SiD DBD Production (DIRAC)

Christian Grefe, Stephane Poss,

Jan Strube (CERN)

Jeremy McCormick (SLAC)

Overview

Samples and setup

Introduction to ILCDIRAC and SiD presence on the grid

Lessons Learned

Physics Processes

Samples provided by Common Generators Group (Barklow, Berggren, Miyamoto): all with the correct beamstrahlung spectrum at each energy

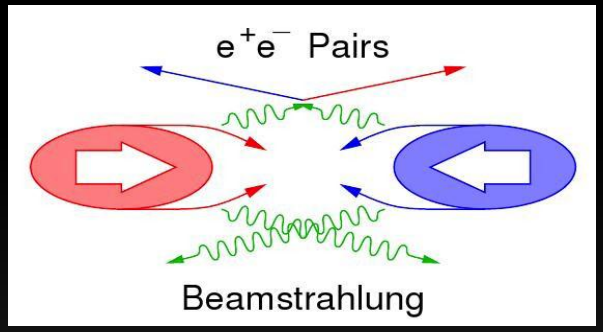
1 TeV (1000 fb^{-1})

- vvH signal (m_H = 125 GeV)
- ttH signal (m_H = 125 GeV)
- WW signal
- 1f - 8f SM background (m_H = 2 TeV)
- γγ → hadrons
- Incoherent pairs

500 GeV (250 fb^{-1})

- Top pairs, m_{top} = 174.0 GeV
- Top pairs, m_{top} = 173.5 GeV
- 6f SM background
- γγ → hadrons

Beam-Induced Background

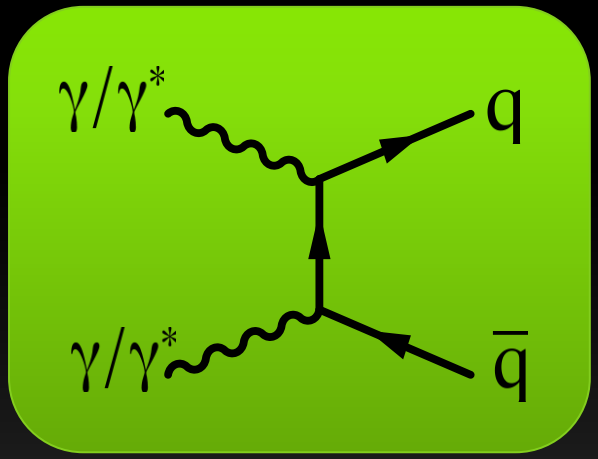


Pair background
1 event per BX
450k particles

Generated by GuineaPig
ascii → hepevt → stdhep

Merged with every "physics" event

MCParticles that don't make hits will be dropped



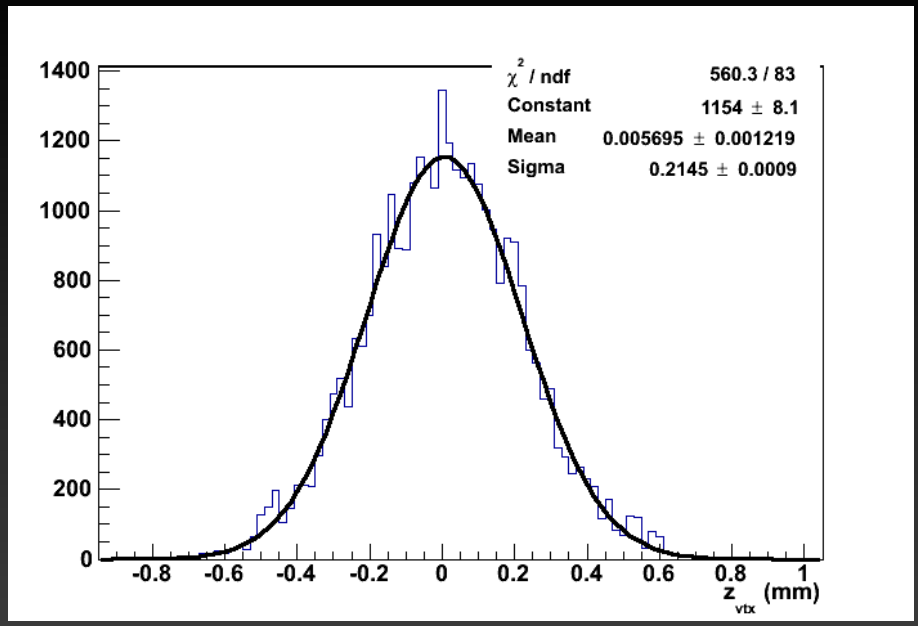
<u>γγ interactions</u>
4.1 events per BX @ 1 TeV
1.7 events per BX at 500 GeV

Generated by Whizard

Luminous Region

- Finite extension: σ_z = 225 μm
  - conservative compromise
- Events from beam-beam interactions (γγ→ hadrons, incoherent pairs) are distributed randomly over the luminous region
- Physics events always at z = 0

