Data Management Challenges
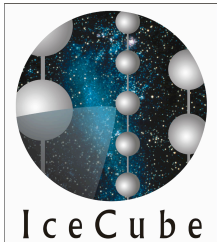
Martin Merck
SAC Review 2009
Madison
Mai 20th-21st 2009

Agenda

- Current Status

- Projected Data and Computing Needs

- Computing Resources

- Waveform Compression Progress

- Simulation Production
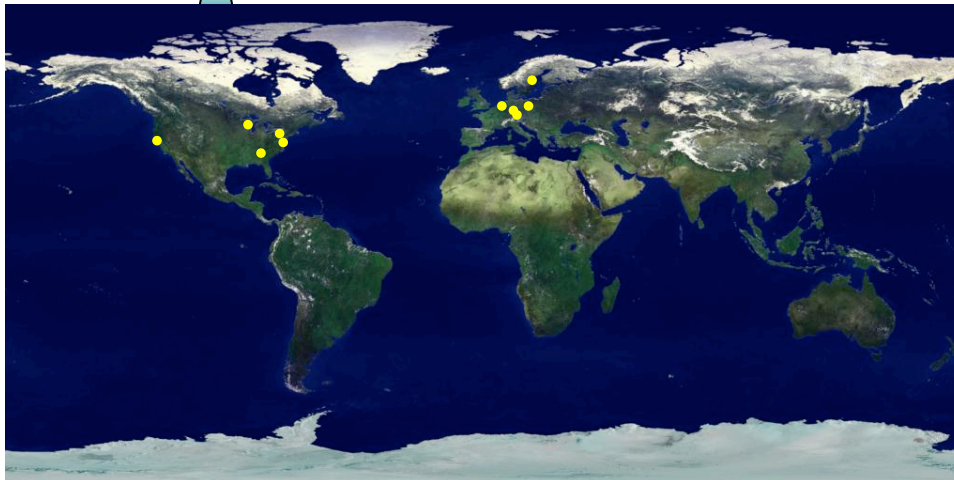
- Data Production Processing

- Outlook

Current Status
Available computing resources

- Datacenter at Madison
  - Central data repository for IceCube
  - Central compute cluster
  - Access and user accounts for the full collaboration
  - Central infrastructure services (DB, WWW, Mail, etc.)

- Collaboration contributed computing resources
  - Computing resources for MC production
  - Temporary storage for MC datasets
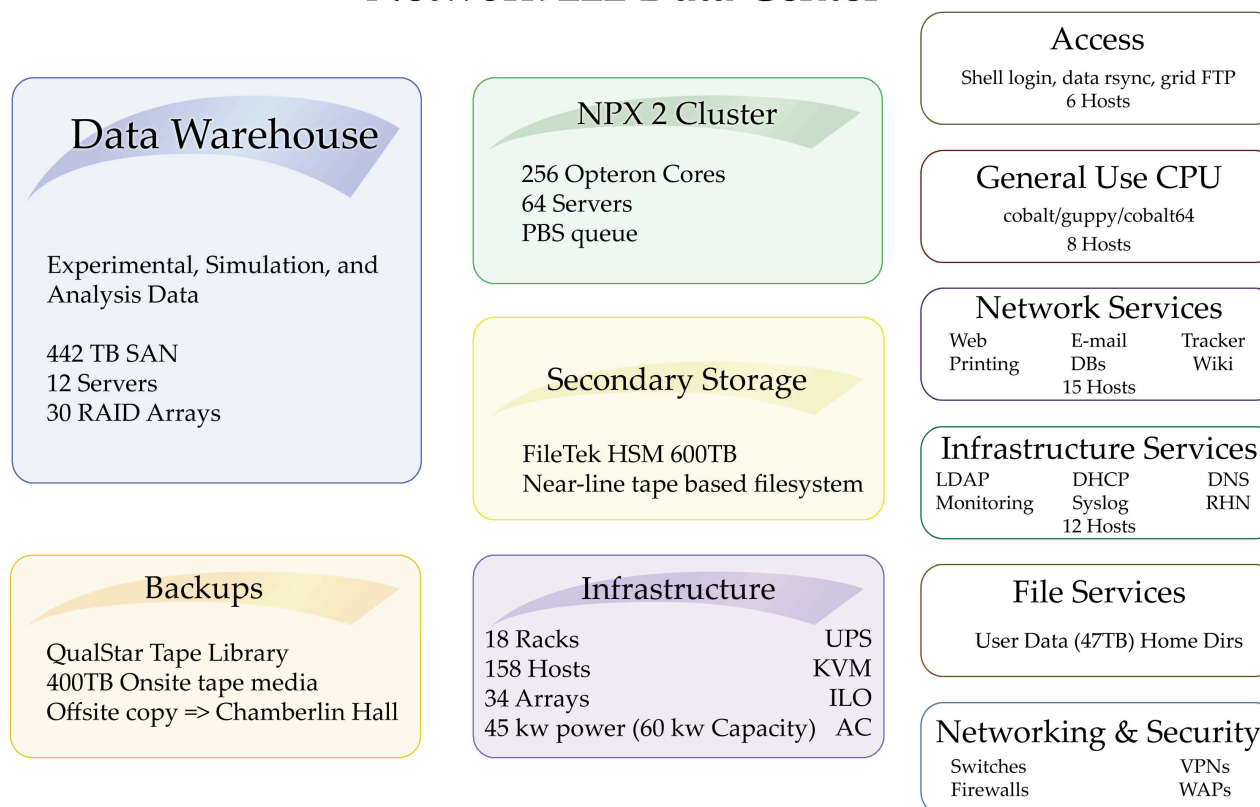 - Secondary data storage for European access
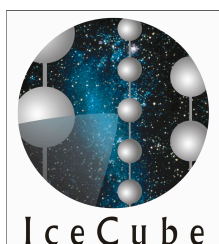
Current Status
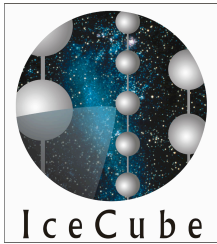## Logical view of the IceCube Datacenter

