

Jan Strube
Rutherford Appleton Laboratory

SID ON THE GRID

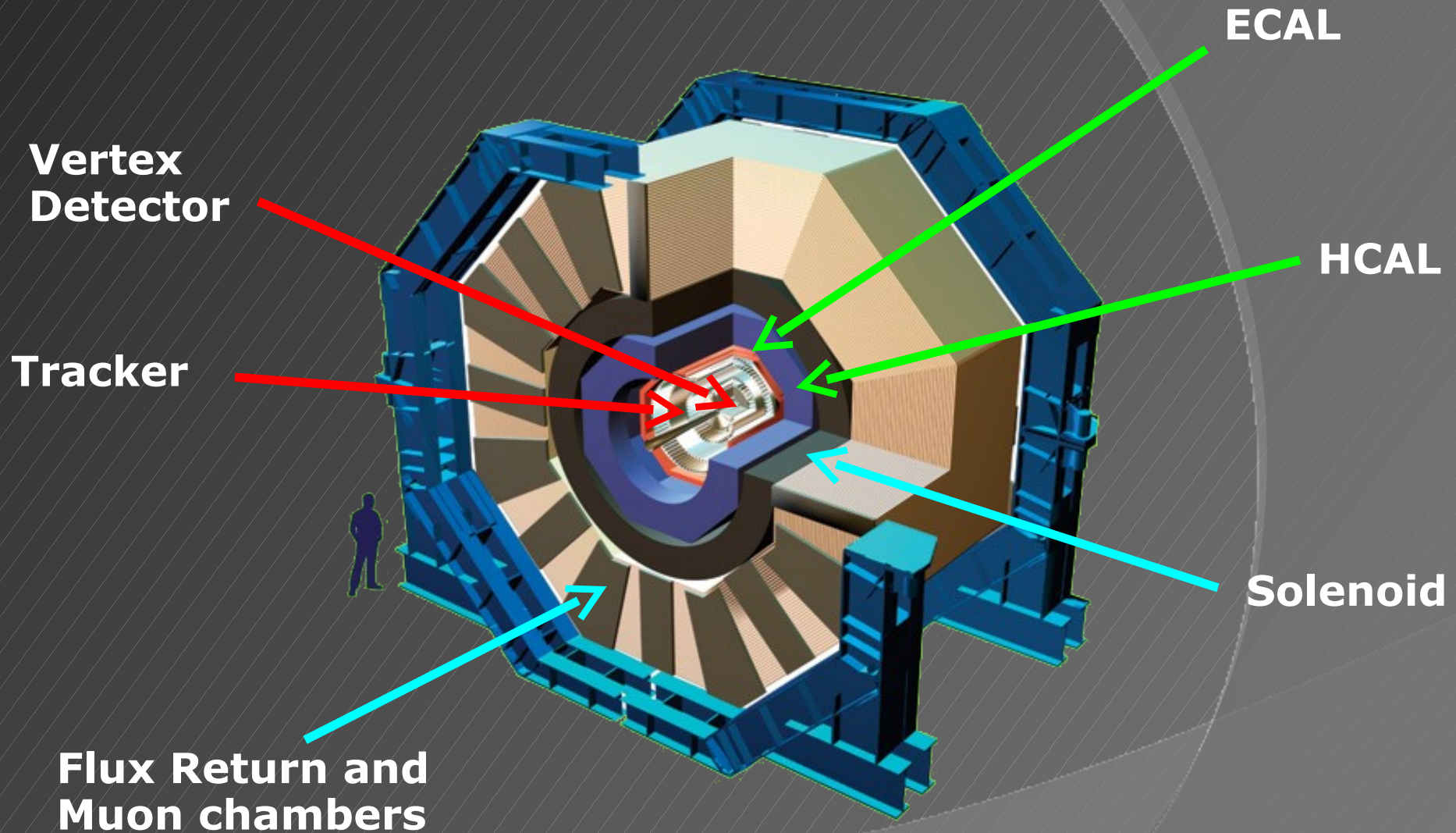
The SiD collaboration

- ⦿ Silicon detector concept
 - Pure Silicon tracking
 - Si/W calorimeter
- ⦿ 5 T field
- ⦿ Highly granular calorimetry
 - Particle Flow Algorithms (PFA) for superior jet energy resolution
- ⦿ ~ 250 people, 80 collaborating institutions