Reconstructed primary vertex position for γγ→ hadrons, pairs



Fitted width: 214 μm

Sample Mixing (T. Barklow)

Samples provided 100% polarized

Samples for SiD grouped by processes and luminosity - weighted

Loss of fraction of files does not cause analysis bias

Mixed to correspond to correct polarization

1 TeV (1 ab^{-1}): ± 80% electron, ∓ 20% positron

500 GeV (250 fb^{-1}): ± 80% electron, ∓ 30% positron

Sample Summary

2 Machine Energies x 2 Polarization settings

SM processes containing 1-8 fermions

1000 fb^{-1} @ 1TeV, 250 fb^{-1} @ 500 GeV

Background from incoherent pairs and hadron events from photon interactions

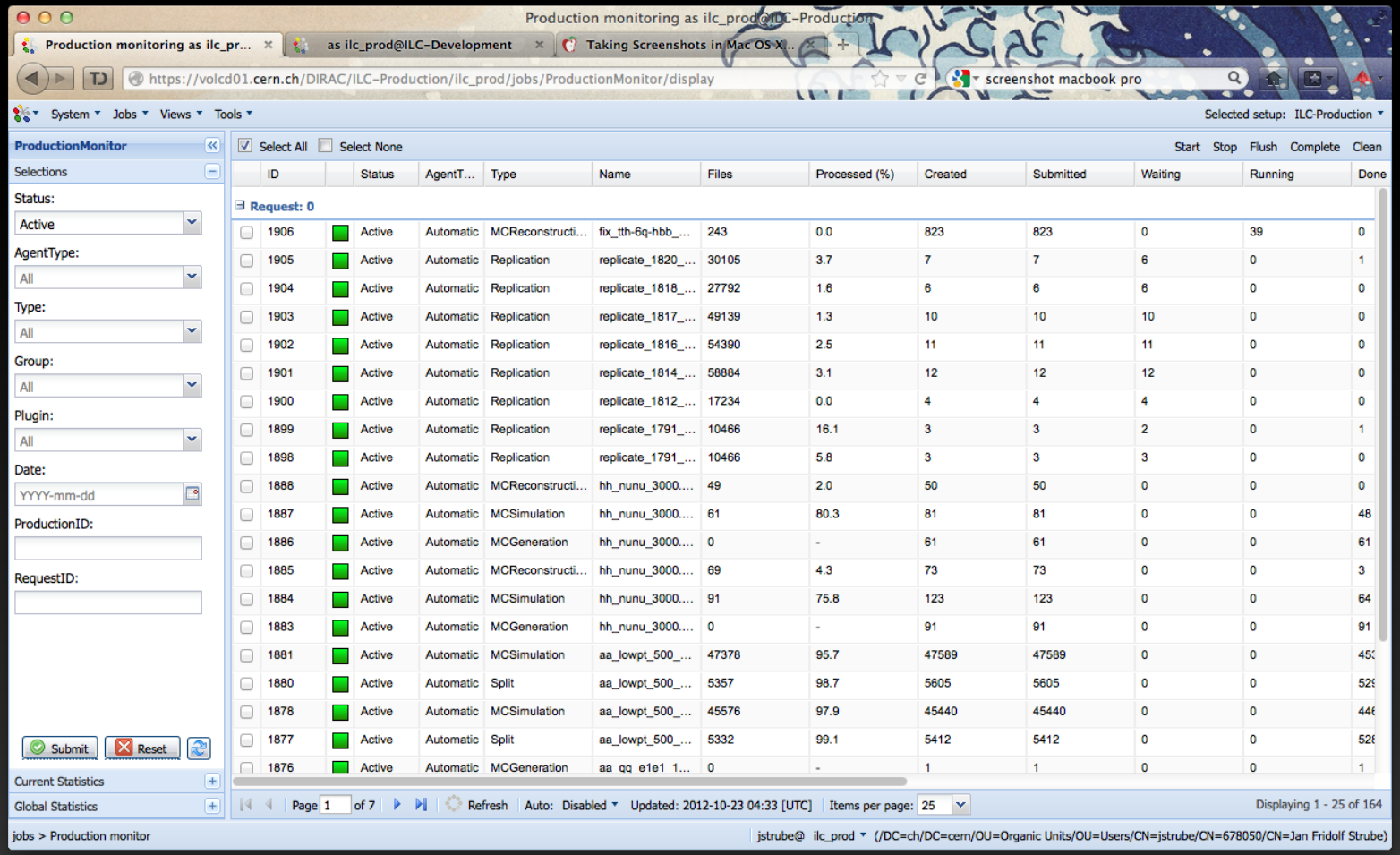
66,421,842 events in 60 categories

Production summary on SLAC confluence

