CHEP2010

Rain and lecturing in Taipei

Sam Skipsey

With thanks to Wahid Bhimji and Stuart Purdie for comments on parallel sessions I was unable to attend.

Itinerary

- What is CHEP?
- Themes overview
  - A Summary of Summaries
  - Common themes
  - Presentation highlights
- Conclusion

CHEP - 2010

- The big Computational High-Energy Physics Conference
- Hosted by Academi(c)a Sinica, Taipei, Taiwan
- Opened by the Vice-President of Taiwan, Vincent Siew (蕭萬長)
    - Good luck getting Nick Clegg in the UK!

A brief slide about Taipei.

- Capital of Taiwan (Republic of China)
- Pronounced “Taibei” (臺北)

- Has an international baseball team.
- Has tons of Taoist, Buddhist and folk Temples.
- Filled with “Night Markets”
- Oh, and typhoon season is around October...

A brief slide about Taipei.

- Capital of T
- Pronounce

We can't stop here; this is Typhoon Country!

An overview of Summaries

- 7 Tracks of Parallel Sessions in total
  - Event Processing; Online Computing; Software Engineering & Data; Distributed Processing & Analysis; Computing Fabrics & Networking Tech.; Grid & Cloud Middleware; Collaborative Tools*
- Each was summarised in 30 minutes on Friday.
- We'll start with a summary of the summaries:

*we implicitly continue the agenda of Newton in pretending rainbows have 7 colours...

Event Processing

- GEANT4 (simulation)
- FAIR plans (GEANT4, event analysis trains)
- Analysis and I/O (GPU, etc)
- Software standardisation
- Alignment and calibration (brilliant pictures)

Event Processing tag cloud

- Fabio Cossutti (INFN) Ryosuke Itoh (KEK) Oliver Gutsche (Fermilab)

Event Processing

Conclusions

- LHC:
  - In general, everything is working remarkably well
  - Current data taking conditions are under control
  - The future will show how the experiments will cope with the increasing PileUp conditions
- Beyond LHC
  - FAIR experiments with non-traditional beam conditions mark a new frontier of challenges
  - New experiments rewrite or develop new frameworks, software standardization helps them ramping up more quickly
- Performance
  - **Multi-Core, GPUs and Vectorization**: buzz words of the processing world
  - Applications show significant speed increase, but usage still very dependent on specific situations
  - New experiments are designing their software for multi-core, multi-thread execution environments and also consider specialized hardware solutions like GPUs

Online Computing

- All LHC experiments ++good with real data taking. (DAQ eff > 90%)
- Emphasis on storage perf (at T0)
- Integration and automation!
- Data quality monitoring - and web access!
- Lessons learned: uniform stack (on and offline)

Online Computing

Summary

- The startup of the LHC experiments has been a tremendous success: DAQ efficiencies well over 90% and over all efficiencies for physics well over 80%
- Sophisticated tools for data quality monitoring allow remote and local experts to react and flag quickly
- Modern web-technologies make experiment info available everywhere
- But…
  - The LHC is improving and in some areas we are already beyond initial design parameters
  - Upgrades are coming: more data, new detectors, faster readout & storage

Software Engineering, Data*

- Heterogeneity of talks (also mentioned in other summaries)
- multithreading cool
- performance monitoring tools
- lots of "new" cool things: go, svn, html5, CVMFS etc
- software recycling for small exps.
- Data archiving and preservation

Software Engineering, Data *

Software Engineering, Data *

Conclusions

- The software frameworks for LHC are in very good shape
  - Processes and tools are in place
  - A lot of efforts on the performance side are underway
    - We need more specialists in this area
- Other experiments should be able to profit from the work
  - We need for more collaboration across experiments
  - Implementation islands are getting bigger

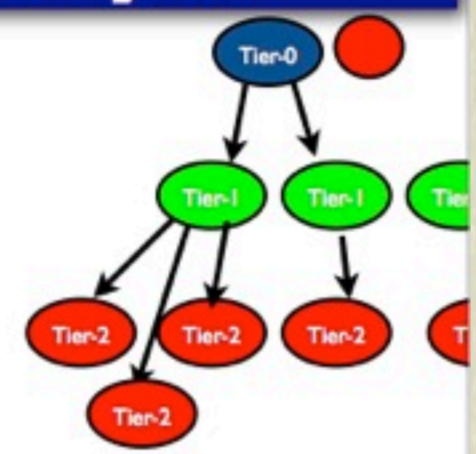
Distributed Processing and Analysis

- Successes
  - First year LHC stuff, organised processing, phenix, ATLAS DDM
- Future Architectures
  - FAIR, SuperB, Belle2, CDF, Fermi Space telescope
- Improvements
  - Global FS, ARC, VMs,

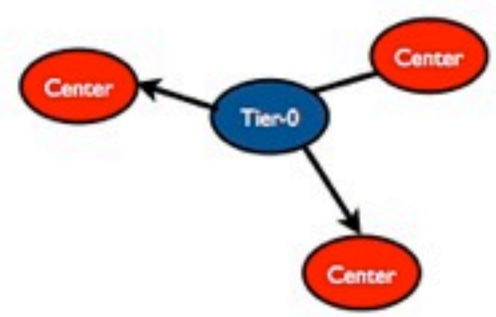
Distributed Processing and Analysis

Distributed Processing and Analysis

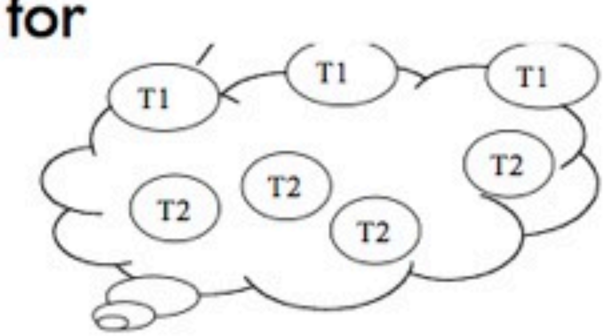
- Basically Breaks into 3 main themes
  - Successfully Processing and Analysis in Distributed Environments
    - Start-up of LHC Computing



  - Architectures for Future Facilities



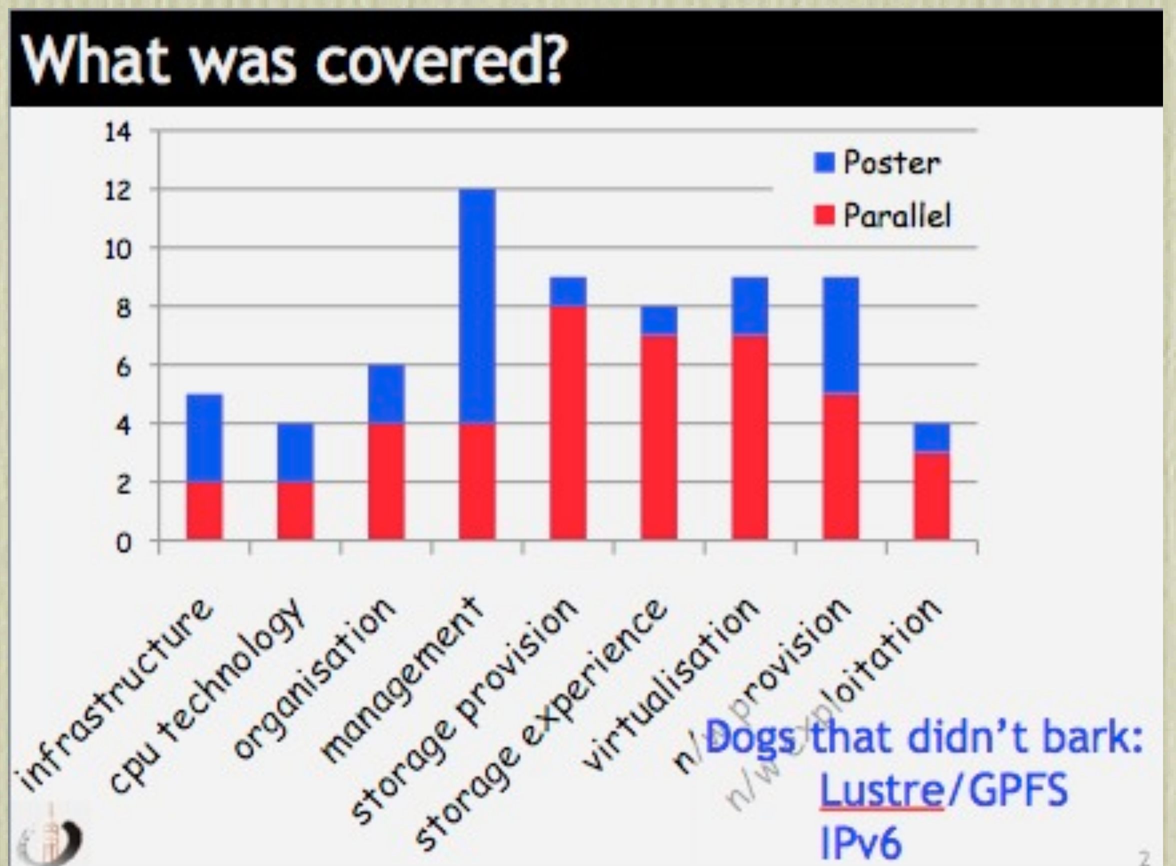
  - Improvements in infrastructure and Services for Distributed Computing
    - VMs, Clouds, Storage
    - Monitoring, Simulation, and Infrastructure

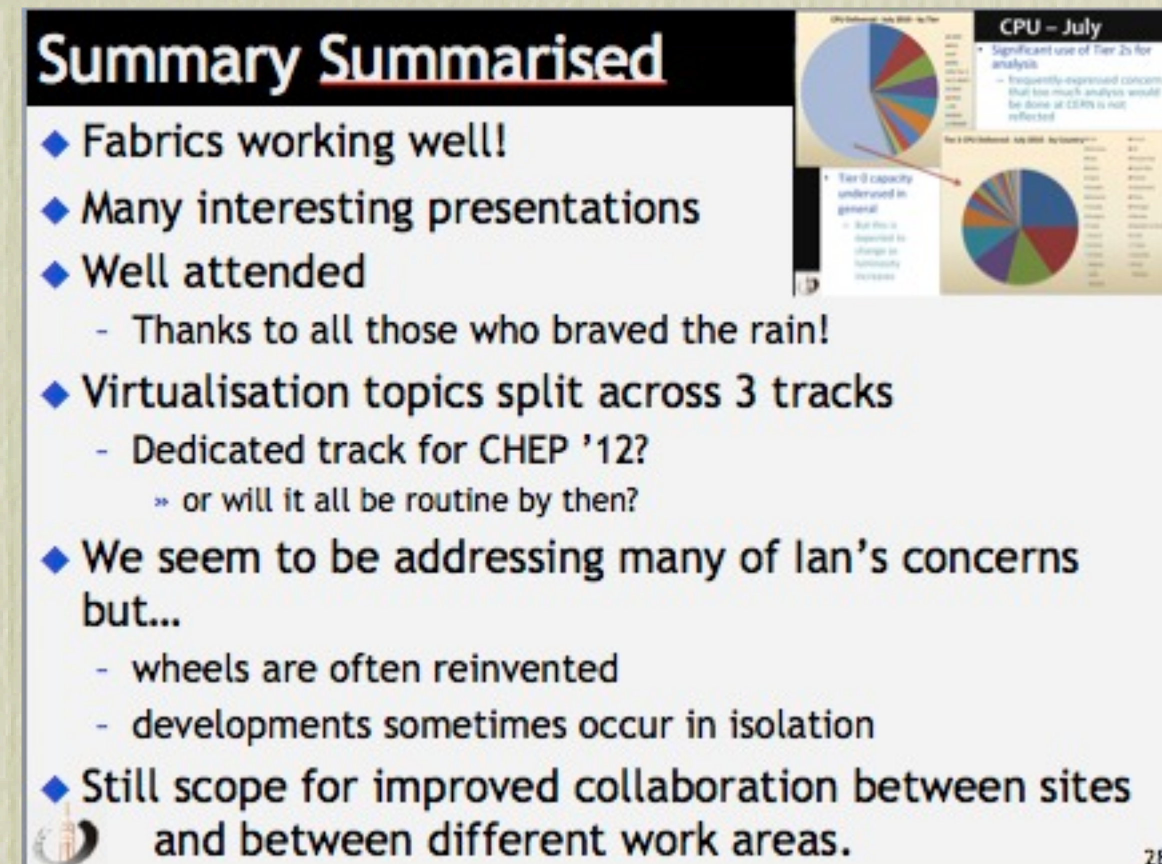


Computing Fabrics and Networking Tech

- Storage, Storage, vms, management, multicore
- No: lustre (as a sole topic), IPv6
- Fabric management (puppet)
- Data management architecture reworking
- NFs4.1, EOS, CPU scaling, Clouds (expensive)

Computing Fabrics and Networking Tech.



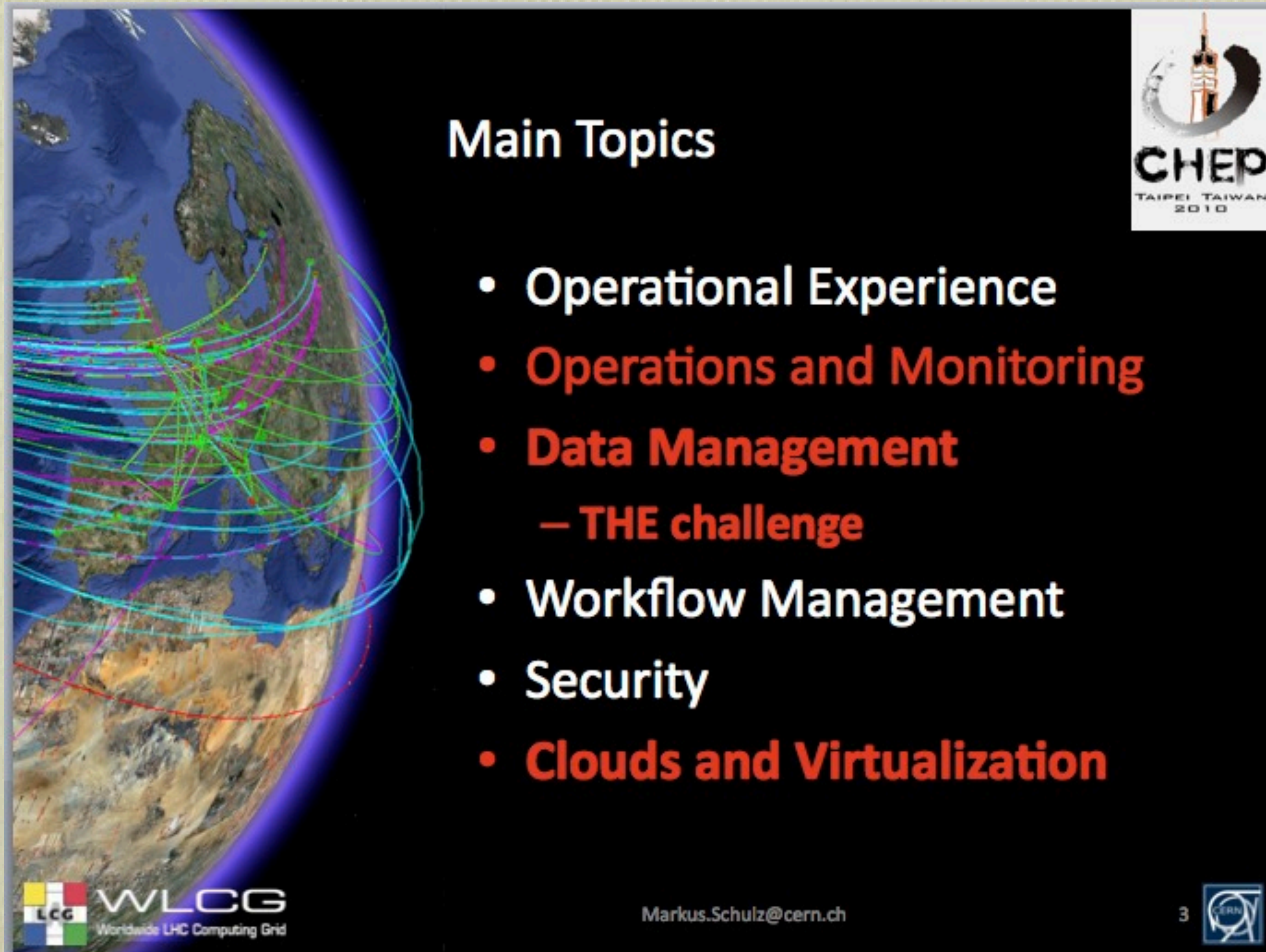


- Tony Cass

Grid and Cloud Middleware

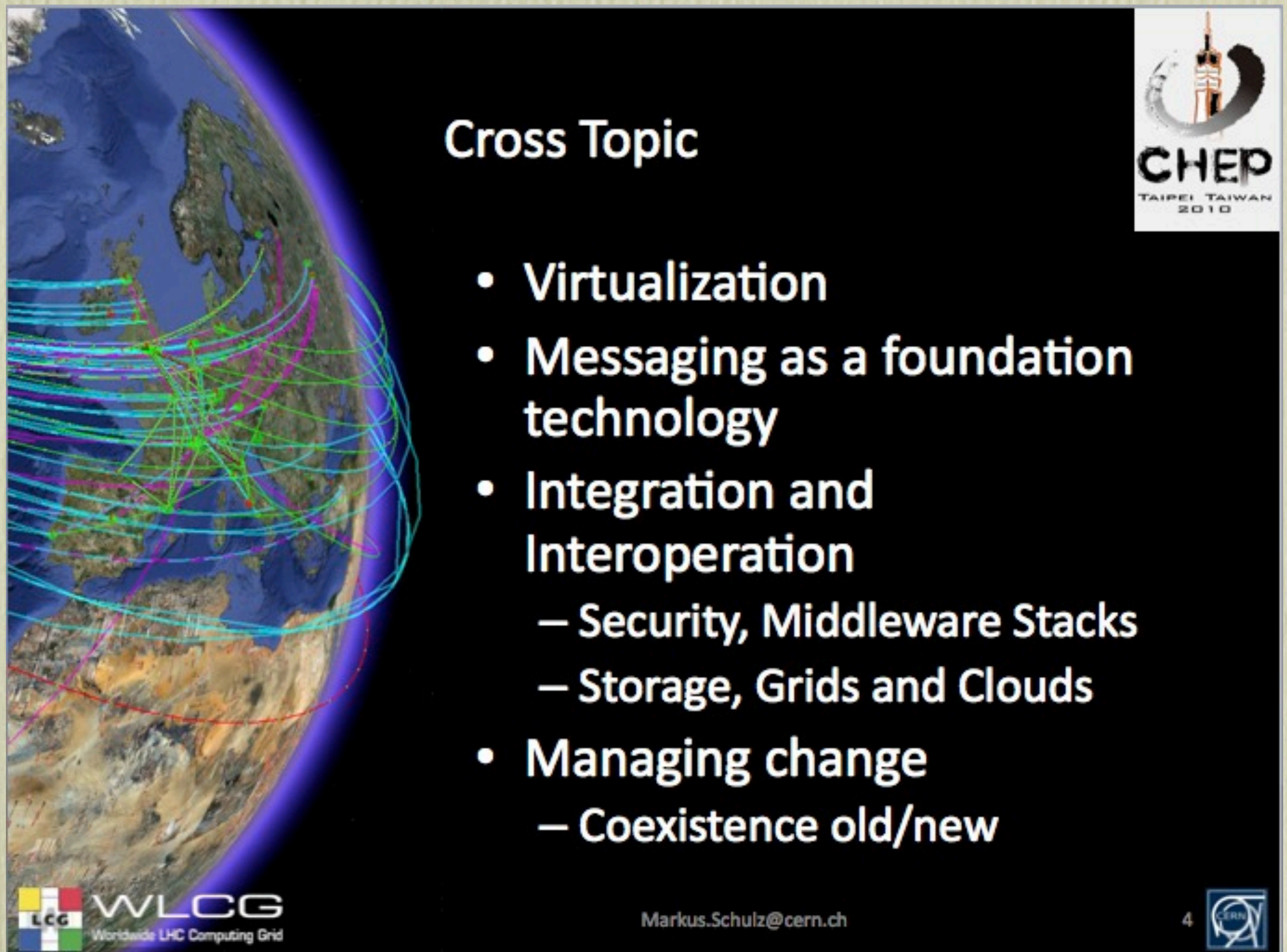
- Operations and Monitoring
- Data Management - EMI plans
- Clouds -H1Data (ceph!), Boinc+CVM+CoPilot
- Virtualisation, messaging, integration
- Pilot jobs (improvement of)

Grid and Cloud Middleware



- Markus Schulz

Grid and Cloud Middleware



Collaborative Tools

- Outreach Plenary
- Web2.0 internal/external communications (ATLAS), CMS collab. infrastructure
- Inspire, ATLAS Live, Glance information system
- EVO
- CERN Lecture archiving system,
- AbiCollab (like google Docs)

Collaborative Tools

Overview

CHEP
TAIPEI TAIWAN
2010

- Increasing areas in the CT field
  - HD videoconferencing systems, Outreach and Inreach activities, Rich Media Content, Information systems, etc.
- Representative examples covering the activities in the HEP community
  - 1st session dedicated to Policies and New initiatives
  - 2nd session dedicated to SW systems and Collaborative Tools
- Plenary Talk (Lucas Taylor, FNAL/CMS)
  - Overview about the importance of the outreach activities for the HEP community
    - Contract with the Society
    - Need of a defined strategy: HQ messages, defined relation with the Media and use of latest multimedia technologies
    - Everyone needs to be involved

- Joao Fernandes – CERN, Philippe Galvez – Caltech, Milos Lokajicek – FZU Prague

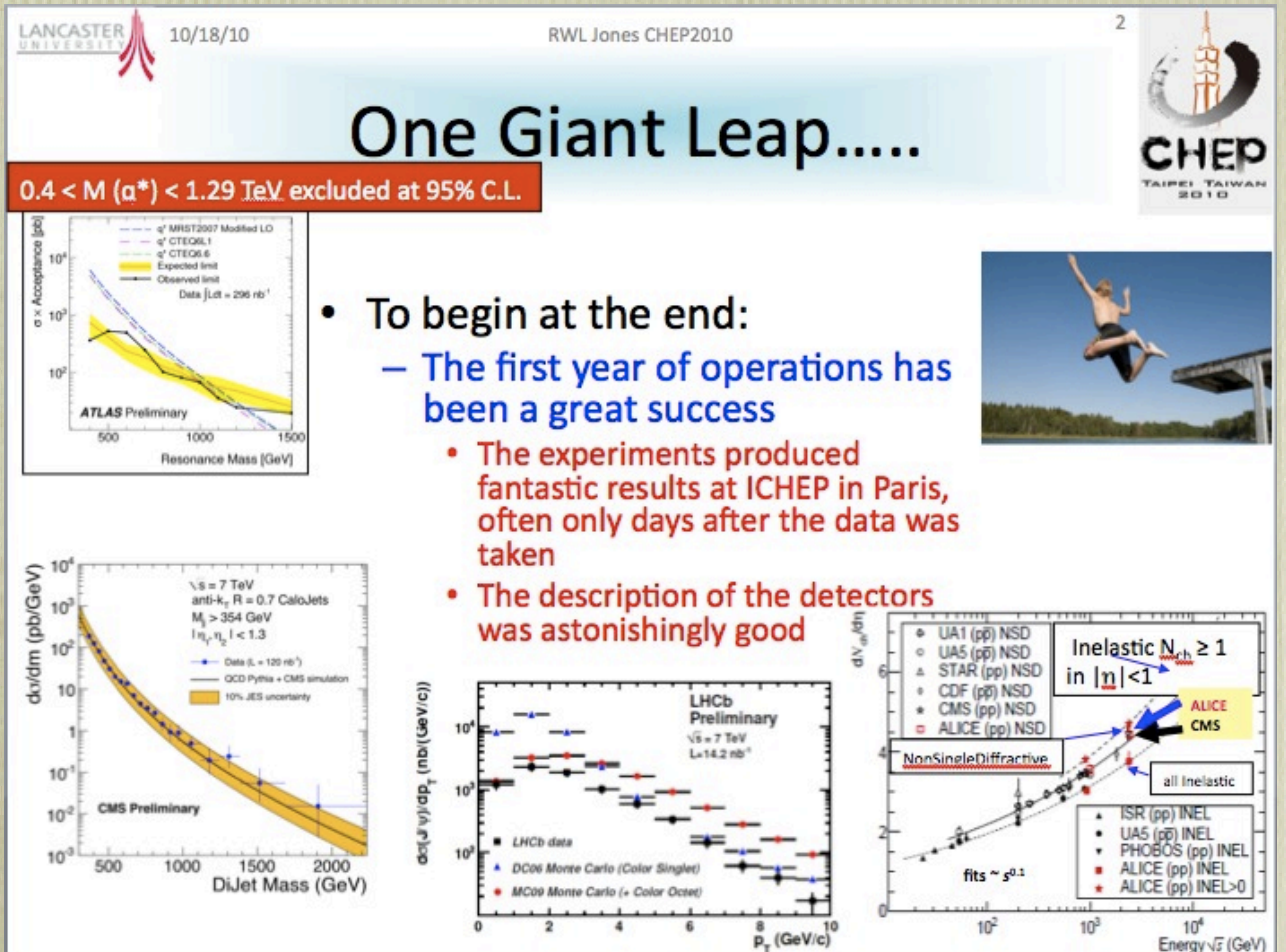
So, the summary of summaries:

- The LHC works!
- Data Management is hard
- Multi-core, GPGPU are exciting
- Virtualisation and Clouds are hot topics.
- Talking to people is good.

