Present and Future Computing Requirements for Daya Bay

Craig E. Tull
Lawrence Berkeley National Lab

NERSC HEP Requirements for 2017
November 27-28, 2012
Rockville, MD

Why θ_{13}?

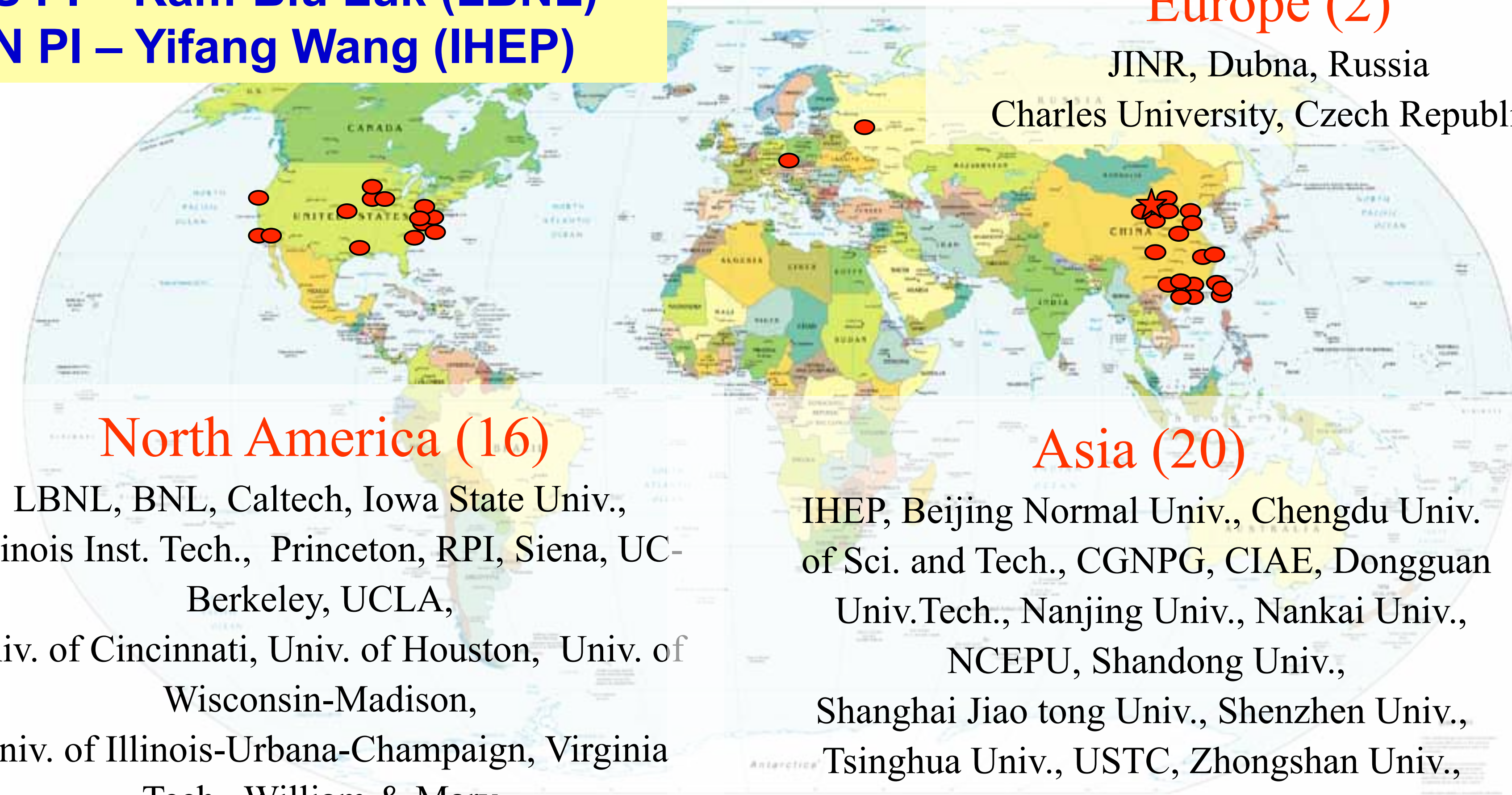
❖ Based on an assumption of three generations, a 3x3 neutrino mixing matrix was proposed – PMNS.