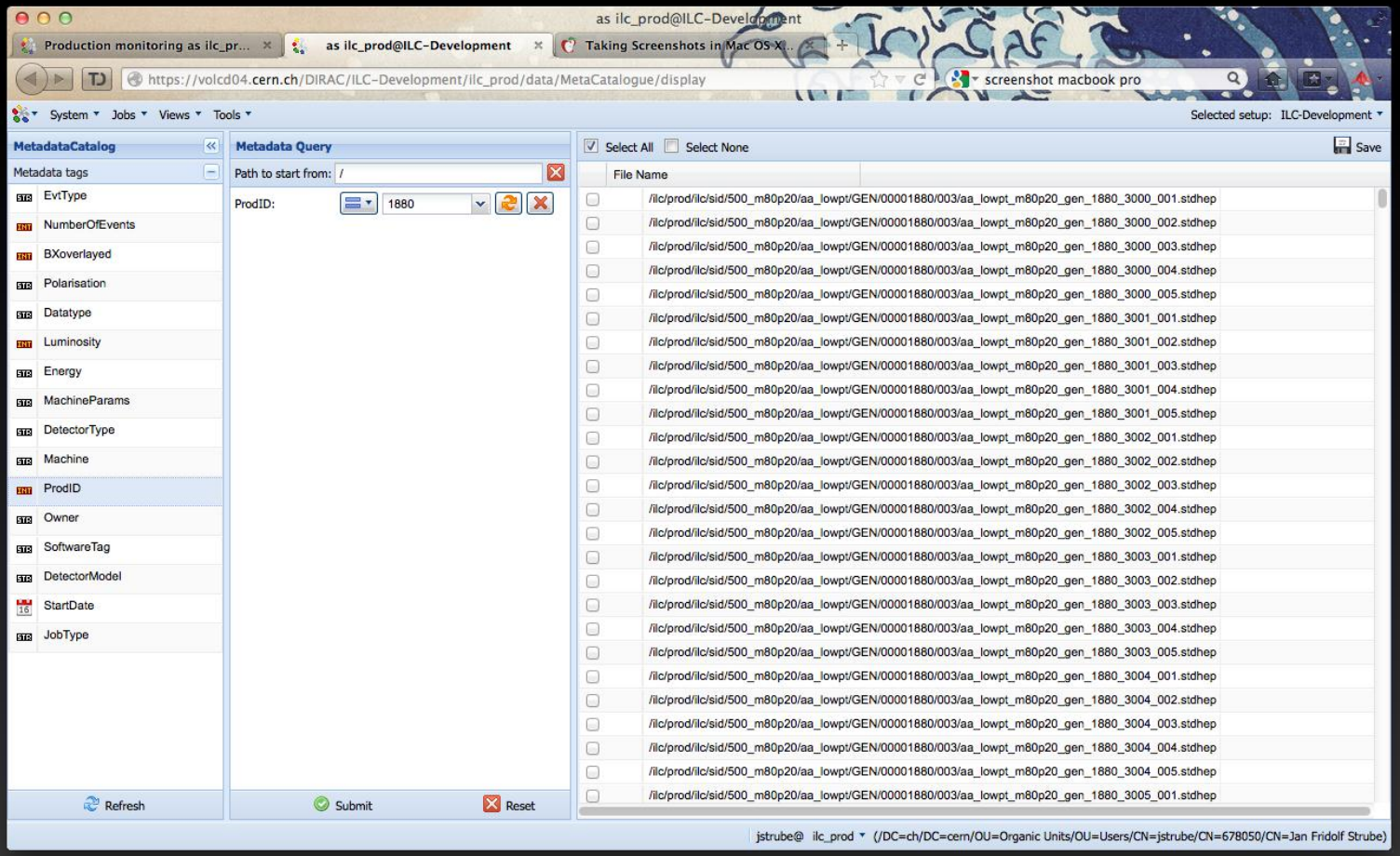
ILCDIRAC

- Dirac system used in LHCb in Production, consists of
  - File catalog (First used in CLIC CDR)
    - Supports meta data (file ancestry, detector model, ...)
  - Job submission, monitoring and bookkeeping
- ILCDIRAC (S. Poss et al.) developed for CLIC CDR production
  - Support for the plethora of ILC software
- Developed and maintained at CERN

Web-based Job Monitor



Meta Data Interface



Supported Software

Software is modularized in Dirac

Sets the context of the program (env vars, dependencies)

Allows to chain different modules together

Currently supported Physics applications:

Whizard, Pythia, Mokka, Marlin, PandoraPFA, SLIC, slicPandora, lcsim, etc.