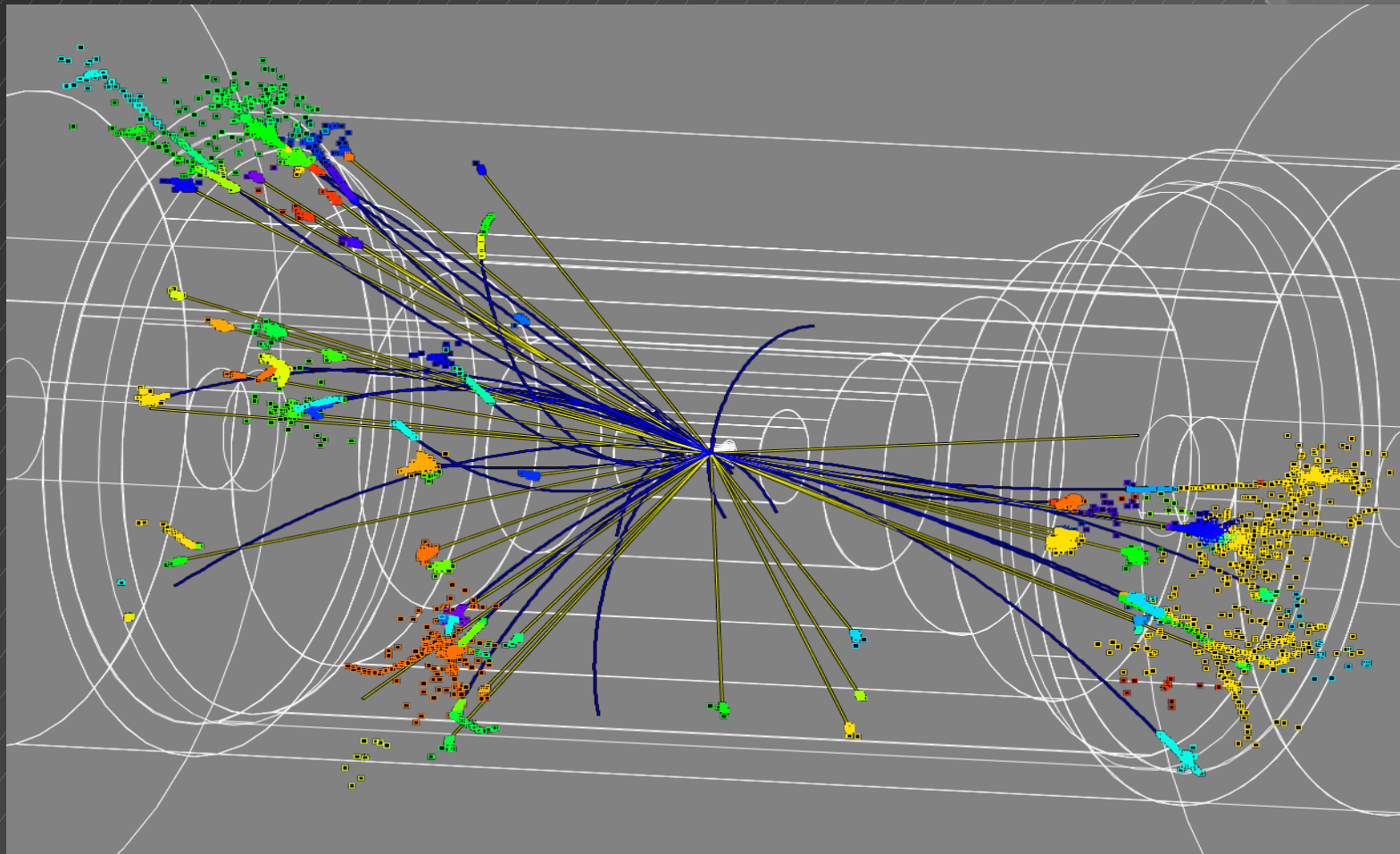
Detector performance

- Calorimeter granularity
- Pixel size
- Material budget, central
- Material budget, forward
- Impact parameter resolution
- Track momentum resolution
- Jet energy resolution
- ~ 200x better than LHC
- ~ 20x smaller than LHC
- ~ 10x less than LHC
- ~ 100x less than LHC
- 3x better than SLD
- 10x better than LEP
- 3x better than CMS
- 2x better than ZEUS