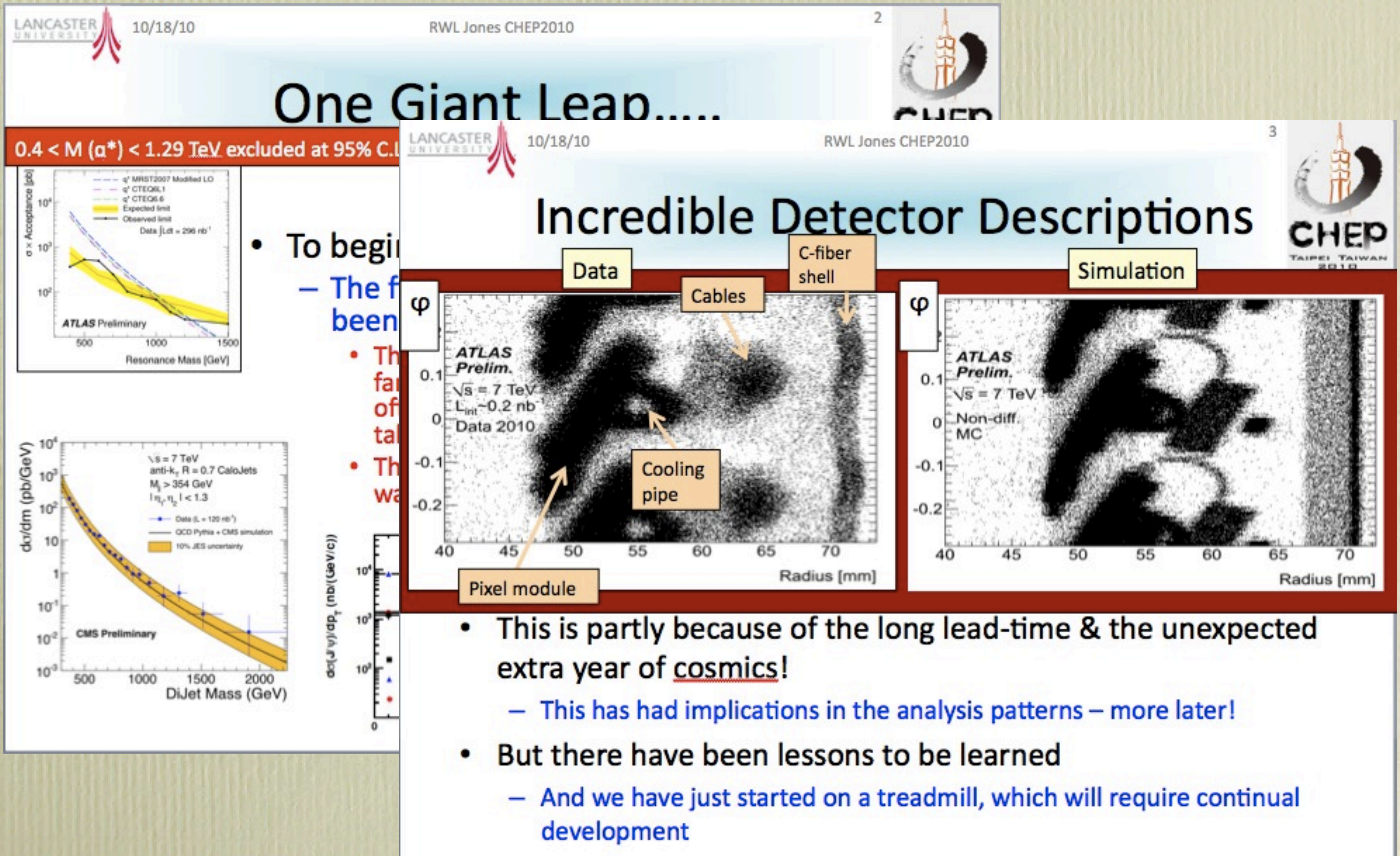
Presentation Spotlights

- A selection of the presentations that were most interesting.
- Subjective! Caveat Auditor!

The LHC works!



The LHC works!



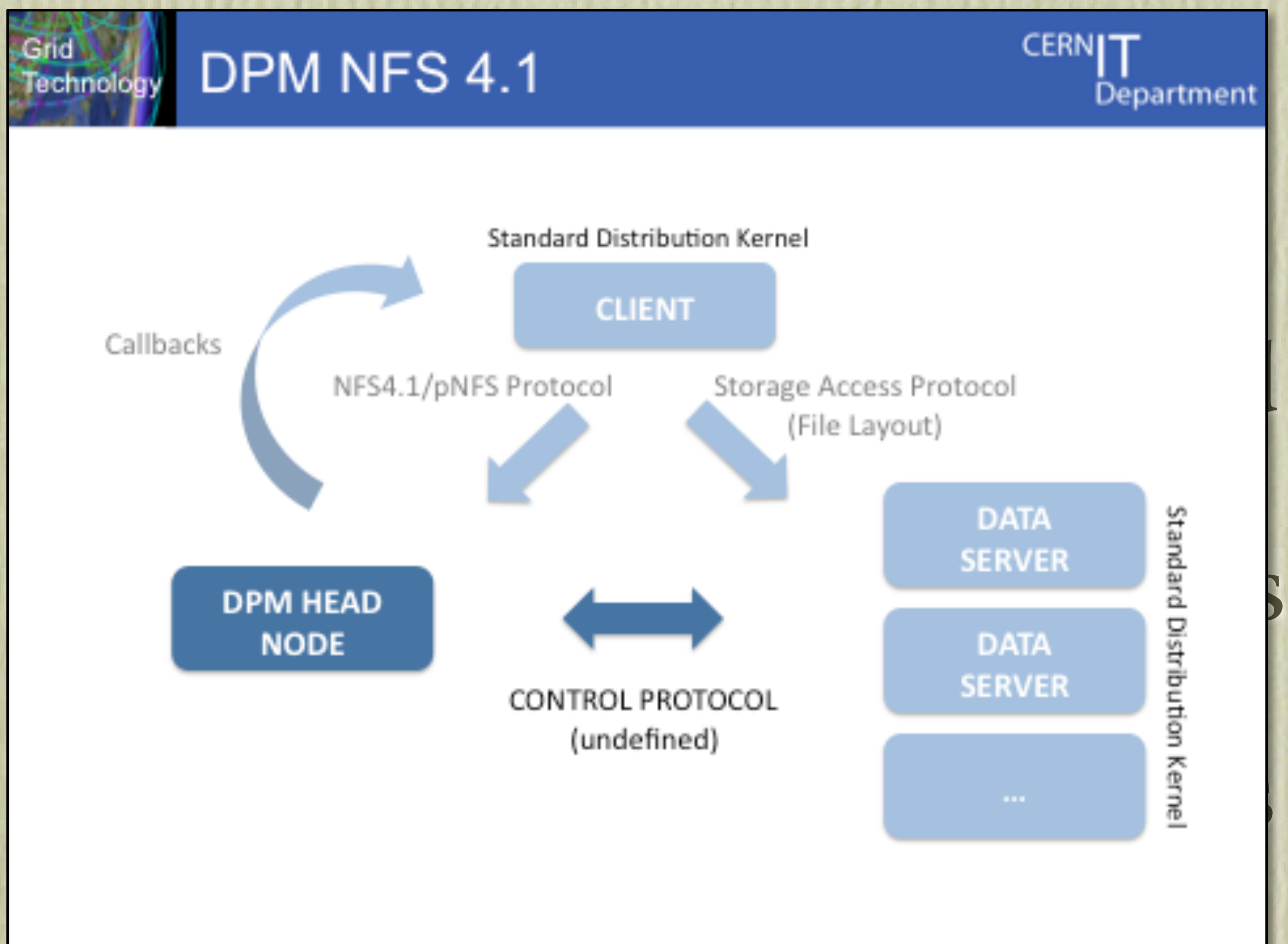
Data Management

- WLCG Data Management meeting 16 June.
  - Many presentations sprung from this.
- New storage technologies evaluated.
  - SSDs, Ceph, CERNVM-FS
- Archiving and long-term storage.

Amsterdam Themes

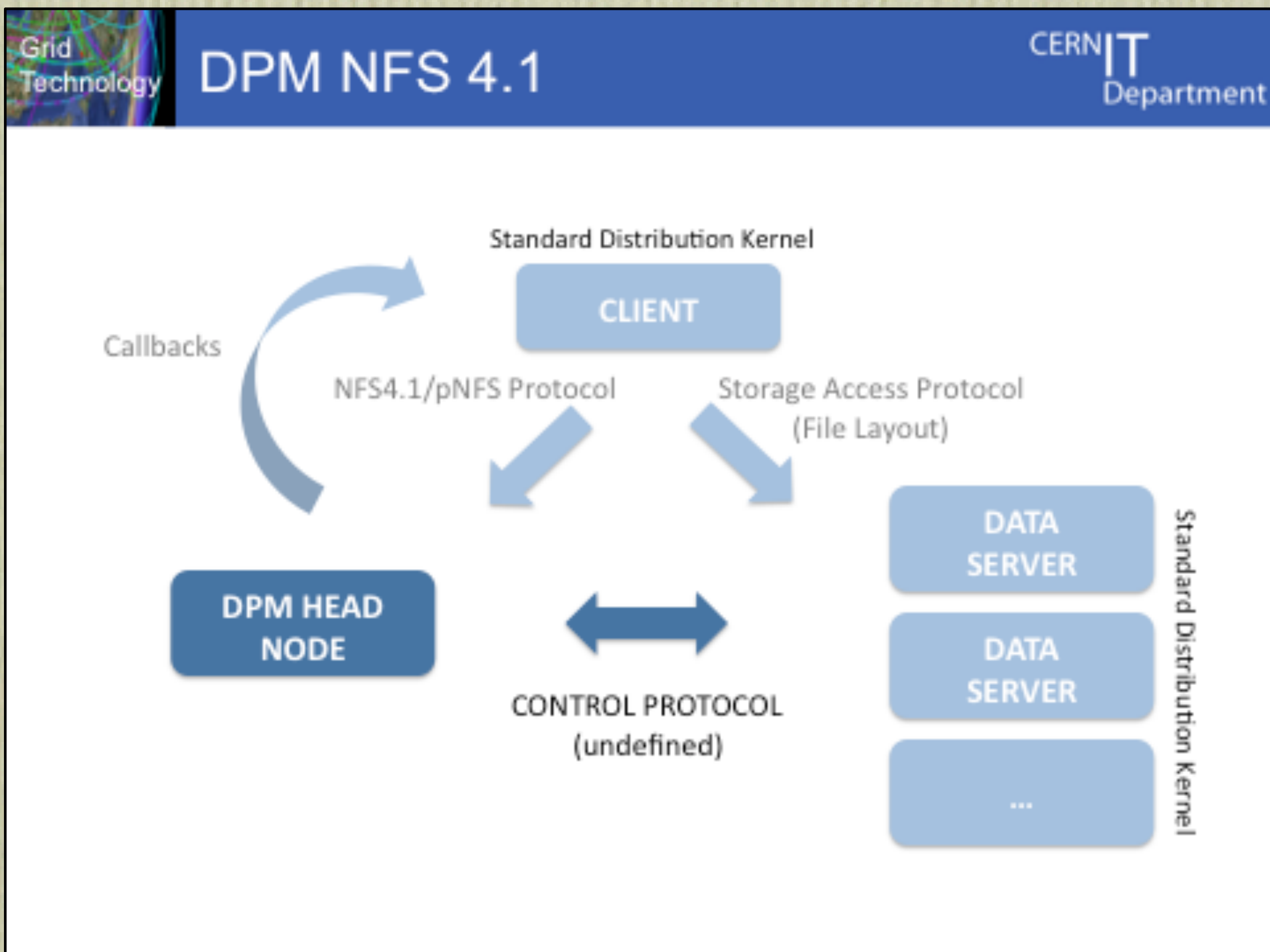
- Storage themes:
  - Dynamic data and Caches
  - Consistency - Messaging
  - Global filesystems (xrootd, mostly)
  - NFS4.1/pNFS
  - Archiving!

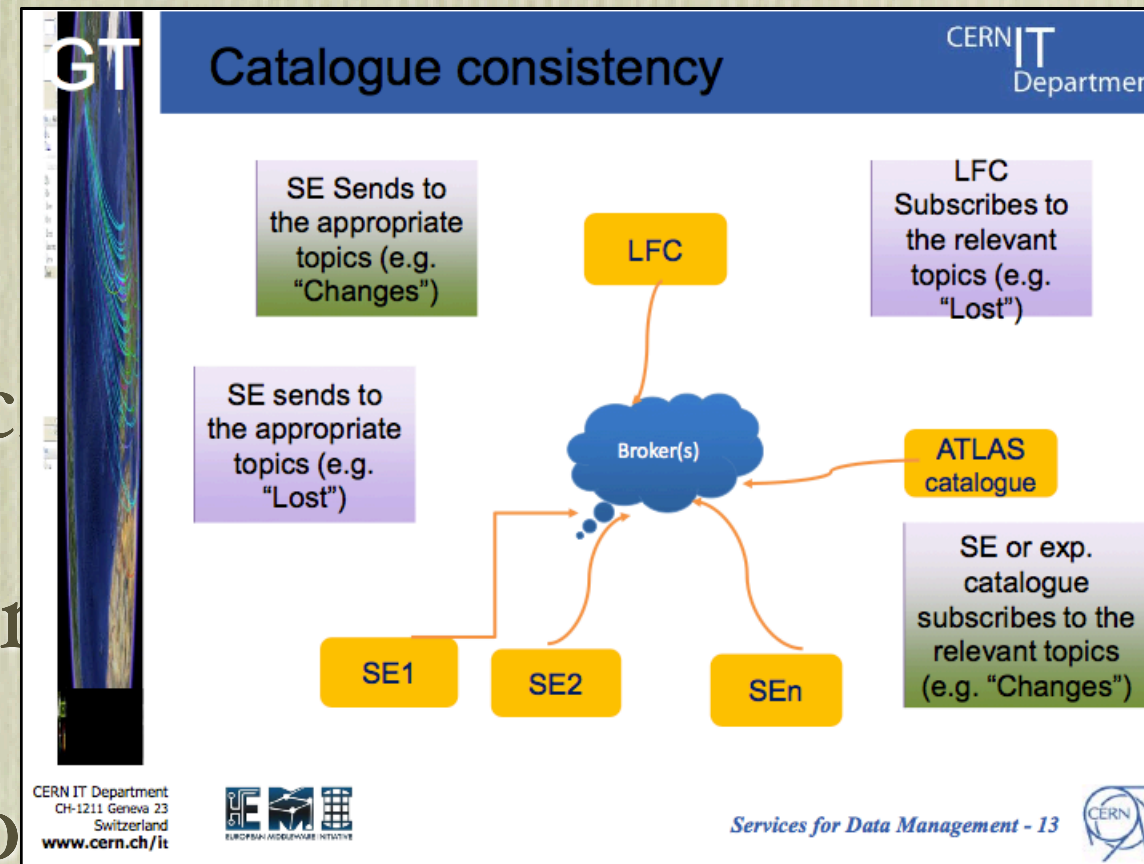
Amsterdam Themes



- l Caches
- ssaging
- s (xrootd, mostly)
- NFS4.1/pNFS
- Archiving!

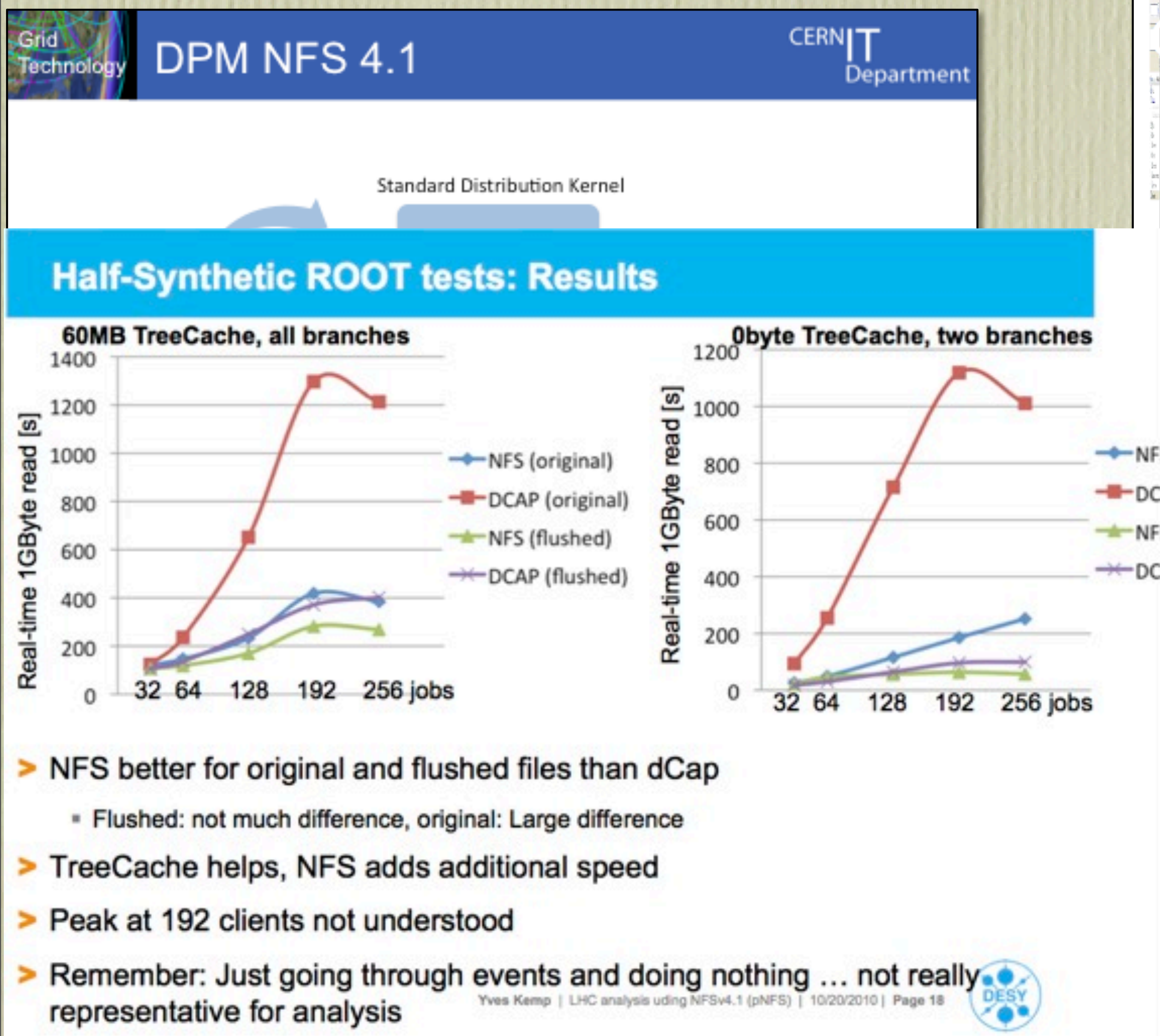
Amsterdam Themes

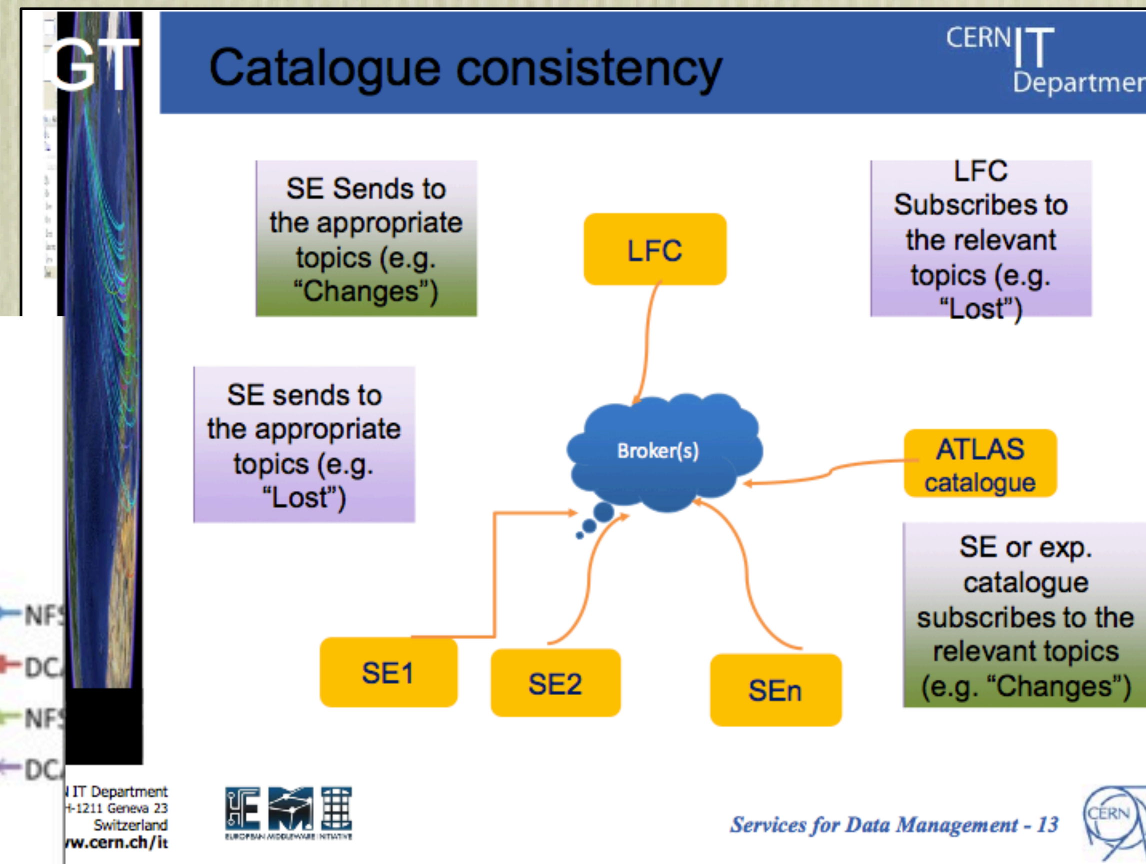




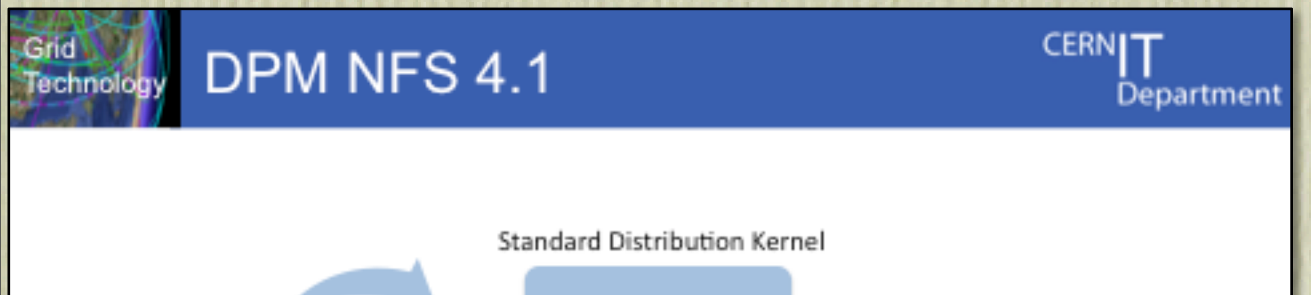
- l Cac
- ssagi
- (xro
- NFS4.1/pNFS
- Archiving!