Mix and match, supply your own steering files

ILC VO

Before Summer 2012:

US colleagues:

Open Science Grid

ILC VO managed at Fermilab

Grid authorization prevents data exchange

European / Asian colleagues:

Worldwide LHC Computing Grid

ILC VO managed at DESY

Now:

Virtual Organizations have been merged

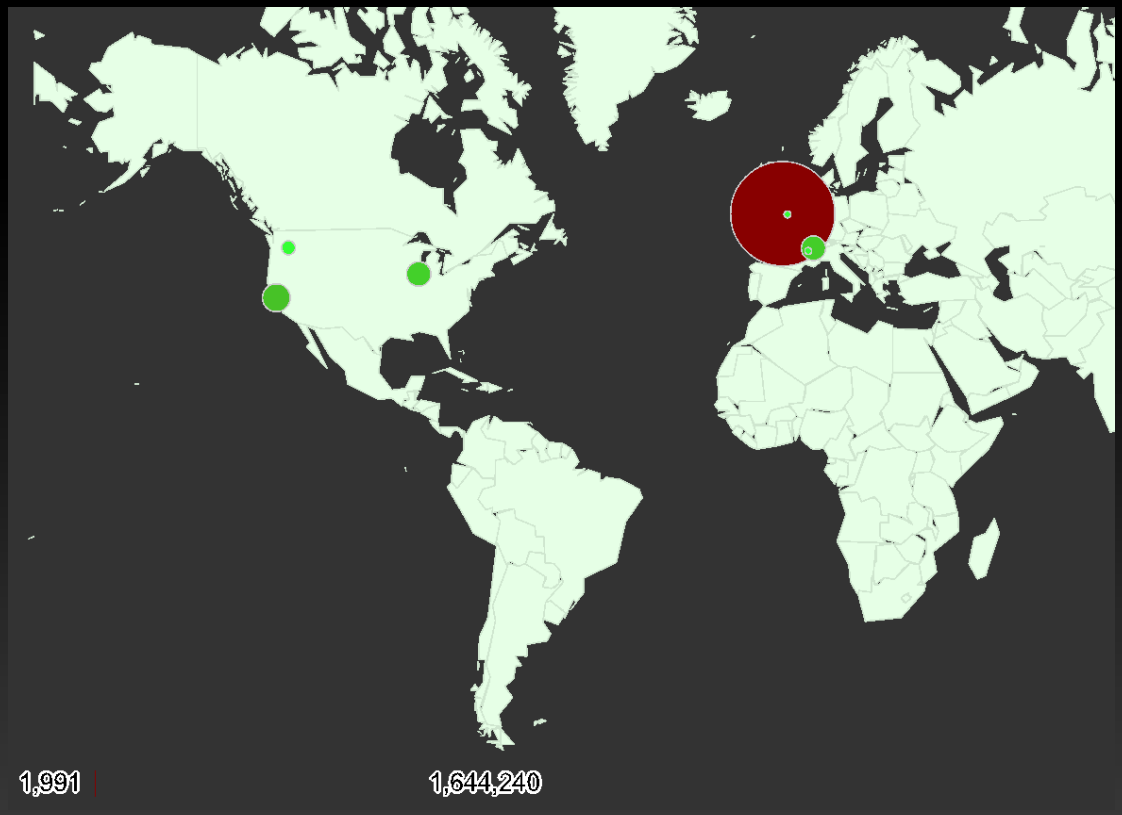
Actively exchanging computing and storage resources

Both, OSG and WLCG sites supported in DIRAC through gLite

Storage Elements

RAL has been largest grid site for SiD since the LOI production
We could not have done the production without being permitted to use 300% of our allocation

PNNL has started making large storage and CPU resources available, able to host one copy of whole production



