GRIDPP UK Particle Physics

CLRC

Technology Summary

John Gordon

GRIDPP UK Particle Physics

Talks

University Multidisciplinary Scientific Computing: Experience and Plans
- Alan Tackett (Vanderbilt University)

PASTA review (technology for the LHC era) - Michael Ernst, FNAL

Building a Computer Centre - Tony Cass (CERN)

CPU Technology Overview - John Gordon (RAL)

Tier1/A Storage Procurement - John Gordon (RAL)

A Pre-Production Update on the NSF TeraGrid - Remy Evard (ANL)

Grid/Fabric interaction discussion - led by Bernd Panzer Steindel (CERN)

European DataGrid Fabric Management - Olof Barring (CERN)

Evaluation of a MOSIX cluster as a group analysis facility at CDF - Andreas Korn (FNAL)

Using free desktop cycles – Frederic Hemmer (CERN)

FBSNG and Disk Farm - parts of large cluster infrastructure - Igor Mandrichenko (FNAL)

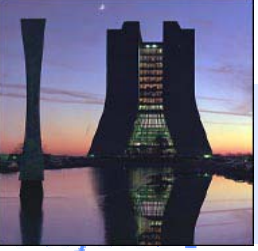
GRIDPP UK Particle Physics

Summary

- Vampire
  - Clusters need not be dedicated to one community.
  - A cluster makes a good general-purpose HPC service
  - Maui scheduler worth further study by others
- PASTA review (technology for the LHC era)
  - Basic technology looks OK for LHC
- Building a Computer Centre
  - Really large resources require professional planning
- CPU Technology Overview
  - Don't just count GHz, consider AMD
- Tier1/A Storage Procurement
  - Commodity disk isn't easy
  - Good contact with suppliers necessary
- A Pre-Production Update on the NSF TeraGrid
  - IA64 not really in use yet
  - Power4 and Alpha not dead yet

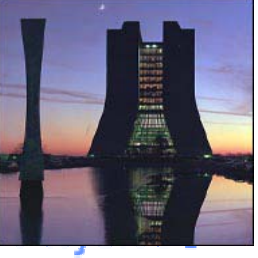
- European DataGrid Fabric Management
  - A lot of tools for installing, configuring and monitoring
- Evaluation of a MOSIX cluster as a group analysis facility at CDF
  - Interesting alternative to batch
  - Not fully scalable yet
- Using free desktop cycles

- FBSNG and Disk Farm
  - Disk farm gives single interface to all disks on a farm
- White boxes
  - CERN now run a single cluster with shares and priorities



Pasta Conclusions

- Tape and Network trends match or exceed our initial needs.
  - Need to continue to leverage economies of scale to drive down long term costs.
- CPU trends need to be carefully interpreted
  - The need for new performance measures are indicated.
  - Change in the desktop market might effect the server strategy.
  - Cost of manageability is an issue.
- Disk trends continue to make a large (multi PB) disk cache technically feasible, but ….
  - The true cost of such an object a bit unclear, given the issues of reliability, manageability and the disk fabric chosen (NAS/SAN, iSCSI/FC etc etc)
  - File system access for a large disk cache (RFIO, DAFS, …) under investigation (urgent !)
- More architectural work is needed in the next 2 years for the processing and handling of LHC data.
  - NAS/SAN models are converging, many options for system interconnects, new High Performance NAS products are (about to be) rolled out (Zambeel, Panasas, Maximum Throughput, Exanet etc)



PASTA has addressed issues exclusively on the Fabric level

- It is likely that we will get the required technology (Processors, Memory, Secondary and Tertiary Storage Devices, Networking, Basic Storage Management)
- Missing: Solutions allowing truly distributed Computing on a Global Scale

  Will the Grid Projects meet our Expectations (in time) ?