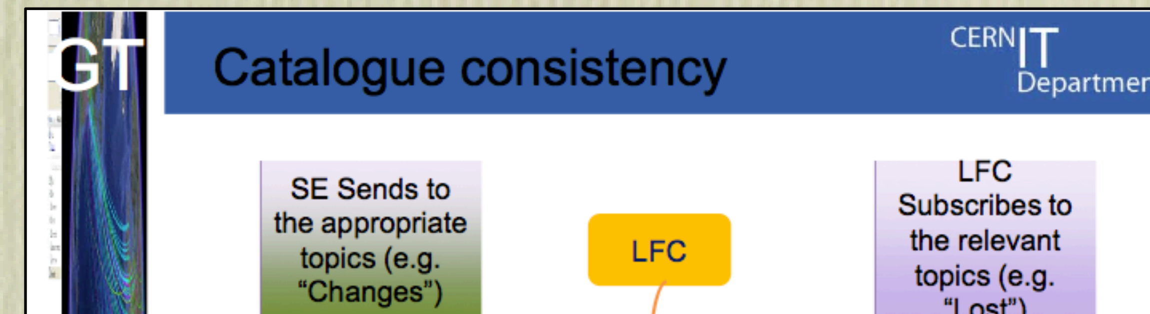
Amsterdam Themes

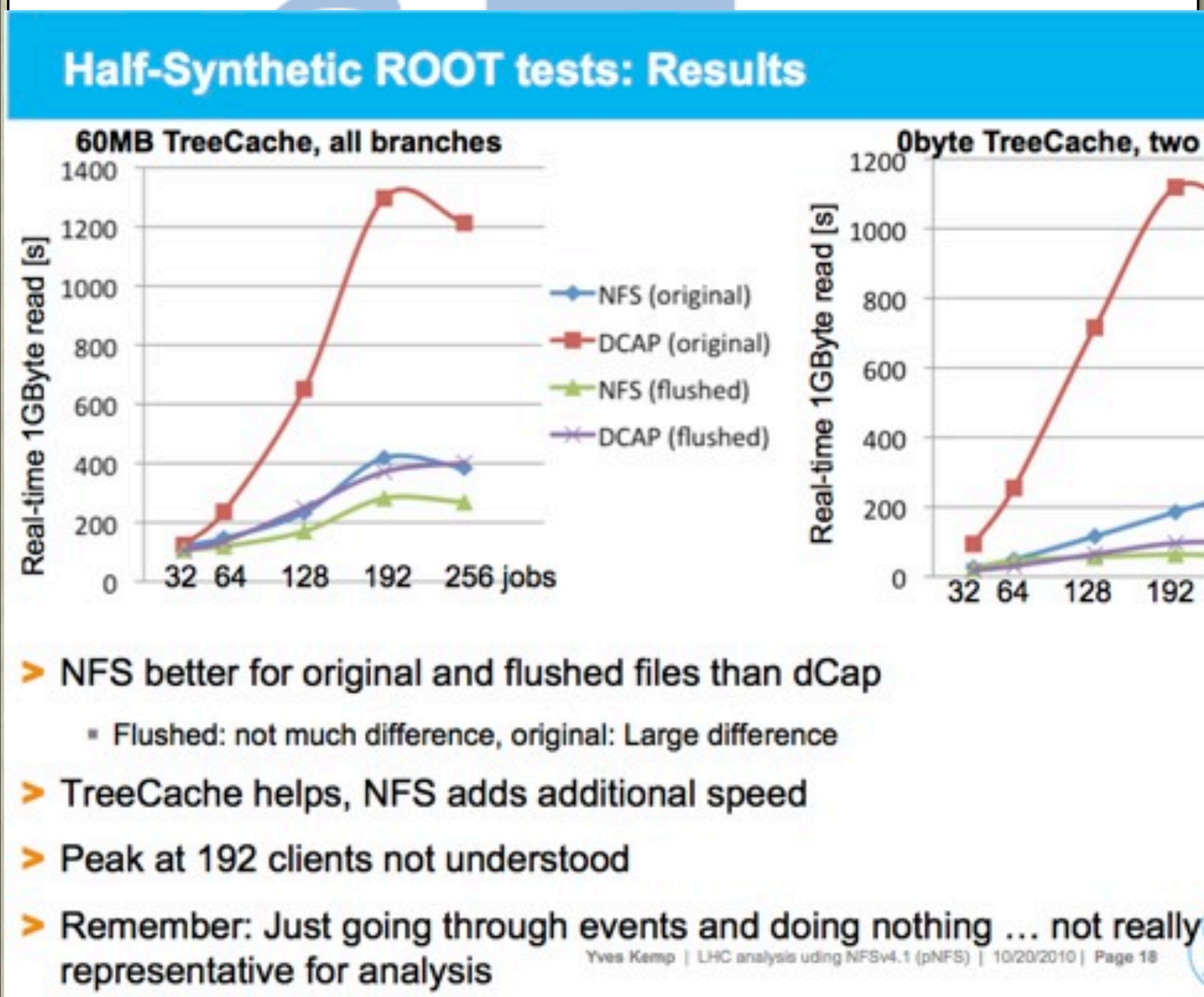


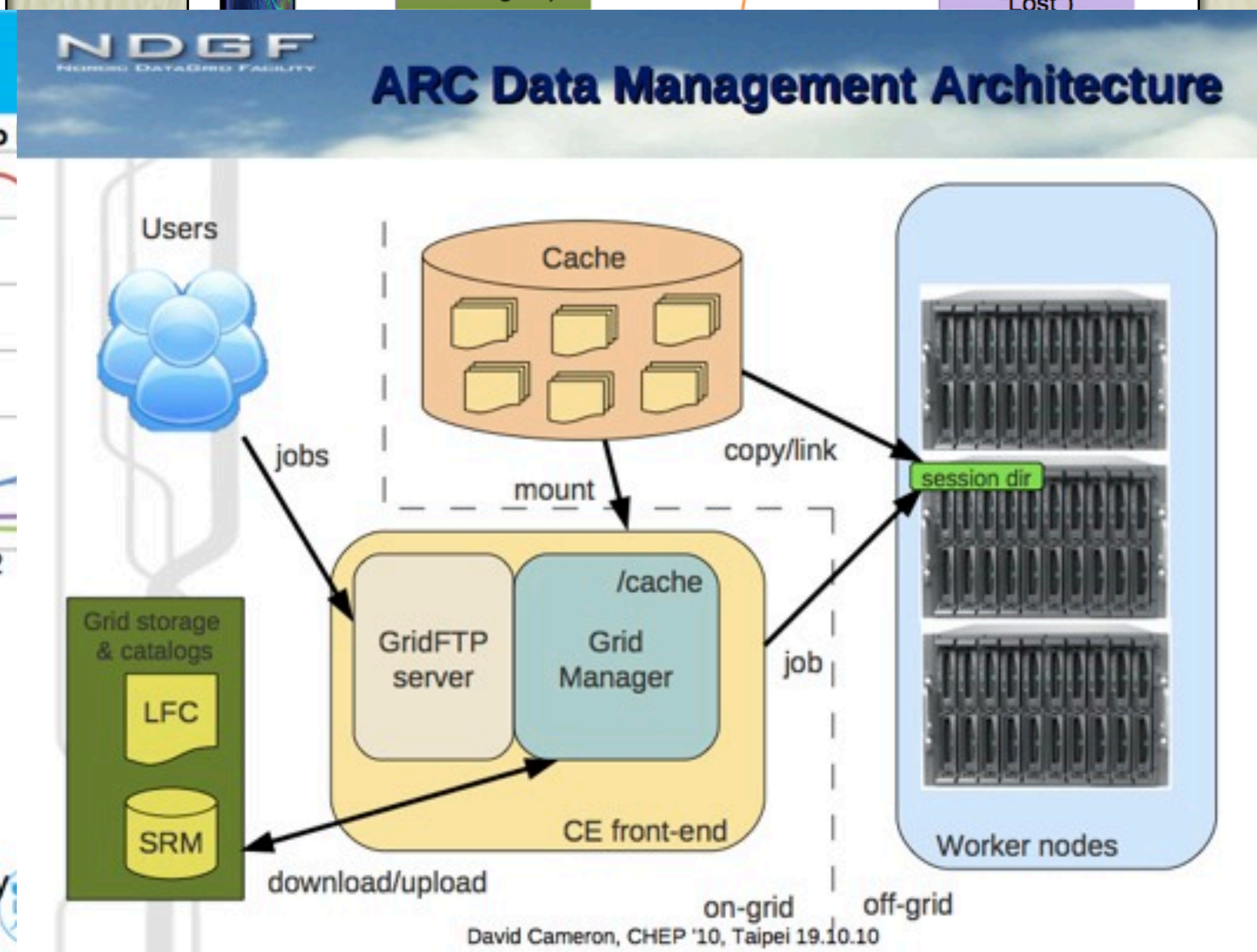


Amsterdam Themes

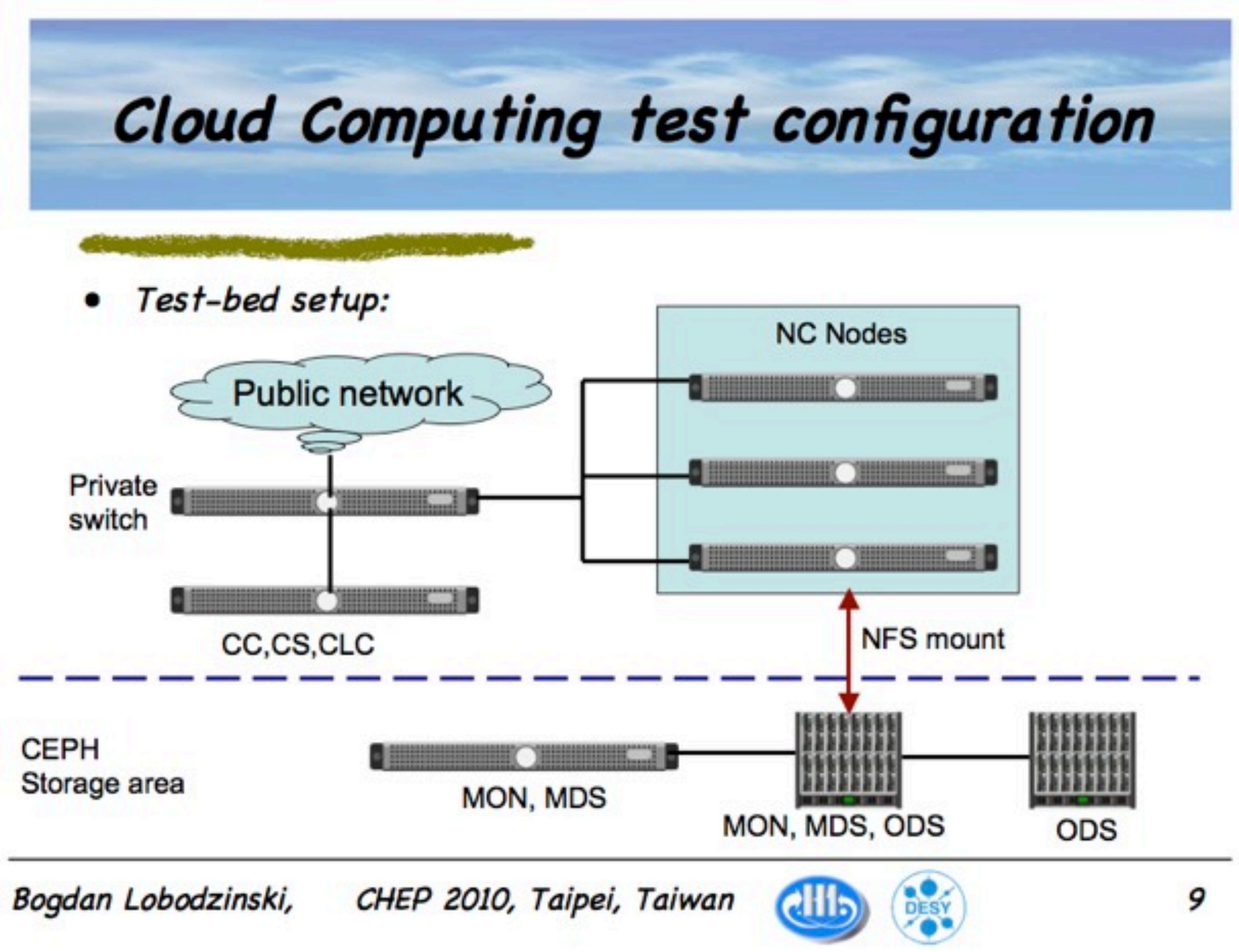




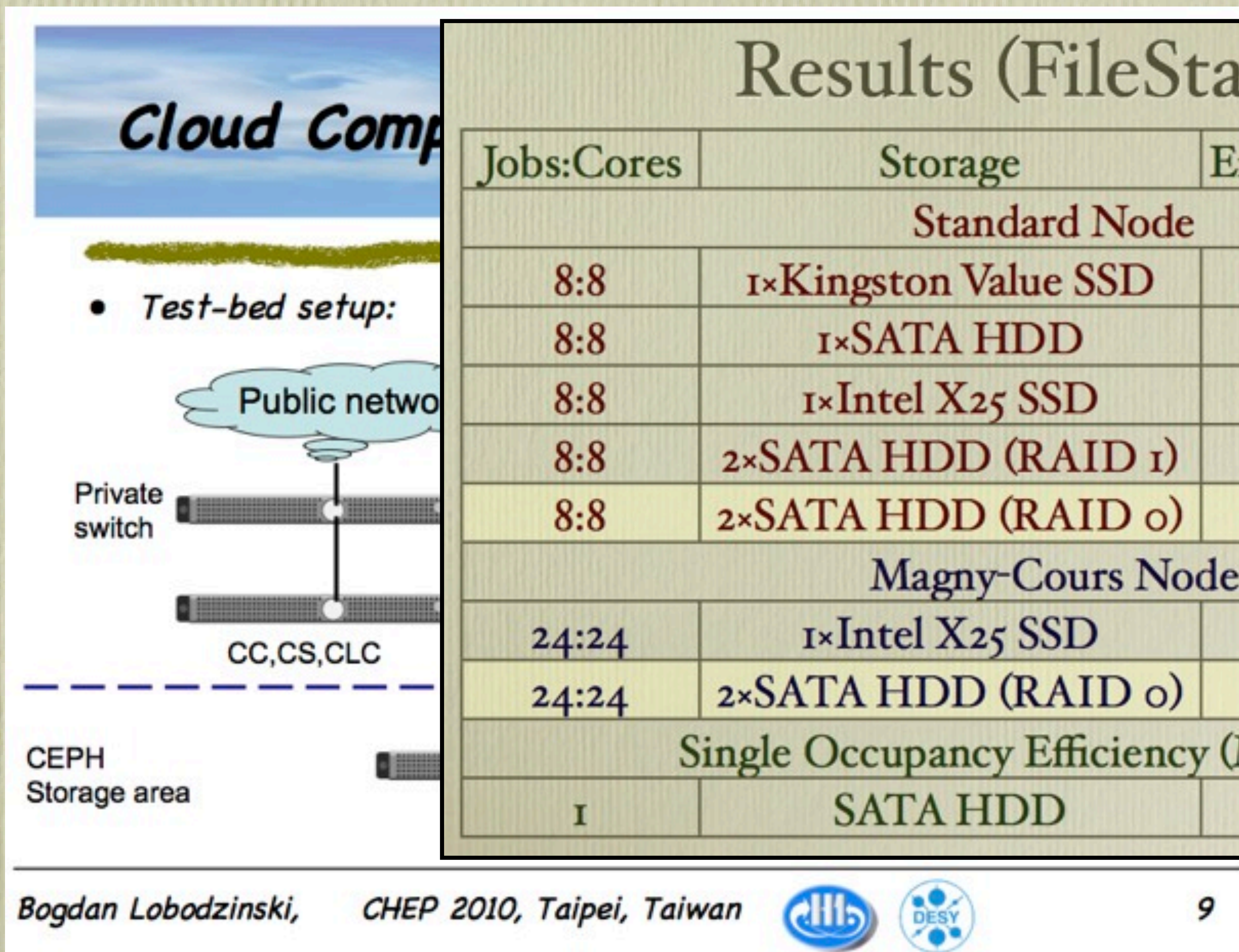




“New” themes


Cloud Computing test configuration
• Test-bed setup:
NC Nodes
Public network
Private switch
CC,CS,CLC
NFS mount
CEPH Storage area
MON, MDS
MON, MDS, ODS
ODS
Bogdan Lobodzinski, CHEP 2010, Taipei, Taiwan
DESY
9

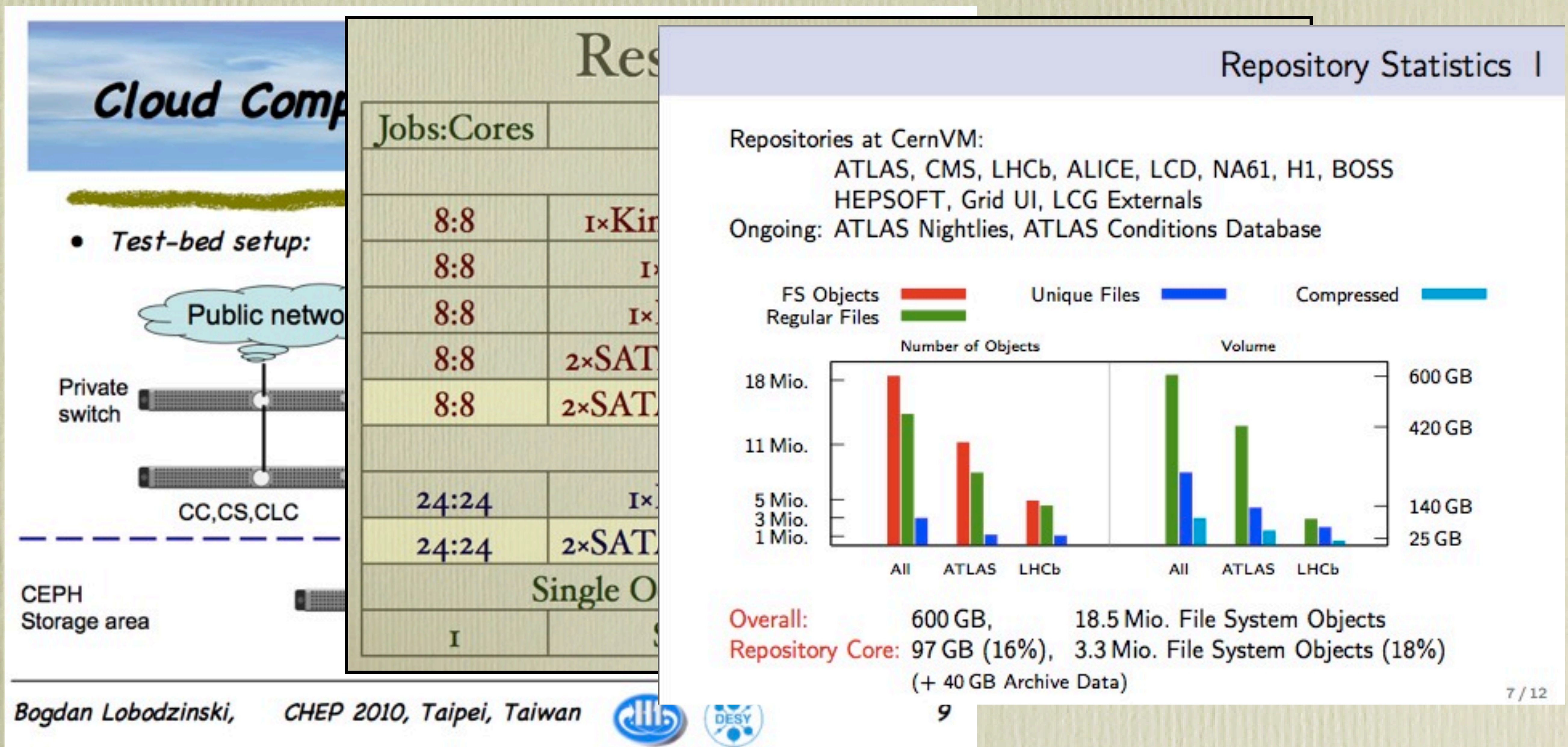
“New” themes



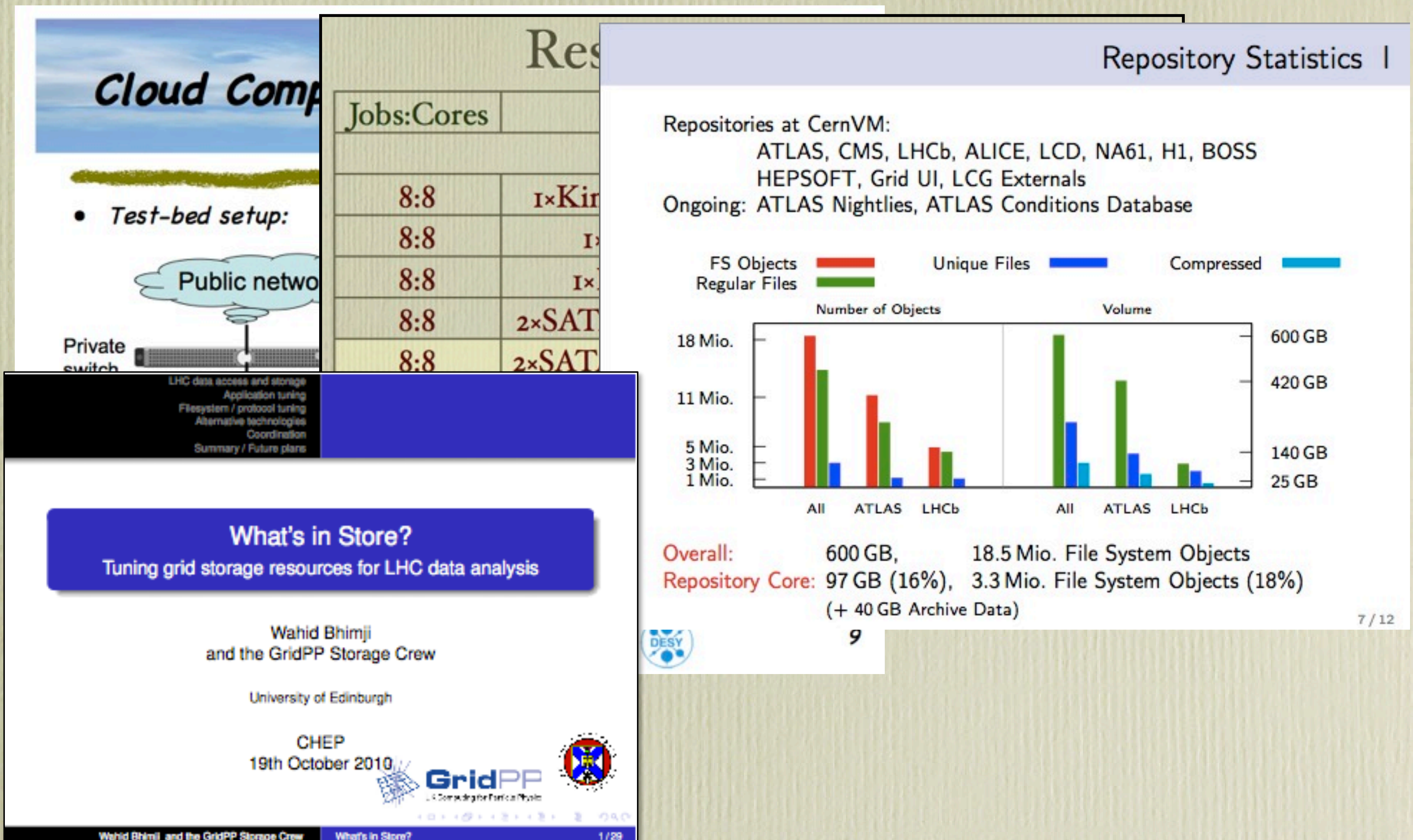
Results (FileStager)

Jobs:Cores	Storage	Efficiency	Throughput
Standard Node			
8:8	1×Kingston Value SSD	60%	4.5
8:8	1×SATA HDD	75%	5.5
8:8	1×Intel X25 SSD	80%	6
8:8	2×SATA HDD (RAID 1)	83%	6.6
8:8	2×SATA HDD (RAID 0)	90%	7
Magny-Cours Node			
24:24	1×Intel X25 SSD	50%	12
24:24	2×SATA HDD (RAID 0)	86%	21
Single Occupancy Efficiency (Measured)			
1	SATA HDD	90%	0.9

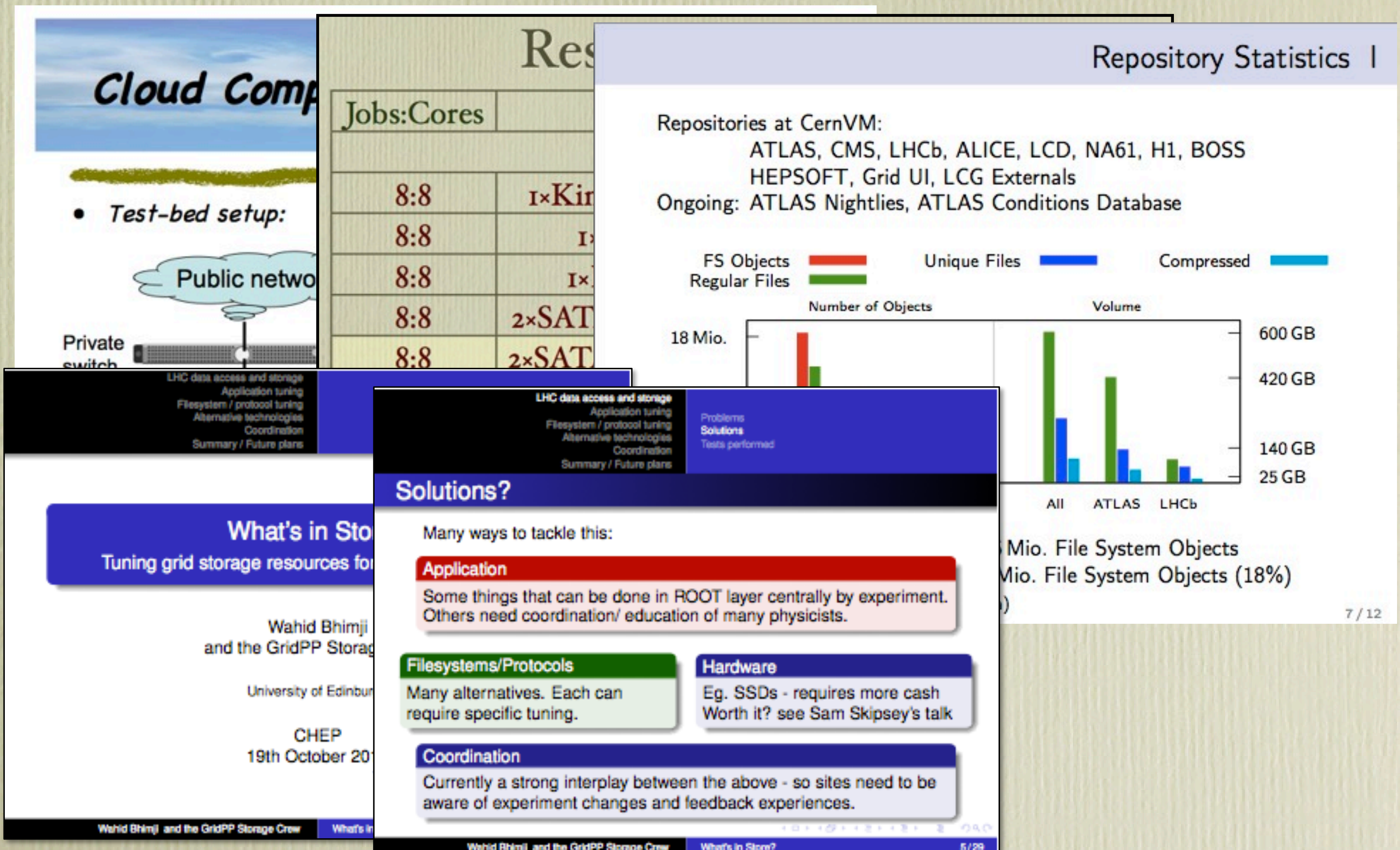
“New” themes


Cloud Comp
Test-bed setup:
Public netwo
Private
switch
CC,CS,CLC
CEPH
Storage area
Bogdan Lobodzinski, CHEP 2010, Taipei, Taiwan
9
Res
Jobs:Cores
8:8
1×Kir
8:8
8:8
8:8
2×SAT
8:8
2×SAT
24:24
24:24
2×SAT
Single O
1
Repository Statistics I
Repositories at CernVM:
ATLAS, CMS, LHCb, ALICE, LCD, NA61, H1, BOSS
HEPSOFT, Grid UI, LCG Externals
Ongoing: ATLAS Nightlies, ATLAS Conditions Database
FS Objects
Regular Files
Unique Files
Compressed
Number of Objects
Volume
18 Mio.
11 Mio.
5 Mio.
3 Mio.
1 Mio.
600 GB
420 GB
140 GB
25 GB
All
ATLAS
LHCb
All
ATLAS
LHCb
Overall: 600 GB, 18.5 Mio. File System Objects
Repository Core: 97 GB (16%), 3.3 Mio. File System Objects (18%)
(+ 40 GB Archive Data)
7 / 12

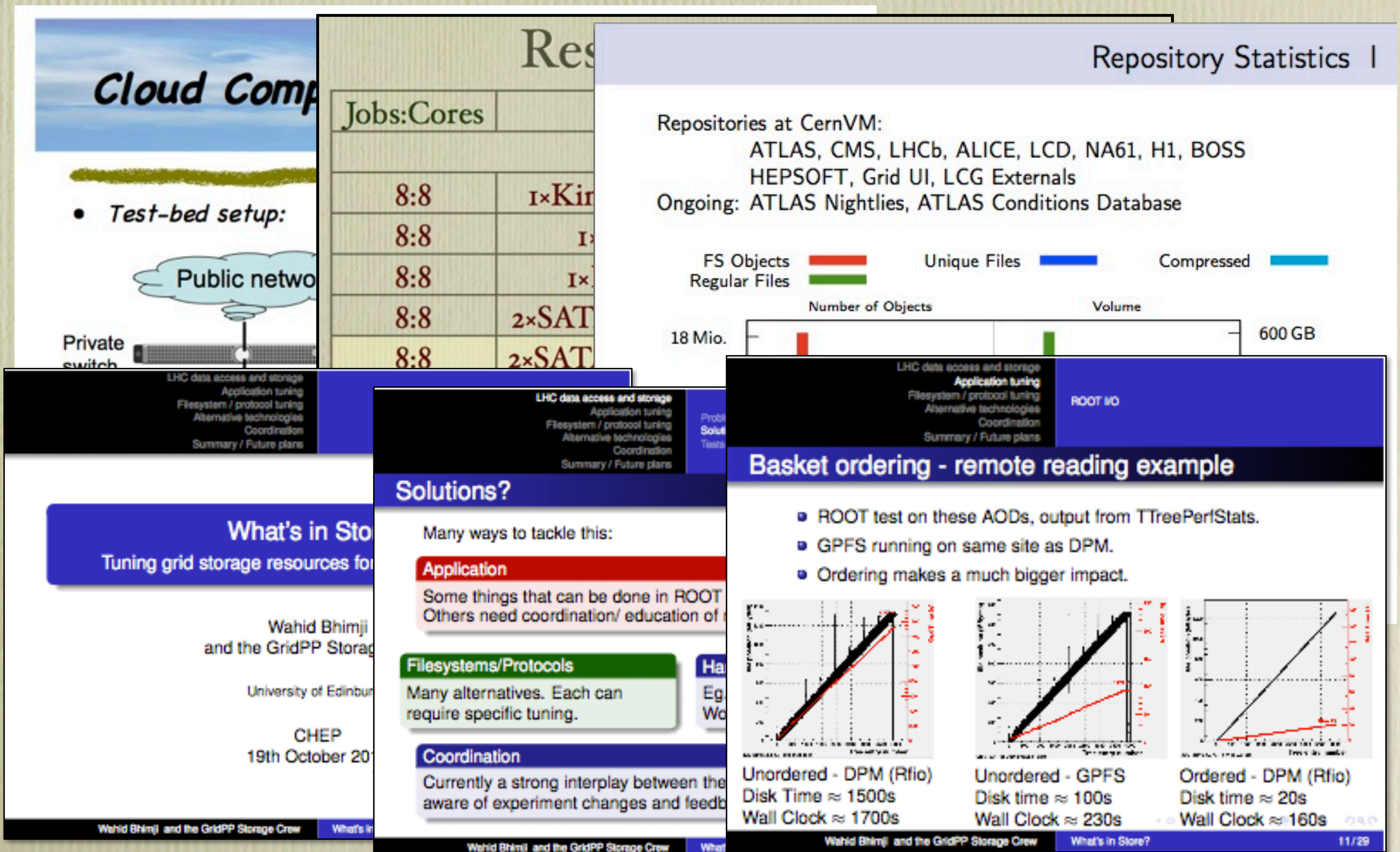
“New” themes


Cloud Comp
Test-bed setup:
Public netwo
Private switch
Res
Jobs:Cores
8:8
8:8
8:8
8:8
8:8
1×Ki
2×SAT
2×SAT
Repository Statistics I
Repositories at CernVM:
ATLAS, CMS, LHCb, ALICE, LCD, NA61, H1, BOSS
HEPSOFT, Grid UI, LCG Externals
Ongoing: ATLAS Nightlies, ATLAS Conditions Database
FS Objects
Regular Files
Unique Files
Compressed
Number of Objects
Volume
18 Mio.
11 Mio.
5 Mio.
3 Mio.
1 Mio.
600 GB
420 GB
140 GB
25 GB
All
ATLAS
LHCb
All
ATLAS
LHCb
Overall: 600 GB, 18.5 Mio. File System Objects
Repository Core: 97 GB (16%), 3.3 Mio. File System Objects (18%)
(+ 40 GB Archive Data)
7 / 12
DESY
9
LHC data access and storage
Application tuning
Filesystem / protocol tuning
Alternative technologies
Coordination
Summary / Future plans
What's in Store?
Tuning grid storage resources for LHC data analysis
Wahid Bhimji
and the GridPP Storage Crew
University of Edinburgh
CHEP
19th October 2010
GridPP
Wahid Bhimji and the GridPP Storage Crew
What's in Store?
1 / 29

