Benchmarking Production

Jan Strube (CERN)
for the SiD Production team

First and Foremost: People

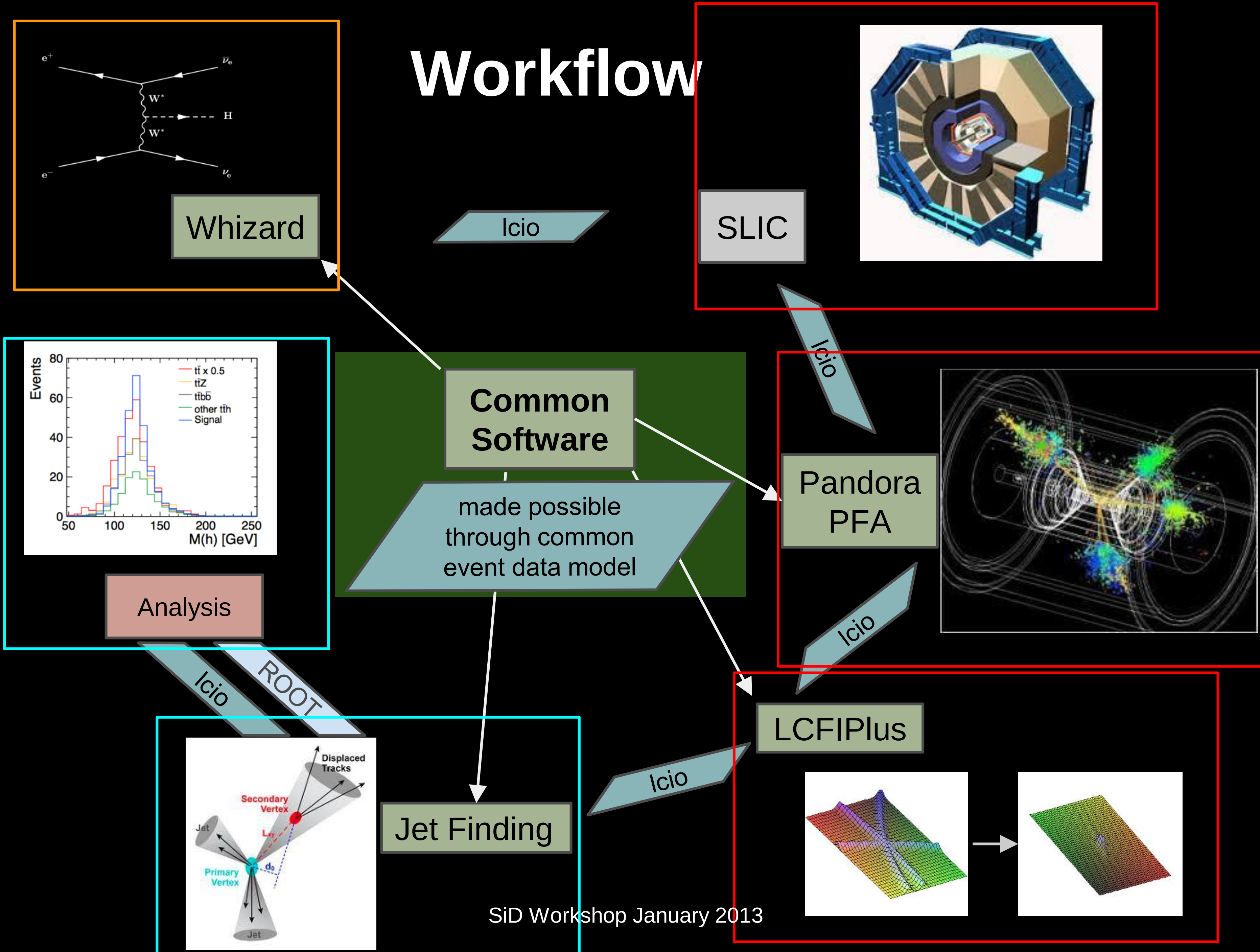
SLAC: Tim Barklow, Norman Graf, Jeremy McCormick

PNNL: David Asner, David Cowley, Brock Erwin, Malachi Schram

CERN: Christian Grefe, Stephane Poss, J. S.

Overview

- Software
- Physics Processes / Backgrounds
- Grid Production
- Lessons Learned / Summary


Workflow
e^+
$\bar{\nu}_e$
W^*
H
W^*
e^-
ν_e
Whizard
lcio
SLIC
Events
80
60
40
20
0
50
100
150
200
250
M(h) [GeV]
$t\bar{t}$ x 0.5
$t\bar{t}Z$
$t\bar{t}b\bar{b}$
other $t\bar{t}h$
Signal
Analysis
Common Software
made possible through common event data model
lcio
Pandora PFA
lcio
LCFIPlus
lcio
ROOT
Displaced Tracks
Secondary Vertex
Jet
L_{xy}
Primary Vertex
d_0
Jet
Jet Finding
lcio
SiD Workshop January 2013

