High Energy Physics and Grid

Oxana Smirnova
FYS225/FKF050
February 27, 2007, Lund

High Energy Physics: quick intro

- High Energy Physics studies elementary particles and forces that constitute the matter
  - Also known as Particle Physics
  - Is the most fundamental of all the physical sciences
  - The ultimate goal: to understand the origin of the Universe
  - We've learned a lot over the past years, yet more is unknown

Standard Model of forces and particles

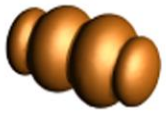
Leptons

Electric Charge

Tau	−1	0	Tau Neutrino
Muon	−1	0	Muon Neutrino
Electron	−1	0	Electron Neutrino

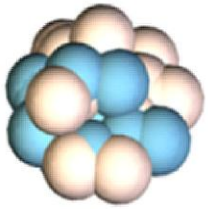
Strong

Gluons (8)



Quarks

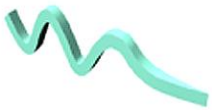
Mesons
Baryons

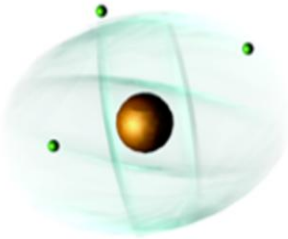


Nuclei

Electromagnetic

Photon





Atoms
Light
Chemistry
Electronics

Quarks

Electric Charge

Bottom	−1/3	2/3	Top
Strange	−1/3	2/3	Charm
Down	−1/3	2/3	Up

each quark: R, B, G 3 colours

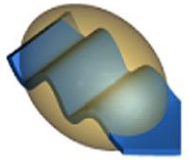
Gravitational

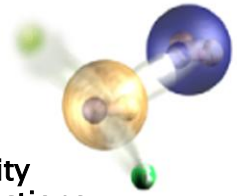
Graviton ?

Solar system
Galaxies
Black holes

Weak

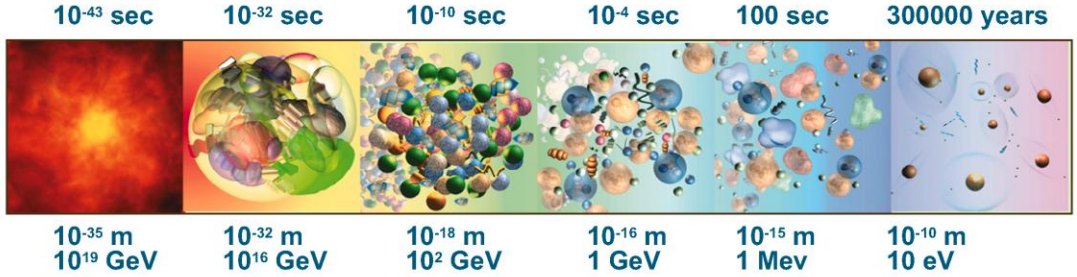
Bosons
(W,Z)

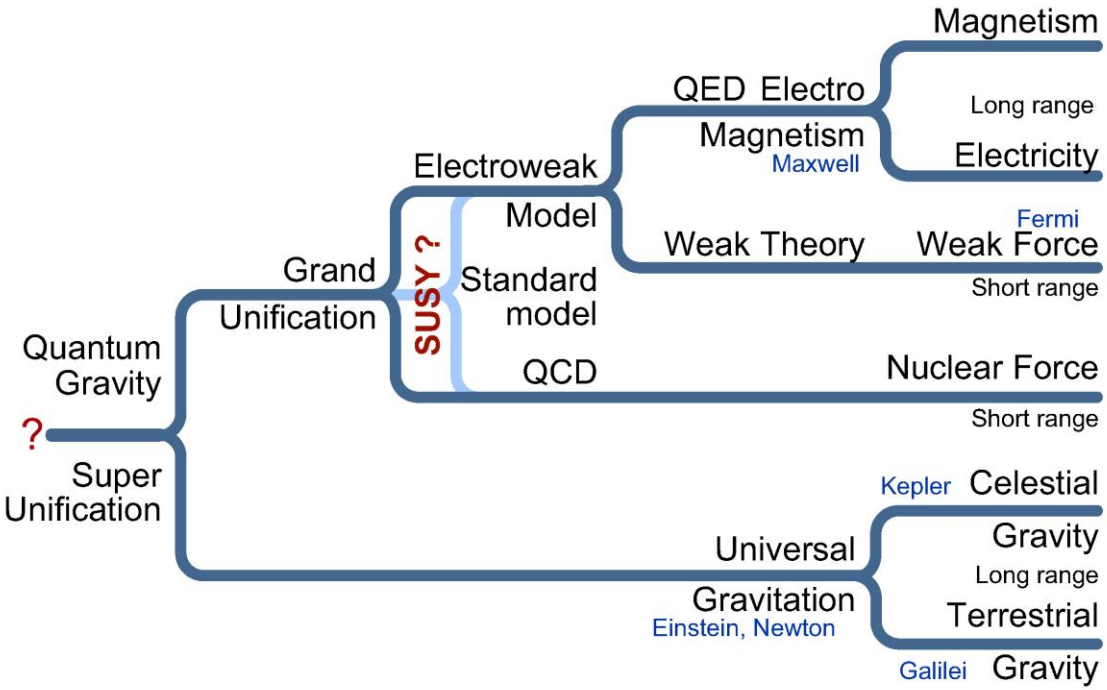




Neutron decay
Beta radioactivity
Neutrino interactions
Burning of the sun

From Big Bang to now: the theory





Still, we know too little about ourselves!