I c e C u b e

IceCube Computer Technology Overview

Network 222 Data Center

Access

Shell login, data rsync, grid FTP
6 Hosts

Data Warehouse

Experimental, Simulation, and
Analysis Data

442 TB SAN
12 Servers
30 RAID Arrays

NPX 2 Cluster

256 Opteron Cores
64 Servers
PBS queue

General Use CPU

cobalt/guppy/cobalt64
8 Hosts

Network Services

Web	E-mail	Tracker
Printing	DBs	Wiki

15 Hosts

Secondary Storage

FileTek HSM 600TB
Near-line tape based filesystem

Infrastructure Services

LDAP	DHCP	DNS
Monitoring	Syslog	RHN

12 Hosts

Backups

QualStar Tape Library
400TB Onsite tape media
Offsite copy => Chamberlin Hall

Infrastructure

18 Racks	UPS
158 Hosts	KVM
34 Arrays	ILO
45 kw power (60 kw Capacity)	AC

File Services

User Data (47TB) Home Dirs

Networking & Security

Switches	VPNs
Firewalls	WAPs

Projected Data and Computing Needs

Increase in data rate

- Raw data files have increased considerably in size
  - Adding SLC readouts increases raw data by a factor of > 2
  - IC40 ~ 325 GB / day
  - IC59 ~ 660 GB / day
  - IC80 ~ 1000 GB /day
  - Compression is useful for limiting satellite transfer needs
  - SLC hits can't be compressed after feature extraction and increase processed data volume
- Need to increase storage for analysis datasets and simulation beyond original plan
  - Some increase from contingency
  - Continuous increase due to technology evolution and standard replacement cycle
- Simulation resources increased due to good GRID usage
  - More MC data to store
 - Use filtering also on MC data (only partly a reduction)

Projected Data and Computing Needs
Increase in data rate

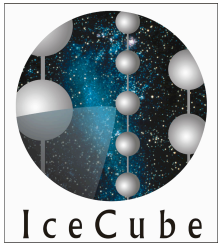|	IC40	IC59	IC86
Trigger rate	1200 Hz	1800 Hz	2700 Hz
Raw data uncompressed	630 GB/day	1300 GB/day	2000 GB/day
Raw data compressed	325 GB/day	660 GB/day	1000 GB/day
Filtered data	35 GB/day	55 GB/day	82 GB/day
Level1 data	50 GB/day	70 GB/day	105 GB/day
Level2 data	85 GB/day	200 GB/day	300 GB/day
Storage 1 Year	60 TB	120 TB	180 TB
Storage 5 years Level2			550 TB
Analysis data			200 TB
Total exp data			930 TB
Simulation data			~ 1.1 TB
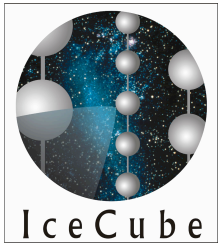
SAC Review Madison Mai 20th-21st 2009