Physics Processes

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

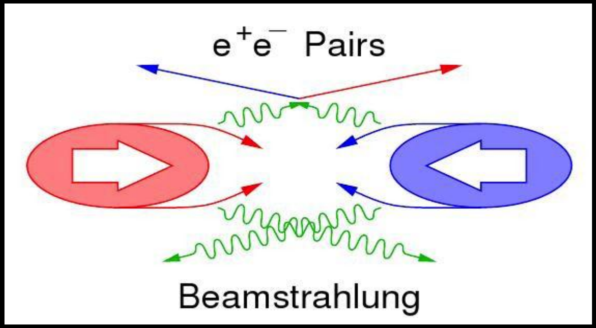
1 TeV (1000 fb^{-1})

- vvH signal (m_H = 125 GeV)
- ttH signal (m_H = 125 GeV)
- WW signal
- 1f - 8f SM background (m_H = 2 TeV)
- $\gamma\gamma \rightarrow$ hadrons
- Incoherent pairs

500 GeV (500 fb^{-1})

- Top pairs, m_{top} = 174.0 GeV
- Top pairs, m_{top} = 173.5 GeV
- 6f SM background
- $\gamma\gamma \rightarrow$ hadrons

Beam-Induced Background

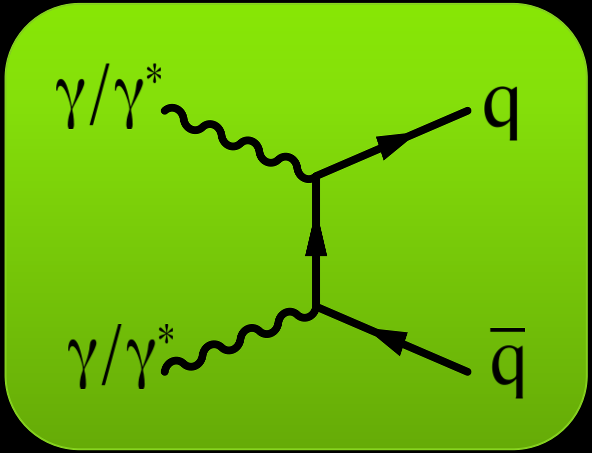


Pair background
1 event per BX
450k particles

Generated by GuineaPig
ascii → hepevt → stdhep

Merged with each "physics" event

MCParticles that don't make hits are dropped



γγ interactions

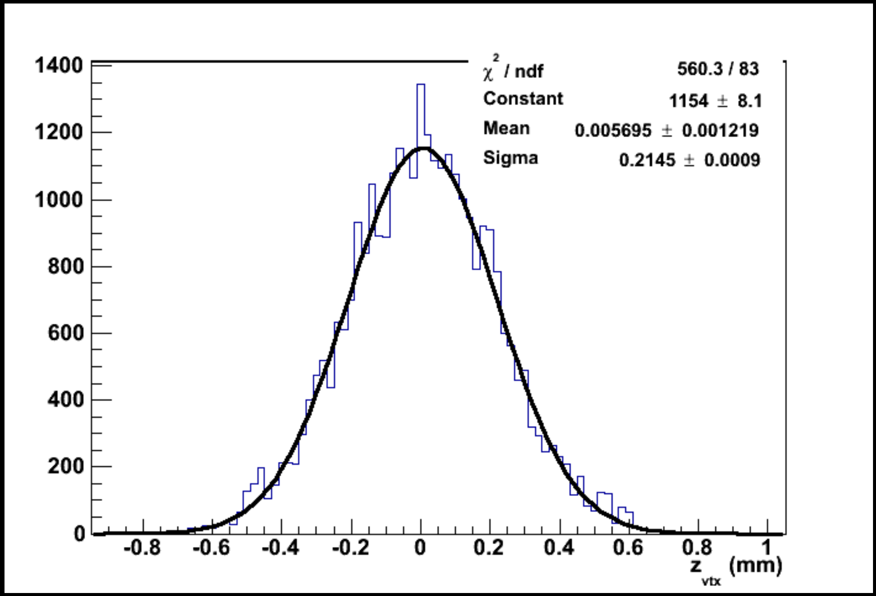
4.1 events per BX @ 1 TeV
1.7 events per BX at 500 GeV