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **The SM has no prediction power on the values of these mixing angles and the CPV phase. It relies on experimental input.**
- **θ_{13} was the last unobserved mixing angle.**
- **Provide knowledge of the basic assumptions:**
  - — **The unitarity of PMNS matrix**
  - — **Three generations of neutrinos**
- **A critical input for other researches, for example:**
  - — **Search for leptonic CP violation**
  - — **Determine the neutrino mass hierarchy**
  - — **Understand the 'effective' neutrino Majorana mass limit**

The Daya Bay Collaboration

US PI – Kam-Biu Luk (LBNL)
CN PI – Yifang Wang (IHEP)



Europe (2)

JINR, Dubna, Russia
Charles University, Czech Republic

North America (16)

LBNL, BNL, Caltech, Iowa State Univ., Illinois Inst. Tech., Princeton, RPI, Siena, UC-Berkeley, UCLA, Univ. of Cincinnati, Univ. of Houston, Univ. of Wisconsin-Madison, Univ. of Illinois-Urbana-Champaign, Virginia Tech., William & Mary

Asia (20)

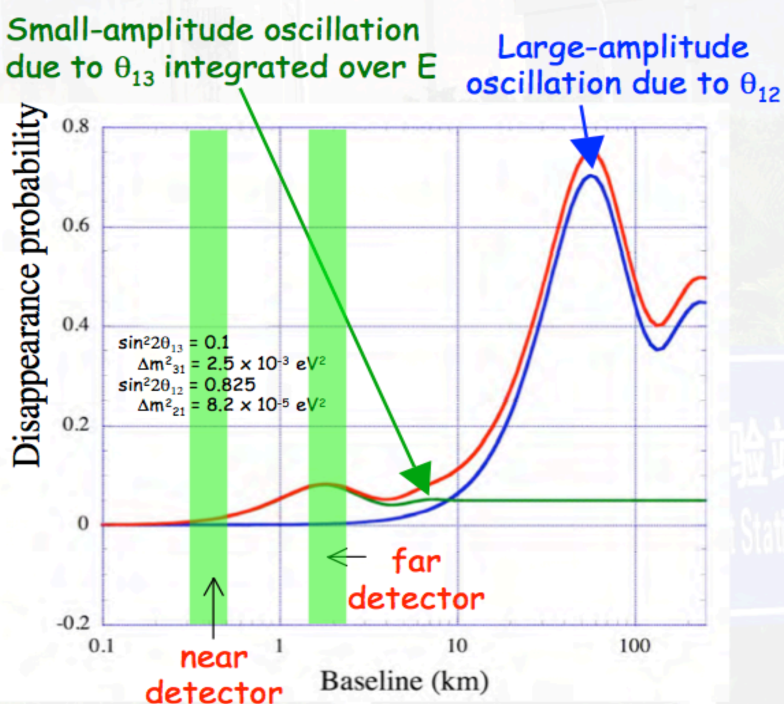
IHEP, Beijing Normal Univ., Chengdu Univ. of Sci. and Tech., CGNPG, CIAE, Dongguan Univ.Tech., Nanjing Univ., Nankai Univ., NCEPU, Shandong Univ., Shanghai Jiao tong Univ., Shenzhen Univ., Tsinghua Univ., USTC, Zhongshan Univ., Univ. of Hong Kong, Chinese Univ. of Hong Kong, National Taiwan Univ., National Chiao Tung Univ., National United Univ.

~230 Collaborators
US/China ~50/50

Reactor Measurement of θ_{13}

- High statistics: powerful nuclear reactors, big detectors, long run-time
- Optimize baselines:

$$P(\bar{\nu}_e \to x) \approx \sin^2 2\theta_{13} \sin^2\left(\frac{\Delta m^2_{31} L}{4E}\right) + \cos^4\theta_{13} \sin^2 2\theta_{12} \sin^2\left(\frac{\Delta m^2_{21} L}{4E}\right)$$