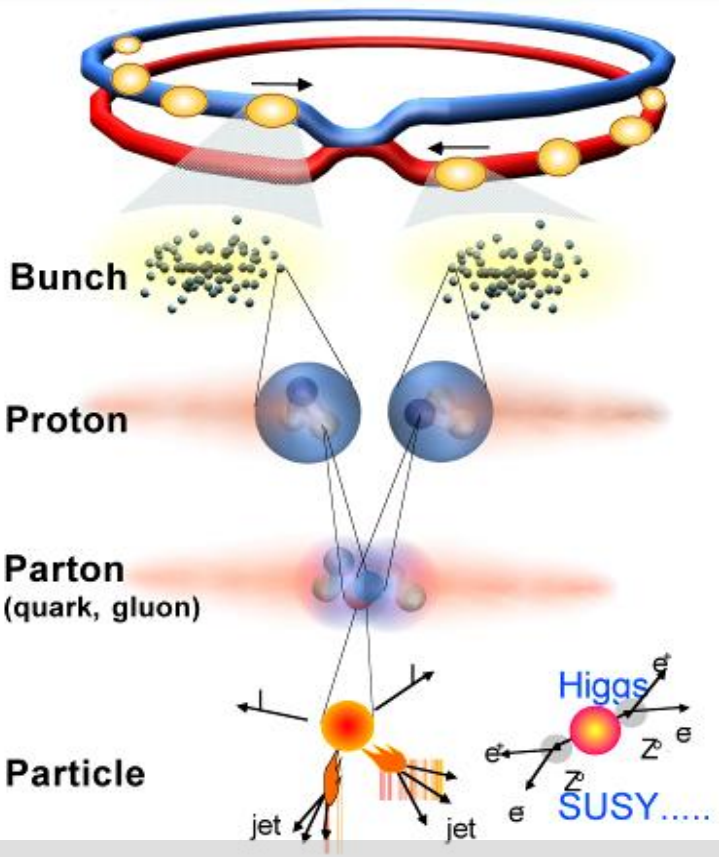
- Can gravity be included in a theory with the other three interactions?
- Why do the particles have the masses we observe, and what is the origin of mass?
- How many space-time dimensions do we live in?
- Are the known elementary particles fundamental or do they possess structure?
- Why is the electrical charge on the electron equal and opposite to that on the proton?
- Why are there three generations of quarks and leptons?
- Why is there overwhelmingly more matter than anti-matter in the Universe?
- Are protons unstable?
- What is the nature of the dark matter that pervades our Galaxy?
- Are there new states of matter at exceedingly high density and temperature?
- Do the neutrinos have mass, and if so why are they so light?

NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

Large Hadron Collider: World's biggest accelerator at CERN

Will start operating in fall 2007

Collisions at LHC



Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34}\ cm^{-2}\ s^{-1}$
Crossing rate	40 MHz
Collisions rate	$\approx 10^{7}$ - 10^{9} Hz
New physics rate	$\approx$ .00001 Hz

Event selection:
1 in 10,000,000,000,000

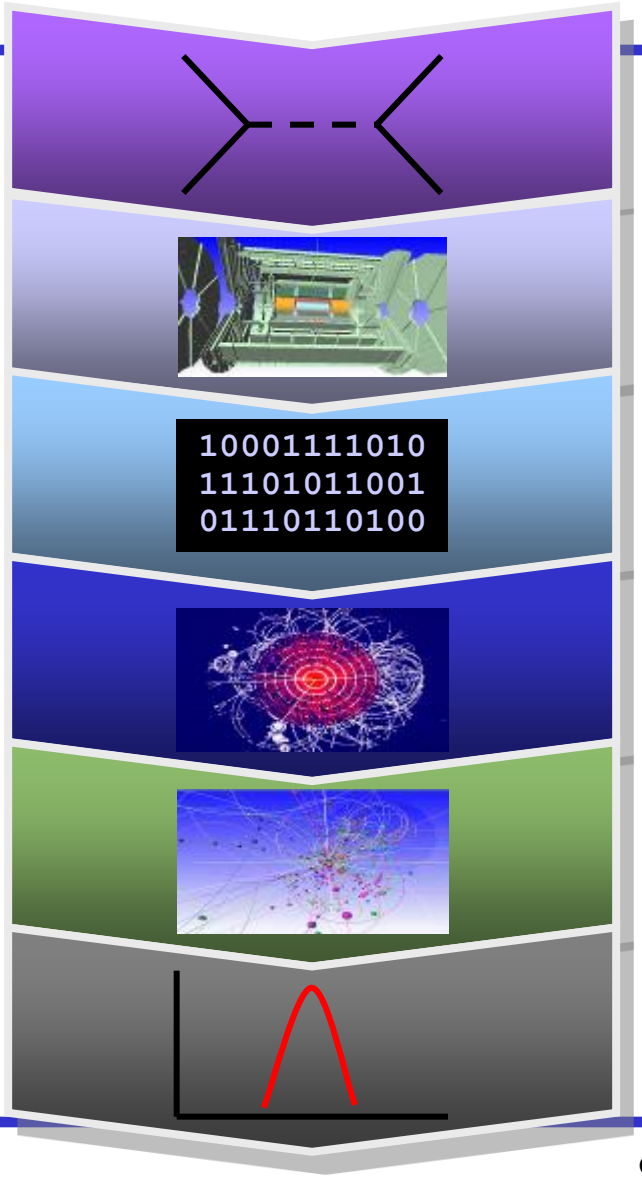
Will help to solve:
- **The mass problem – discover Higgs particle?**
- **The Dark Matter of the Universe – discover supersymmetric particles?**
- **Matter vs Antimatter – CP violation?**

NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

ATLAS detector at LHC

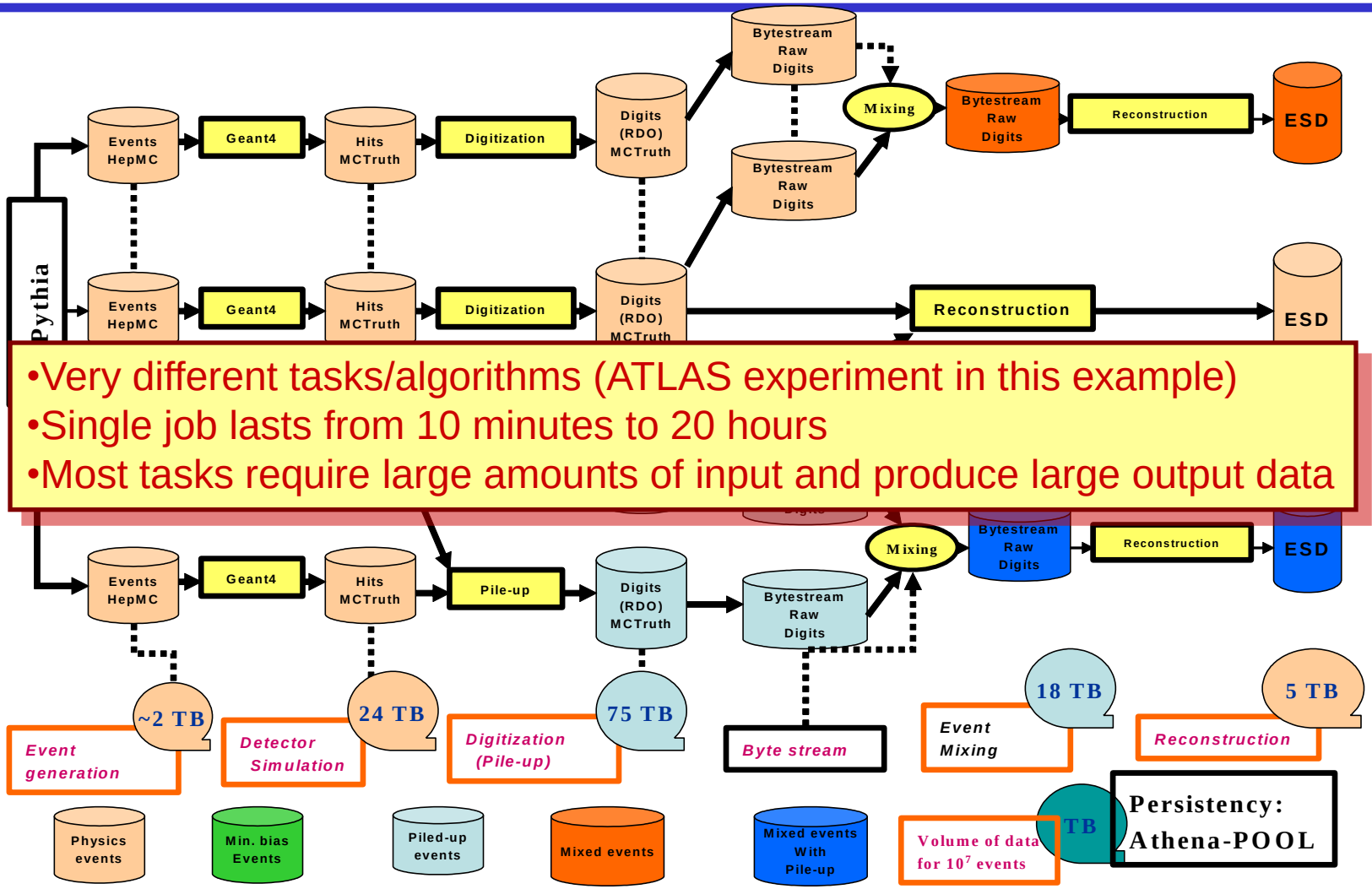
“Full chain” of HEP data processing

- Event generation (*Pythia*)
- Detector simulation (*Geant4*)
- Hit digitization
- Reconstruction
- Analysis data preparation
- Analysis, results (*ROOT*)



Slide adapted from Ch.Collins-Tooth and J.R.Catmore

Monte Carlo data production flow (10 Mevents)



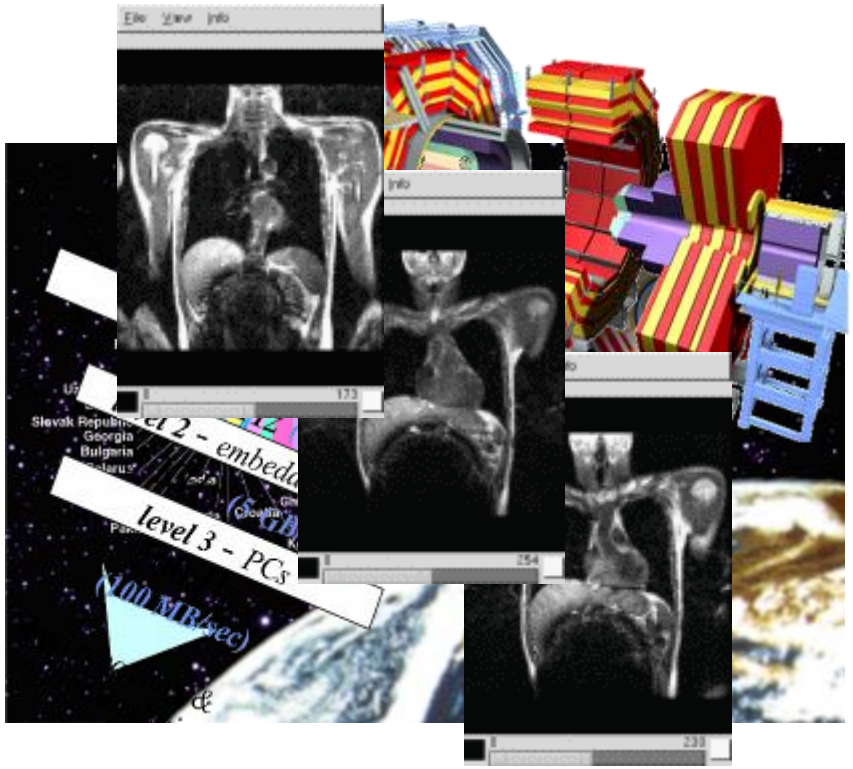
- Very different tasks/algorithms (ATLAS experiment in this example)
- Single job lasts from 10 minutes to 20 hours
- Most tasks require large amounts of input and produce large output data

Software for HEP experiments

- Massive pieces of software
  - Written by very many different authors in different languages (C++, Java, Python, Fortran)
  - Dozens of external components
  - Occupy as much as ~10 GB of disk space each release
- Frequent releases
  - Every experiment can release as often as once a month during the preparation phase (which is *now* for LHC)
- Difficult to set up outside the lab
  - Experiments can not afford supporting different operating systems and different computer configurations
- For a small university group it is very difficult to manage different software sets and maintain hardware
  - ALICE, ATLAS, H1, PHENIX etc – all in many versions
  - Solution: use external computing resources