"New" themes


Cloud Comp
Test-bed setup:
Public netwo
Private switch
Res
Jobs:Cores
8:8
8:8
8:8
8:8
8:8
1×Ki
2×SAT
2×SAT
Repository Statistics I
Repositories at CernVM:
ATLAS, CMS, LHCb, ALICE, LCD, NA61, H1, BOSS
HEPSOFT, Grid UI, LCG Externals
Ongoing: ATLAS Nightlies, ATLAS Conditions Database
FS Objects
Regular Files
Unique Files
Compressed
Number of Objects
Volume
18 Mio.
600 GB
420 GB
140 GB
25 GB
All
ATLAS
LHCb
Mio. File System Objects
Mio. File System Objects (18%)
7 / 12
What's in Sto
Tuning grid storage resources fo
Wahid Bhimji
and the GridPP Storag
University of Edinbur
CHEP
19th October 20
Wahid Bhimji and the GridPP Storage Crew
Solutions?
Many ways to tackle this:
Application
Some things that can be done in ROOT layer centrally by experiment.
Others need coordination/ education of many physicists.
Filesystems/Protocols
Many alternatives. Each can require specific tuning.
Hardware
Eg. SSDs - requires more cash
Worth it? see Sam Skipsey's talk
Coordination
Currently a strong interplay between the above - so sites need to be aware of experiment changes and feedback experiences.
Wahid Bhimji and the GridPP Storage Crew
What's in Store?
5/29

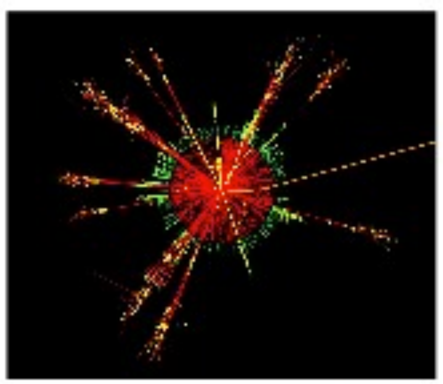
“New” themes


Cloud Comp
Test-bed setup:
Public netwo
Private
Res
Jobs:Cores
8:8
8:8
8:8
8:8
8:8
Repository Statistics I
Repositories at CernVM:
ATLAS, CMS, LHCb, ALICE, LCD, NA61, H1, BOSS
HEPSOFT, Grid UI, LCG Externals
Ongoing: ATLAS Nightlies, ATLAS Conditions Database
FS Objects
Regular Files
Unique Files
Compressed
Number of Objects
Volume
18 Mio.
600 GB
LHC data access and storage
Application tuning
Filesystem / protocol tuning
Alternative technologies
Coordination
Summary / Future plans
What's in Sto
Tuning grid storage resources fo
Wahid Bhimji
and the GridPP Storag
University of Edinbur
CHEP
19th October 20
Wahid Bhimji and the GridPP Storage Crew
Solutions?
Many ways to tackle this:
Application
Some things that can be done in ROOT
Others need coordination/ education of
Filesystems/Protocols
Many alternatives. Each can
require specific tuning.
Coordination
Currently a strong interplay between the
aware of experiment changes and feedb
ROOT I/O
Basket ordering - remote reading example
ROOT test on these AODs, output from TTreePerfStats.
GPFS running on same site as DPM.
Ordering makes a much bigger impact.
Unordered - DPM (Rfio)
Disk Time ≈ 1500s
Wall Clock ≈ 1700s
Unordered - GPFS
Disk time ≈ 100s
Wall Clock ≈ 230s
Ordered - DPM (Rfio)
Disk time ≈ 20s
Wall Clock ≈ 160s
What's in Store?
11/29

Multi-Core, GPGPU and all that Jazz

CHEP 2010

How to harness the performance potential of current Multi-Core CPUs and GPUs



Sverre Jarp

CERN openlab

IT Dept.

CERN

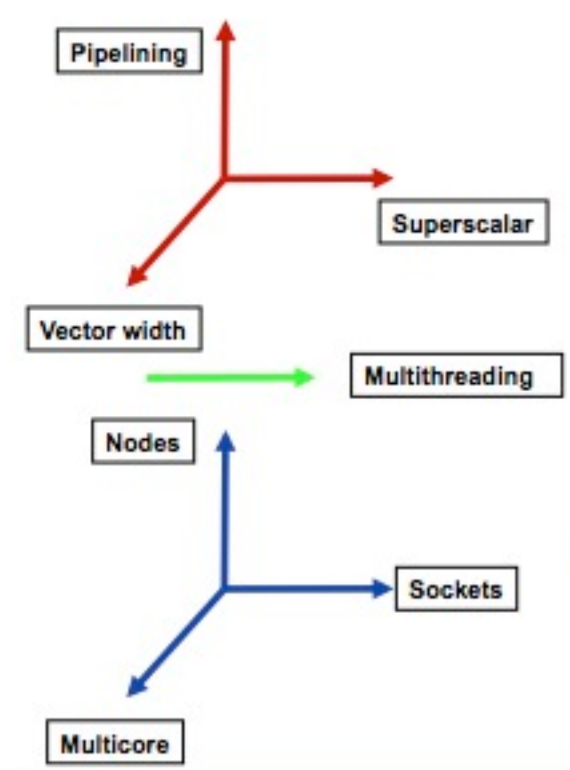
Taipei, Monday 18 October 2010

CHEP 2010, Taipei

Today: Seven dimensions of multiplicative performance

CERN openlab

- **First three dimensions:**
  - Pipelined execution units
  - Large superscalar design
  - Wide vector width (SIMD)
- **Next dimension is a "pseudo" dimension:**
  - Hardware multithreading
- **Last three dimensions:**
  - Multiple cores
  - Multiple sockets
  - Multiple compute nodes



SIMD = Single Instruction Multiple Data

Sverre Jarp - CERN

CPU scaling

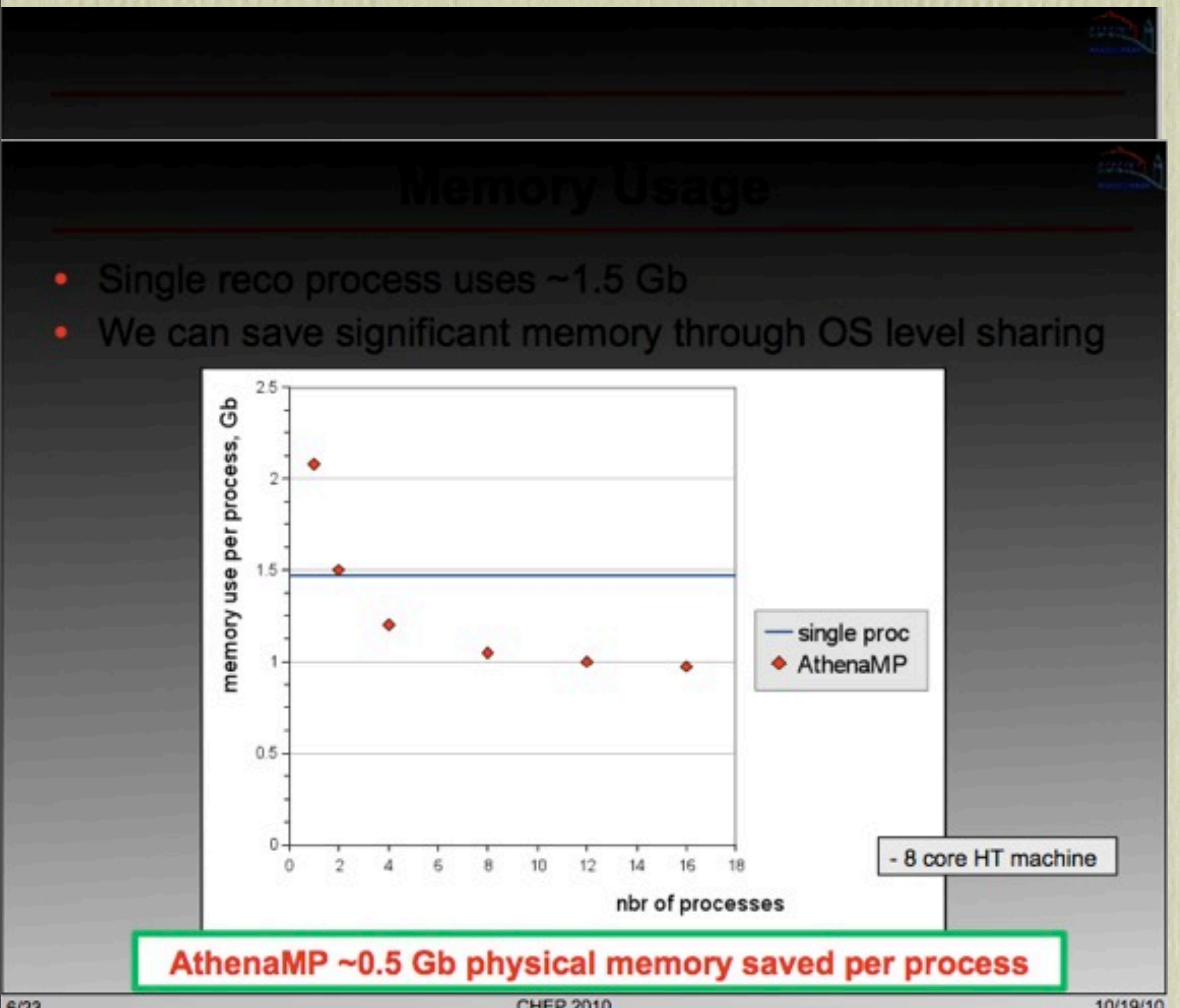
Parallelizing Atlas Reconstruction and Simulation:
Issues and Optimization Solutions for Scaling on Multi- and Many-CPU Platforms

Charles Leggett[1]

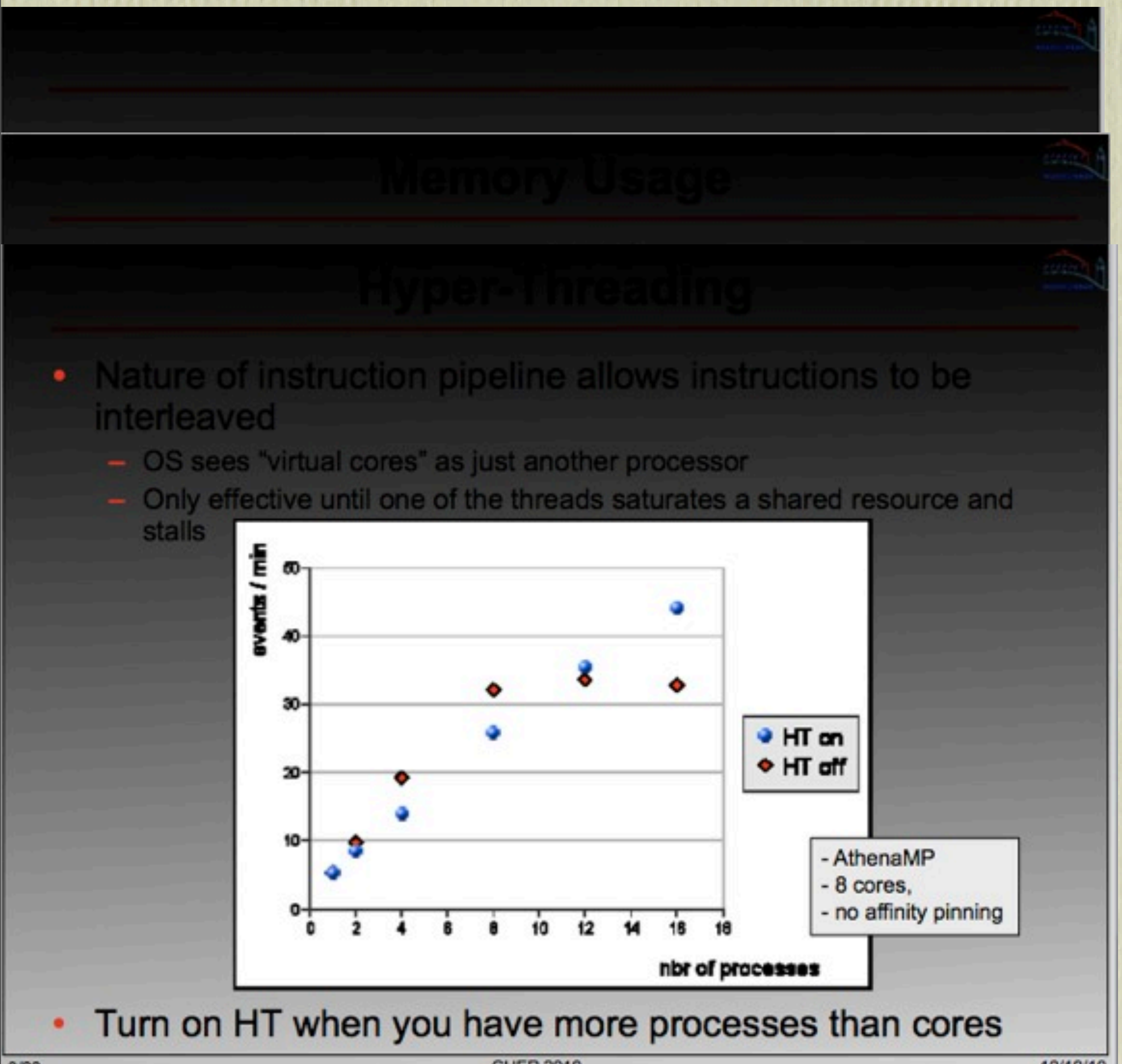
Sebastien Binet[2], Paolo Calafiura[1], Keith Jackson[1], David Levinthal[3], Mous Tatarkhanov[1], Yushu Yao[1]

[1]Lawrence Berkeley National Lab, [2]LAL, [3]Intel

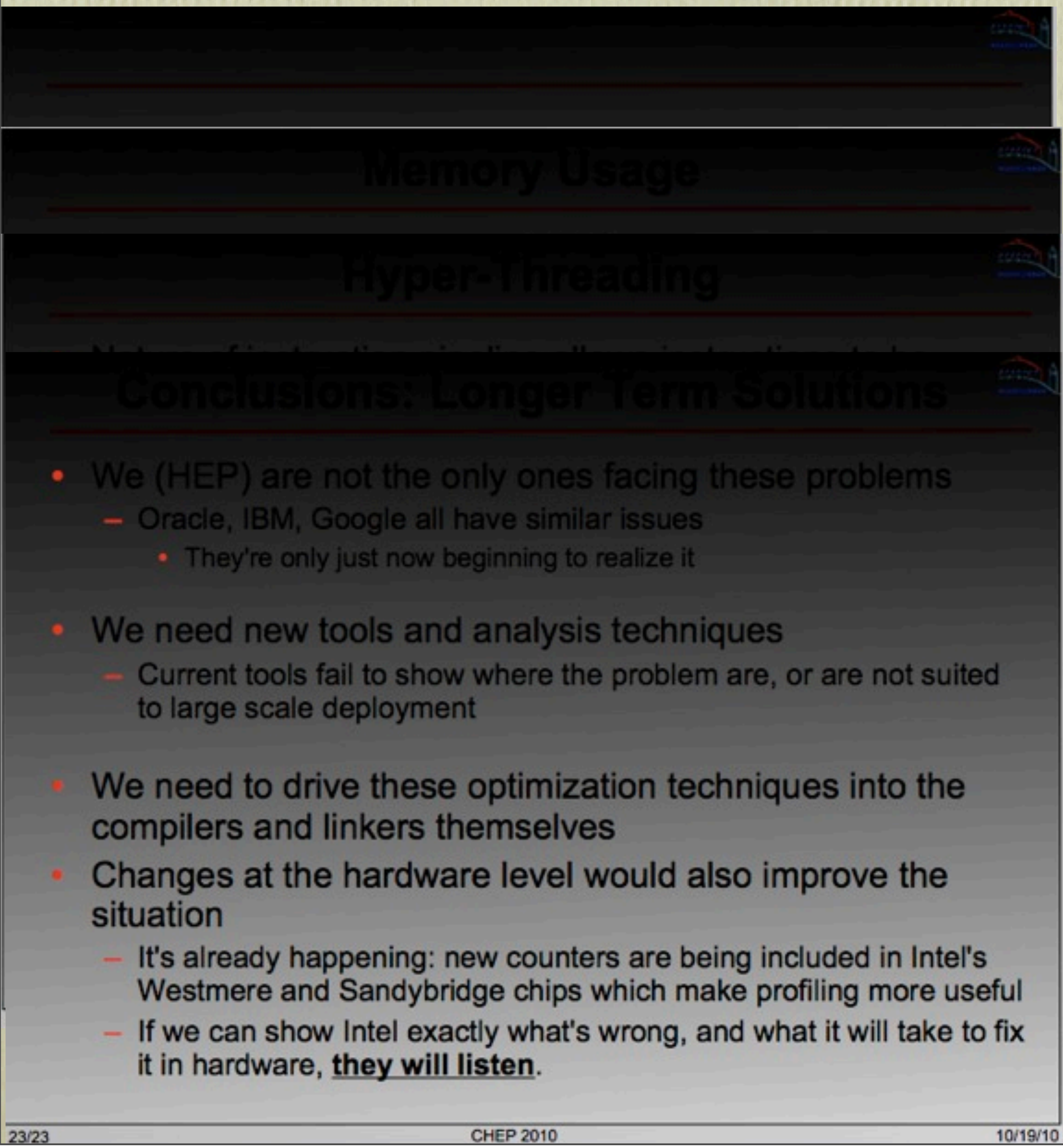
CPU scaling


Memory Usage
Single reco process uses ~1.5 Gb
We can save significant memory through OS level sharing
memory use per process, Gb
2.5
2
1.5
1
0.5
0
0
2
4
6
8
10
12
14
16
18
nbr of processes
single proc
AthenaMP
- 8 core HT machine
AthenaMP ~0.5 Gb physical memory saved per process
6/23
CHEP 2010
10/19/10

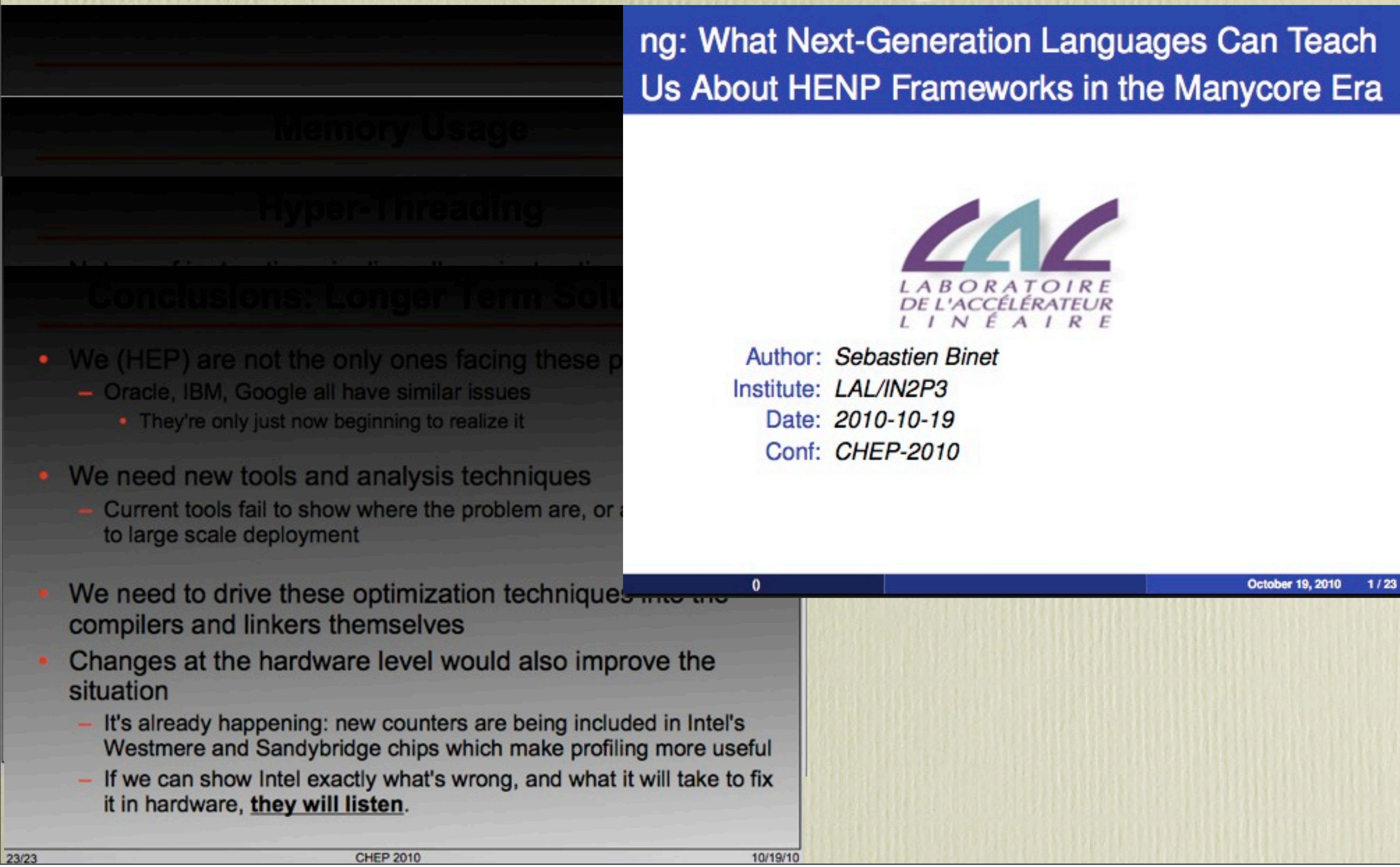
CPU scaling


Memory Usage
Hyper-Threading
Nature of instruction pipeline allows instructions to be interleaved
– OS sees "virtual cores" as just another processor
– Only effective until one of the threads saturates a shared resource and stalls
events / min
nbr of processes
HT on
HT off
- AthenaMP
- 8 cores,
- no affinity pinning
Turn on HT when you have more processes than cores
8/23
CHEP 2010
10/19/10

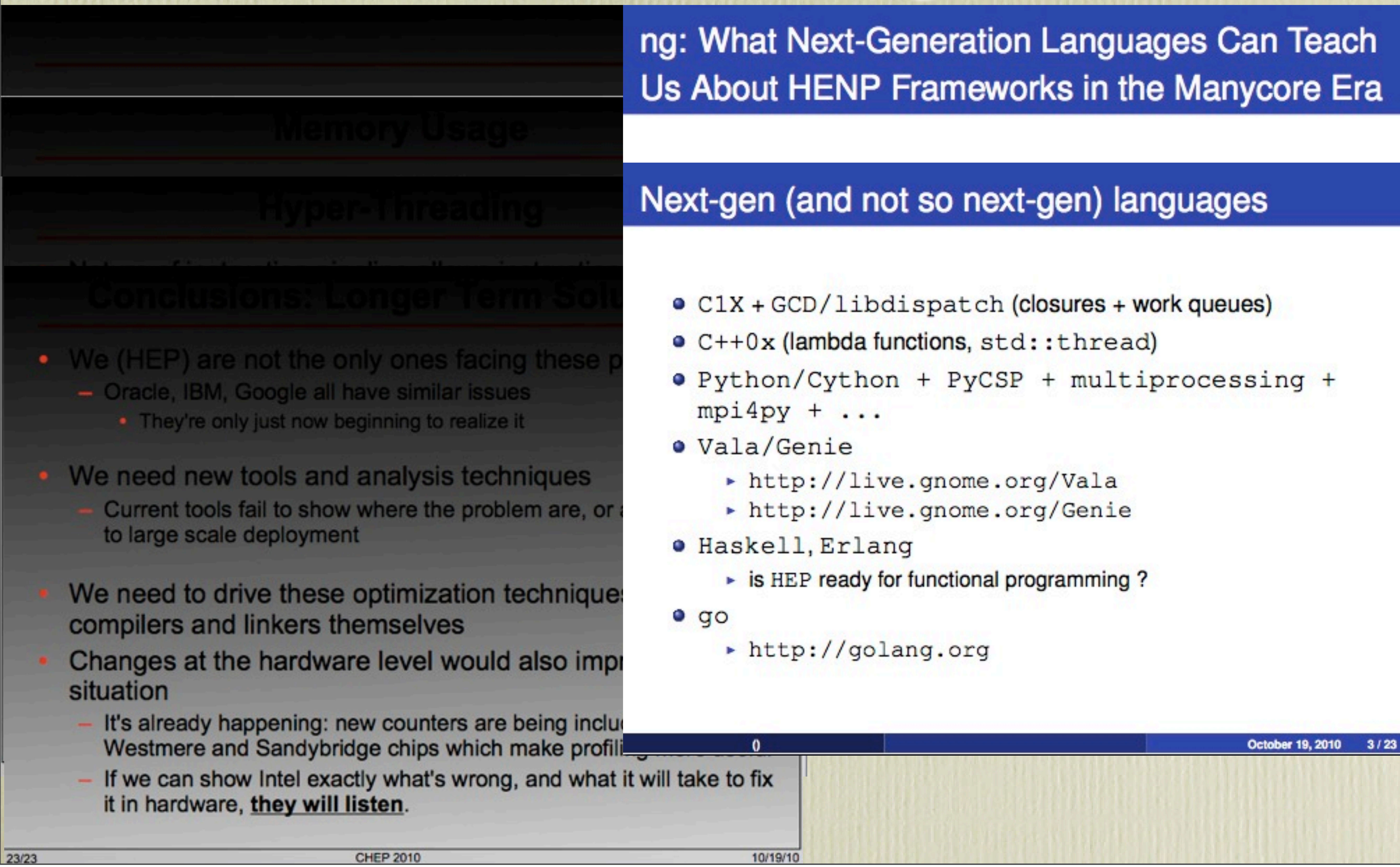
CPU scaling


Memory Usage
Hyper-Threading
Conclusions: Longer Term Solutions
We (HEP) are not the only ones facing these problems
Oracle, IBM, Google all have similar issues
They're only just now beginning to realize it
We need new tools and analysis techniques
Current tools fail to show where the problem are, or are not suited to large scale deployment
We need to drive these optimization techniques into the compilers and linkers themselves
Changes at the hardware level would also improve the situation
It's already happening: new counters are being included in Intel's Westmere and Sandybridge chips which make profiling more useful
If we can show Intel exactly what's wrong, and what it will take to fix it in hardware, they will listen.
23/23
CHEP 2010
10/19/10