SiD – the detector



A typical ILC event



Data Rates

- ⦿ 1 BX + several underlying events every 25 ns
- ⦿ 100s of particles in every event
- ⦿ Event size: few MB (including MC)
- ⦿ 200 Hz trigger rate

LHC

- ⦿ 2860 BX every 200 ms
- ⦿ 10s of particles in every event
- ⦿ Event size: ~50 kb
- ⦿ 5 Hz trigger rate

ILC

The LOI effort

- ⊙ International Detector Advisory Group (IDAG) had the mandate to investigate the feasibility of 4(3) detector concepts.
- ⊙ Benchmark analyses
 - Produce several years worth of data
 - + speculative signals
 - + Backgrounds we may(!) never have to worry about
 - + at different collision energies
 - 73 Million events to produce, simulate process

Data Flow

Event Generation (whizard)

- particles

stdhep

Simulation (GEANT4)

- Detector hits

lcio

Reconstruction (org.lcsim, MarlinReco)

- Tracks, clusters, jets, flavor tagging

lcio

Analysis (AIDA, ROOT)

- plots

LCIO

- ⦿ Linear Collider IO
- ⦿ Common Event Data Model for the ILC experiments
- ⦿ Allows sharing of tools and reconstruction algorithms between the experiments
 - This, and the fact that our ReconstructedParticles are geometry-independent

Production setup 2007

- ⦿ FTP (wget, curl)
- ⦿ bsub/qsub
- ⦿ NFS
 - Huge advantage in accessibility of log files
 - POSIX access
- ⦿ Limited by CPU availability

Production setup this time

- ⦿ Technical terms (space token, ...)
- ⦿ Castor
- ⦿ Get grid certificate
- ⦿ Sign up with VO
 - Not more than one, though!
- ⦿ lcg_cp, lcg-cp?
 - Google is not always your friend
 - GridPP pages are a much better entry point (now)
- ⦿ File Checksums
 - not something I worried about on NFS or FTP
- ⦿ Limited by disk throughput
 - (Only used Tier1 storage)

Production with User Tools

- ⊙ Started out with glite-wms-job-submit
 - Gradually wrapping this in scripts

→ This doesn't scale for production

- ⊙ Ganga

- I like python, quite easy to get started
 - + Keeps track of all your jobs
 - Keeps track of **all** your jobs

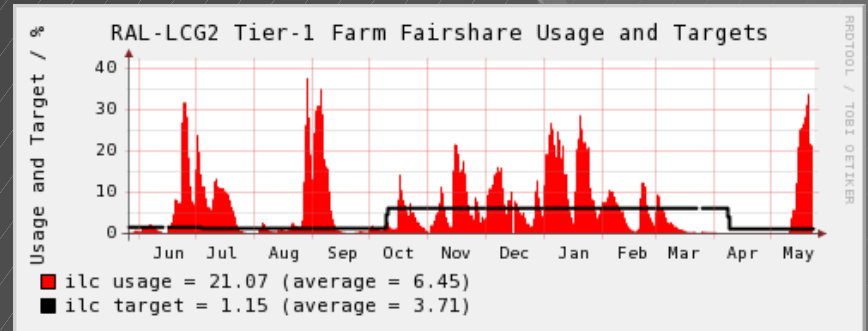
→ This doesn't scale for production, either...