- Reduce systematic uncertainties:
  - detector-related: 'identical' detectors, careful calibration
  - reactor-related: relative measurement with near and far detectors

$$\frac{R_{Far}}{R_{Near}} = \left(\frac{L_{Near}}{L_{Far}}\right)^2 \left(\frac{N_{Far}}{N_{Near}}\right)\left(\frac{\varepsilon_{Far}}{\varepsilon_{Near}}\right)\left(\frac{P_{Far}(L_{Far})}{P_{Far}(L_{Far})}\right)$$

$\bar{\nu}_e$ flux | $1/r^2$ | number of protons | detection efficiency | yield $\sin^2 2\theta_{13}$

- Reduce background: shield and veto

Daya Bay Nuclear Power Complex

Detecting Reactor $\bar{\nu}_e$

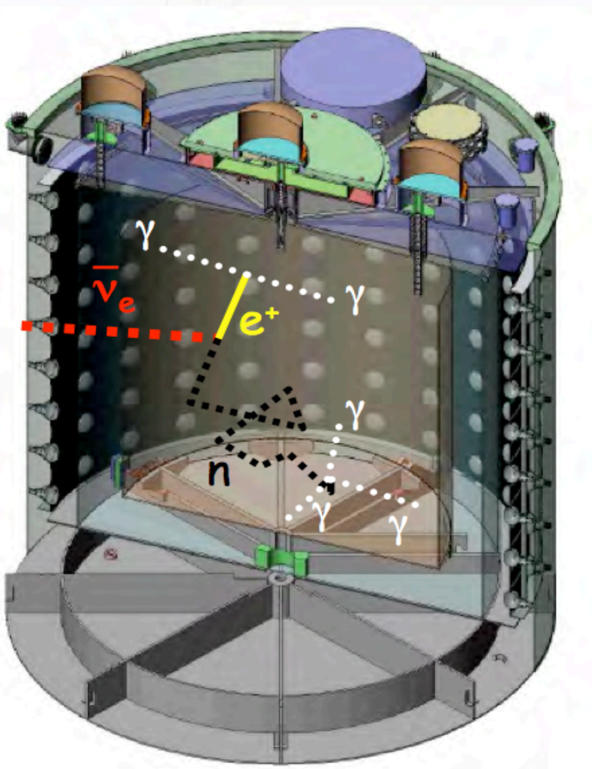
- Use the inverse β-decay reaction:

$$\bar{\nu}_e + p \rightarrow e^+ + n \quad \text{(prompt signal)}$$

~180μs → $+ p \rightarrow D + \gamma(2.2\ \text{MeV})$ (delayed signal)

→ $+ Gd \rightarrow Gd^*$

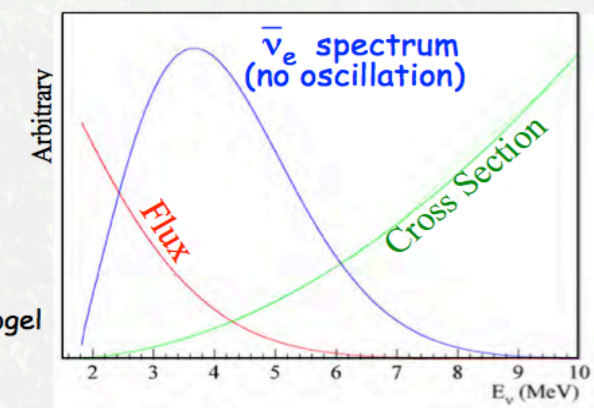
~30μs for 0.1% Gd → $Gd + \gamma\text{'s}(8\ \text{MeV})$ (delayed signal)



- Time- and energy-tagged signal is a good tool to suppress background events.
- Energy of $\bar{\nu}_e$ is given by:

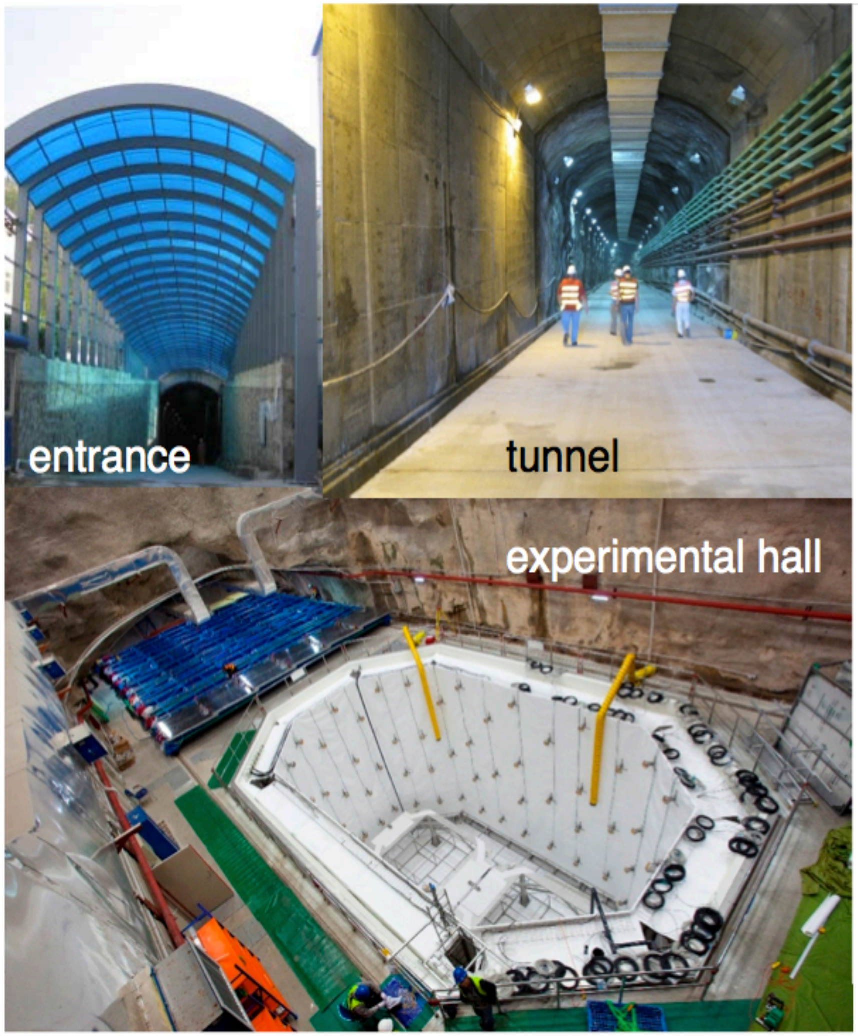
$$E_{\bar{\nu}} \approx T_{e+} + T_n + (m_n - m_p) + m_{e+} \approx T_{e+} + 1.8\ \text{MeV}$$