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



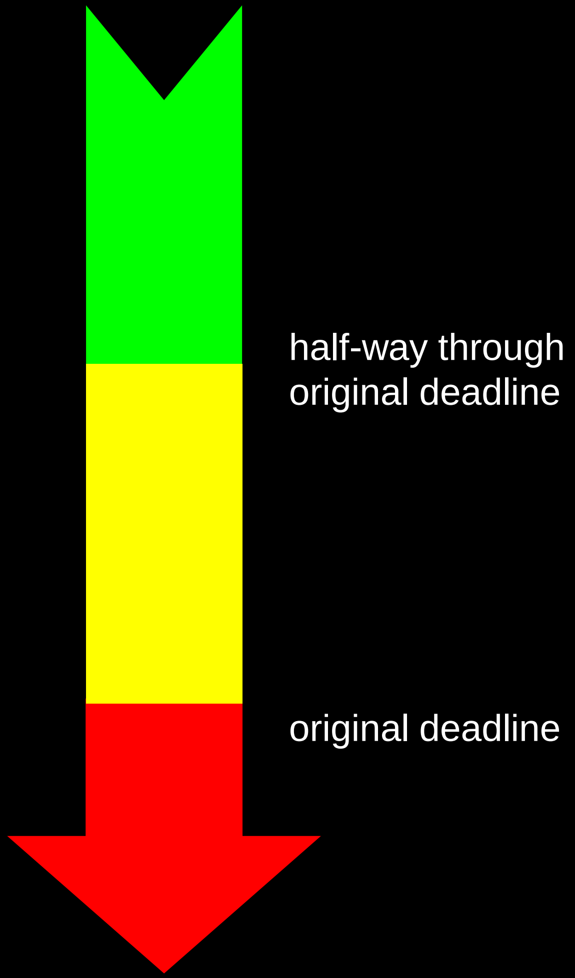
Fitted width: 214 μm

Sample Mixing (T. Barklow)

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

A Timeline of Events

20-Oct-2011	Start
10-Feb-2012	Beam Parameters chosen
15-Mar-2012	MC Integration complete
12-Apr-2012	MC Generation done
26-Jul-2012	Simulation Budget
7-Aug-2012	Production Start
3-Sep-2012	Z smearing chosen
22-Oct-2012	LCWS 2012
22-Oct-2012	Second Iteration Start
19-Nov-2012	Production Done
21-Jan-2013	Current Deadline



Production in Numbers

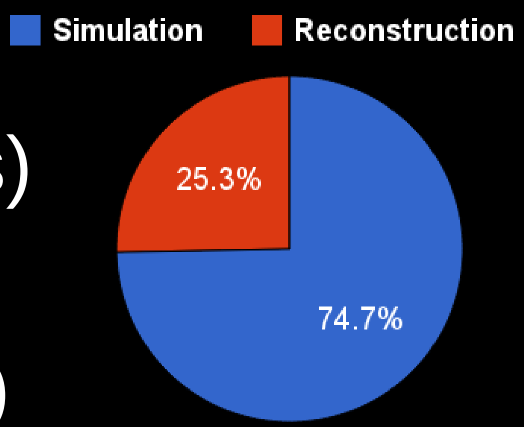
Production summary on SLAC confluence

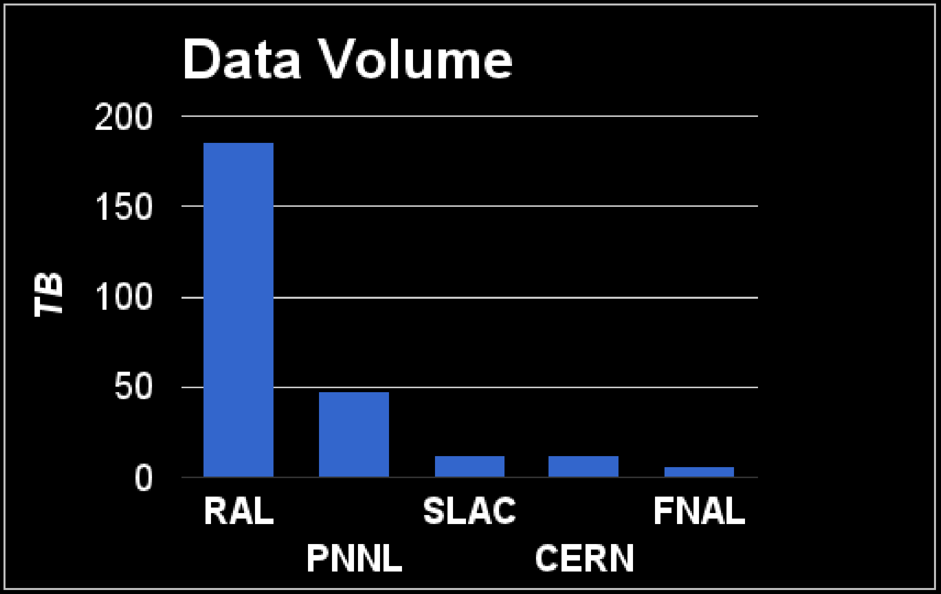
50,746,683 events at 1 TeV

(+ 4.7 million gghadrons)

6,550,022 events at 500 GeV

(+ 4.4 million gghadrons)





Country	Total CPU Time (years)
UK	100.2
CH	68.2
FR	15.0
US	28.2

SiD on the Grid

Open Science Grid (OSG) resources are new addition to the SiD resource pool

dedicated resources: PNNL, SLAC, CERN
temporary quota increase: FNAL, RAL Tier 1
opportunistic use: all others

Worldwide LHC Computing Grid (WLCG) resources have been established during LOI and CLIC CDR efforts