Projected Data and Computing Needs

Increase in data rate

SAC Review Madison Mai 20th-21st 2009
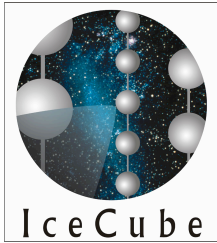
	Tape storage	Data accessible from online disk			
	Experimental raw data [TB]	Experiment [TB]	Simulation {TB}	Analysis [TB[	Total [TB]
2010	900	250	300	90	**640**
2011	1400	350	450	120	**920**
2012	1900	450	650	160	**1260**
2013	2400	550	850	250	**1650**
2014	2900	650	1000	350	**2000**

Projected Data and Computing Needs

Increase in data rate

SAC Review Madison Mai 20th-21st 2009

Increase in disk storage

Legend: —◇— Increase in disk storage

Y-axis: 0, 500, 1000, 1500, 2000, 2500

X-axis: 2009, 2010, 2011, 2012, 2013, 2014
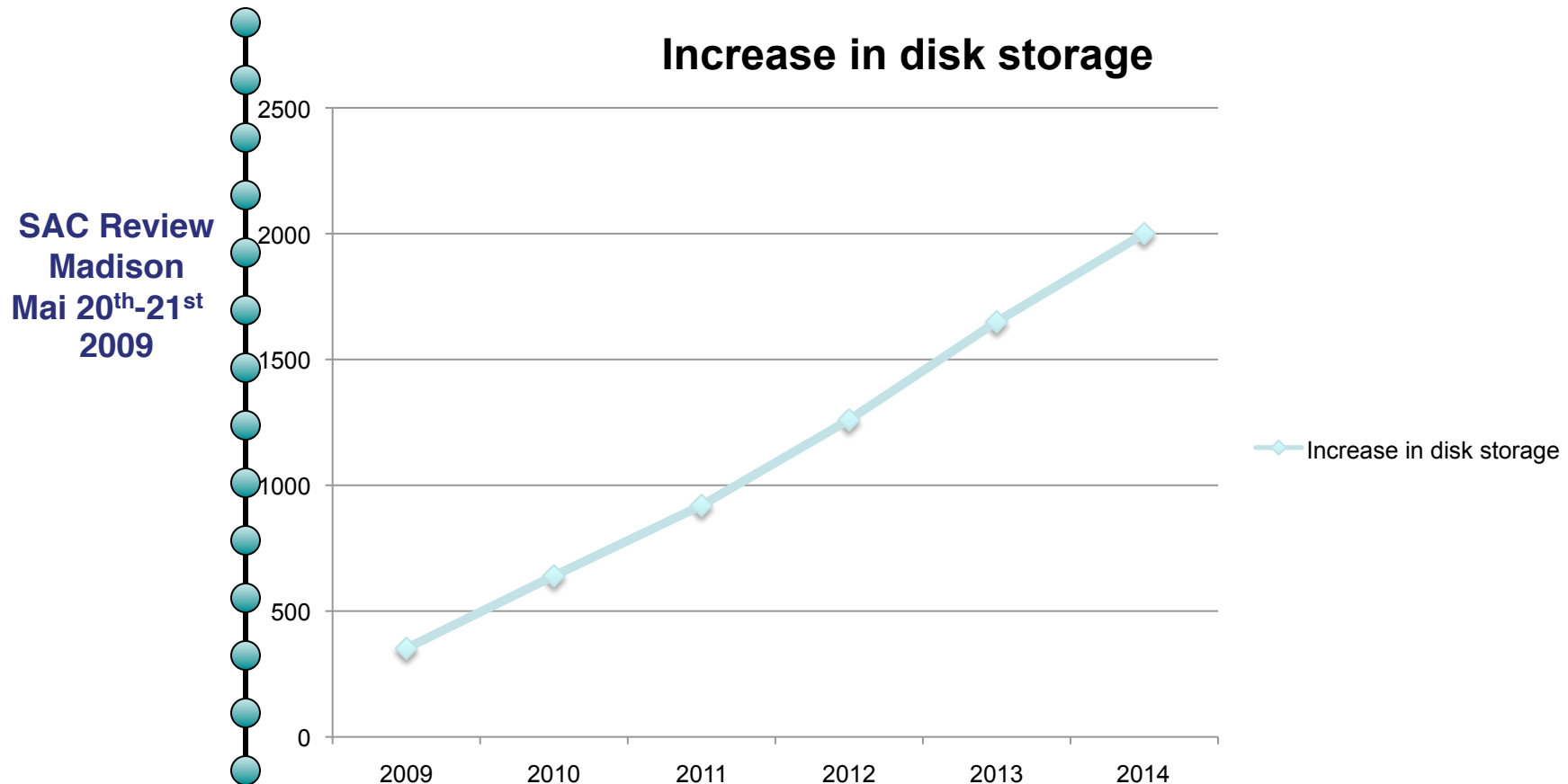
Computing Resources
HPC compute cluster