CPU scaling

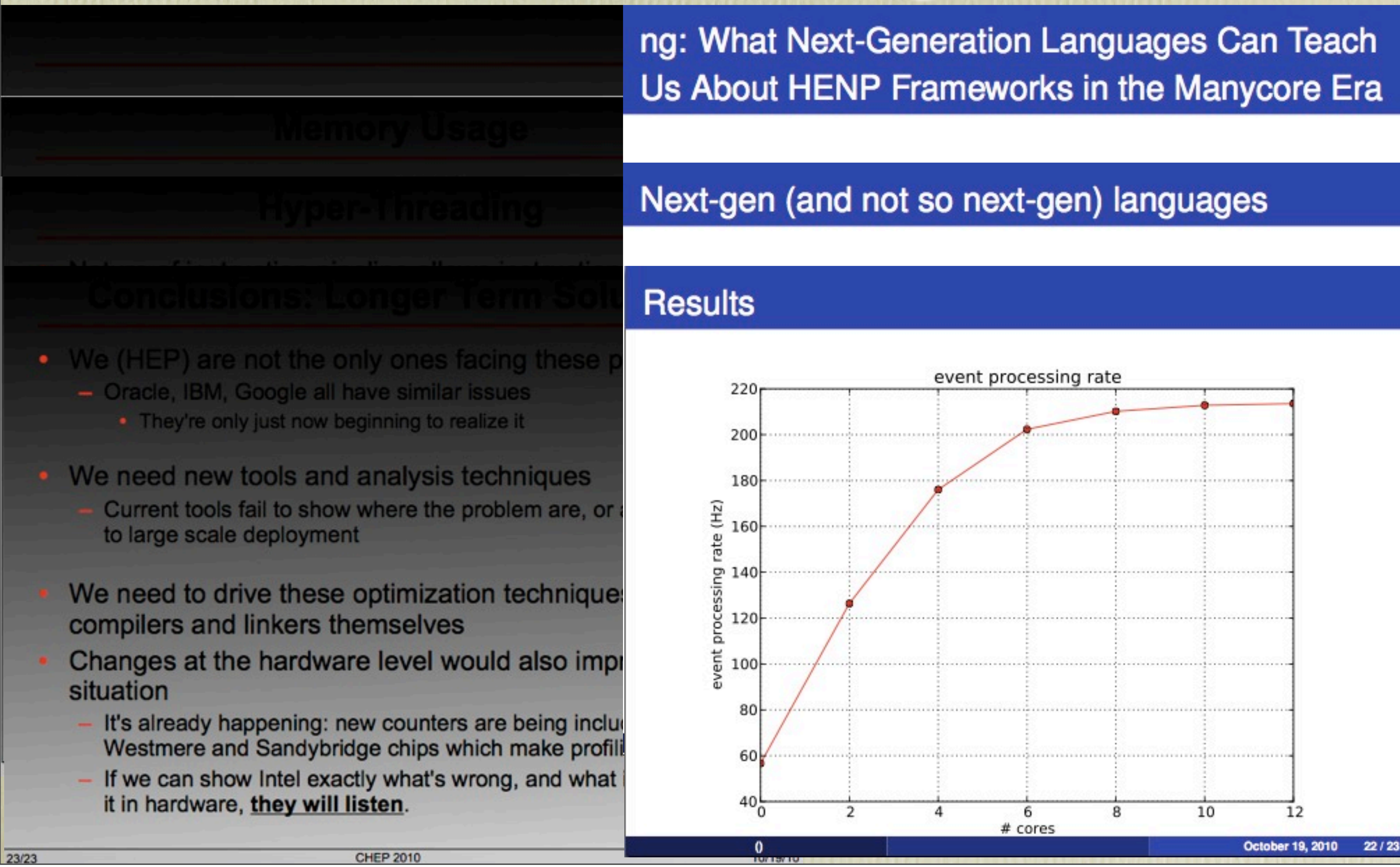

Memory Usage
Hyper-Threading
Conclusions: Longer Term Solu
We (HEP) are not the only ones facing these p
Oracle, IBM, Google all have similar issues
They're only just now beginning to realize it
We need new tools and analysis techniques
Current tools fail to show where the problem are, or
to large scale deployment
We need to drive these optimization techniques
compilers and linkers themselves
Changes at the hardware level would also improve the situation
It's already happening: new counters are being included in Intel's Westmere and Sandybridge chips which make profiling more useful
If we can show Intel exactly what's wrong, and what it will take to fix it in hardware, **they will listen**.
23/23
CHEP 2010
10/19/10
ng: What Next-Generation Languages Can Teach Us About HENP Frameworks in the Manycore Era
LABORATOIRE DE L'ACCÉLÉRATEUR LINÉAIRE
Author: Sebastien Binet
Institute: LAL/IN2P3
Date: 2010-10-19
Conf: CHEP-2010
October 19, 2010
1 / 23

CPU scaling

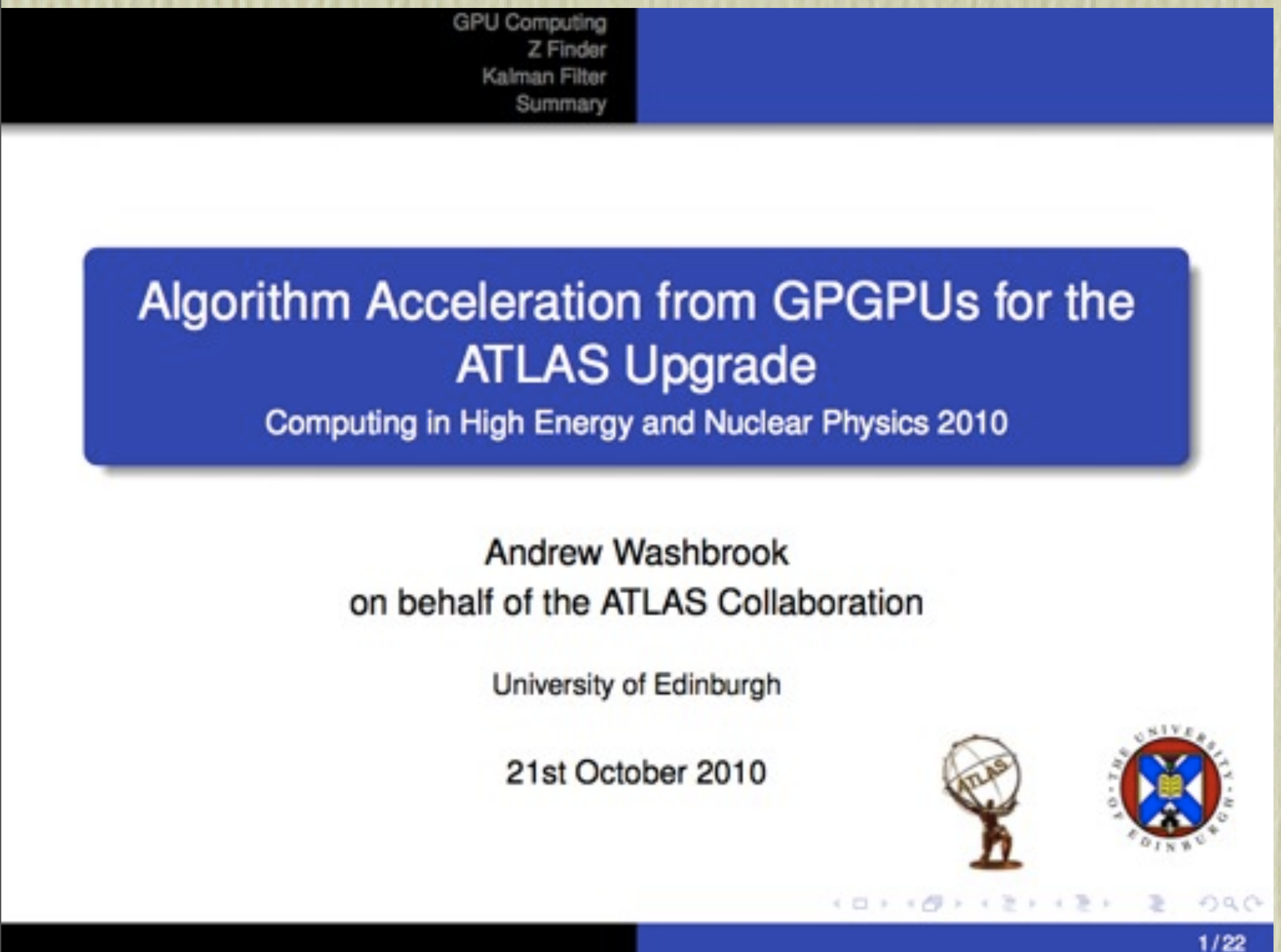


Memory Usage
Hyper-Threading
Conclusions: Longer Term Solu
We (HEP) are not the only ones facing these p
Oracle, IBM, Google all have similar issues
They're only just now beginning to realize it
We need new tools and analysis techniques
Current tools fail to show where the problem are, or
to large scale deployment
We need to drive these optimization technique
compilers and linkers themselves
Changes at the hardware level would also impr
situation
It's already happening: new counters are being inclu
Westmere and Sandybridge chips which make profili
If we can show Intel exactly what's wrong, and what it will take to fix it in hardware, they will listen.
23/23
CHEP 2010
10/19/10
ng: What Next-Generation Languages Can Teach Us About HENP Frameworks in the Manycore Era
Next-gen (and not so next-gen) languages
C1X + GCD/libdispatch (closures + work queues)
C++0x (lambda functions, std::thread)
Python/Cython + PyCSP + multiprocessing + mpi4py + ...
Vala/Genie
http://live.gnome.org/Vala
http://live.gnome.org/Genie
Haskell, Erlang
is HEP ready for functional programming ?
go
http://golang.org
()
October 19, 2010
3 / 23

CPU scaling


Memory Usage
Hyper-Threading
We (HEP) are not the only ones facing these p
Oracle, IBM, Google all have similar issues
They're only just now beginning to realize it
We need new tools and analysis techniques
Current tools fail to show where the problem are, or
to large scale deployment
We need to drive these optimization technique
compilers and linkers themselves
Changes at the hardware level would also impr
situation
It's already happening: new counters are being inclu
Westmere and Sandybridge chips which make profili
If we can show Intel exactly what's wrong, and what
it in hardware, they will listen.
23/23
CHEP 2010
ng: What Next-Generation Languages Can Teach
Us About HENP Frameworks in the Manycore Era
Next-gen (and not so next-gen) languages
Results
event processing rate
event processing rate (Hz)
cores
220
200
180
160
140
120
100
80
60
40
0
2
4
6
8
10
12
October 19, 2010
22 / 23

GPGPU use

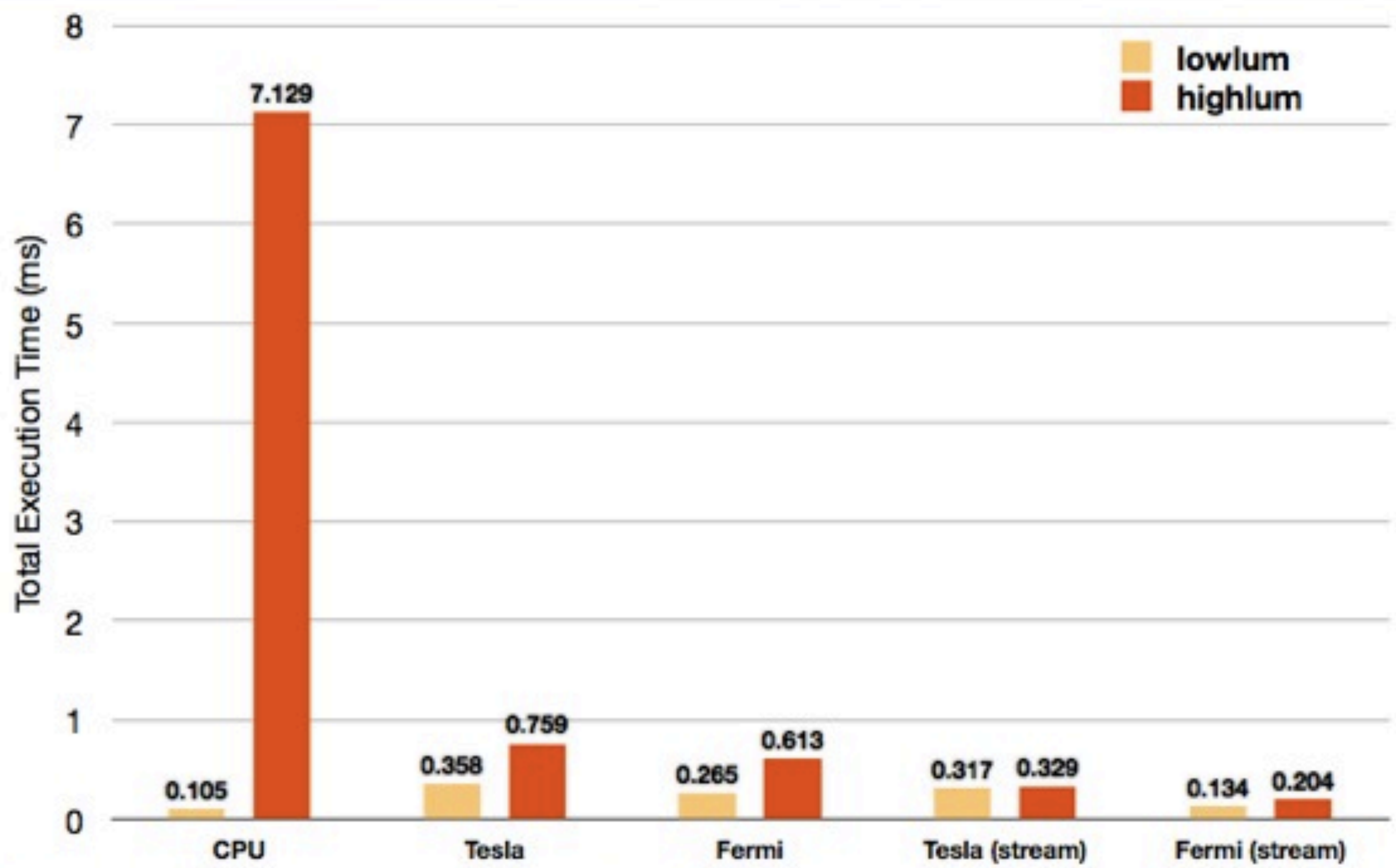

GPU Computing
Z Finder
Kalman Filter
Summary
Algorithm Acceleration from GPGPUs for the ATLAS Upgrade
Computing in High Energy and Nuclear Physics 2010
Andrew Washbrook
on behalf of the ATLAS Collaboration
University of Edinburgh
21st October 2010
1 / 22

GPGPU use

Algorith

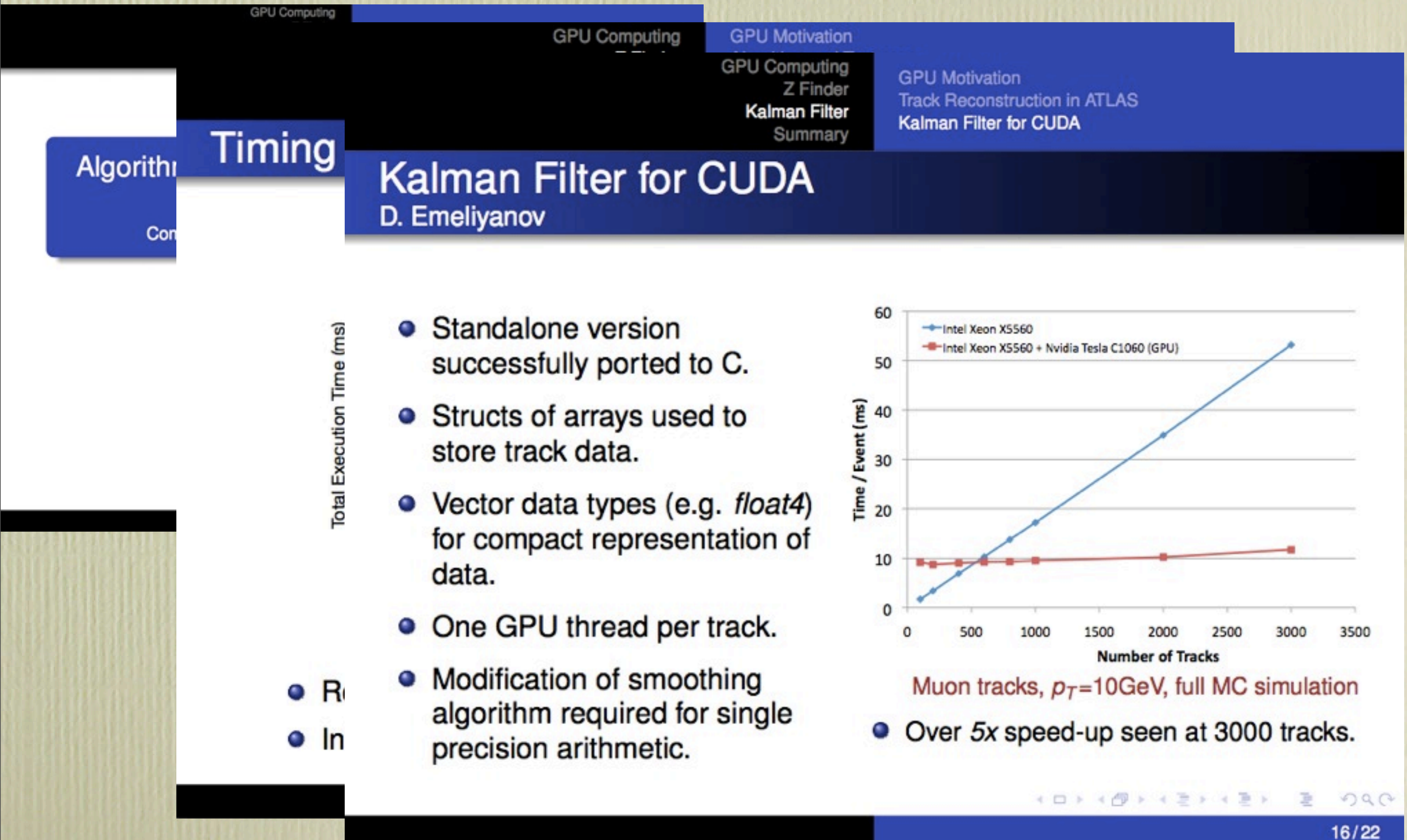
Con

Timing Results



- Results for spacepoint pairs show up to *35x* speed-up (Fermi).
- Initial results for spacepoint *triplets* also show speed-up.

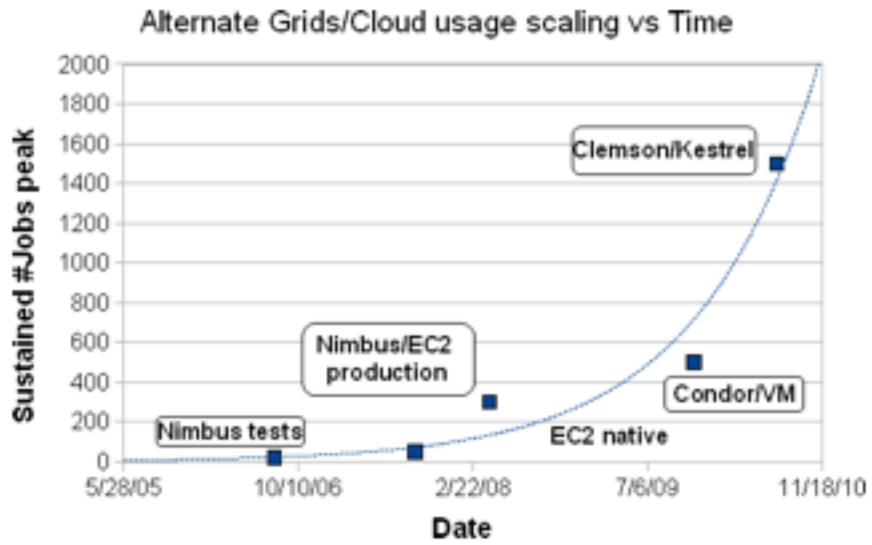
GPGPU use


GPU Computing
GPU Computing
GPU Motivation
Algorithr
Con
Timing
Total Execution Time (ms)
R
In
GPU Computing
Z Finder
Kalman Filter
Summary
GPU Motivation
Track Reconstruction in ATLAS
Kalman Filter for CUDA
Kalman Filter for CUDA
D. Emeliyanov
Standalone version successfully ported to C.
Structs of arrays used to store track data.
Vector data types (e.g. float4) for compact representation of data.
One GPU thread per track.
Modification of smoothing algorithm required for single precision arithmetic.
Intel Xeon X5560
Intel Xeon X5560 + Nvidia Tesla C1060 (GPU)
Time / Event (ms)
Number of Tracks
0
10
20
30
40
50
60
0
500
1000
1500
2000
2500
3000
3500
Muon tracks, pT=10GeV, full MC simulation
Over 5x speed-up seen at 3000 tracks.
16/22

Virtualisation and Clouds

- Virtual machines promising
  - but there are security and config implications
- Use of Clouds - works, but you still have to pay more (and aren't they just VMs in a datacenter?).

Testing scale & usage growth



Today, 1,000 jobs stable on Cloud or Hybrid ("virtualized Grids") possible (some challenges with stability / scalability at times)

10 k to 100 k jobs needed for STAR – Should be "routine" by beginning of 2011

Promising ... OSG ~ 13 M jobs at times … some way to go …

Testing scale & usage growth

Models – 10-100 job scale series

Model	Scale, Effic. & Scope	Observations	Key features & observations
Amazon/EC2 amazon webservices Xen	**Eff. > 99%** Medium instance *Raw simulation workflow (event generator)*	▪ IO rather inadequate for large scale efforts (5 MB/sec per node) ▪ Good for simulations and simple workflows: little "I", not that much "O" in IO ▪ Normal instances unlikely suitable for HPC/HTC or large data mining (did not test the HPC instance. Has anyone?)	▪ **Amazon has a concept of VM repository (needed)** ▪ **Amazon AAA rudimentary (lacking?)** ▪ **Amazon has a simple and competitive pricing model:** $0.09 / hour - 300 jobs, week long cost ~ $5,600. A year long CPU @ 100 jobs saturation ~ 79k$
Nimbus/EC2 NIMBUS Xen	**Eff. ~ 85% (1st) and ~ 97% (2nd)** *Raw simulation workflow (event generator)*	▪ Initial drop of efficiency due to batch system "inside" ▪ More "natural" to handle submission (Cloud / Virtualization behind, Grid-like in front) ▪ **First REAL usage for Physics**	▪ **Contextualization needed at startup (GK not known a-priori, batch need to know topology)** ▪ **OSG stack inside, GK+WN - virtual space looks like "another OSG site" – good attempt to unify**
Virtual Org. Cluster (VOC) CLEMSON [ACAT 2010] KVM	**Eff 100% (?)** Did not lose a single job *Raw simulation workflow (event generator)*	▪ Looked like a "grid" site – submit job to it (pull or push), VM appeared on demand due to a "subscribe" mechanism for bacth ▪ Transfer limited to University / National lab line ▪ Required IPs (some scale problem; thought of overlaid network …)	▪ **Contextualization remains a site specific (time) overhead** ▪ **Interface is standard Grid – user is agnostic of technology** ▪ Performance improved by **caching image locally OR directing changes to local disk** – impossible on EC2. Final overhead < 1% (no NFS read)

BROOKHAVEN NATIONAL LABORATORY

Testing scale & usage growth

Models – 10-100 job scale series

Models – larger scales …

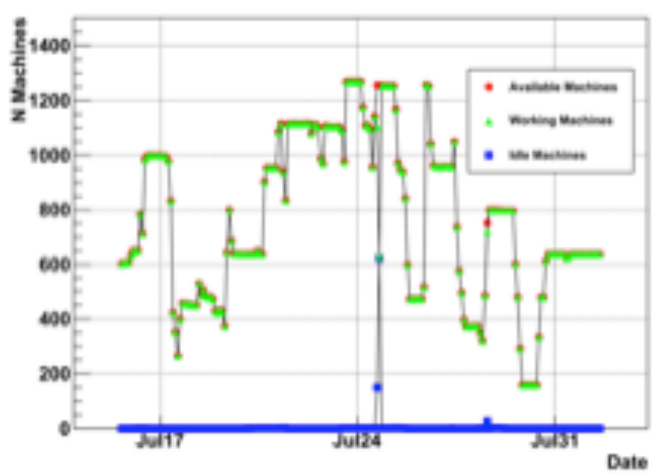
Model	Scale, Effic. & Scope	Observations	Key features & observations
Condor/VM	Scale 500+ jobs **Eff. unclear [80,85%]** *Raw simulation workflow (event generator)*	▪ 10% of the VM never started, 15% stopped (crashed), 5% net loss for long simulation jobs (VM reboot every 24 hours). Need to be able to extend lease? ▪ Data transfer mechanism was through common SE and separate from workflow ▪ *Results used for an analysis PoP*	▪ **Interface remains grid-like but only starts VM; No real job get "inside" – "pull model" (via cron)** ▪ **As many VMs as one wants: nearly no contextualization (apart from SE) reduce overheads on local staff, condor steering** ▪ **IP space is local – no connection to outside – transfer of data out hard but SE & Cloud may be the path …**
Clemson/Kestrel	Scale 1000 jobs **Eff. Unclear, cruising average 90%** *Full simulation with track reconstruction, detector efficiency correction using the full STAR framework*	▪ Job are "lost" if problem (communication, dead program, dying VM) ▪ FIRST time we mixed VMs from Clemson + CERN (true "Cloud" idea) ▪ Full database access as a service within the VM ▪ **Second massive usage of cloud targeting a conference, physics publication**	▪ **VM packaged everything** ▪ **Globus and MyProxy inside for transfer out – no real problems** ▪ **Job do not get "inside' – simple command trigger a script (external workload manager needed)** ▪ **Jobs start/stop were managed using a common Jabber-like client**