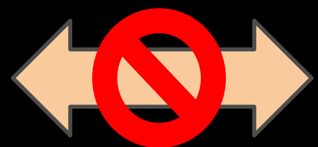
ILC Virtual Organization

<u>Before Summer 2012:</u>

<u>US colleagues:</u>

Open Science Grid

ILC VO managed at Fermilab



Grid authorization prevents data exchange

<u>European / Asian colleagues:</u>

Worldwide LHC Computing Grid

ILC VO managed at DESY

<u>Now:</u>

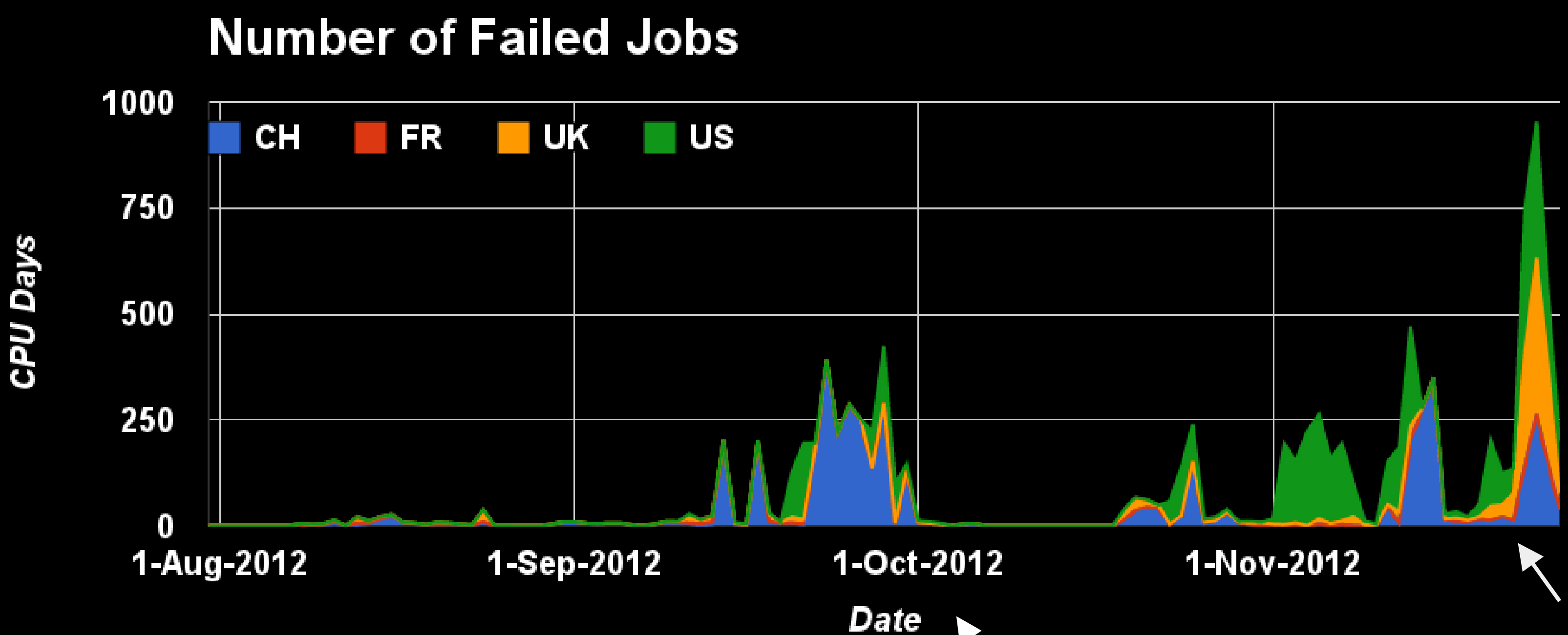
Virtual Organizations have been merged

Actively exchanging computing and storage resources

Both, OSG and WLCG sites supported in DIRAC through gLite