- HPC compute cluster
  - 64 x 4 core Servers
  - 4 x 16 core blades
  - 320 Opteron cores (2.3 – 2.4 GHz)
- High memory per core
  - 4 GB
- Planned extension
  - 32 x 8 core Servers
  - 256 Nethalem cores
- 3 year replacement cycle
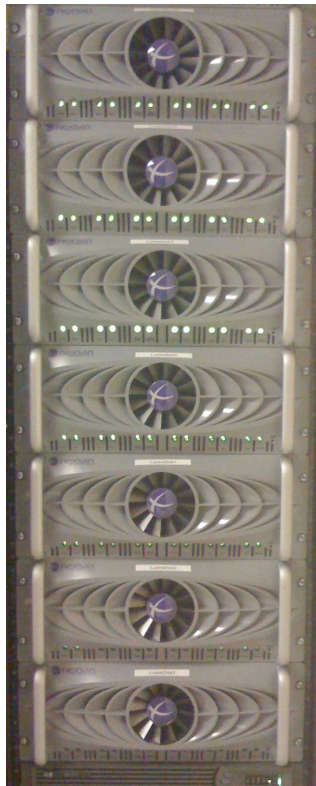- keep 2 generations of clusters for processing and user analysis
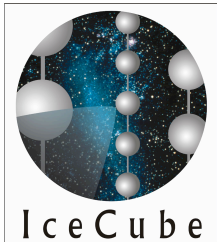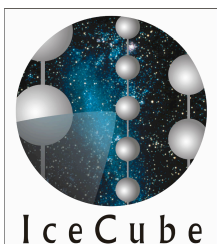
Computing Resources
The central data repository