GRIDPP UK Particle Physics

Topics

- CPU
- Disk
- Tape
- Network
- Operating Systems
- Infrastructure

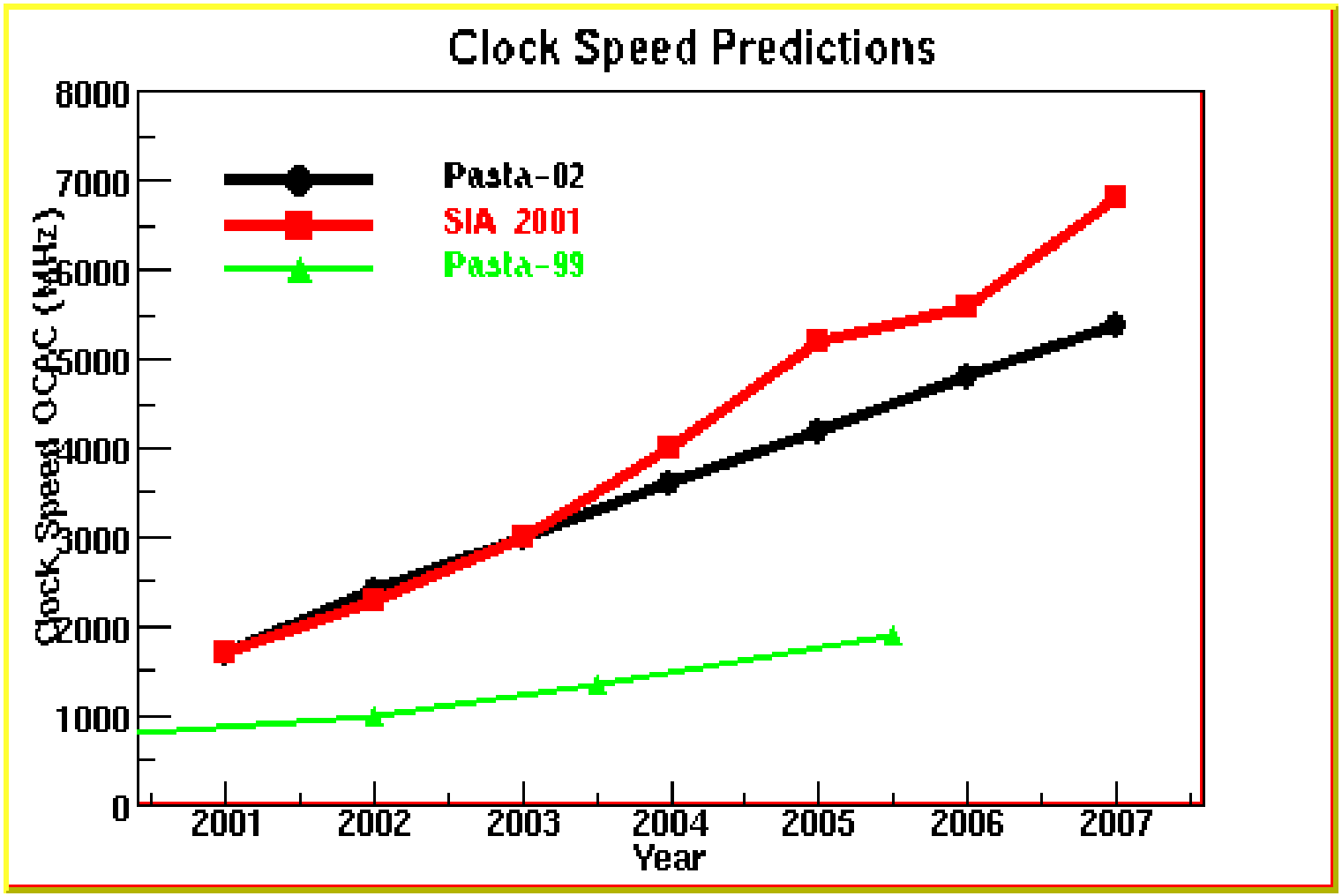
GRIDPP UK Particle Physics

CPU

- Pasta
- Intel widely used
- AMD used for floating point
- HEP doesn't need low latency switches but HPC does.
- Power4, Alpha not dead yet
- White boxes still cheapest capital cost
  - But racking benefits are felt to be worth it by many