Total Storage: 127 TB -- 2,681,083 Files

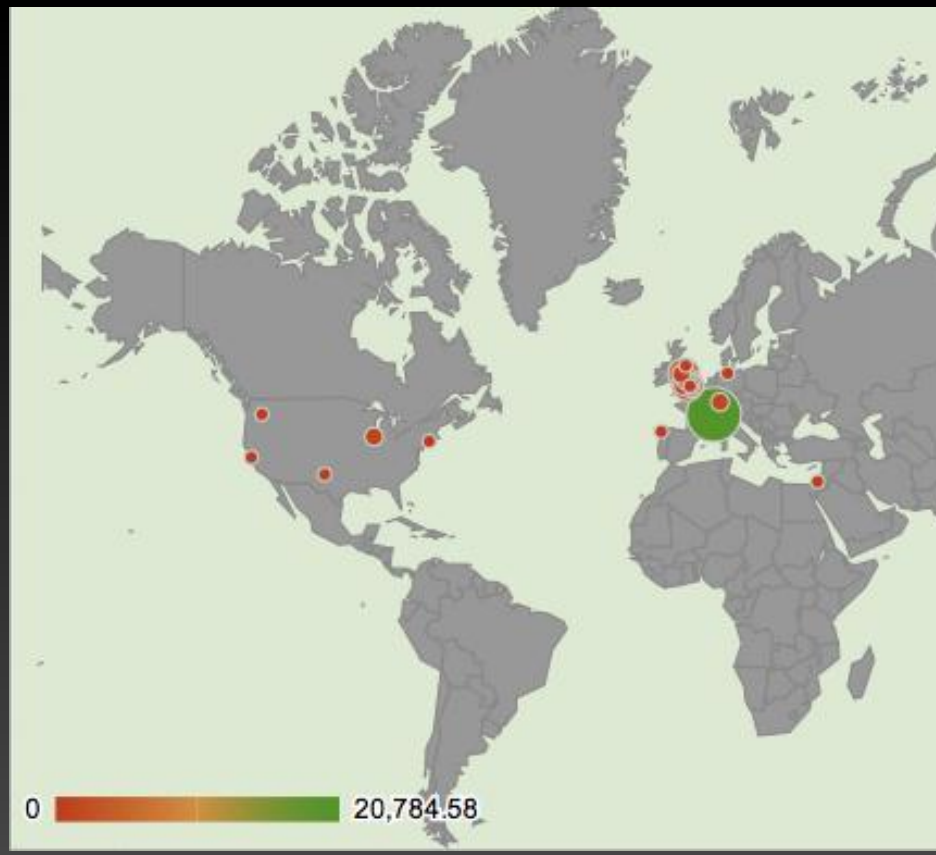
Computing Elements

During LOI: 70% of production at SLAC farm, other 30% used to gain experience at various grid sites across Europe

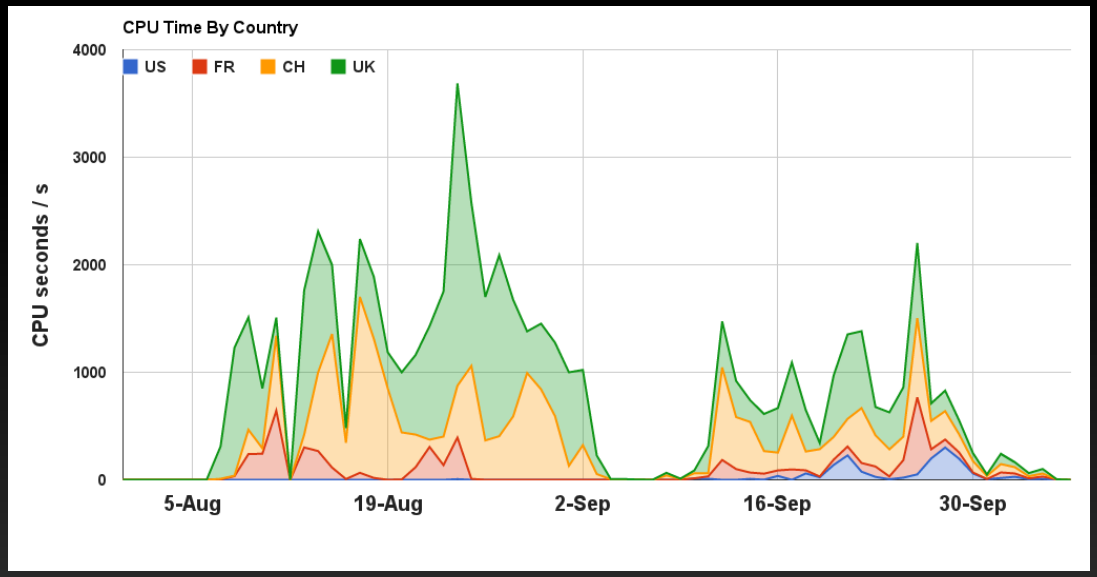
Tremendous benefit from LOI and CDR experience

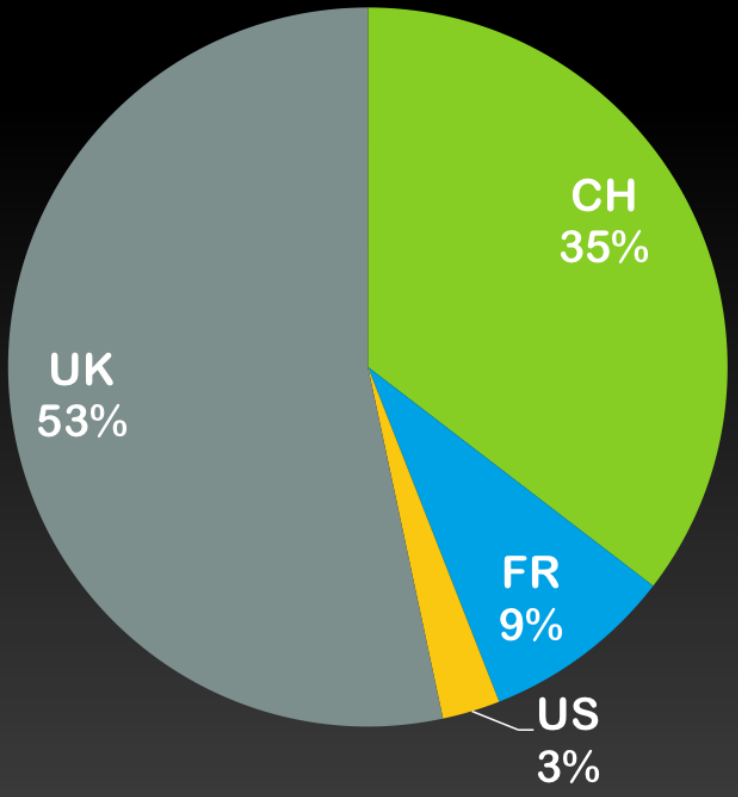
Production now simply scales with additional sites