- 404 TB SAN
  - 15 high capacity units
  - 76% used
  - exp/sim/ana
- 45 TB user space
- Tape library with 1.3 PByte tape capacity
- Needs continuous expansion by replacement with newest technology in a 4 year cycle (assuming 50% disk capacity growth per year)
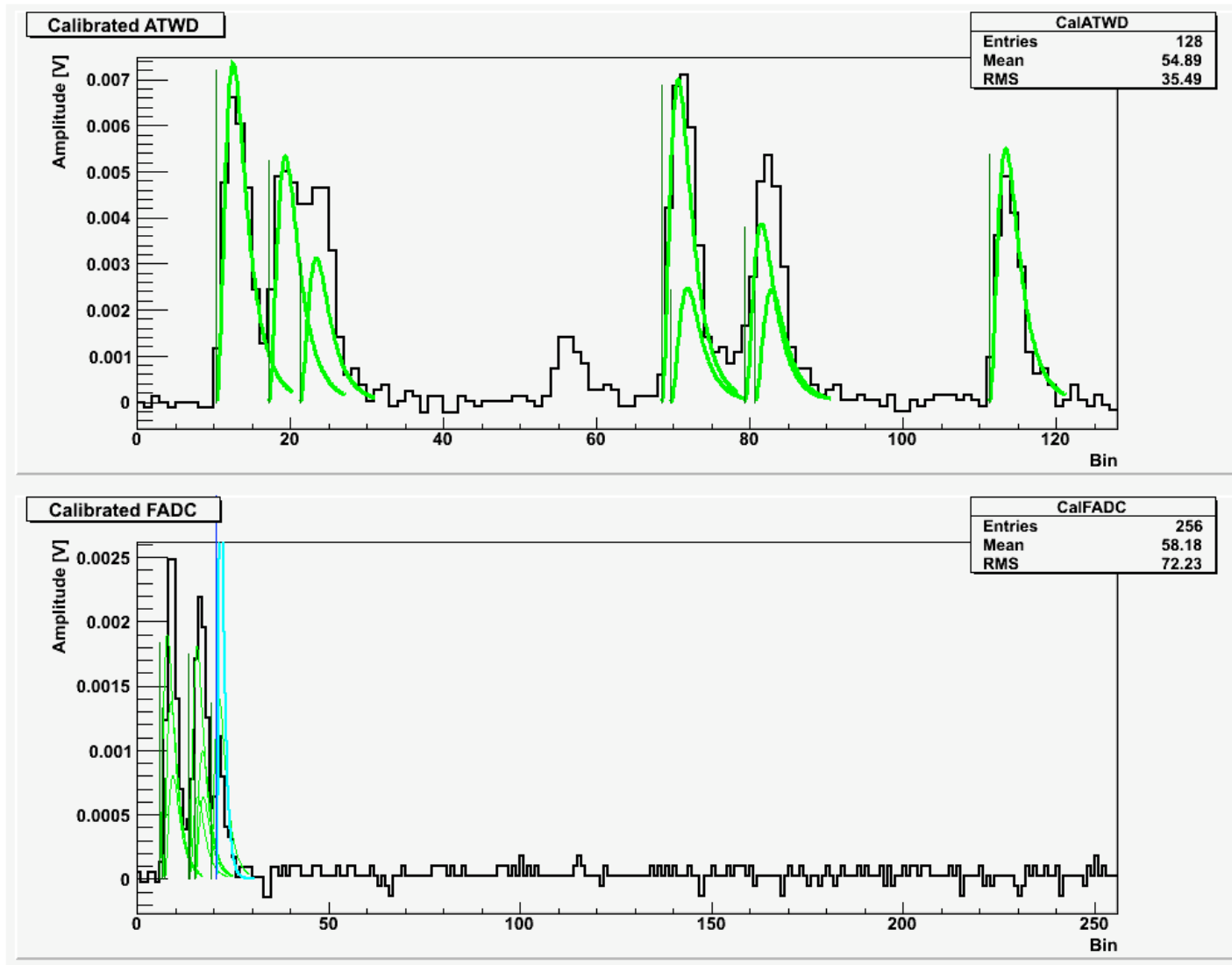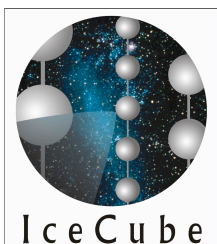
Waveform Compression Progress
Feature Extraction