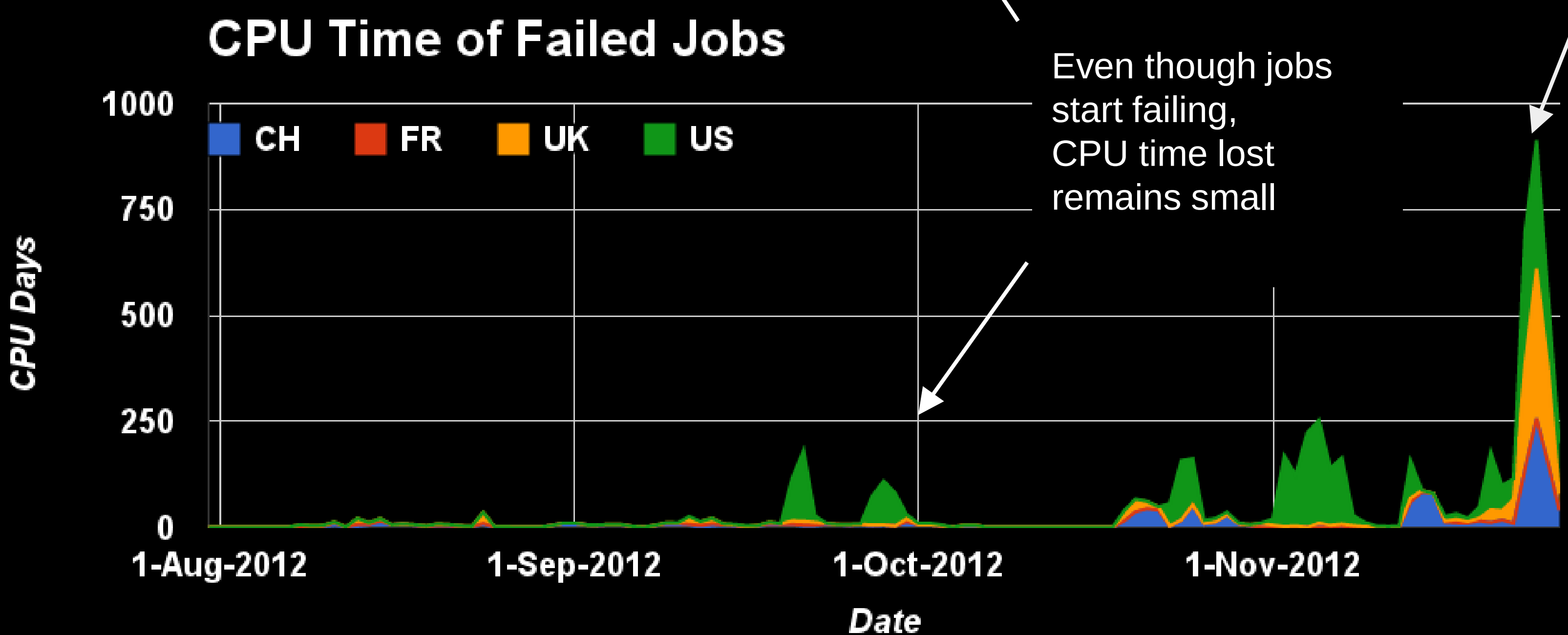
ILCDIRAC

- 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

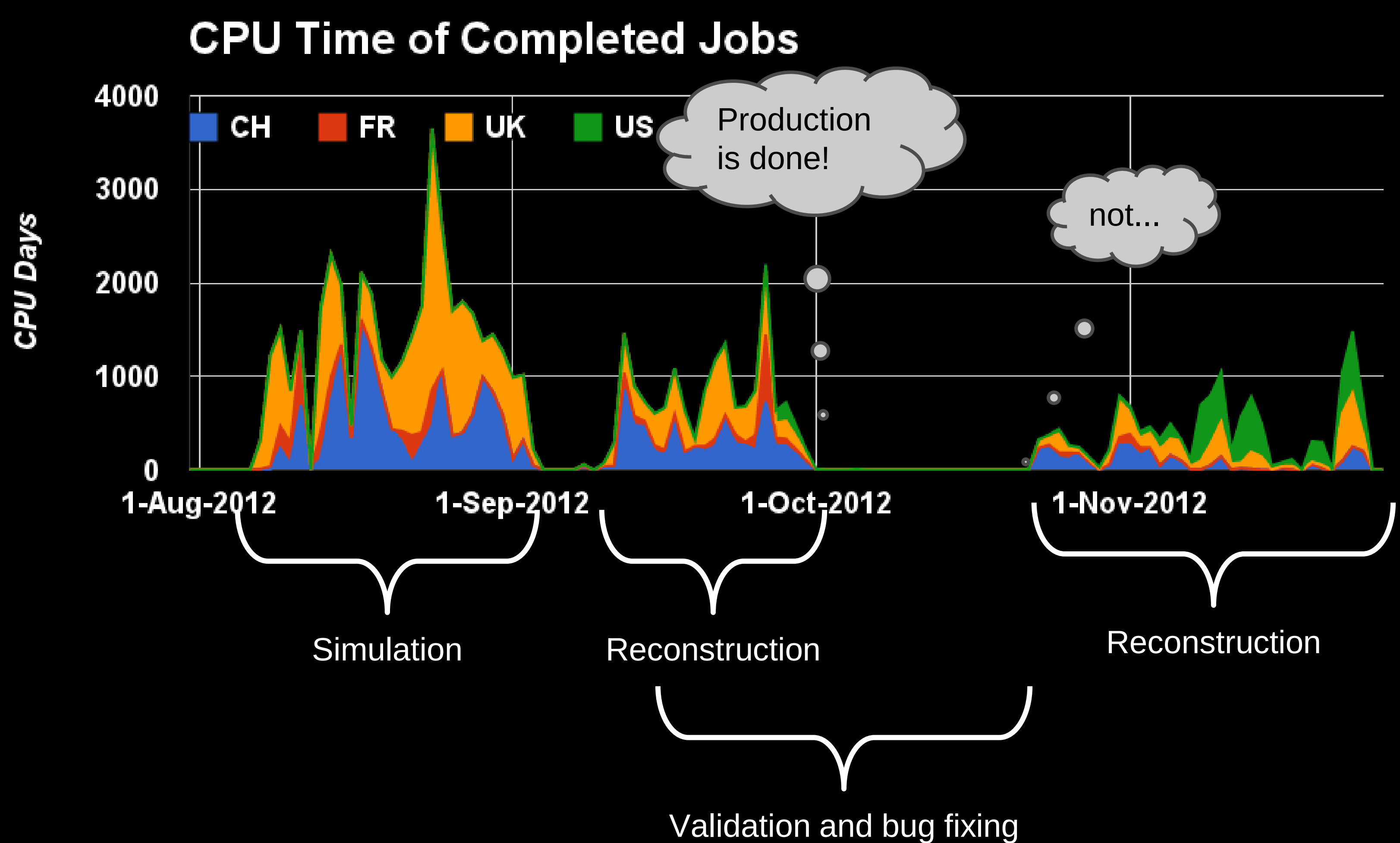


Fire at RAL

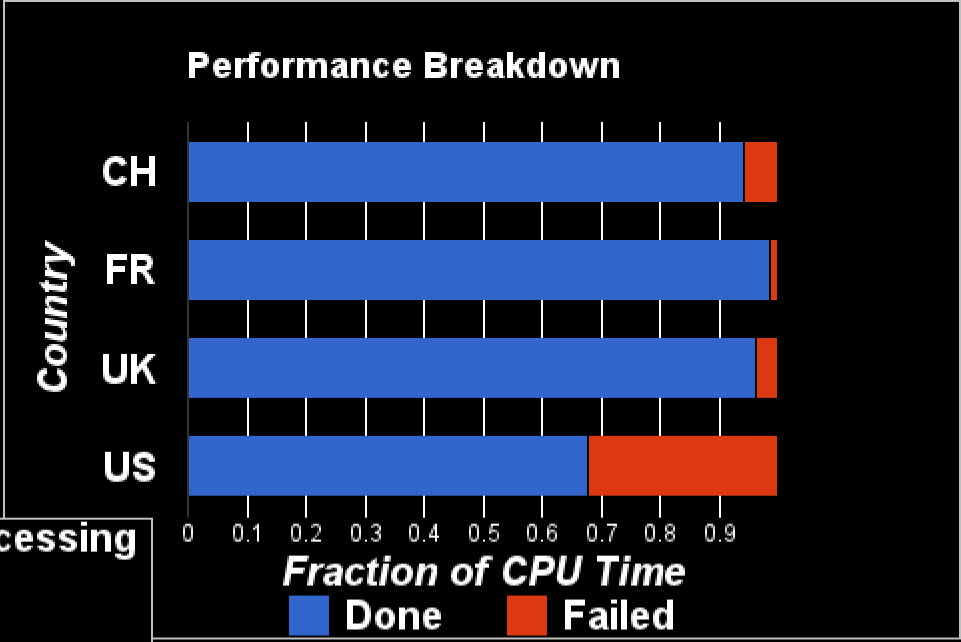
Even though jobs start failing, CPU time lost remains small