Differences in setup

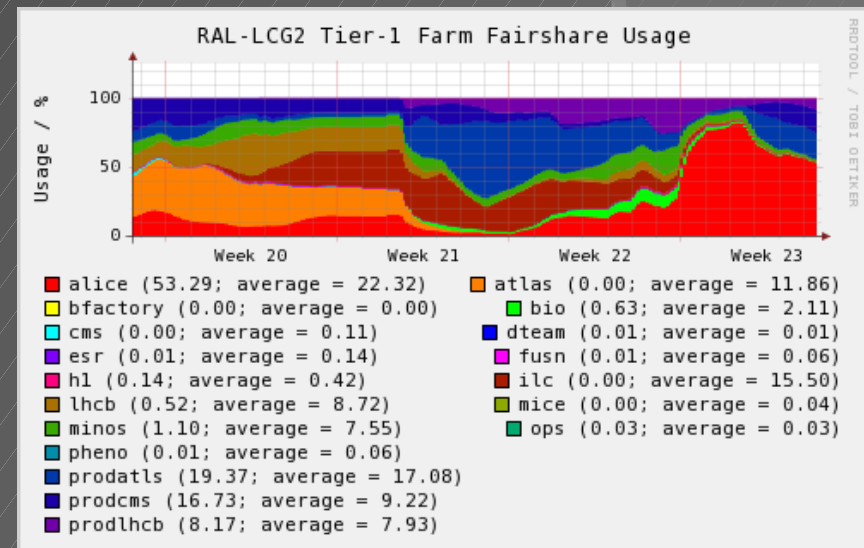
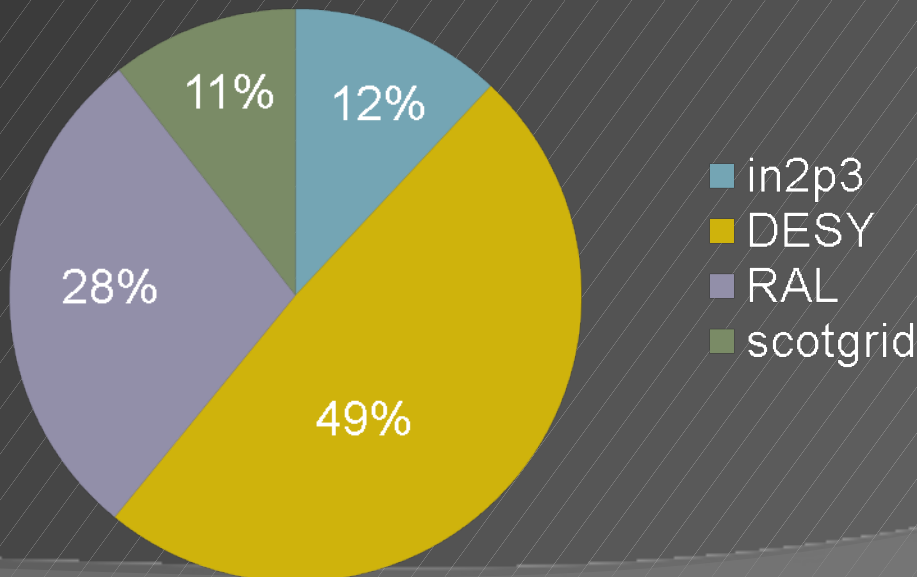
- ⊙ File catalog vs. File system as database
 - For small teams the latter has lower overhead
- ⊙ Production tools
 - LHC experiments use their own
- Grid computing means distributed computing
 - Bookkeeping across multiple sites
 - *Grid computing requires more discipline*
 - *File Checksums, job status, file copy status(!), software validation on each site*

Production Stats

Production Step	# events
Simulation (SLiC)	42,611,000
Reconstruction	31,991,000
Flavor Tagging	20,446,000



CPU times (for the jobs for which I have this info)



Summary

- ⦿ Castor reliability
- ⦿ Allocating storage for our VO
- ⦿ Enforcing sandbox quotas
 - Yeah, that was me. Sorry!
- ⦿ Not all sites published accurate capabilities for the queues
 - Jobs get stuck, this is worse than crashing
- ⦿ Fairshare
 - Several dozen times above allocation
- ⦿ Gridpp Support team
 - Thank you!
- ⦿ Docs improved quite a bit

Things that didn't go so well

Things that went quite well

Conclusion

- ⦿ The SiD LOI production effort couldn't have been carried out without grid resources
- ⦿ Moving a production that can be carried out by one person to the grid incurs overhead
 - Learning curve
 - Distributed Computing
- ⦿ Reward for figuring all of this out
SiD has been validated!

Thank you

- ⦿ Marcel Stanitzki and Norman Graf for material
- ⦿ The whole GridPP team for helping us get started
 - Glenn Patrick, Stephen Burke, Shaun De Witt
- ⦿ The LHC for not clogging the queues with data