5/19/09

Waveform Compression Progress
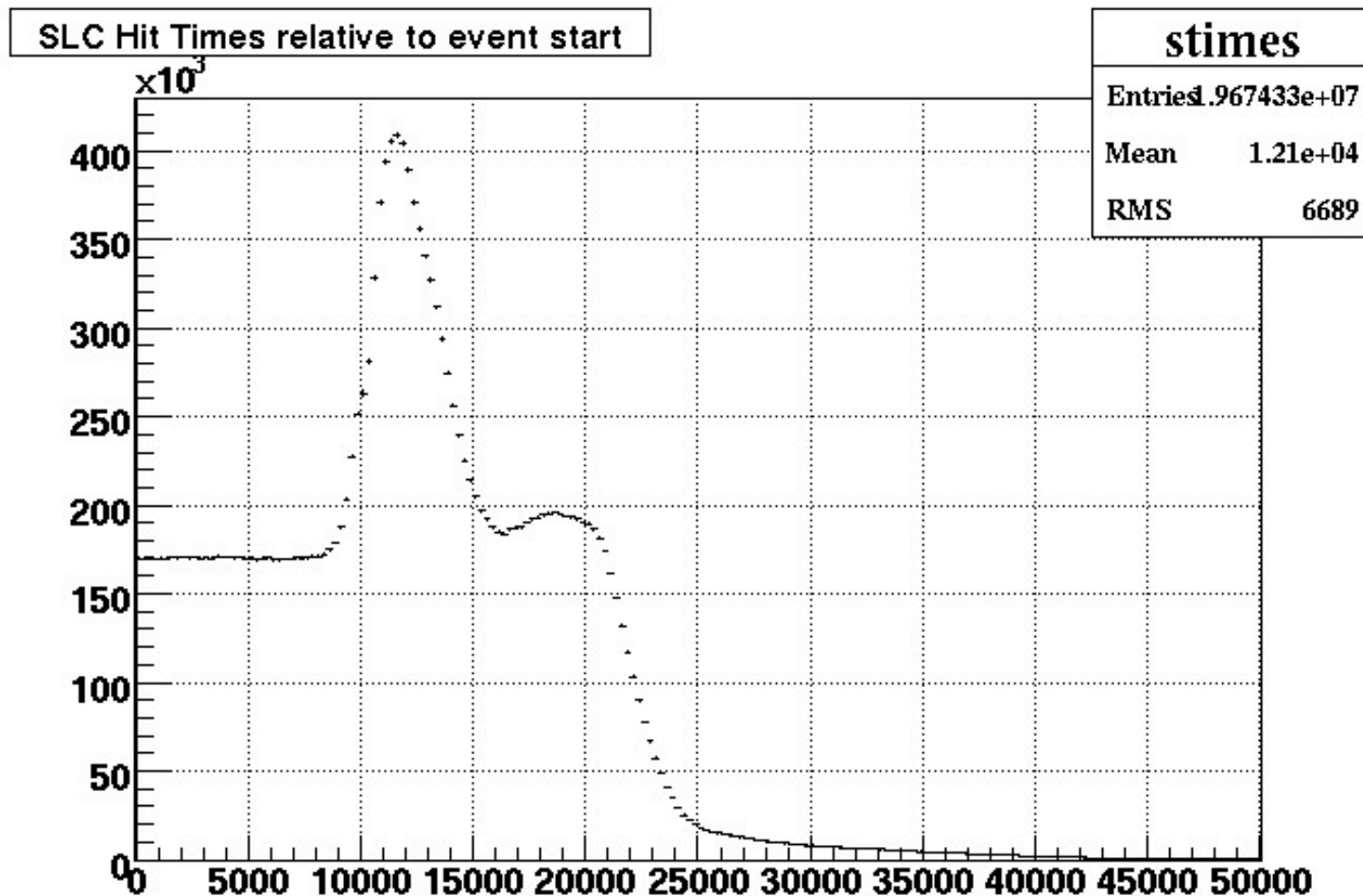## Feature Extraction

5/19/09

Waveform Compression Progress
SLC hit reduction

SAC Review Madison Mai 20th-21st 2009

Waveform Compression Progress
Feature Extraction

- Identified several problems with Feature Extraction
  - Splitting of pulses due to waveform templates and selected unfolding algorithm
  - High storage usage due to usage of floating point numbers
  - Some bugs identified during the process
- High increase of SLC hits
  - Need of event cleaning due to high number of noise hits
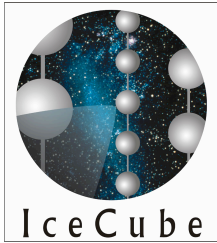
Waveform Compression Progress
Feature Extraction

- Started rewrite of FeatureExtractor
  - Analysis of several approaches
    - Better unfolding algorithms
    - New templates
    - Several bug fixes
  - Needs still some time
    - New ideas needed
- Work on SLC hit cleaning in progress
  - Can reuse some algorithms from AMANDA
  - Some more studies needed to determine potential savings
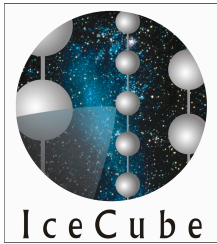
Simulation Production
Available GRID resources

- Produce MonteCarlo dataset for the collaboration
  - Signal generation
    - ν_μ, ν_e, ν_τ
    - WIMPs
    - Monopoles
  - Background generation
    - Muons from cosmic ray air showers
      - Need at least 4000 cores for real time simulation
- Distributed production leveraging GRID resources
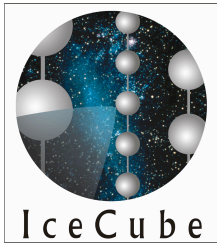
I'M AFRAID HE CAN'T COME TO THE PHONE, HE'S IN THE MIDDLE OF A MONTE CARLO SIMULATION

Deals Department

BRMM BRMM!

RALLY!