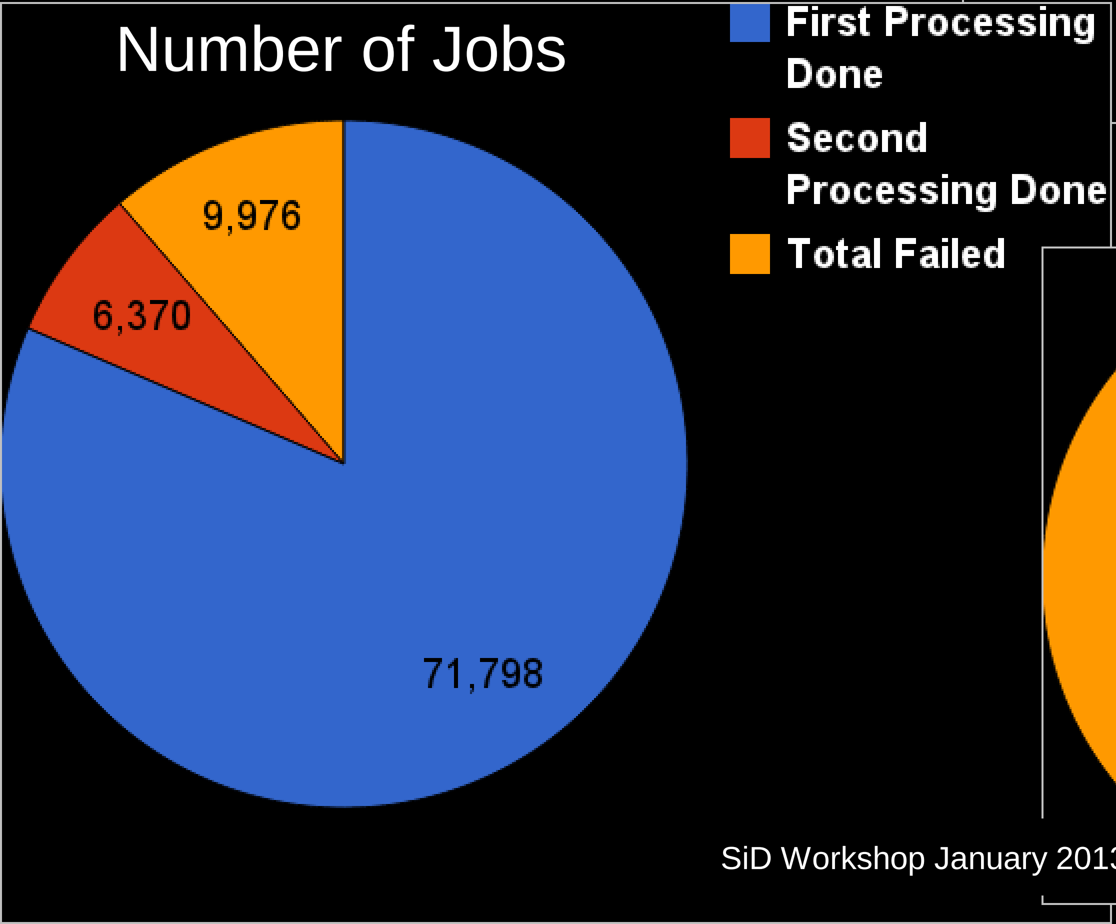


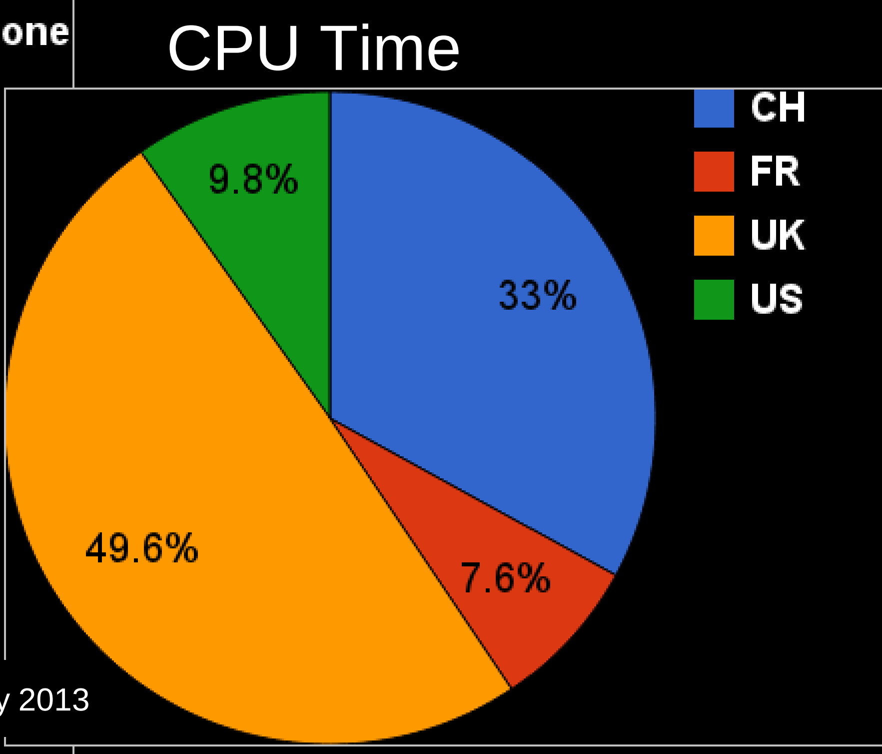
Production Timeline



Performance Summary







Main Reasons of Lost Time

Problem	Analysis	Resulting Damage
Fraction of events had single muon as "physics event"	Bad Default Behavior Changing software versions in production setting w/o validation	~ 4 million events to be reproduced (simulation and reconstruction)
ECAL hits have wrong energy	Advertised Software version is buggy, although fix exists Changing software versions in production setting w/o validation	Event mixing and reconstruction of all events was re-done
Installation of OSG resources at PNNL / FNAL	Trying to break new ground while running a major production	Large number of failed jobs / drain of resources for support
Fire in UPS room at RAL	Lack of redundancy	3 days lost

Conclusions / Lessons Learned

- CPU time is not wall time. Faster processing does not necessarily mean faster time to finish.
- Every time software gets changed without validation, somewhere a kitten dies