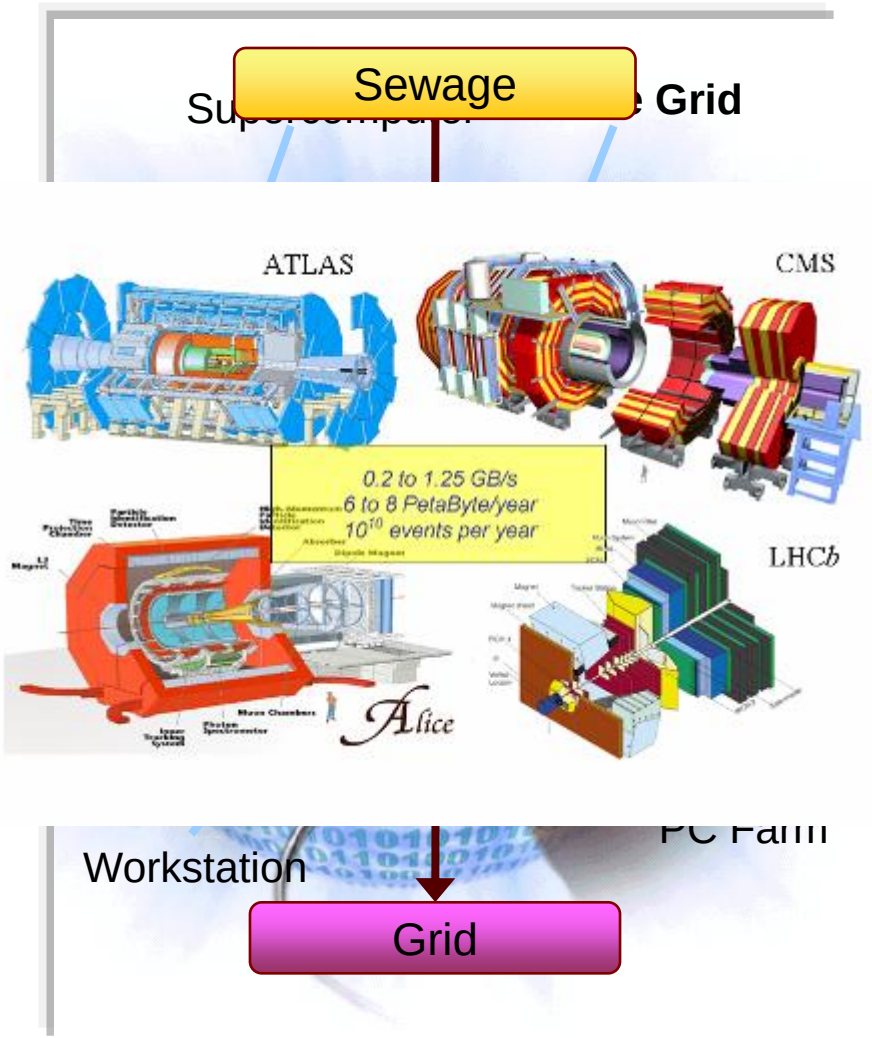
HEP computing specifics

- Data-intensive tasks
  - Large datasets, large files
  - Lengthy processing times
  - Large memory consumption
  - High throughput is necessary
- Very distributed resources
  - Distributed computing resources of *modest* size
  - Produced and processed data are hence distributed, too
  - Issues of coordination, synchronization and authorization are outstanding
- HEP is by no means unique in its demands, but we are first, we are many, and we *badly* need it