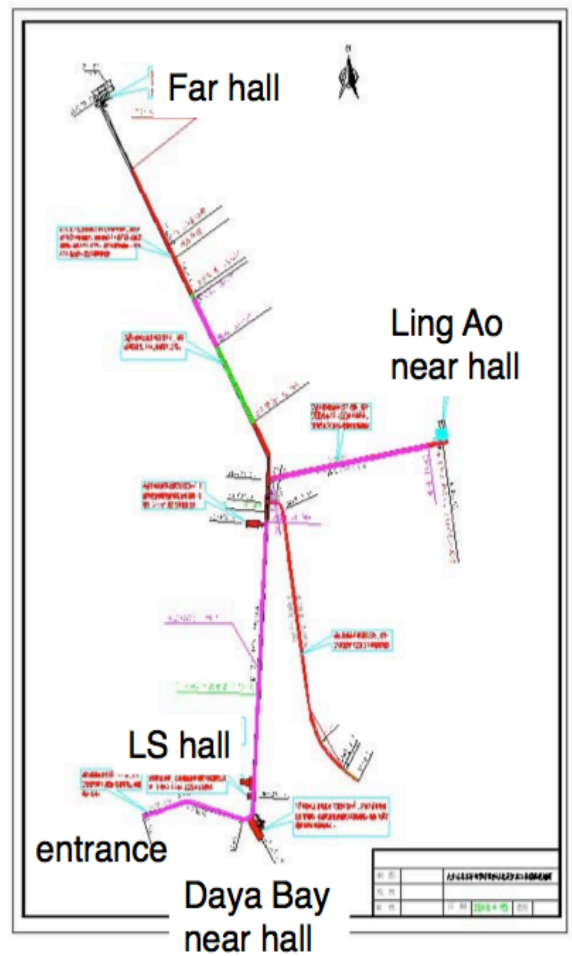
10-40 keV



From Bemporad, Gratta and Vogel

Daya Bay Underground Laboratory

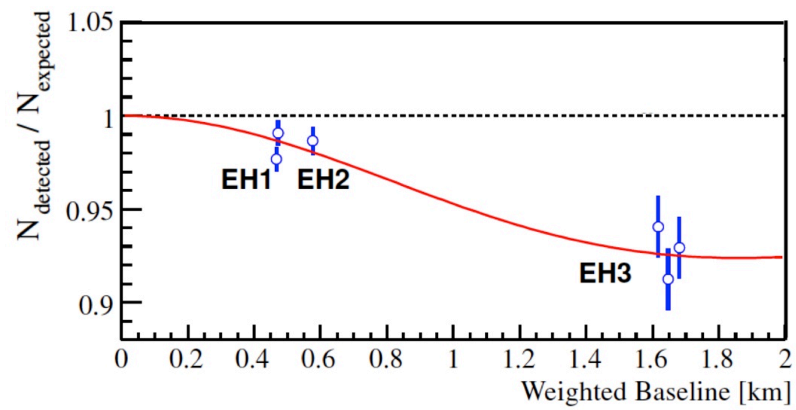




Rate-only Analysis Result (March 8, 2012)

- **First θ_{13} data:**
  - —54.8 days
  - —18k files
  - —16 TB
- **KUP:**
  - —23.5 k CPU-hours
- **Production:**
  - —17.5 k CPU-hours
- **20 days between close of last file and PRL.**

"Physics ready software and computing on Day 1"



$\sin^2 2\theta_{13} = 0.092 \pm 0.016\ (stat) \pm 0.005\ (syst)$

$\sin^2 2\theta_{13} = 0$ excluded at 5.2σ

Our years of preparation payed off in spectacular fashion.

Computational Strategies

- Daya Bay experiment operates 24/7
  - —6 ADs in operation since 24 Dec 2011
  - —8 ADs since 12 Oct 2012 (<400 GB/day)
- Keep Up Processing (KUP) provides real-time feedback to physicists (PDSF and/or Carver)
- Production Physics Analysis Runs:
  - —NuWa Release + DBI Roll-back date + Dataset
  - —3-5 per year using full dataset (70 TB now)
- Individual Analyses:
  - —ROOT or NuWa + ROOT
  - —Use most recent Physics Production
  - —Ongoing, chaotic, small➔large

Daya Bay Application

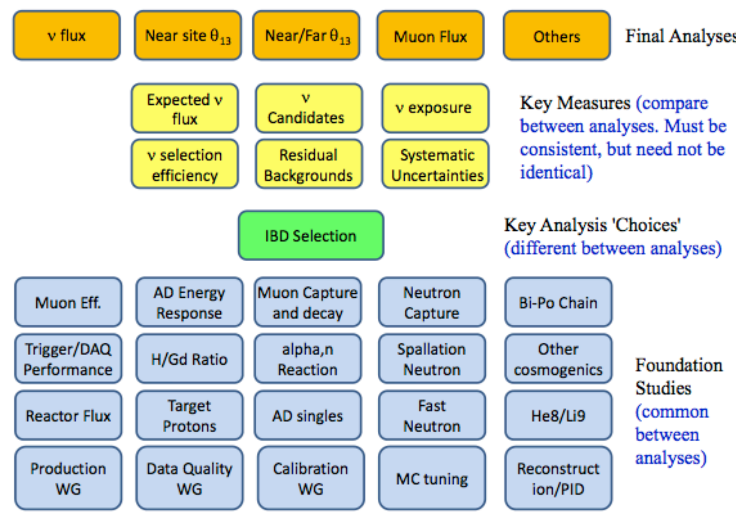
- We operate in a software ecosystem.
- NuWa (Gaudi)
  - Component simulation and analysis framework.
- DBI (DataBase Interface)
  - Condition DB, QA DB, Tag DB
- Spade
  - Moves data from detector to warehouse.
- P2 (psquared)
  - Manages execution of file processing.
- ODM/PQM
  - Near-time feedback for data collectors.
- Geant4 (NuWa)
  - Simulation engine for detector/physics studies
- ROOT
  - Interactive data exploration & libraries.