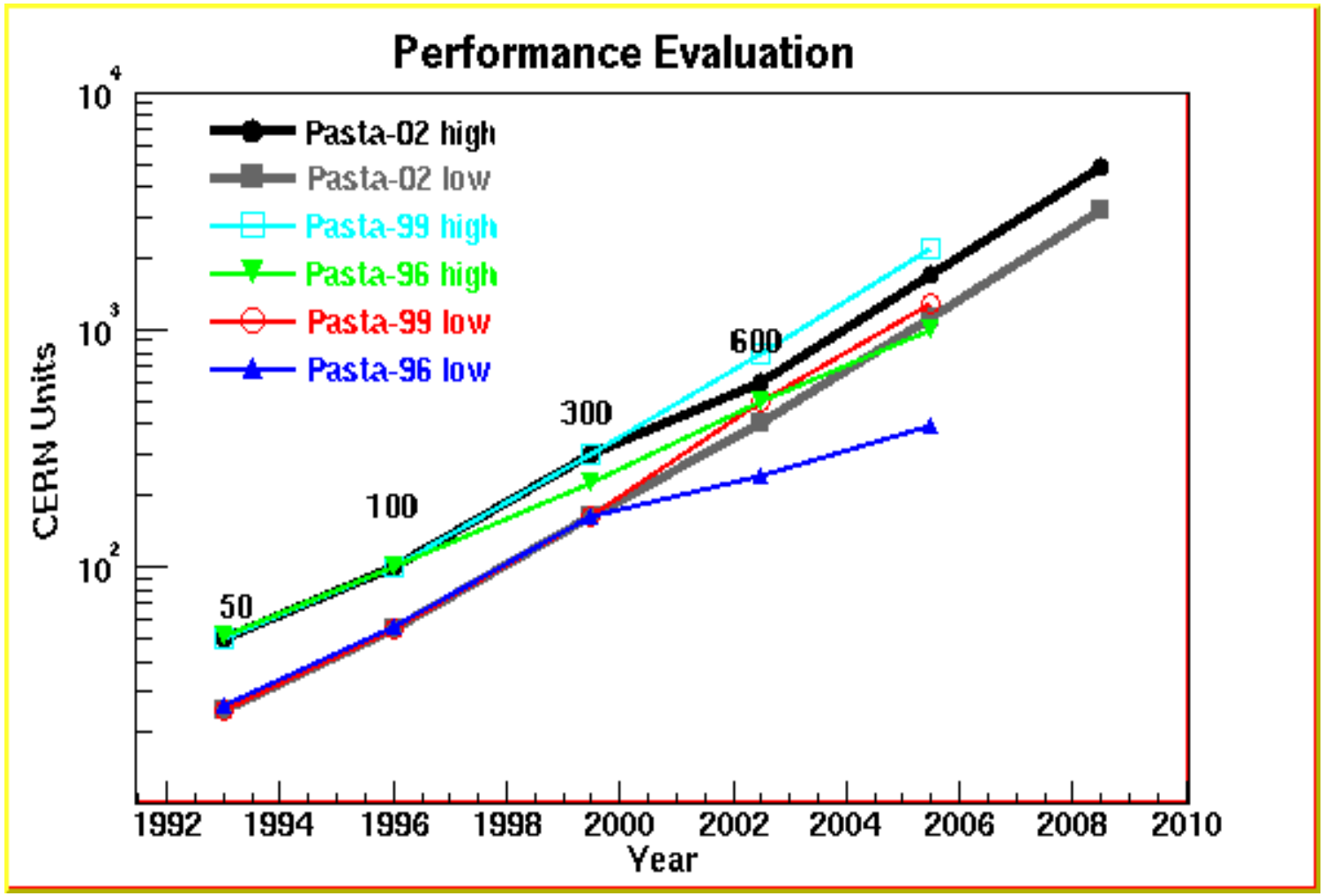
GRIP UK Particle Physics

Pasta



GRIP UK Particle Physics

Pasta



GRIDPP UK Particle Physics

Disk

- Bigger and bigger - 350GB
  - I thought things were good until I heard RHIC☺
- Not that much faster
- Still a debate on commodity vs SAN (and in between)
- Still not as cheap as tape

GRIDPP UK Particle Physics

Tape

- Some communities (HEP, environment, bio, astronomy) still require tape for foreseeable future.
- Capacity and bandwidth still increasing
  - Tapes roughly match disks in size but we still read/write files
  - Objects on tapes?
- Tape LTO look useful but STK still dominates

GRIDPP UK Particle Physics

Network

- TeraGrid
  - Techniques exist for very fast reliable bulk transfer
  - But the general Internet does not deliver this yet
- Pasta
  - The basic speed will be there but topology is already important

GRIDPP UK Particle Physics

Transatlantic Net WG (HN, L. Price)
Bandwidth Requirements [*]

	2001	2002	2003	2004	2005	2006
CMS	100	200	300	600	800	2500
ATLAS	50	100	300	600	800	2500
BaBar	300	600	1100	1600	2300	3000
CDF	100	300	400	2000	3000	6000
D0	400	1600	2400	3200	6400	8000
BTeV	20	40	100	200	300	500
DESY	100	180	210	240	270	300
CERN BW	155-310	622	2500	5000	10000	20000

[*] Installed BW. Maximum Link Occupancy 50% Assumed

GRIDPP UK Particle Physics

Operating Systems

- Linux
  - hardly discussed – a given
  - Agreeing the release is not trivial though
- Mosix
  - Interesting features
  - Version 0.98.0!!
  - not ready for large scale use

GRIDPP UK Particle Physics

Infrastructure

- Many Large Clusters exist
  - Some were designed and some just grew.
- Large clusters need serious thought about environment (cooling, power, noise, safety)
- Management
  - EDG have a rich set of management tools for installation, configuration and monitoring
  - Installation has been solved in many places
    - Many solutions
  - Configuration is not so well solved
  - Application Installation is not solved

GRIDPP UK Particle Physics

What Else?

- Many Technology Issues not addressed at this workshop
  - Users – managing large communities
  - Software certification – control over what software runs
  - Security – discussed in many other places.
  - Taxonomy – do we agree what words mean?
    - NAS, SAN, authentication, authorisation, backup

- I am sure we will discuss them in future