Several limitations of the ILCDIRAC CDR setup have been lifted



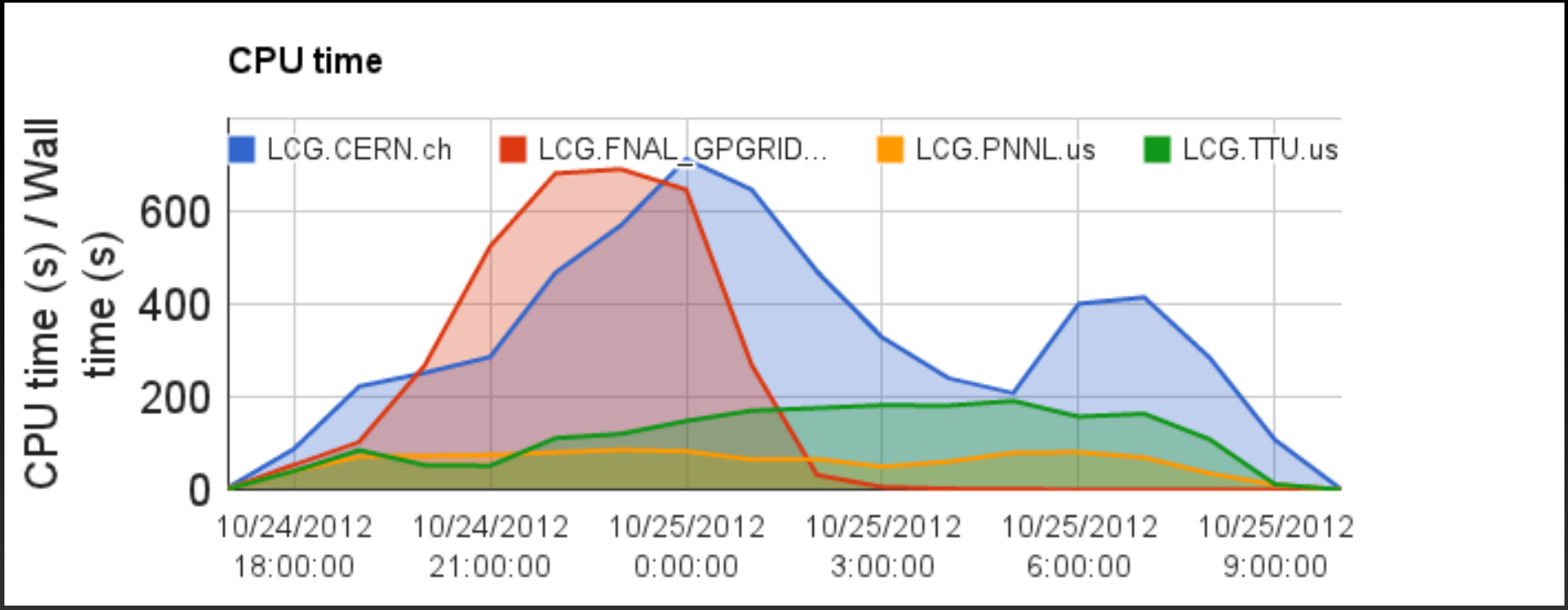
CPU Time by Country





Country	CPU Time (years)
UK	85.8
CH	56.9
FR	13.9
US	4.2

Production update



US sites are now configured to take jobs

Increasing efficiency very time-intensive, PNNL only US site with manpower to liaise