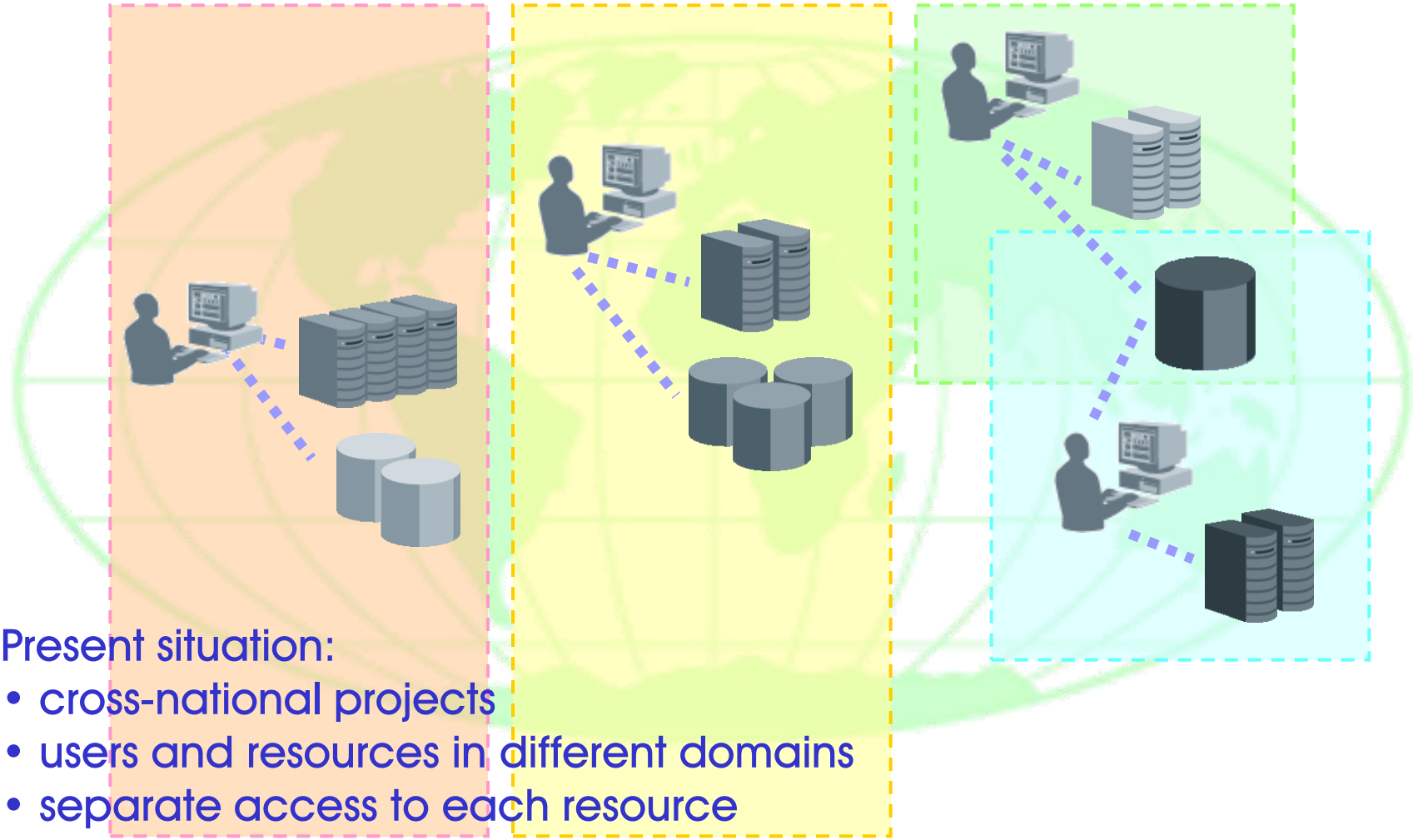


Grid to the rescue

- Distributed supercomputer, based on commodity PCs and fast WAN
- Access to the great variety of resources by a single pass – certificate
- A possibility to manage distributed data in a synchronous manner
- A new commodity



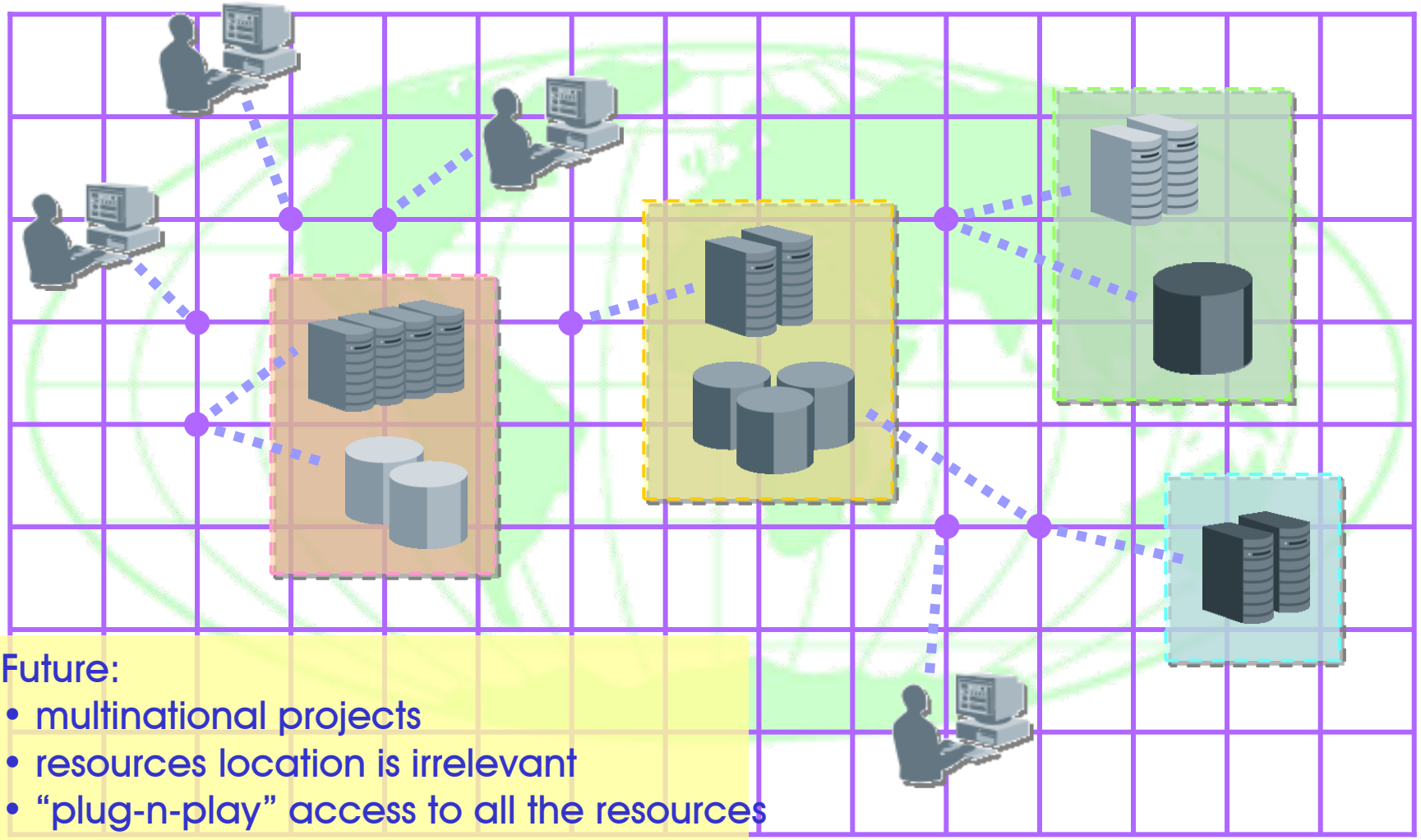
From distributed resources …



Present situation:

- cross-national projects
- users and resources in different domains
- separate access to each resource