Goddard Cartoon ©PharmaVentures; all rights reserved

Simulation Production
Available GRID resources

	Institute	Cores		Disk Space	Farm Type	FTE
		Guaranteed	Available	TB		In-kind
US	UW GLOW (US)	140	800	190	Grid	
	UW NPX2 (US)	100	256		Batch	
	UW CHTC (US)	100	700		Batch	
	UMD (US)	140	278	5	Batch	0.40
	PSU (US)	100	560		Batch	0.40
	LBNL PDSF (US)	50	700	2	Batch	0.20
	UDEL (US)	40	136	50	Batch	0.40
	LONI (US)	200			Batch	0.20
Germany	Aachen (DE)	90	200	15	Grid	0.40
	Dortmund (DE)	150	300	30	Grid	0.40
	Dortmund (DE)	100		20	Batch	
	Mainz (DE)	230	400	26	Grid	0.40
	Wuppertal (DE)	64	128	17	Grid	0.40
	Wuppertal (DE)	150		30	Batch	
	DESY (DE)	400	700	100	Batch	0.40
	DESY (DE)	100	200	20	Grid	
Sweden	SweGrid (SE)	100	400		Grid	0.20
Belgium	Brussels (BE)					0.20
	Totals	2,254	5,758	505		4.00

Data Production Processing

- **Steps of central processing**
  - Level1: unpacking an first guess reconstructions
  - Level2: high level reconstructions
- **Optimize use of limited resources**
  - Provide high level reconstructions of tracks, cascades, air showers, EHE events, etc for all types of analysis
- **Provide analysis ready data in the shortest time possible**
- **Fast feedback to detector operations on analysis level data quality**
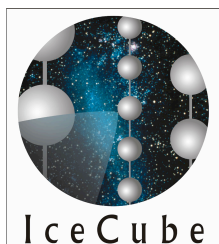
Data Production Processing
IC40 reconstructions

Reconstruction	Percentage of events	CPU time / day
Linefit	100%	
SPEFitSingle	100 %	1.0 day
SPEFit32	66 %	21.3 days
SPEFit32Paraboloid	35 %	4.8 days
SPEFit32Umbrella	66 %	0.6 days
SPEFit32Bayesian	9 %	3.0 days
SPEFit16Split (4 times)	9 %	3.3 days
SPEFit16SplitBayesian (4 times)	9 %	3.8 days
MPEFit	66 %	1.1 days
MPEFitParaboloid	35 %	8.6 days
SPEFit32PhotorecTrack	35 %	1.6 days
SPEFit32PhotorecEnergy	66 %	0.1 days
MPEFitPhotorecEnergy	35 %	0.1 days
SPE32FitMuE	66 %	0.05 days
MPEFitMuE	66 %	0.05 days
CombinedSPEFitSingle	22 %	0.5 days
CombinedSPEFitParaboloid	22 %	1.7 days
JamsSPEFit32	22 %	16 days
CascadeLlh (4 recos)	30 %	1.2 days
CascadeLlhAmOnly (4 recos)	10 %	0.3 days
CascadeLlhCombined (2 recos)	10 %	0.15 days
EHE (3 recos)	2 %	
Shower front (3 recos)	5 %	0.02 days

5/19/09
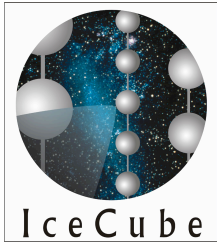
Summary

- Data storage is impacted by higher then expected data volumes.
  - Importance of SLC hits has significantly increased with the extension to low energy physics
- Simulation and data production centrally managed and providing all datasets for scientific analysis
  - Relying heavily on distributed and GRID resources to achieve the goals of MC production
  - Centralized data processing leveraging the high speed access to the data in the Madison datacenter
- Urgent need to expand storage and computing
  - Faster processing needed
  - More reserve for unexpected reprocessing
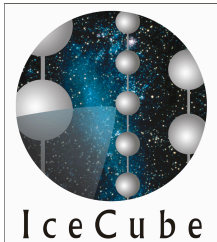
Summary

- Need to raise awareness in the collaboration of best use of limited resources
  - Use online filtering as level3 trigger.
    - e.g.: Don't define raw data as hardware triggered data but software filtered data.
      (Only write filtered data to tape)
  - Earlier reduction of datasets with harder cuts
    - Efficiency needs to be defined as loss at final cut level and not only based on signal efficiency
  - Limit use of expensive reconstructions to reduce CPU requirements
    - More intelligent event cleaning
    - Pattern recognition and event categorization
- Lots of ideas, but time and manpower needed to implement
  - Competing against exciting discovery potential of physics analysis