HPC Characteristics

- Run-time throughput latencies critically important
  - —Data arrive in 20 minutes, analyzed in 2 hours.
- Analysis codes are straight forward
  - —Gives rise to very high IO/CPU ratios
  - —Typical HEP OO code (C++, Python, Java)
  - —Analysis Blinding:
    - AD mass (Measured at fill)
    - Neutrino Flux (Derived from NPP thermal power & simulation)
    - Reactor core distances (Surveyed)
- Event parallelism ➔ no communication scaling
- We are constrained by I/O performance of file system to 400-800 jobs and queue latencies.

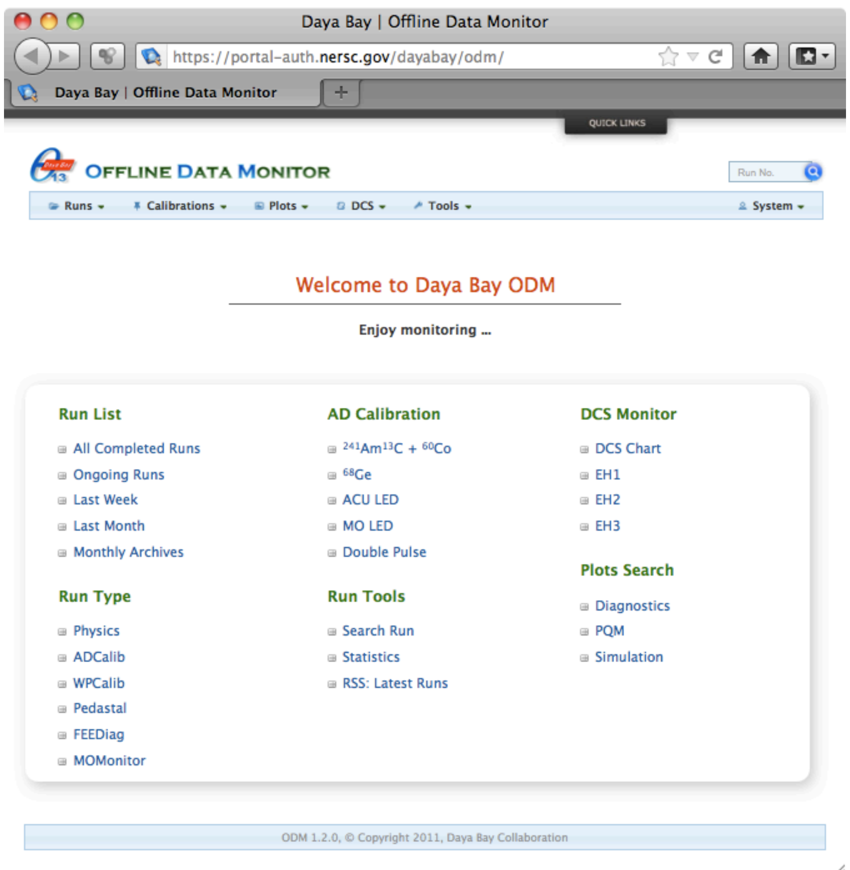
NuWa allows Multiple Independent Analyses

