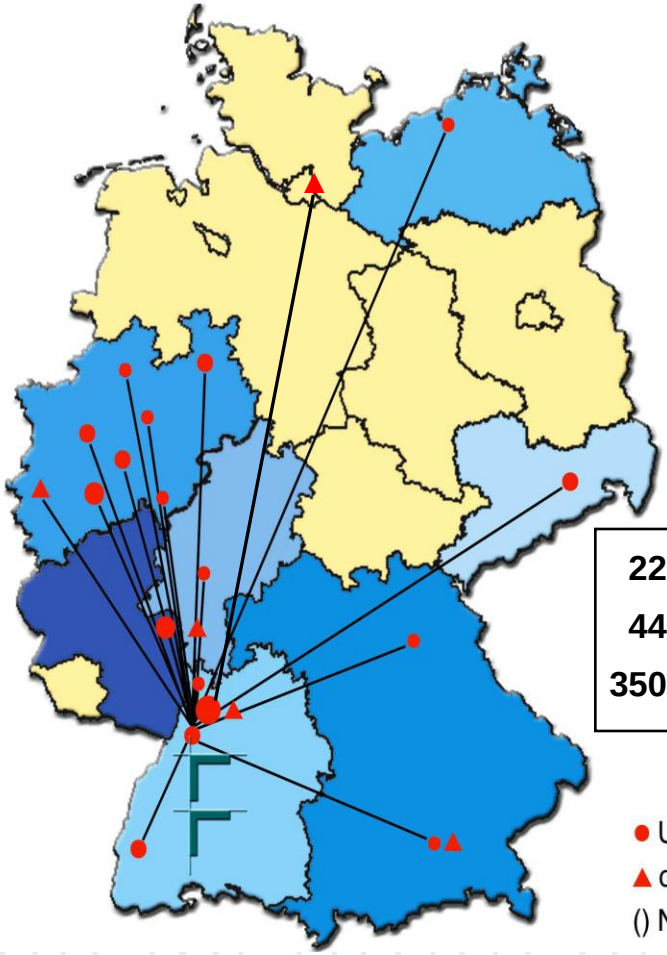
RDCCG **GridKa** (part of WLCG 2001 – 2015+)



GridKa, the Tier-1 computer center for LHC



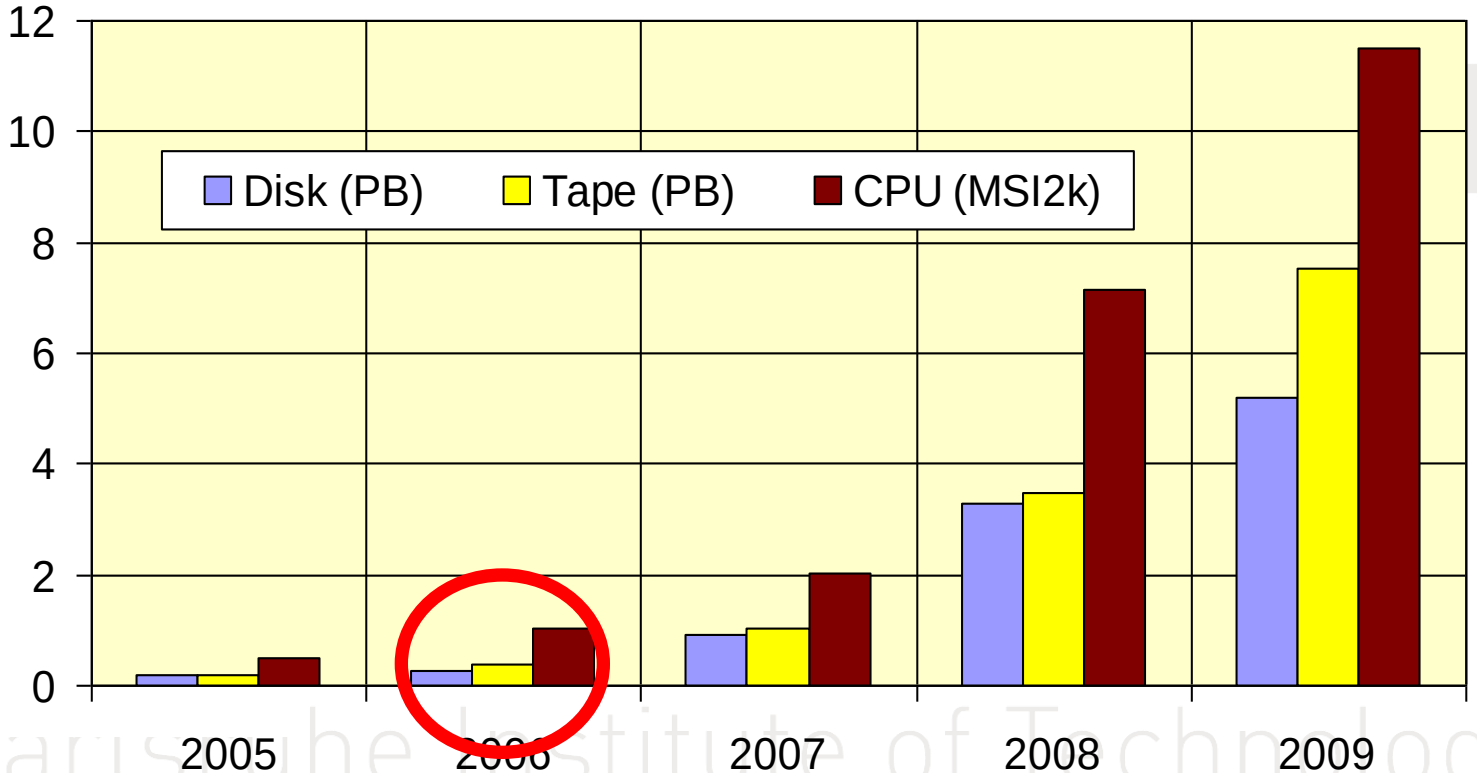
Aachen (4)●
Bielefeld (2)●
Bochum (2)●
Bonn (3)●
Darmstadt (1)▲
Dortmund (1)●
Dresden (2)●
Erlangen (1)●
Frankfurt (1)●
Freiburg (2)●
Hamburg (1)▲
Heidelberg (1)▲(6)●
Karlsruhe (2)●
Mainz (3)●
Mannheim (1)●
München (1)●(5)▲
Münster (1)●
Rostock (1)●
Siegen (1)●
Wuppertal (2)●