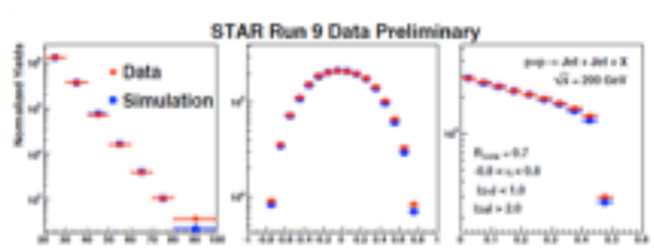
STAR framework btw is a single purpose framework for simu, reco, user analysis

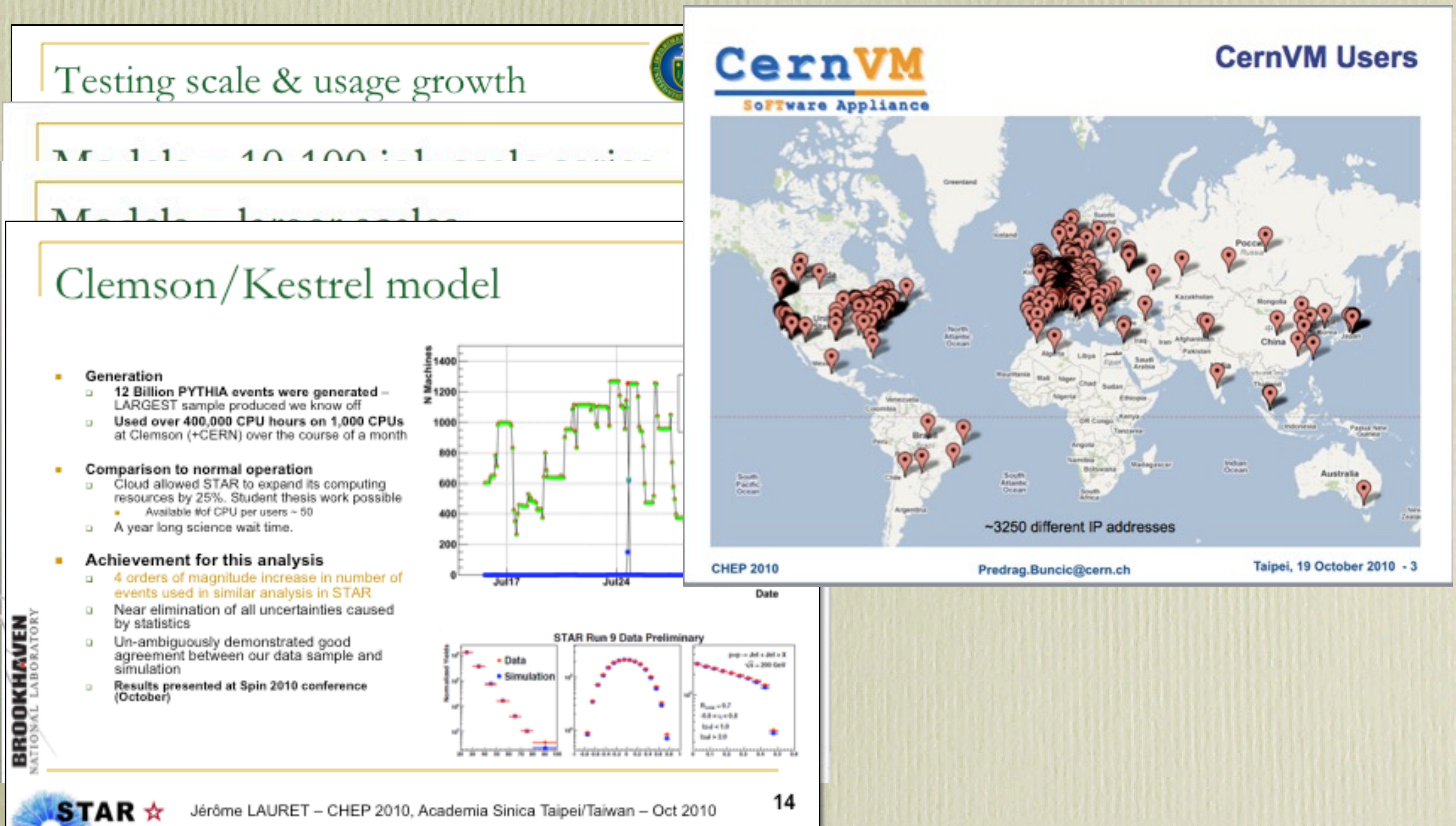
Testing scale & usage growth

Clemson/Kestrel model

- **Generation**
  - **12 Billion PYTHIA events were generated** – LARGEST sample produced we know off
  - **Used over 400,000 CPU hours on 1,000 CPUs** at Clemson (+CERN) over the course of a month
- **Comparison to normal operation**
  - Cloud allowed STAR to expand its computing resources by 25%. Student thesis work possible
    - Available #of CPU per users ~ 50
  - A year long science wait time.
- **Achievement for this analysis**
  - 4 orders of magnitude increase in number of events used in similar analysis in STAR
  - Near elimination of all uncertainties caused by statistics
  - Un-ambiguously demonstrated good agreement between our data sample and simulation
  - **Results presented at Spin 2010 conference (October)**



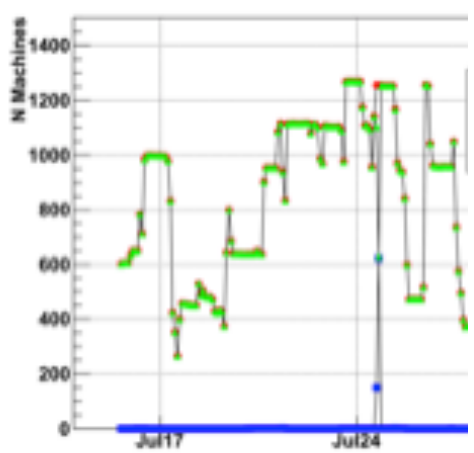


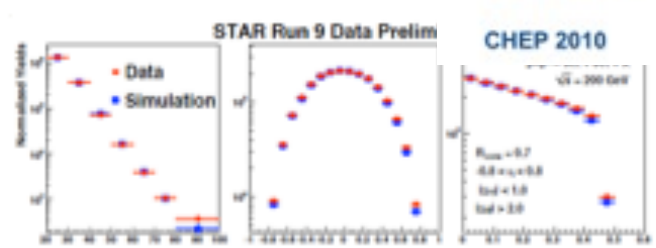

Testing scale & usage growth
Clemson/Kestrel model
Generation
12 Billion PYTHIA events were generated – LARGEST sample produced we know off
Used over 400,000 CPU hours on 1,000 CPUs at Clemson (+CERN) over the course of a month
Comparison to normal operation
Cloud allowed STAR to expand its computing resources by 25%. Student thesis work possible
Available #of CPU per users ~ 50
A year long science wait time.
Achievement for this analysis
4 orders of magnitude increase in number of events used in similar analysis in STAR
Near elimination of all uncertainties caused by statistics
Un-ambiguously demonstrated good agreement between our data sample and simulation
Results presented at Spin 2010 conference (October)
N Machines
Jul17
Jul24
Date
STAR Run 9 Data Preliminary
Data
Simulation
BROOKHAVEN NATIONAL LABORATORY
STAR
Jérôme LAURET – CHEP 2010, Academia Sinica Taipei/Taiwan – Oct 2010
14
CernVM
SoFTware Appliance
CernVM Users
~3250 different IP addresses
CHEP 2010
Predrag.Buncic@cern.ch
Taipei, 19 October 2010 - 3

Testing scale & usage growth

Clemson/Kestrel model

- **Generation**
  - **12 Billion PYTHIA events were generated** – LARGEST sample produced we know off
  - **Used over 400,000 CPU hours on 1,000 CPUs** at Clemson (+CERN) over the course of a month
- **Comparison to normal operation**
  - Cloud allowed STAR to expand its computing resources by 25%. Student thesis work possible
    - Available #of CPU per users ~ 50
  - A year long science wait time.
- **Achievement for this analysis**
  - 4 orders of magnitude increase in number of events used in similar analysis in STAR
  - Near elimination of all uncertainties caused by statistics
  - Un-ambiguously demonstrated good agreement between our data sample and simulation
  - **Results presented at Spin 2010 conference (October)**

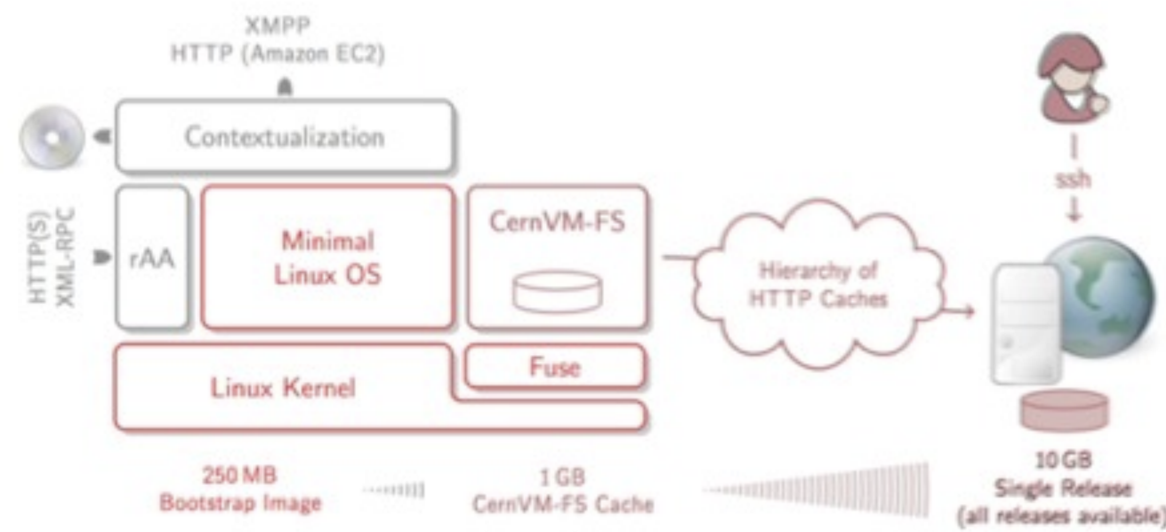




BROOKHAVEN NATIONAL LABORATORY

CernVM Users

CernVM Part #1: Minimal OS image

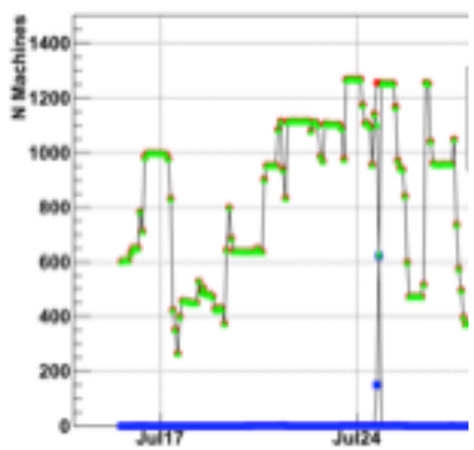


- Just enough OS to run LHC applications
- Built using commercial tool (rBuilder by rPath)
  - Top-down approach - starting from application and automatically discovering dependencies
- Small images (250MB), easy to move around

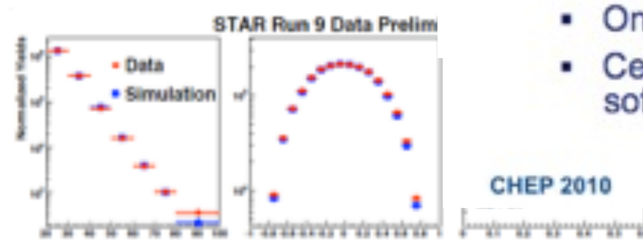
Testing scale & usage growth

Clemson/Kestrel model

- **Generation**
  - **12 Billion PYTHIA events were generated** – LARGEST sample produced we know off
  - **Used over 400,000 CPU hours on 1,000 CPUs** at Clemson (+CERN) over the course of a month
- **Comparison to normal operation**
  - Cloud allowed STAR to expand its computing resources by 25%. Student thesis work possible
    - Available #of CPU per users ~ 50
  - A year long science wait time.
- **Achievement for this analysis**
  - 4 orders of magnitude increase in number of events used in similar analysis in STAR
  - Near elimination of all uncertainties caused by statistics
  - Un-ambiguously demonstrated good agreement between our data sample and simulation
  - **Results presented at Spin 2010 conference (October)**



STAR Run 9 Data Prelim

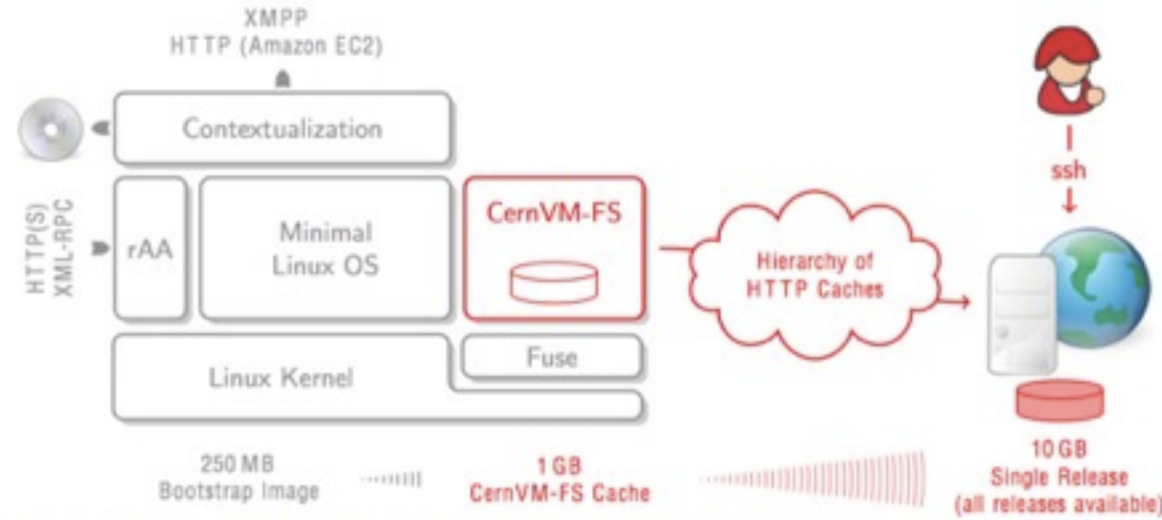


BROOKHAVEN NATIONAL LABORATORY

CernVM **CernVM Users**

CernVM **Part #1: Minimal OS image**

CernVM SoFTware Appliance **Part #2: CernVM-FS**



- Experiment software is changing frequently and we want to avoid need to frequently update, certify and redistribute VM images with every release
- Only a small fraction of software release is really used
- CernVM-FS: Read-only, network (HTTP) file system optimized for efficient software delivery. See: J.Blomer (PS06-5-434)

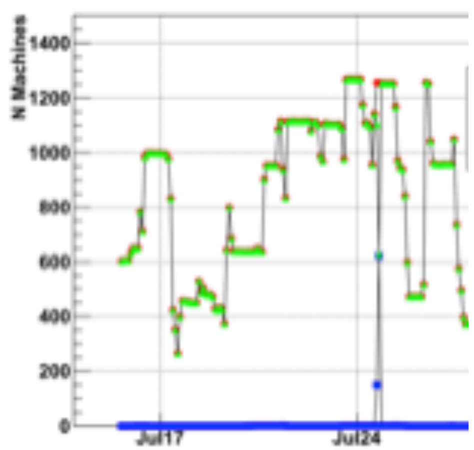
Testing scale & usage growth

Models – 10-100 job scale ...

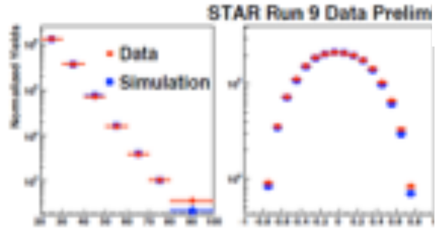
Models – larger scales

Clemson/Kestrel model

- **Generation**
  - **12 Billion PYTHIA events were generated** – LARGEST sample produced we know off
  - **Used over 400,000 CPU hours on 1,000 CPUs** at Clemson (+CERN) over the course of a month
- **Comparison to normal operation**
  - Cloud allowed STAR to expand its computing resources by 25%. Student thesis work possible
    - Available #of CPU per users ~ 50
  - A year long science wait time.
- **Achievement for this analysis**
  - 4 orders of magnitude increase in number of events used in similar analysis in STAR
  - Near elimination of all uncertainties caused by statistics
  - Un-ambiguously demonstrated good agreement between our data sample and simulation
  - **Results presented at Spin 2010 conference (October)**



STAR Run 9 Data Prelim



BROOKHAVEN NATIONAL LABORATORY

STAR ☆ Jérôme LAURET – CHEP 2010, Academia Sinica Taipei/Taiwan – Oct

CernVM Users

Part #1: Minimal OS image

Part #2: CernVM-FS

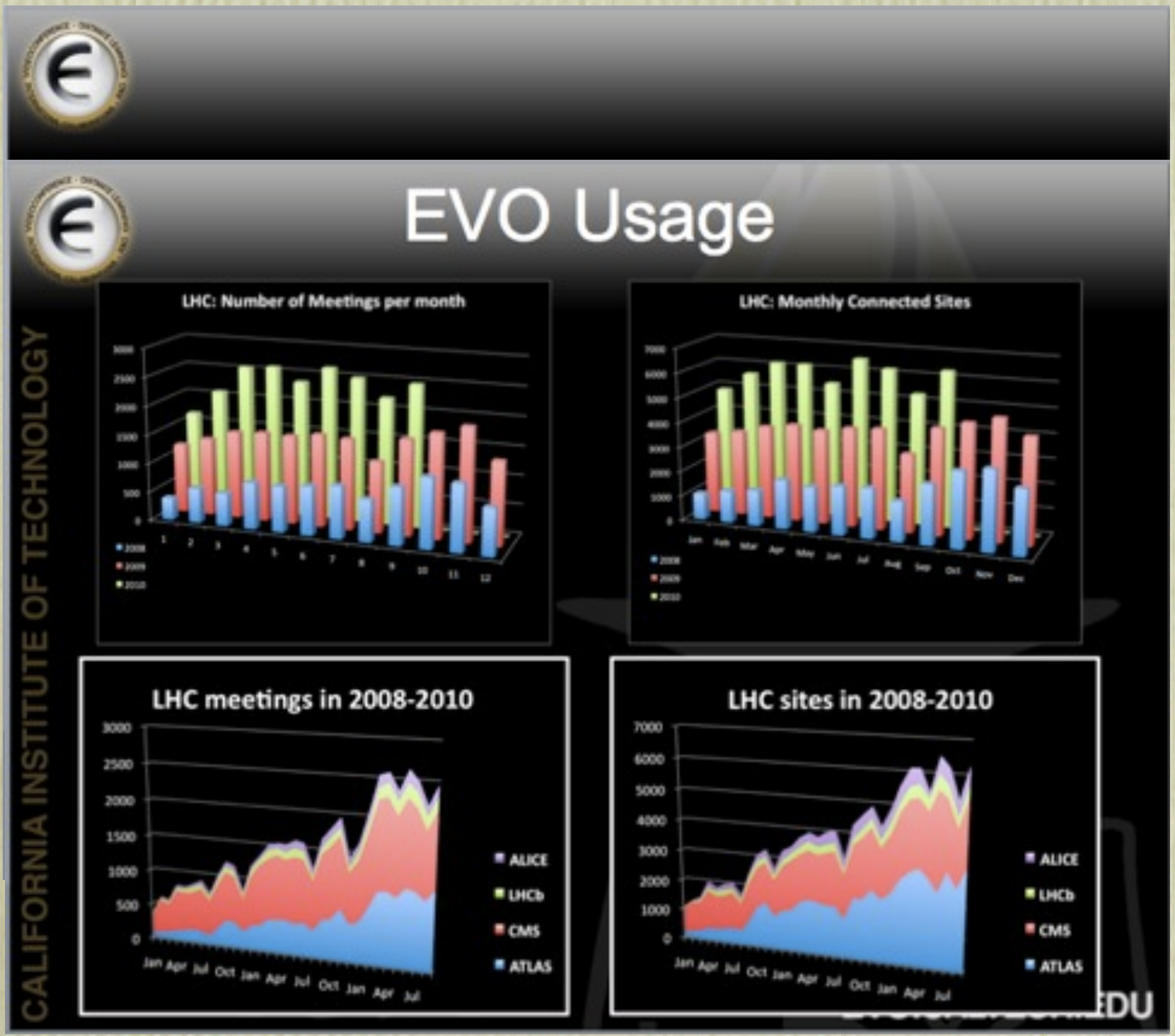
As easy as 1,2,3

1. Login to Web interface
2. Create user account
3. Select experiment, appliance flavor and preferences

CHEP 2010 Predrag.Buncic@cern.ch Taipei, 19 October 2010 - 18

Talking to people.

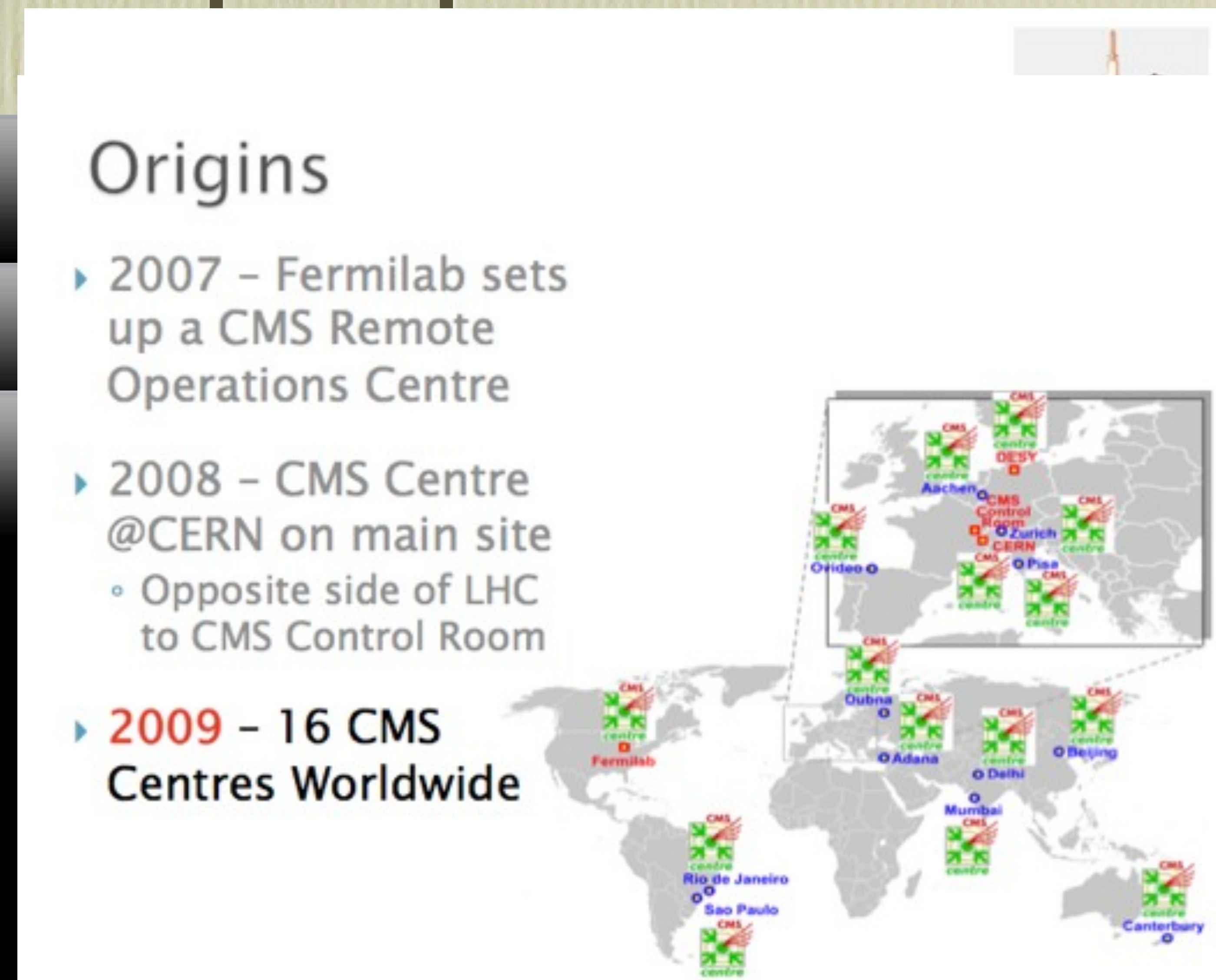
Talking to people.



Talking to people.

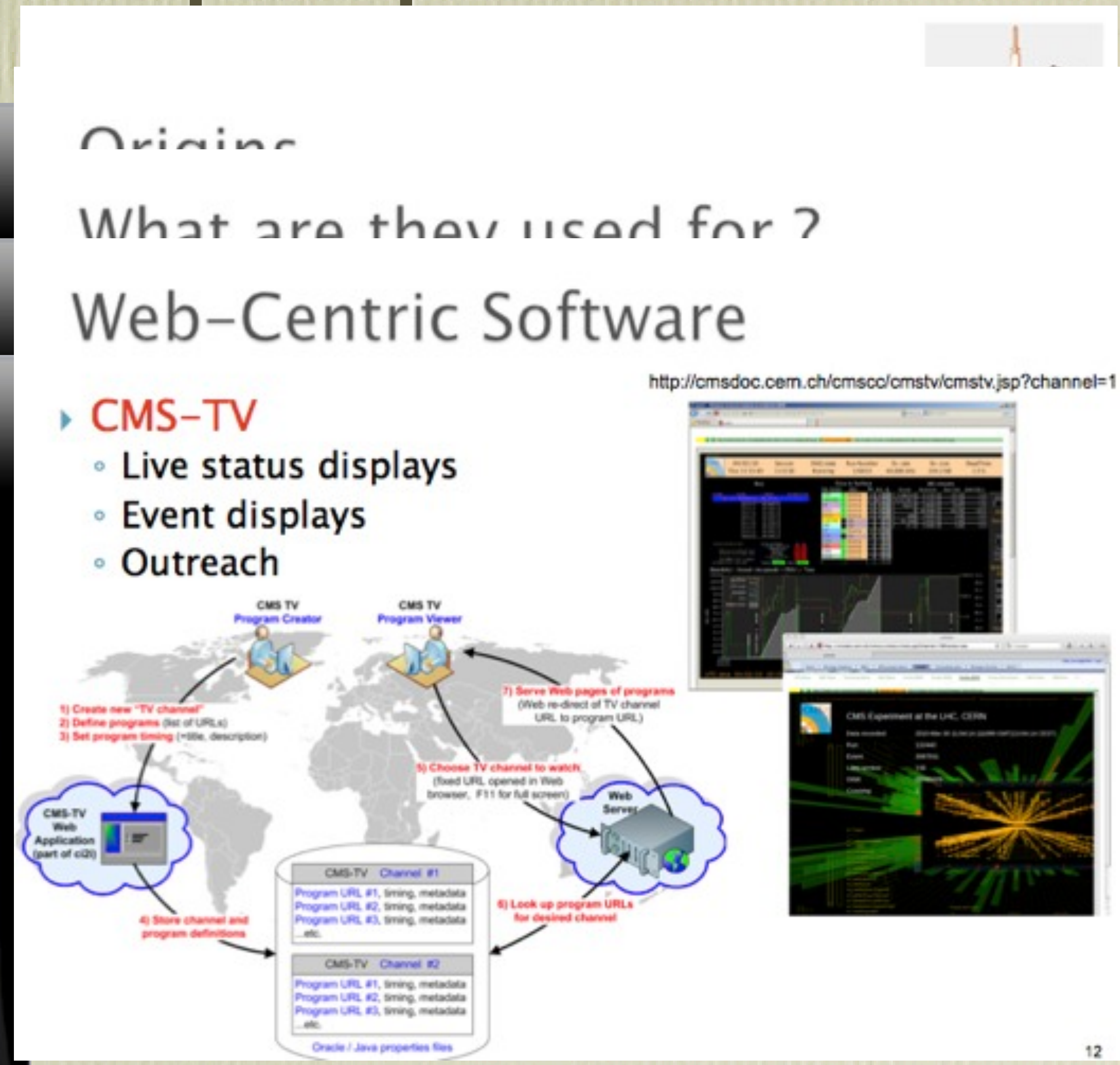
Talking to people.