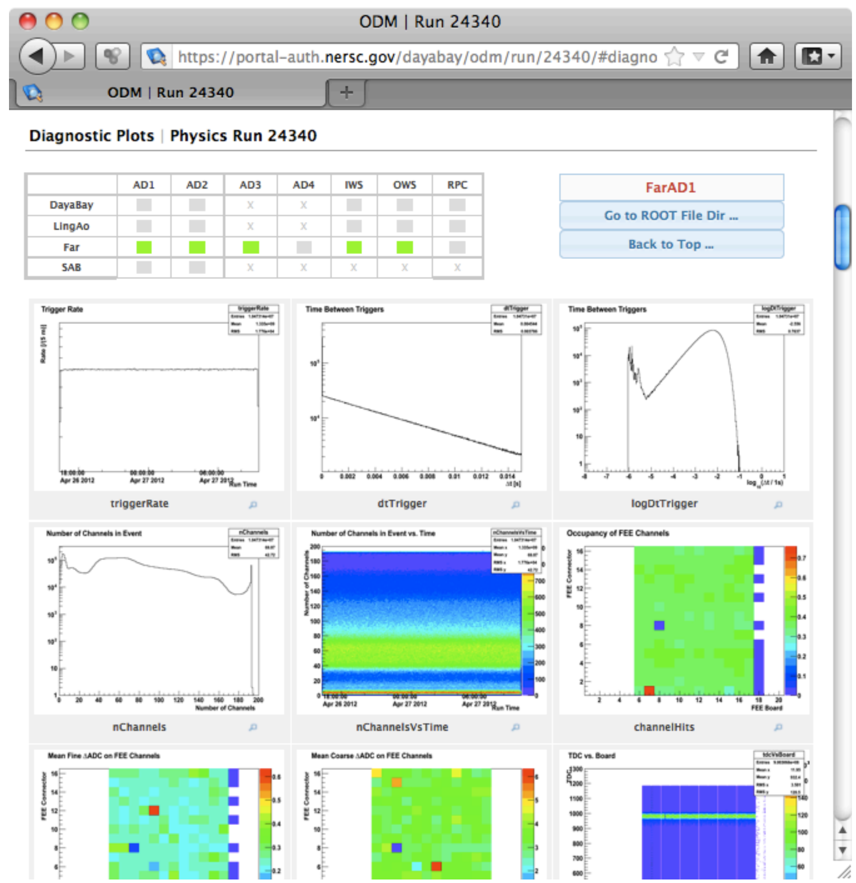
- **Share foundation, but can differ in:**
  - —Energy calibration & reconstruction
  - —Candidate selection/efficiency
  - —Background studies
- **March result rate-only analysis (blind analysis)**



ODM - Offline Data Monitor

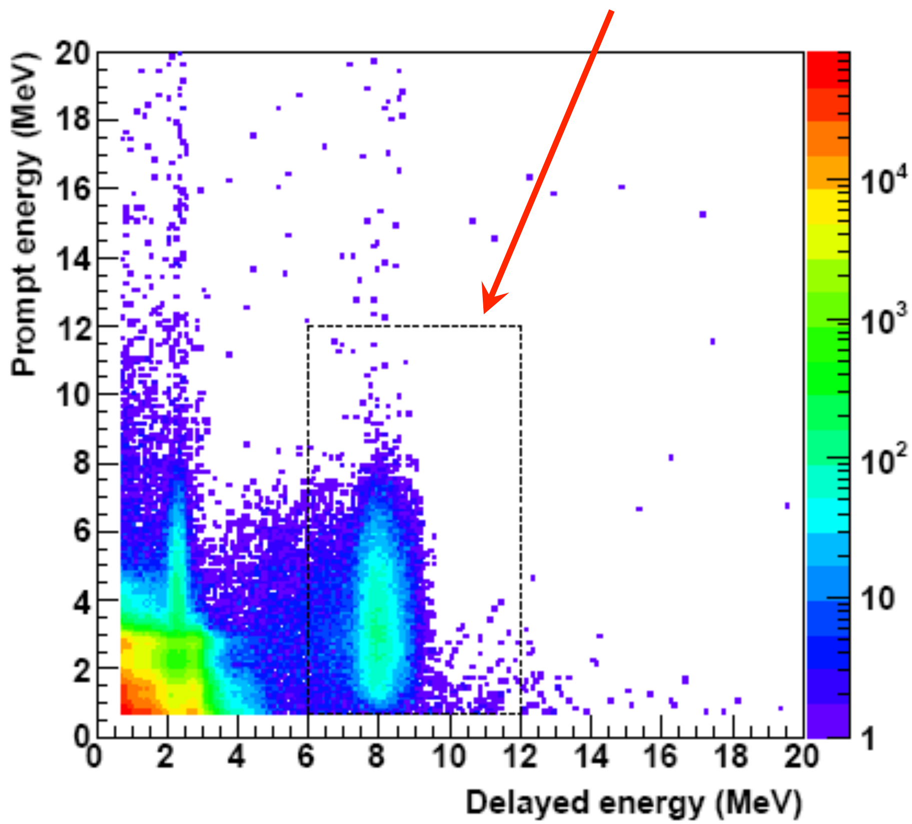
- Bring together information from many different sources.
- Run on PDSF/Carver & NERSCs science data gateway.
- Results within 2 hours of DAQ file close.