22 Research Centers
44 Working Groups
350 Scientists

● University
▲ other research institutions
() Number of working groups

FZK has signed the WLCG MoU and is able to deliver resources for the LHC experiment up to 2022 as the Tier-1 center



resources for WLCG

Interactive European Grid (int.eu.grid)

- sponsored by the European Union
- 13 full partners coming from 13 countries
- 2 FTE at FZK/IWR
- objectives
  - a huge parallel computer (MPI) across Europe
    - no special network → high latency, small bandwith
  - interactive access to the Grid
  - user friendly access
- applications
  - medicien
  - ultrasonic-CT (FZK)
  - cerebral aneurysm
  - high energy physics
  - catastrophe precaution
  - nuclear fusion

int.eu.grid – key aspects of FZK activities

- NA2: Applications
  - interactive access to Grid resources from within MATLAB
  - gridification of ultrasonic applications
  - user friendly access to the Migratiing Desktop (DataGrid project)
- SA1: infrastructure
  - developer Center: https://savannah.fzk.de
  - dokumentations Center: https://wiki.fzk.de/i2g
  - automatic releasebuilds
  - coordination of the deployment

CampusGrid

A fast and easy-to-use intra-site Grid infractructure for HPC

the globus toolkit®
www.globustoolkit.org

WSRF
OGSA

Vision:
The Scientists at KIT use *the Grid* as convenient and natural as they use email and the Web today.

R & D Project

CampusGrid

Opus[IB]

BladeCenter

eserver pSeries

AIX L

Initial situation (of HPC at FZK):

- a wide range of computer architectures, operating systems, schedulers etc.
- Application of various scientific domains (proteomics, climate, material sciences, …)
- User requirements: single-sign-on, global view of the data, easy access

Parts of the CampusGrid Project

watercooled Infiniband cluster with 32 SUN V20z and 64 FSC RX220 nodes (more than 190 cores and >1 TByte main memory)