Talking to people.

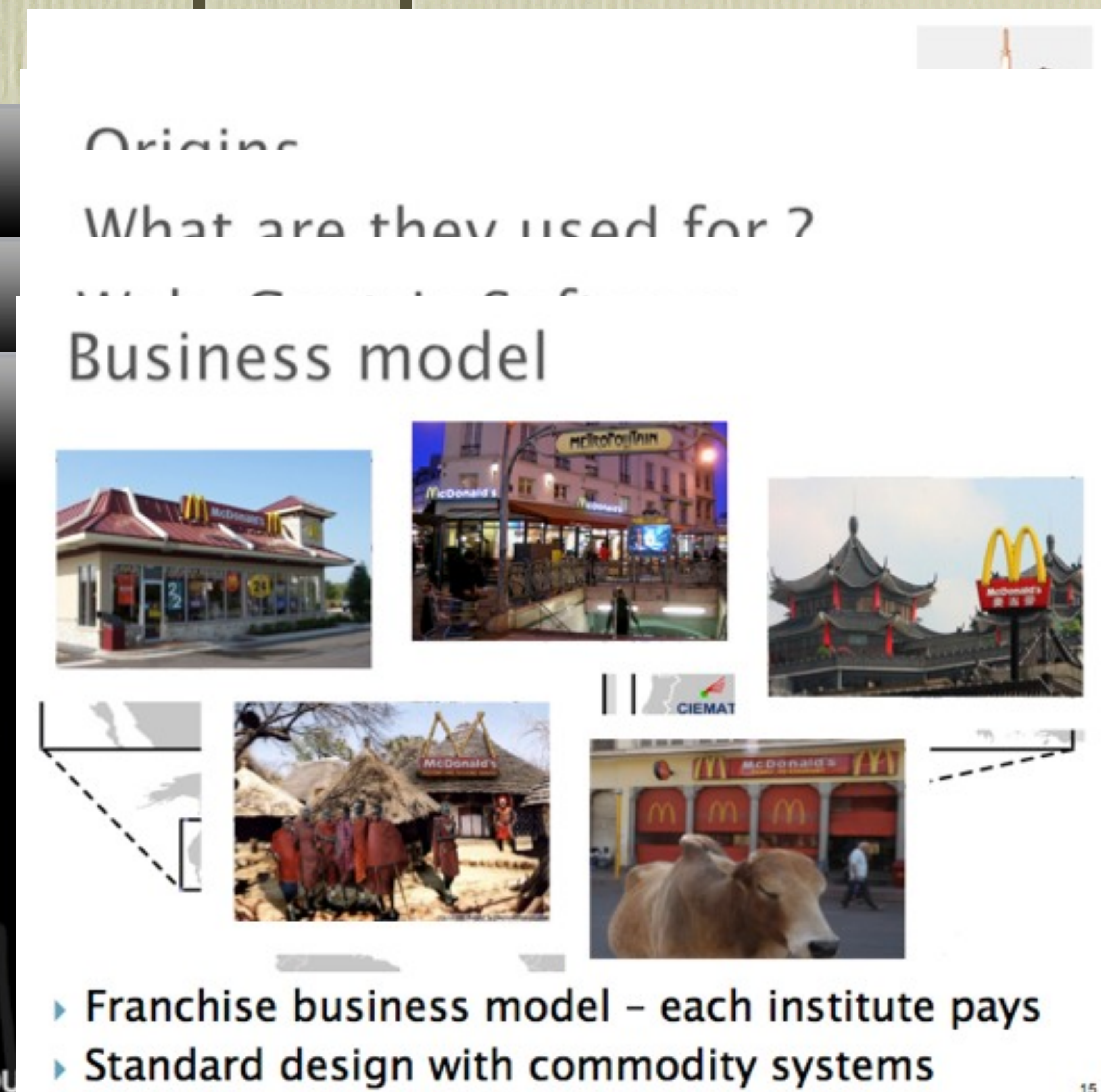


Talking to people.

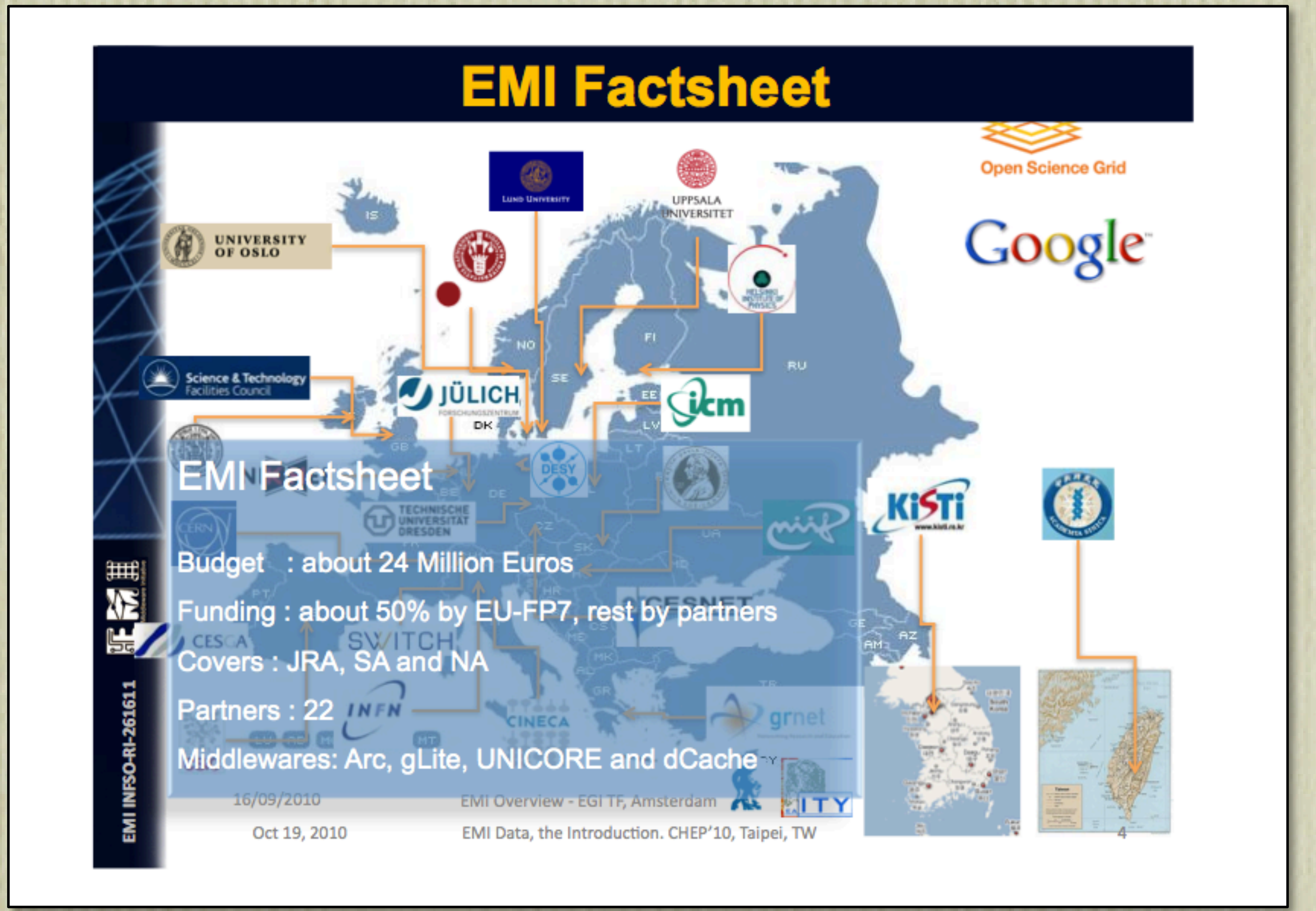
Talking to people.



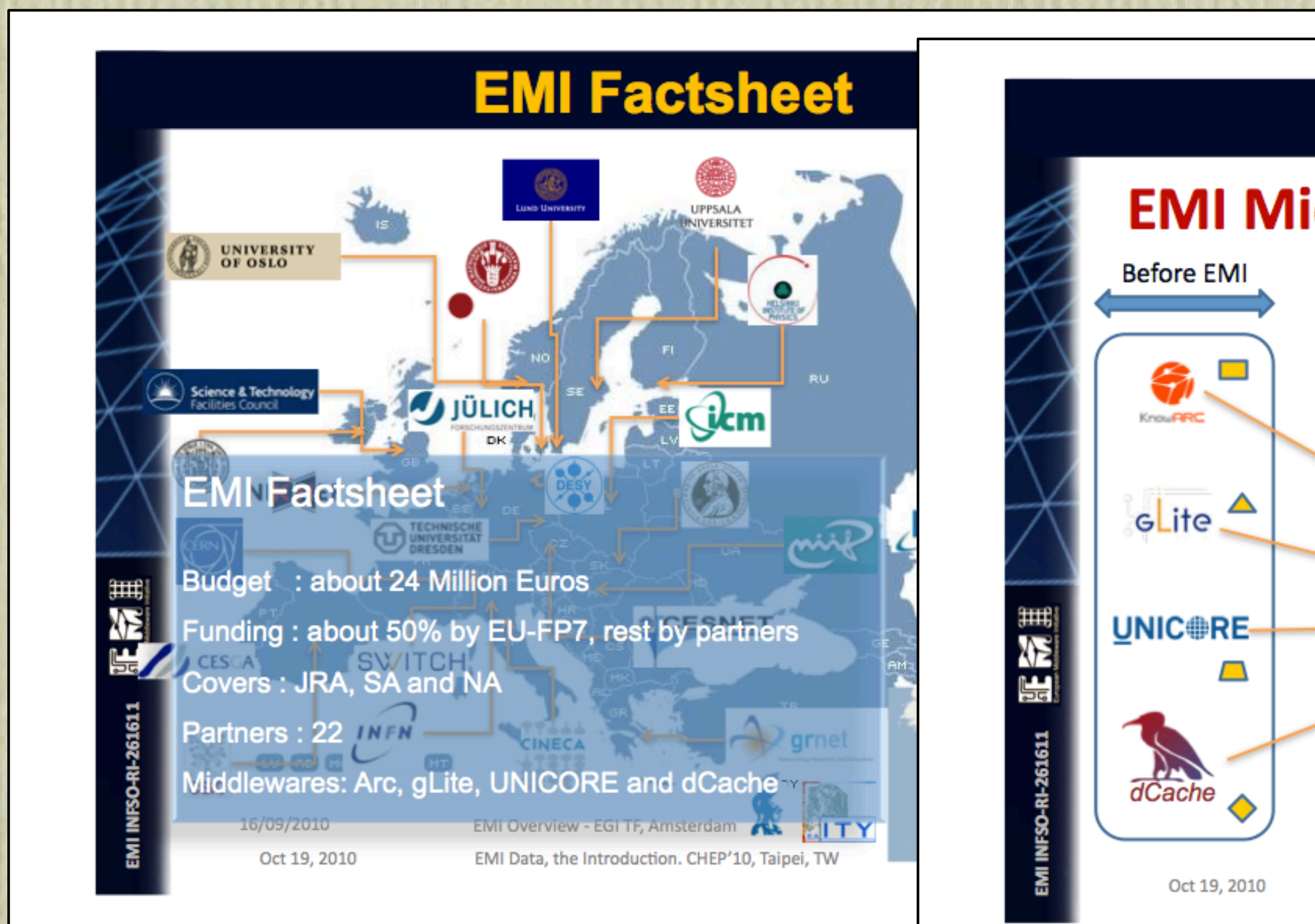
Talking to people.

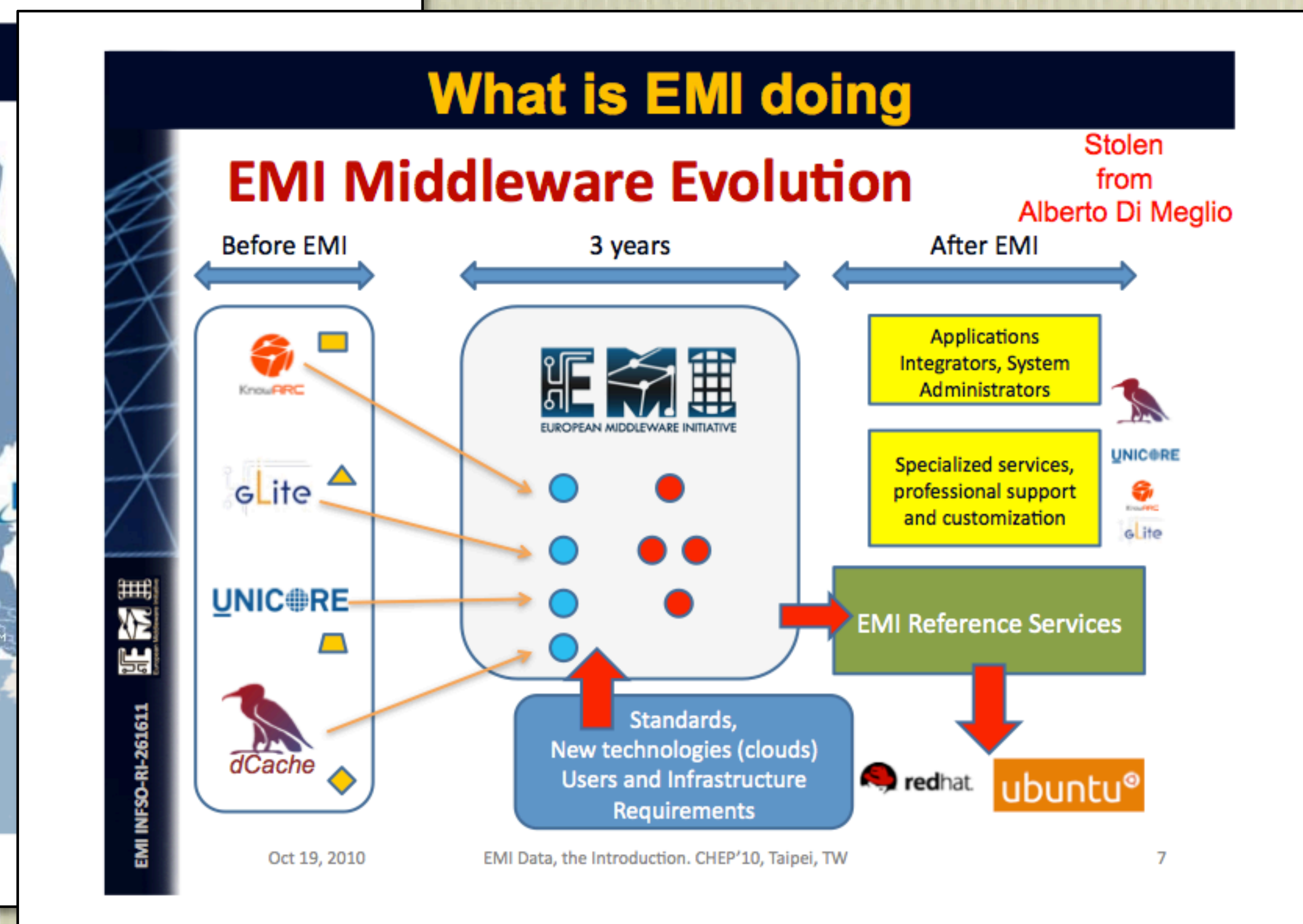


And another thing: EMI

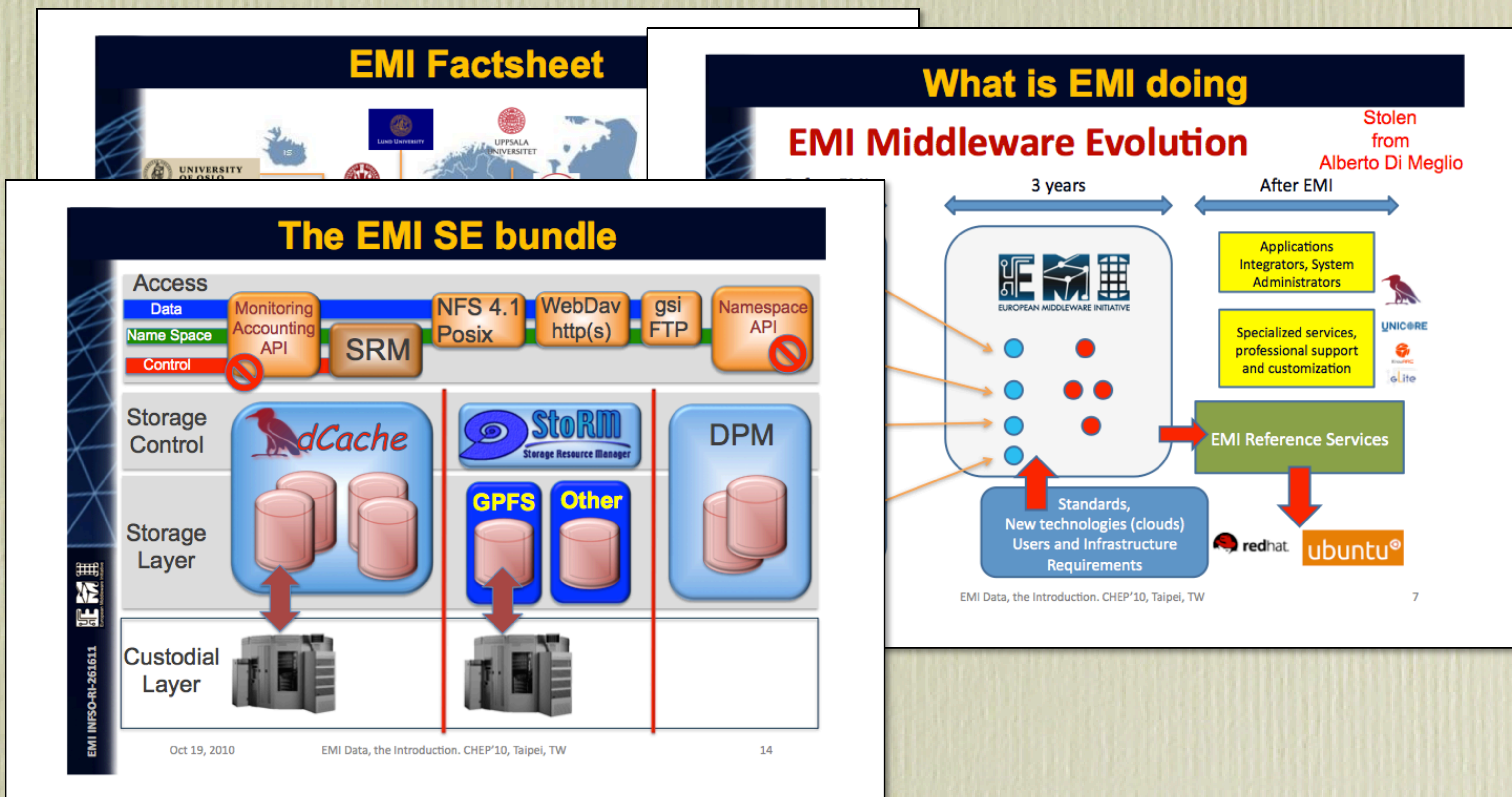


And another thing: EMI

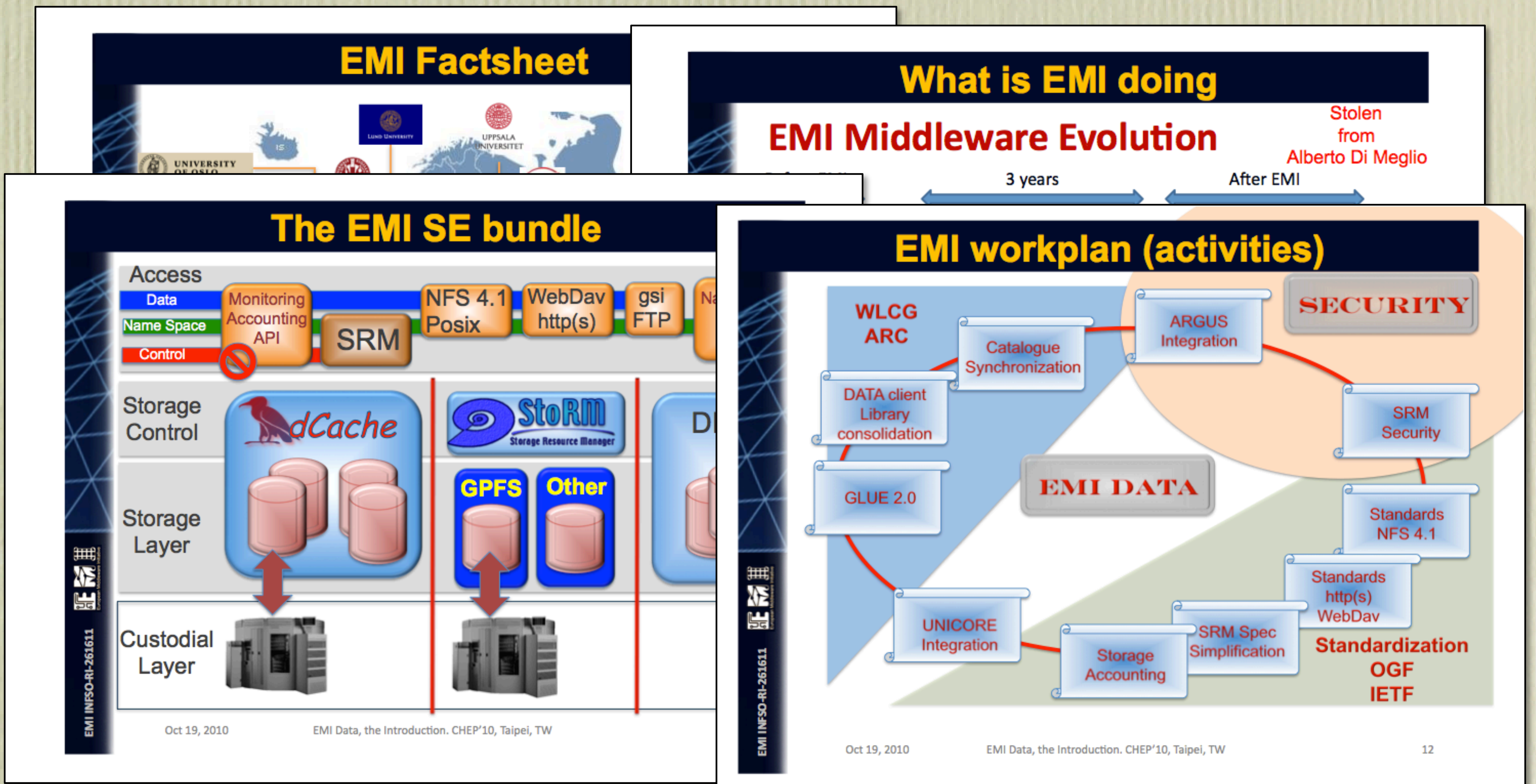




And another thing: EMI



And another thing: EMI

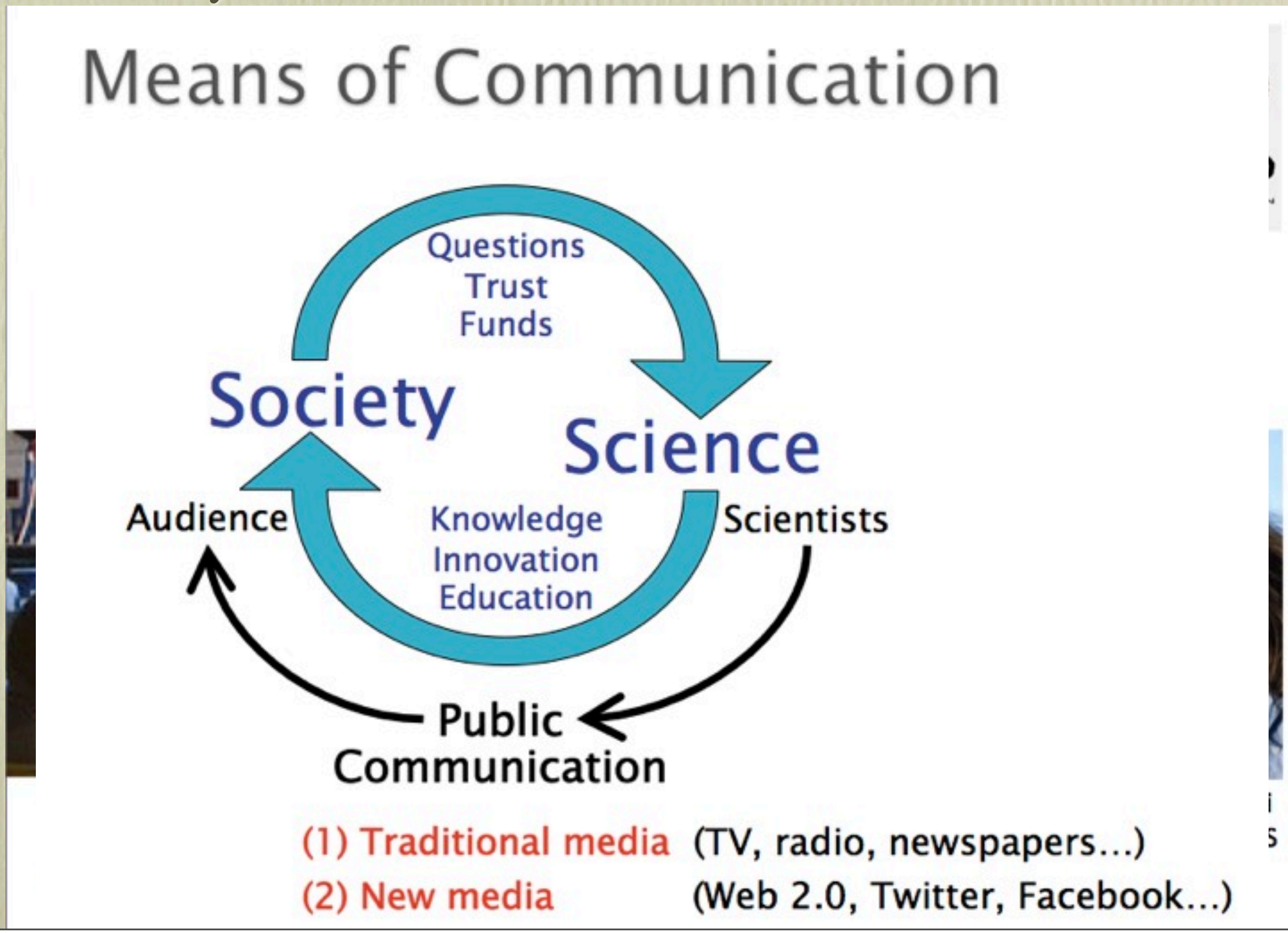


The best plenary

- Lucas Taylor's CERN Outreach talk.