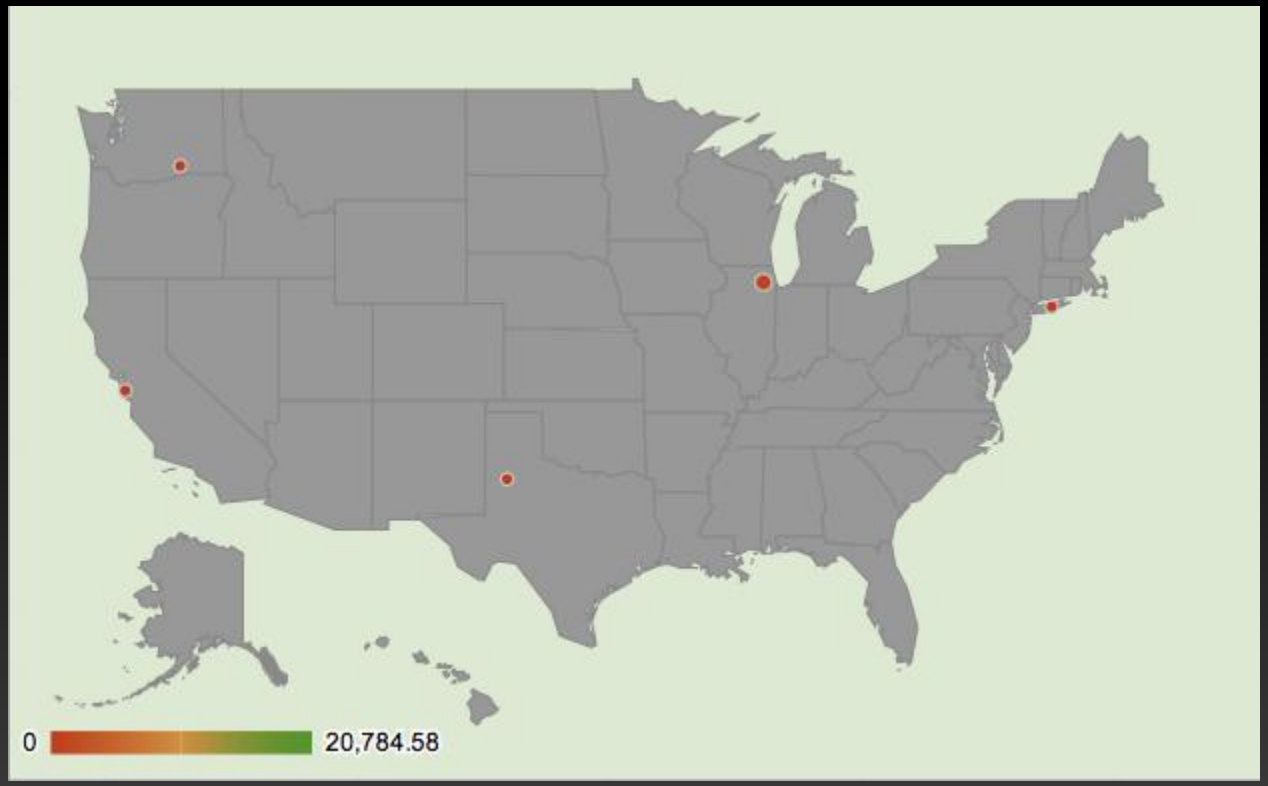
Conclusions / Lessons Learned

- ILCDIRAC production system was further improved in terms of efficiency and error handling
- Bookkeeping is most time consuming part of production
  - Automate as much as possible
- Site configuration problems (wrong defaults) can lead to deterioration of the overall performance
  - Optimal performance during production requires continuous monitoring
  - Every site is different
- Even with the best tools, running a production is a manpower-intensive task