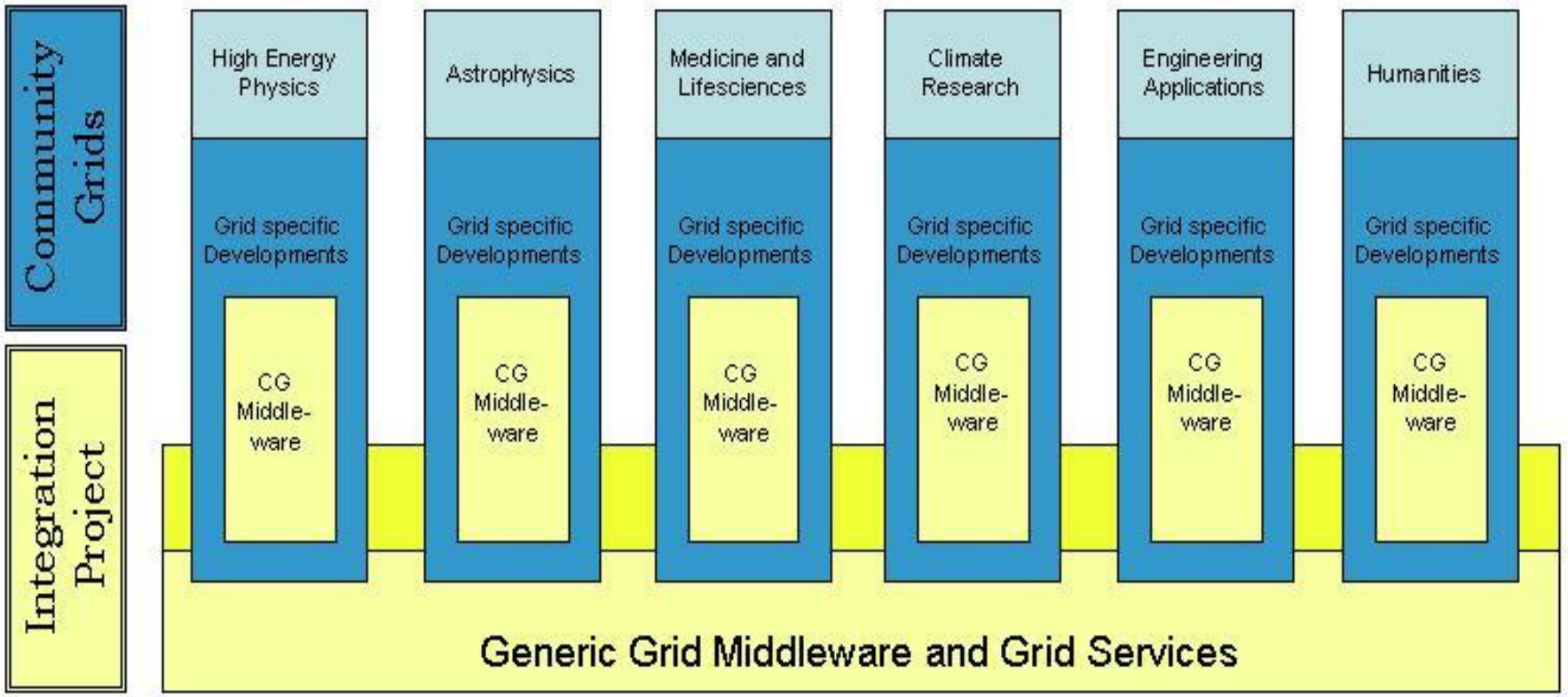
2 Infinicon 9100, Infiniband switches, MPI latency 4.0 µs between nodes

D-Grid (the German Grid initiative)

- starting September 2005 to build a sustainable Grid infrastructure, IWR is one of the founder of the project idea
- six Community Grid projects an the D-Grid Integration Project (DGI), IWR leads the DGI part
- funded by BMBF, the Federal Ministry of Education and Research
- astronomy, climate research, high energy physics, engineering research, medical research, humanities
- gLite, GT4, UNICORE
- SRM/dCache for the data access
- heterogeneous CPU access (AIX, Linux, Solaris, SuperUX)

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung


Community Grids
Integration Project
High Energy Physics
Astrophysics
Medicine and Lifesciences
Climate Research
Engineering Applications
Humanities
Grid specific Developments
CG Middle-ware
Generic Grid Middleware and Grid Services

Conclusion

Are there possibilities for HPC users in a Grid environment?

- Using DEISA (www.deisa.eu) because it's the EU Grid project for HPC.
- Accessing HPC cluster infrastructure by Grid based middleware like GT4, gLite or UNICORE (D-Grid).
- Building a local HPC cluster and integrate the hardware into a global Grid infrastructure (D-Grid).
- Running applications in the CampusGrid environment to become much more familiar in using Grid based heterogeneous infrastructure.

Why Grid?

- In case of growing problem size, you will find hardware resources in the Grid.