The best plenary

- Lucas Taylor's CERN Outreach talk.



The best plenary

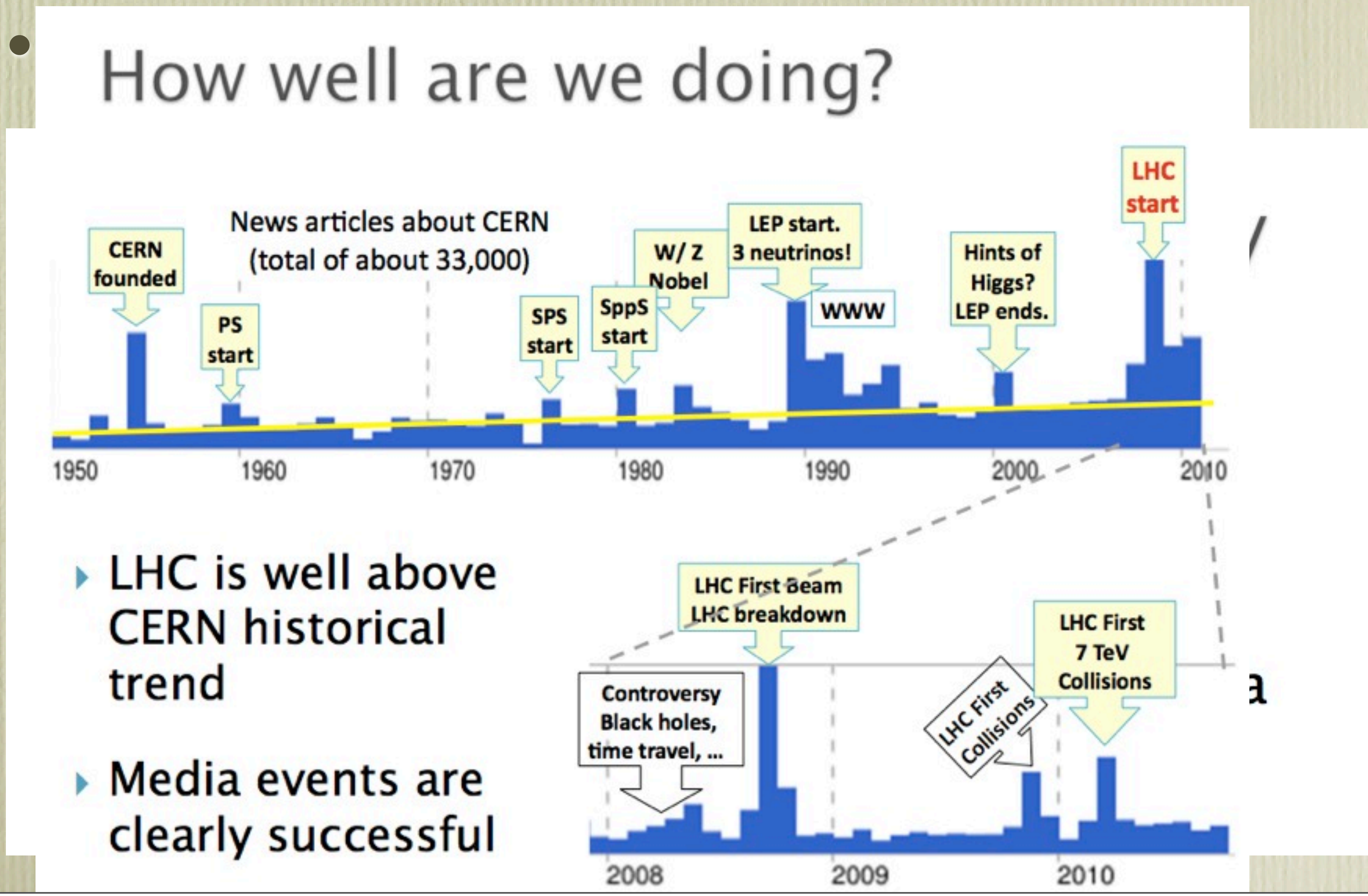
- Lucas Taylor's CERN Outreach talk

The best plenary

- Lucas Taylor's CERN Outreach talk.

The best plenary

•

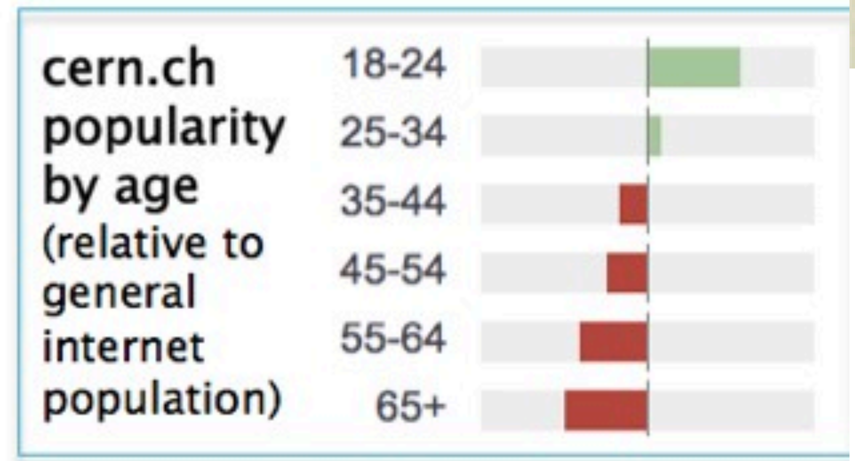


The best plenary

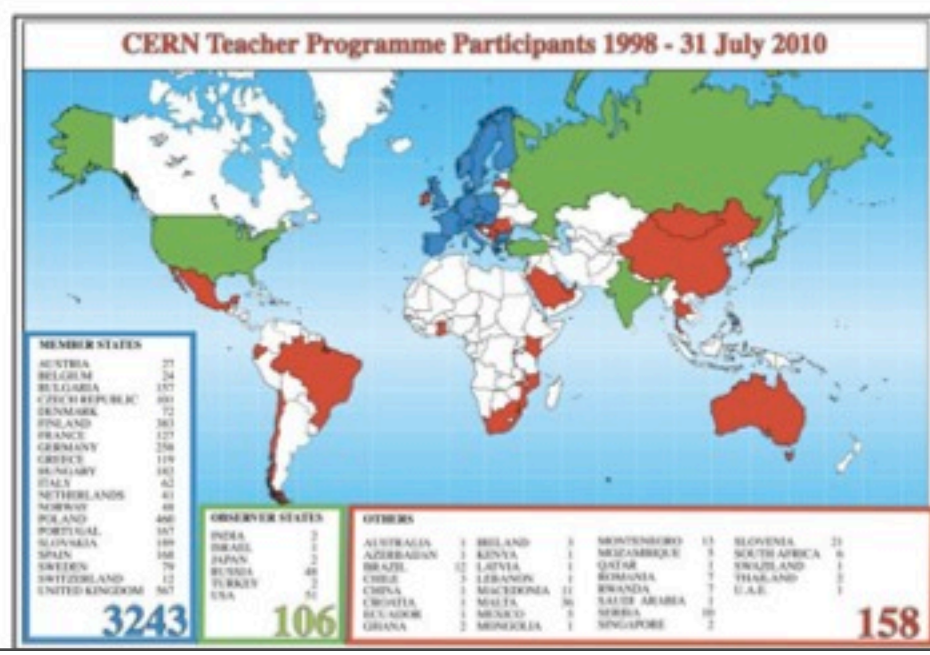
- How about with younger people ?

EUROBAROMETER

Poll says not enough effort is made to reach young people



- 3,500 teachers went through CERN Teachers Programme and now teach O(100,000) school students at any time



The best plenary

How well are we doing?

Language monitoring of online and print media

http://www.languagemonitor.com/news/top-words-of-2009/

Top Phrases of 2009	Top Words of 2009	Top Names of 2009
1. King of Pop	**1. Twitter**	1. Barack Obama
2. Obama-mania	2. Obama	2. Michael Jackson
3. Climate Change	3. H1N1	3. Mobama
4. Swine	4. Stimulus	4. Large Hadron Collider
5. Too Large to Fail	5. Vampire	5. Neda Agha Sultan
6. Cloud Computing	**6. 2.0 (next gen.)**	6. Nancy Pelosi
7. Public	7. Deficit	7. M. Ahmadinejad
8. Jai Ho!	8. Hadron	8. Hamid Karzai
9. Mayan Calendar	9. Healthcare	9. Rahm Emmanuel
10. God Particle	10. Transparency	10. Sonia Sotomayor

with nothing at all LHC-related in 2008

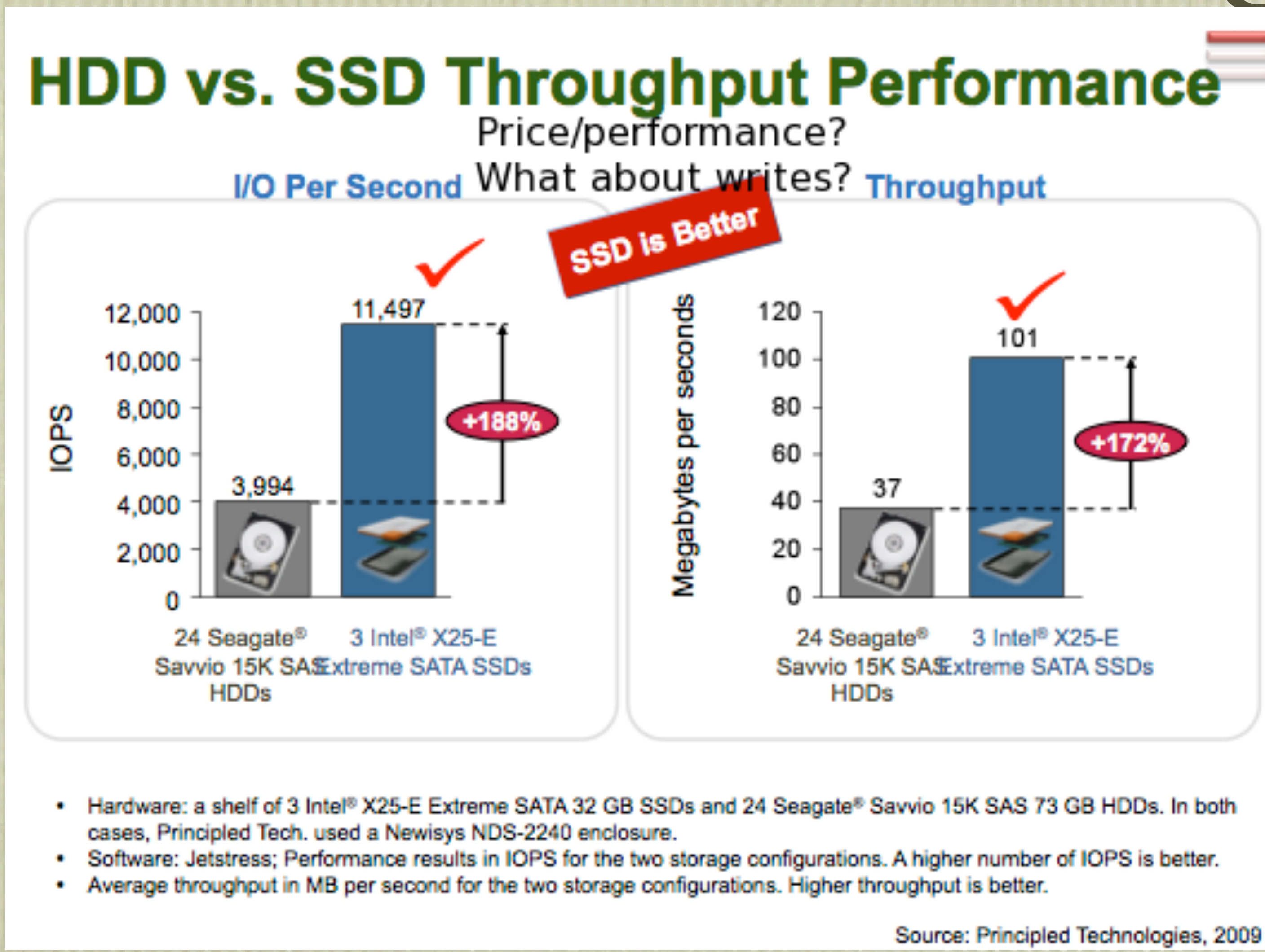
The best plenary

•

The worst plenary (a moment of indulgence).

- ACER Marketing SSD talk.
- An object lesson in how marketing people will misuse graphs to attempt to mislead you.
- SLC (expensive) SSDs used for performance comparisons.
- MLC (cheap) SSDs used for price comparisons.
- Write performance carefully not explored much (15K HDDs and RAID0 both beat even some SLC SSDs at this).

The worst plenary (a moment of indulgence).



alk.

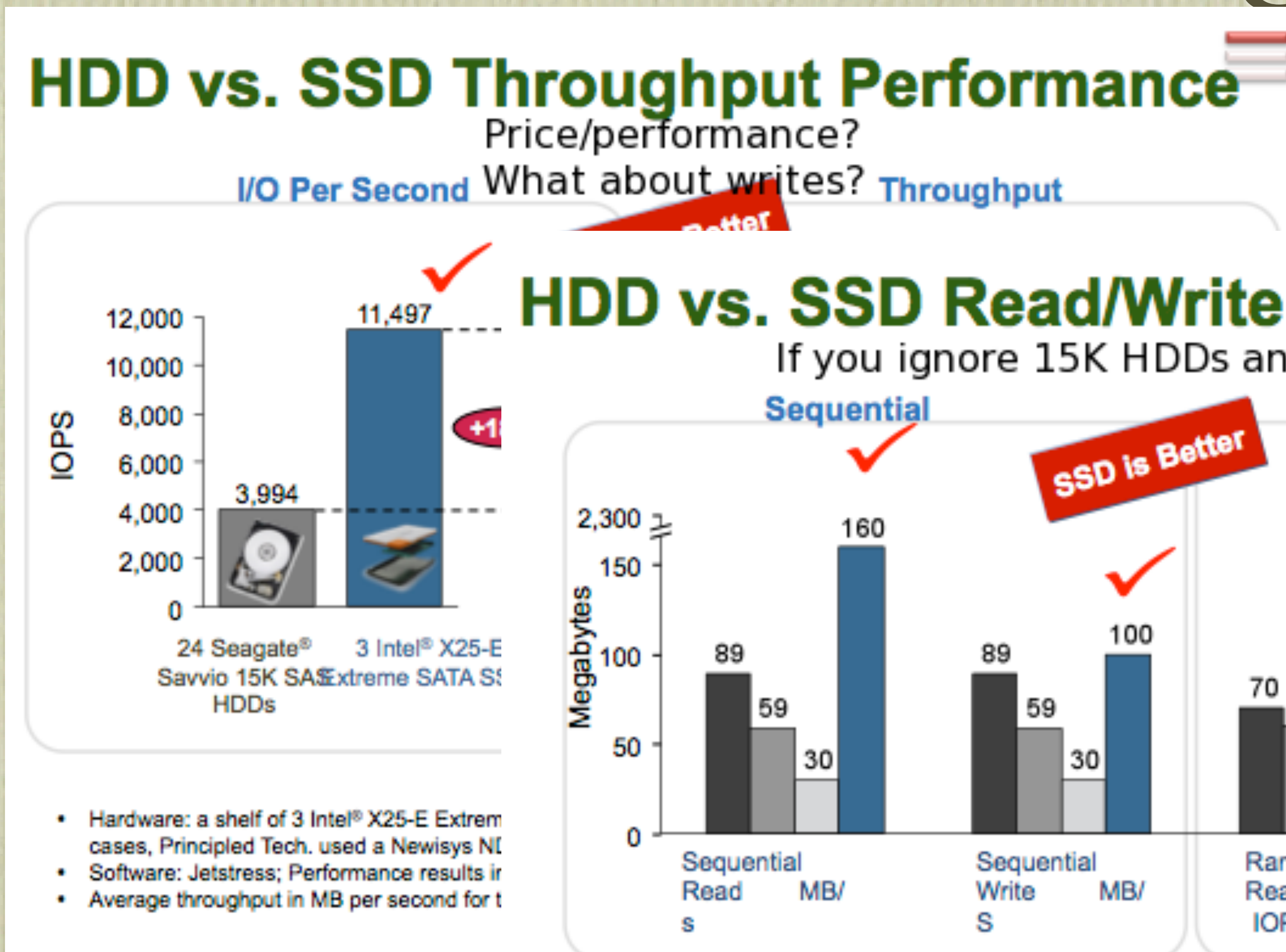
marketing people will
t to mislead you.

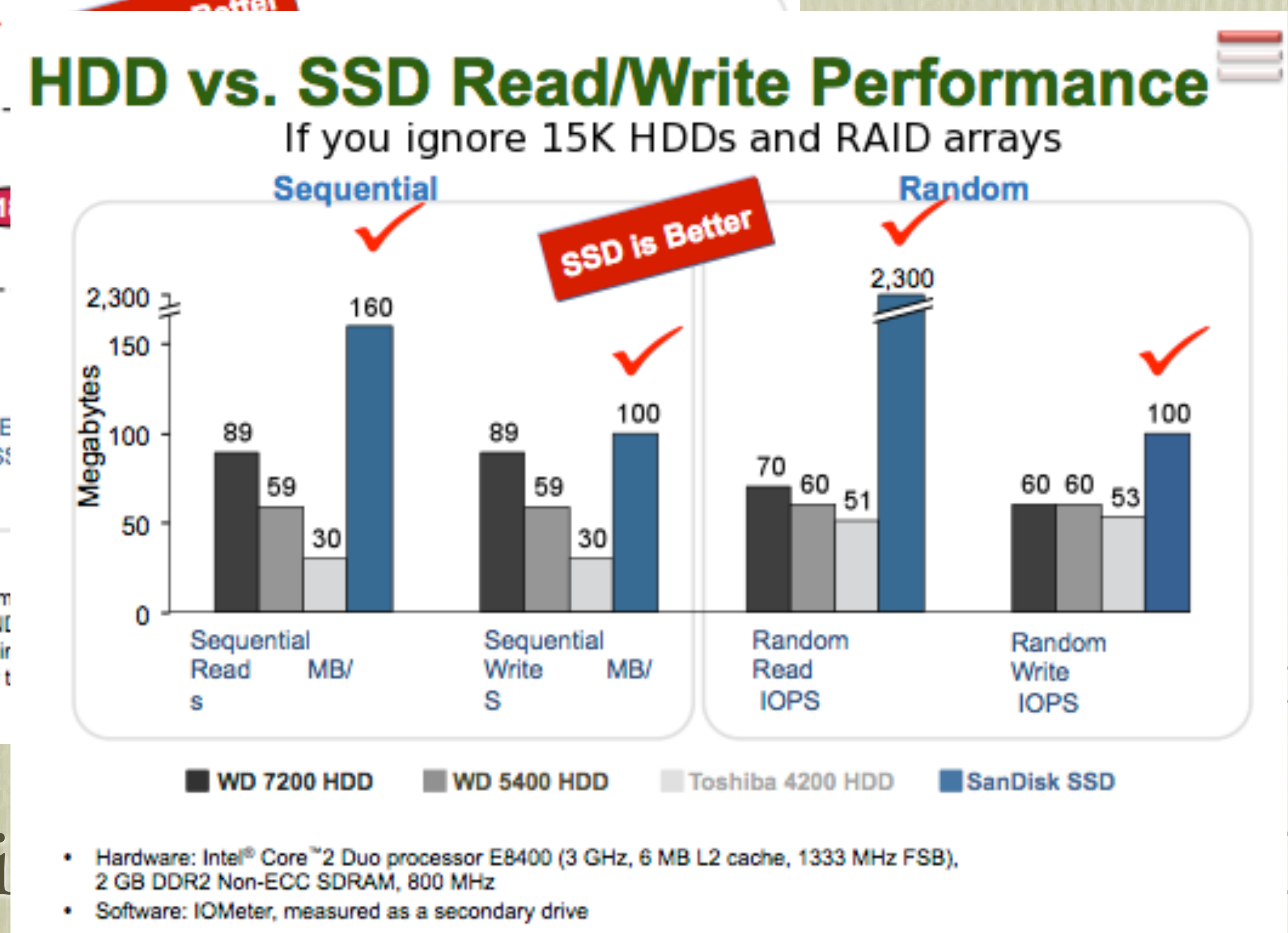
sed for performance

l for price comparisons.

- Write performance carefully not explored much (15K HDDs and RAID0 both beat even some SLC SSDs at this).

The worst plenary (a moment of indulgence).





alk.

eople will you.

rmance

mparisons.

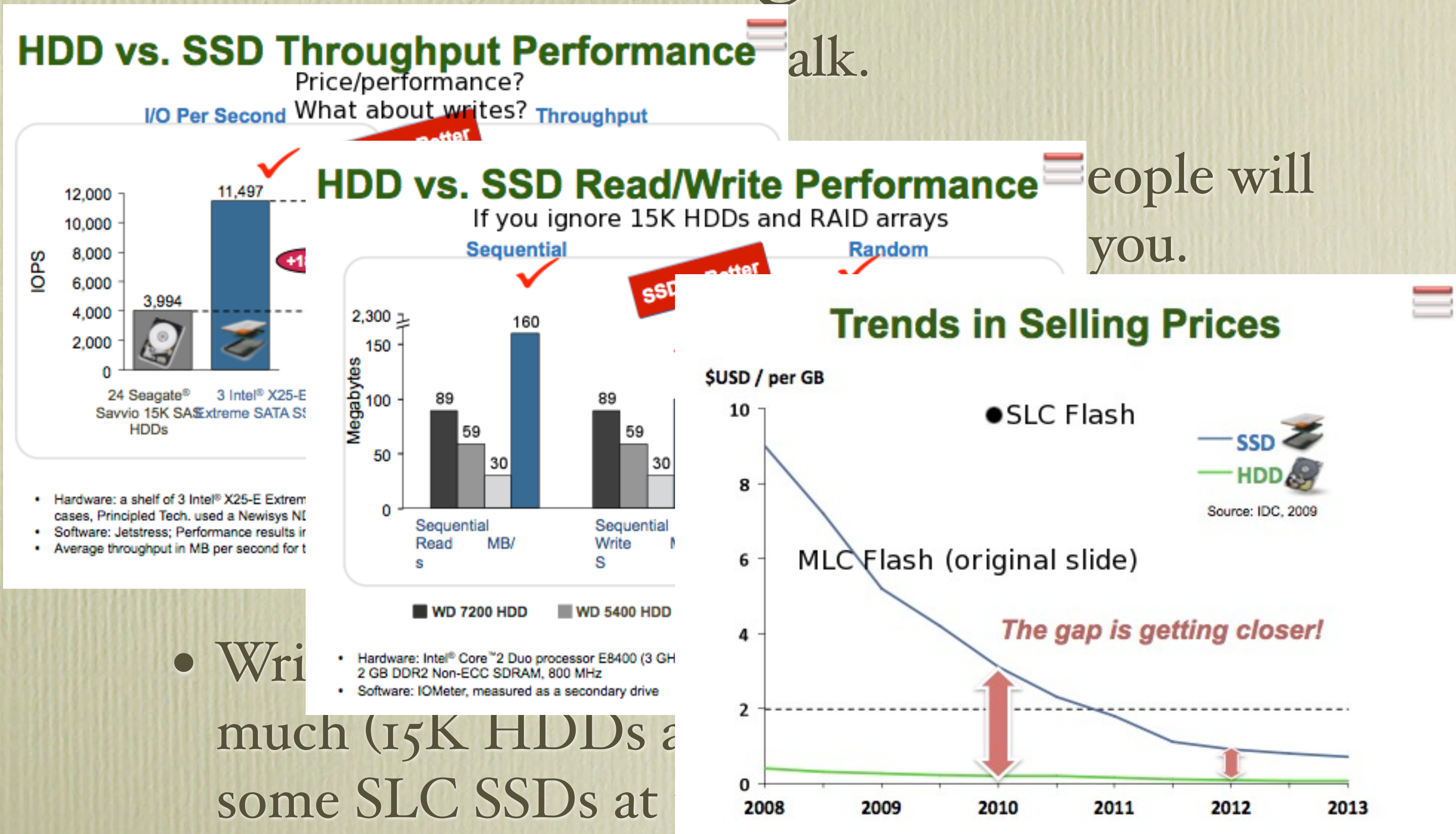
- Wri lored much (15K HDDs and RAID0 both beat even some SLC SSDs at this).

The worst plenary (a moment of indulgence).

alk.

eople will

you.



- Wri
much (15K HDDs a
some SLC SSDs at

The summarised Summary, concluded.

- Data Management is hard.
  - So we have to be cleverer about how we do it.
- But the LHC works.
- And lots of cool tech is waiting for us in the future, if we can use it.

Conclusion

- 謝謝你們。你們要訊嗎?
- (Thank you. Do you have any questions?)

List of talks referenced.

- EMI, the Future of the European Data Management Middleware (PS15-1-463)
- Standard Protocols in DPM (PS15-3-443)
- LHC Data Analysis Using NFSv4.1 (pNFS): A Detailed Evaluation. (PS35-4-289)
- ng: What Next-Gen Languages Can Teach Us About HENP Frameworks in the Manycore Era [114]
- Parallelizing Atlas Reconstruction and Simulation: Issues and Optimization Solutions for Scaling on Multi- and Many-CPU Platforms (PS18-3-126)
- Algorithm Acceleration from GPGPUs for the ATLAS Upgrade [273]
- Adaptive Data Management in the ARC Grid Middleware (PS21-1-156)
- When STAR Meets the Clouds – Virtualization & Grid Experience (PS44-1-322)
- Tests of Cloud Computing and Storage System Features for Use in the H1 Collaboration Data Preservation Model (H1 Collaboration) (PS44-2-346)
- Establishing Applicability of SSDs to LHC Tier-2 Hardware Configuration (PS35-3-288)
- Distributing LHC Application Software and Conditions Databases Using CernVM File System (PS06-5-434)
- CernVM: Minimal Maintenance Approach to the Virtualization (PS29-4-432)
- ATLAS, CMS, New Challenges for the Public Communication (PL-08)
- One Small Step: A View of the LHC Experiments' Offline Systems After One Year of Data-Taking (PL -02)
- How to Harness the Performance Potential of Current Multi-Core CPUs and GPUs (PL-04)
- CMS Centres Worldwide - A New Collaborative Infrastructure (PS34-2-267)