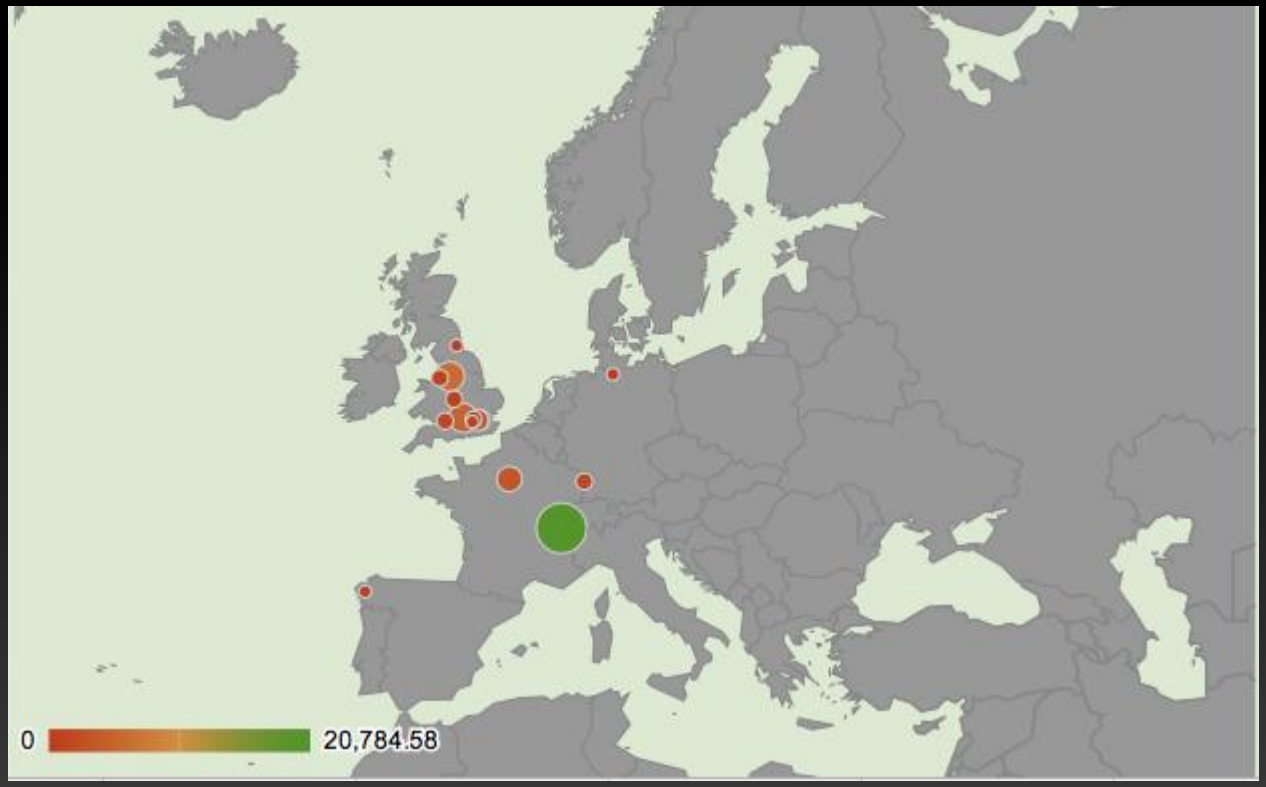
Summary

- More than 66 million events have been processed in about 6 weeks
  - Re-processing stage currently on-going, estimated 3 weeks to finish
- The Fermilab and DESY ILC VOs have been merged
  - ILCDIRAC supports sites in OSG and in WLCG
  - Talk to us if you have idle resources, default setup is likely to cause problems
- All data produced in the context of the SiD DBD and all tools are freely available to anybody
  - Theorists, or Students can perform analysis on realistic simulation of the detector concept at a 1 TeV ILC
  - Talk to us if you would like to participate

Backup

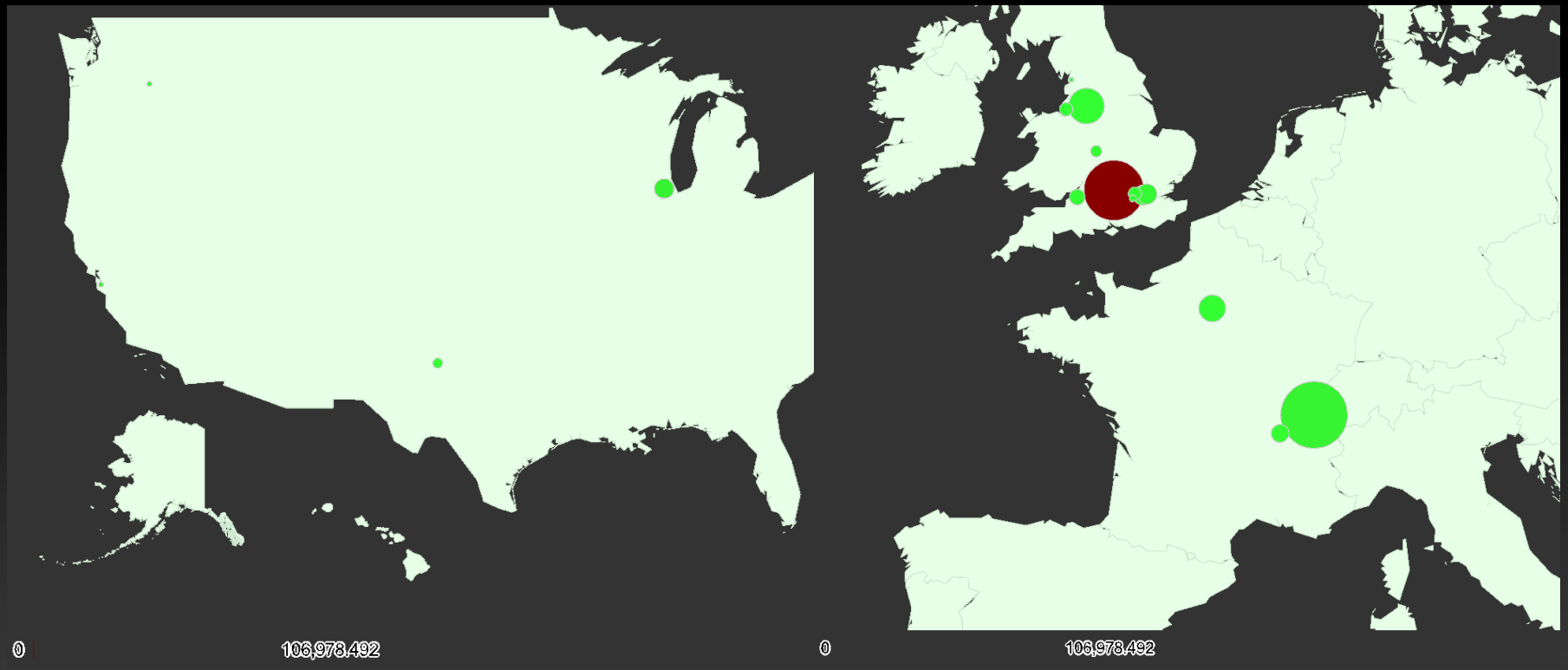


SiD Production Status LCWS 2012



SiD Production Status LCWS 2012

Global Distribution of Resources



Size of circles proportional to CPU time

Color: Amount of storage used during production