… to World Wide Grid



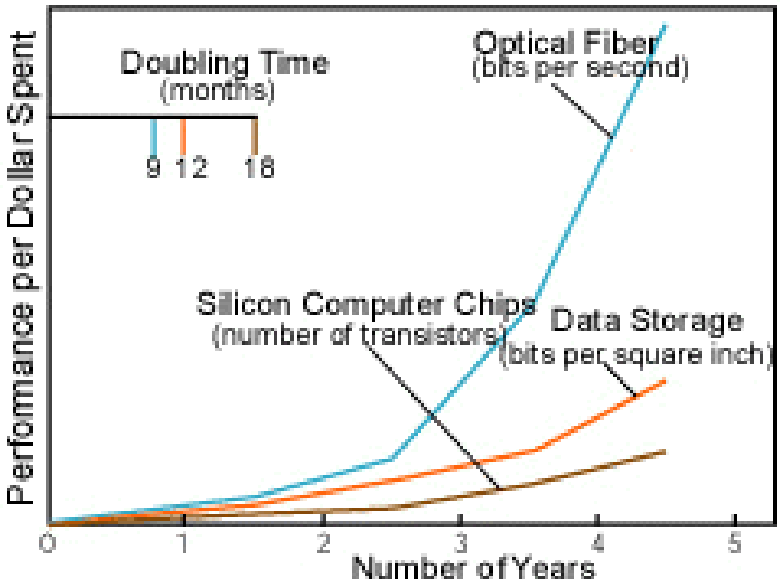
Future:
- multinational projects
- resources location is irrelevant
- “plug-n-play” access to all the resources

Origin of the word “Grid”

- Coined by **Ian Foster** and **Carl Kesselman** around 1997
- Refers to computing grids as analogy of power grids
  - Many producers
  - Competing providers
  - Simple for end-users
- Spelled “grid” or “Grid”
  - Except in French: “Grille de calcul”

Grid as a result of IT progress

- Network vs. computer performance:
  - Computer speed doubles every 18 months
  - Network speed doubles every 9 months
- 1986 to 2000:
  - Computers: 500 times faster
  - Networks: 340000 times faster
- 2001 to 2010 (projected):
  - Computers: 60 times faster
  - Networks: 4000 times faster

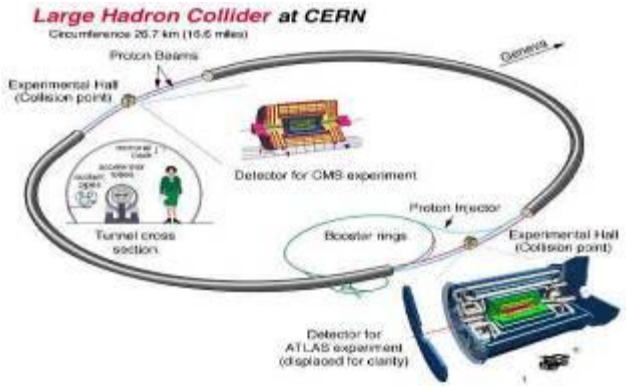


Bottom line: CPUs are fast enough; wide area networks are very fast – gotta make use of it!