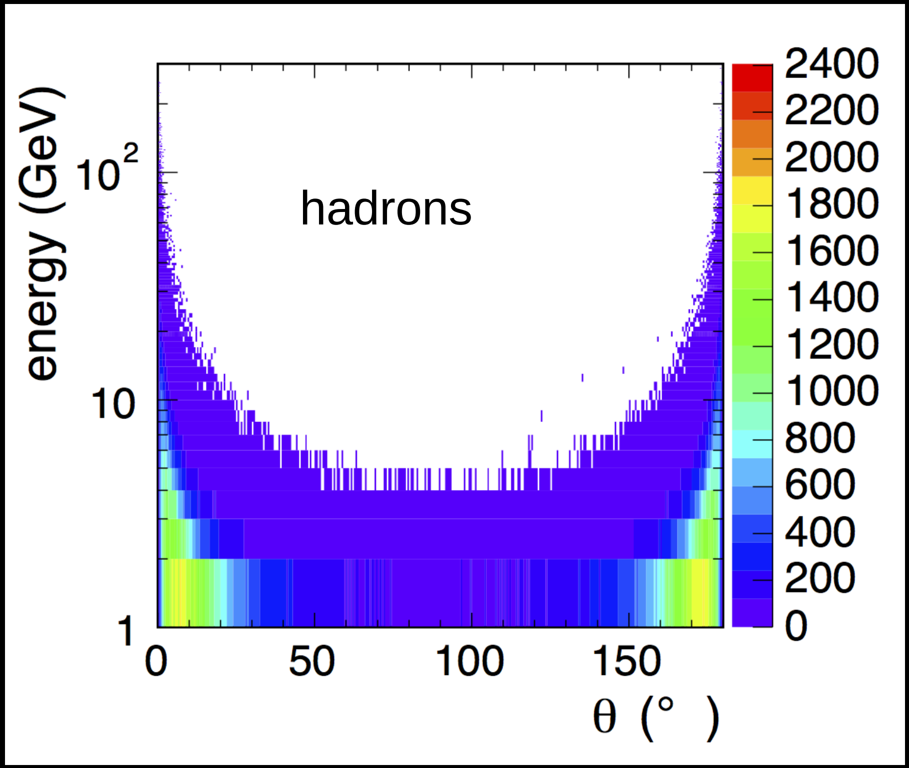
- A large part of the production was run under high pressure. Combined with lack of experience, this led to some loss of time.
- On the other hand, mistakes were spotted because of hard work and diligence of individuals
- Even with the best tools, running a production is a manpower-intensive task

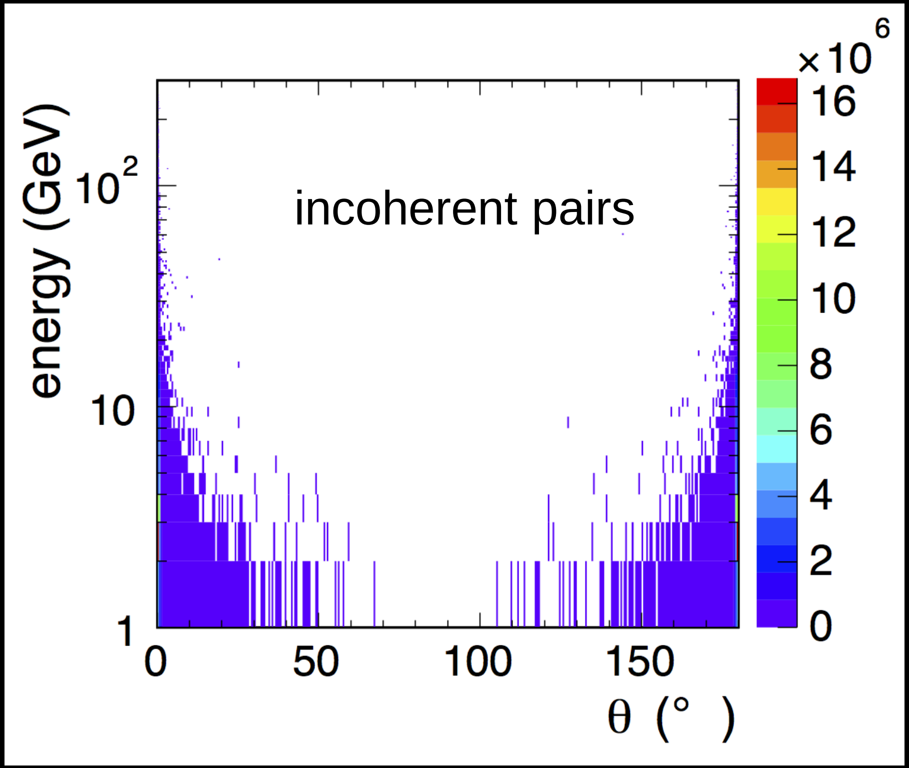
Summary

- More than 57 million events have been processed in about 6 weeks
  - Re-processing stage took a bit under 4 weeks (3 estimated)
- The Fermilab and DESY ILC VOs have been merged
  - ILCDIRAC supports sites in OSG and in WLCG
  - FNAL quota back to pre-DBD, but significant resources now available at PNNL
- Development of ILCDIRAC, event mixing tools and grid experience developed at CERN during CLIC CDR was invaluable
- This was not a smooth experience. We need to establish (and follow) procedures for how to treat our software and our data with more respect.
- The DBD production got done by the effort of dedicated individuals

Backup

Angular distribution of background





Incoherent pairs affect mostly occupancies and tracking efficiencies

Hadrons have enough energy to reach the calorimeter

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