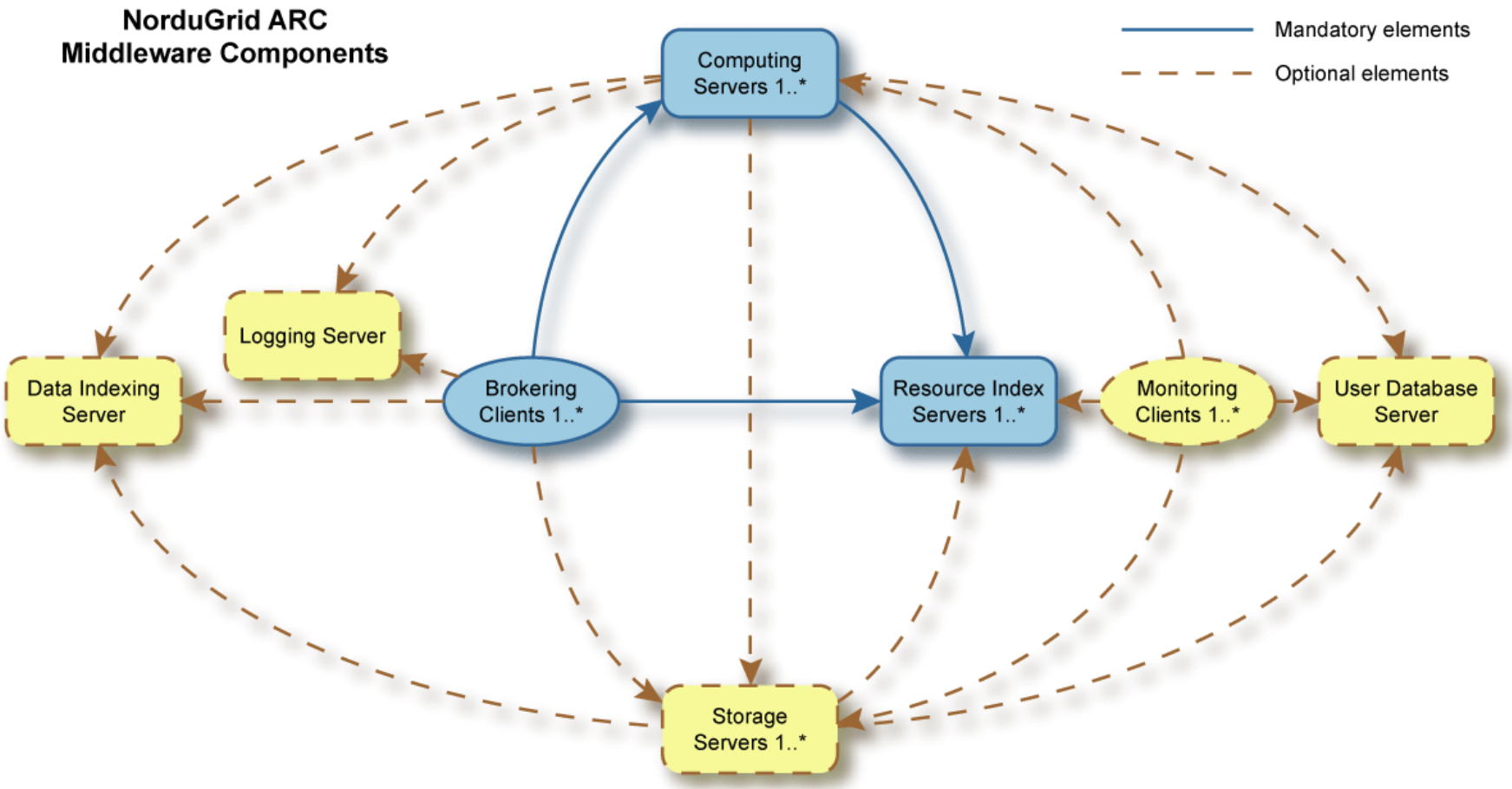
Slide adapted from the Globus Alliance

How did Grid appear in Lund

- Back in 2001, High Energy Physics Institutes from Scandinavia wanted to share their computing resources and jointly contribute to CERN/LHC computing
  - We needed Grid
  - The Grid hype just begun…
  - … and we created a NorduGrid project (Lund, Uppsala, Copenhagen, Oslo, Helsinki and many others)
- No production ready grid software (*middleware*) was available or seen on the horizon in fall 2001
- In February 2002, NorduGrid boldly decided to develop own Grid middleware
  - Was baptized ARC, for Advanced Resource Connector
  - Since May 2002 ARC is extensively used in ATLAS production and other scientific computing projects

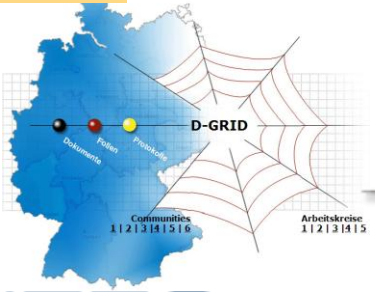


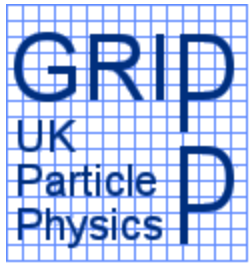
ARC in a nutshell

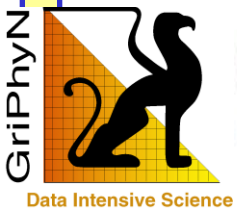


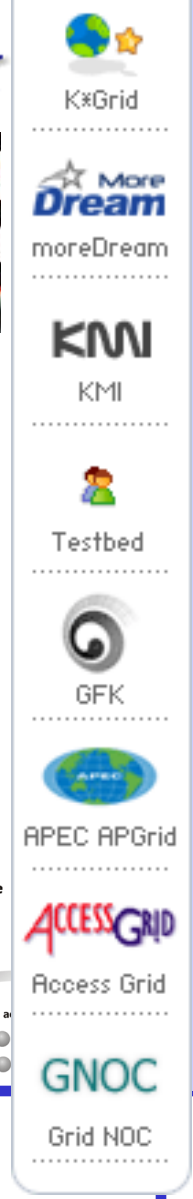
Goal: no single point of failure

ARC is not the only one…









www.nordugrid.org

Slide adapted from Vicky White, FNAL

Grids in ATLAS

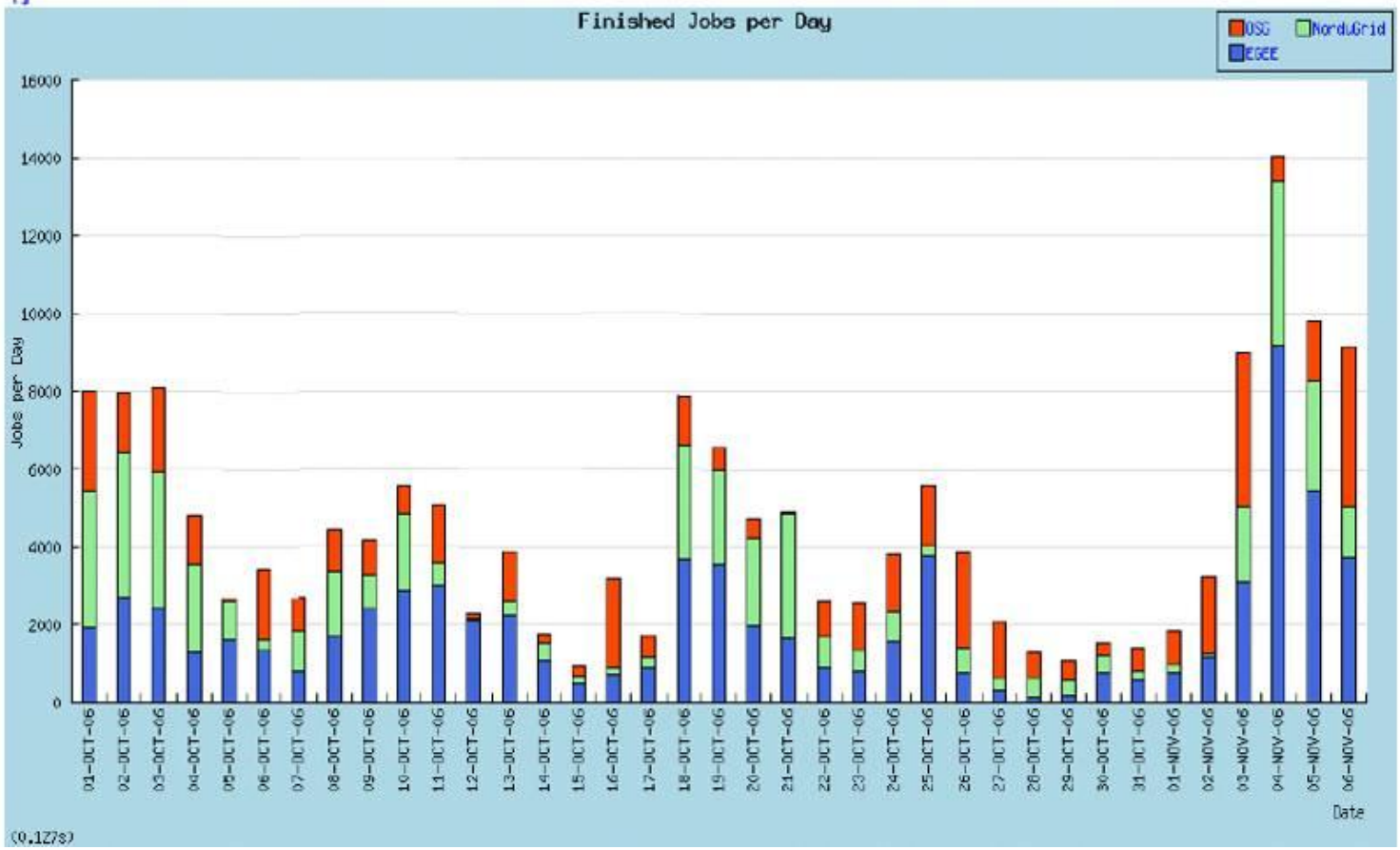
- All ATLAS production today is done via Grid
- There are 20+ Grid flavors out there
  - Almost all are incompatible with each other
  - Almost all are tailored for a specific application and/or specific hardware
- ATLAS makes use of only 3 Grid flavors:
  - gLite – formerly known as LCG
  - ARC – formerly known as NorduGrid
  - OSG – formerly known as Grid3
  - They are still incompatible with each other
    - A lot of *interoperability* efforts take place

ARC and ATLAS

- Via NorduGrid/ARC, Nordic countries contribute to ATLAS Data Challenges since 2002
  - Resources donated to ATLAS by national Grid projects, enthusiastic owners
    - Highly heterogeneous (OS: Fedora Core N, Red Hat, Debian, Gentoo,...; LRMS: PBS/Torque, Condor, SGE…)
    - No common policies enforced
    - Loosely coupled
  - Currently, ca 10% of ATLAS production tasks
    - Only 2 persons in charge of the production
    - Highest resource usage efficiency, reliability
  - Accumulated ~40TB of ATLAS data in ~50 locations
    - Includes e.g. Ljubljana – still, indistinguishable for jobs and outside users

ATLAS Monte Carlo production with NorduGrid/ARC

Jobs/day

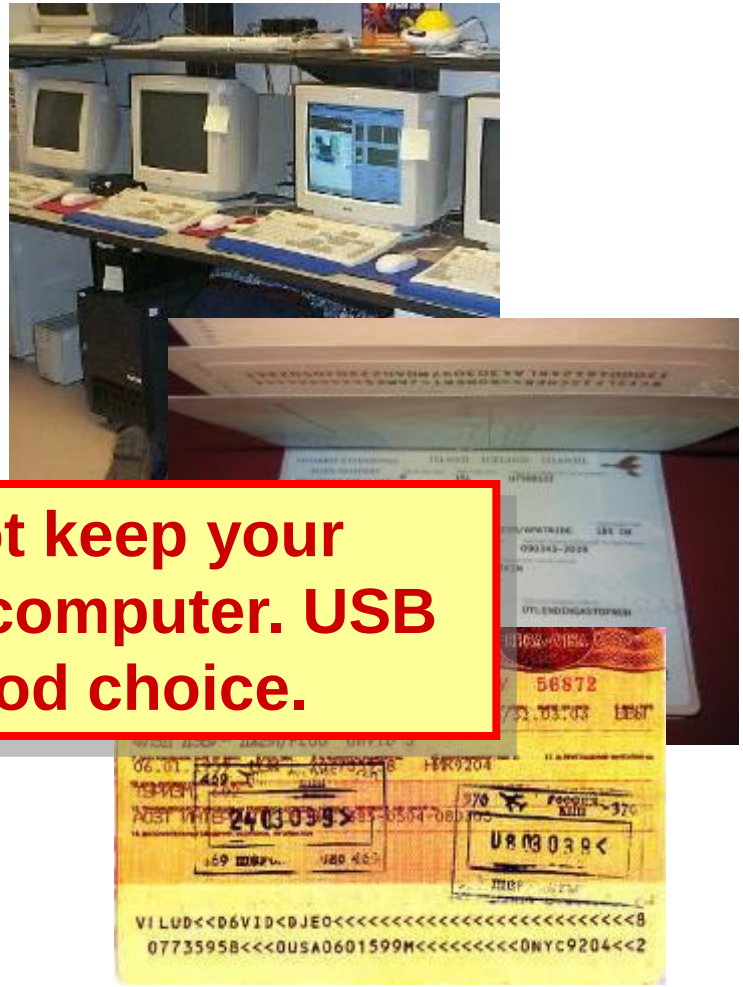


Next step: Nordic Tier1

- WLCG: Worldwide LHC Computing Grid
  - A CERN project aiming to provide HEP computing infrastructure
  - Tiered structure: Tier0 at CERN, a dozen of regional Tier1s, many local Tier2s etc
- WLCG Tier1 is primarily a set of services:
  - 24/7 on-call support system
  - Infrastructure: network, power, cooling, safety etc
  - Authorization, specific software for entire multinational VOs
  - Job submission interface
  - Data indexing service
  - Storage resource management interface
  - File transfer services between Tiers
  - Experiment-specific interfaces ("*VOBoxes*")
  - Database service (ORACLE)
  - Other: information system, monitoring, logging etc
- Resource commitments (processing power, storage, bandwidth) are expressed in WLCG MoU

Where do I start on my way to Grid?

- Ask local sysadmin to install a Grid client
  - NorduGrid standalone client is the easiest: you can install it yourself
- Apply for a “passport”: the Grid certificate (grid-cert-request)
  - NorduGrid Certification Authority is appropriate for SU employees
- Ask a
  Grid
  colla
- Appl
  appropriate Virtual Organization (VO) member
- Read the manual



Rule of thumb: do not keep your credentials on a public computer. USB memory key is a good choice.

Conclusion

- High Energy Physics is the major consumer of Grid technologies
  - Every HEP researcher sooner or later will have to learn grid basics
  - HEP community invests massive efforts into grid development
  - If grid won't help, it is unclear what would be the "backup solution"
    - The data will eventually be processed, the question is – how soon and how accurate
- Many other sciences are on-looking
  - Bioinformatics and radioastronomy appear to be the next in line
    - Huge data volumes, trivially parallel processing, distributed user base
- See you in the Grid-space!