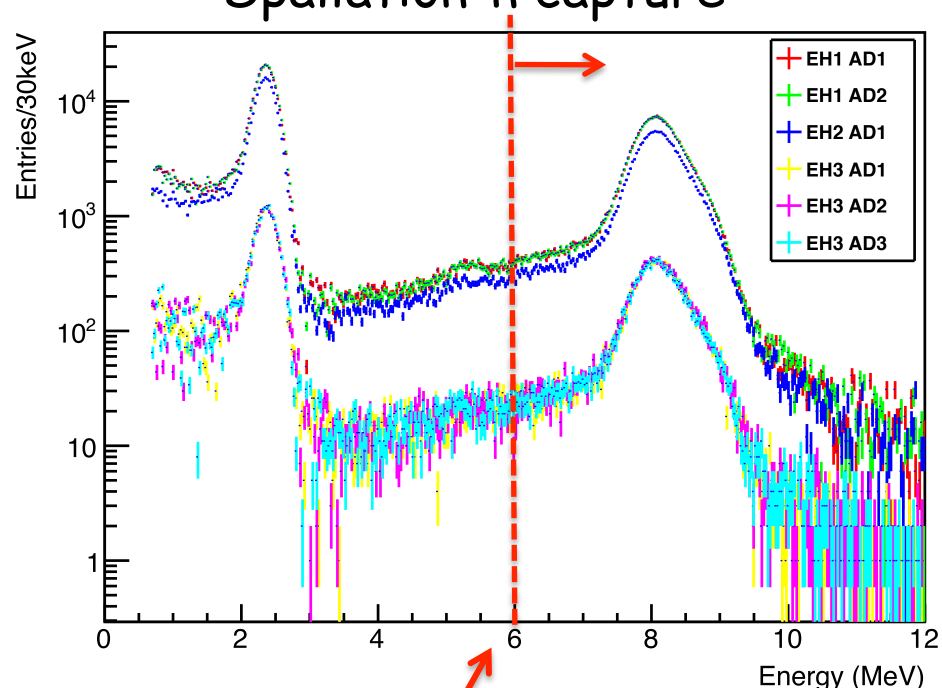


Prompt/Delayed Energy

Clear separation of antineutrino events from most other signals



Spallation-n capture



Uncertainty in relative E_d efficiency (0.12%) between detectors.

Current HPC Usage

- Machines: PDSF, Carver, (IHEP, RACF, DBay)
  - —PDSF 1M CPU-hours, Carver 16k CPU-hours
- Keep Up Processing (KUP): <50 cores 24/7
- 3-5 Production Processing per year
  - —Raw Data grows linearly (70 TB + 120 TB/yr)
  - —Concurrency limited by file system & queues
    - Typically 200-800 jobs for ~3-4 weeks
  - —2-4 GB/core RAM depending on process
  - —Output larger than input (x 1.5)
- Simulation:
  - —Geant4 (NuWa): 0 TB input, 5-20 TB output
  - —Mixing (NuWa): 5-20 TB in & out, many inodes

PDSF Overview

- A computing facility at NERSC primarily used by the HEP and NS communities
- In continuous operation since 1996
- Commodity Linux Cluster
- Interconnected with 10GE, 1GE and FDR IB.
- ~2300 compute cores
- ~1.5 PB of globally accessible storage (GPFS and xRootD)
- Tier-1 site for STAR, Tier-1 for Daya Bay, Tier-2 for ALICE (w/LLNL) and Tier-3 for ATLAS
  - —IceCube, KamLAND, CUORE, Majorana, (BaBAR, CDF, Planck, SNFactory, SNO)

Current HPC Usage

- Software/Services: Chos (PDSF), Modules, DTNs, Science Data Gateway (NEST), NX, htar, GridFTP
  - —dybinst installs full software environment
- Data Resources:
  - —NGFS (478 TB)
    - All/raw data, most up-to-date processing
  - —HPSS (540 TB)
    - All raw data, published processing, Offline DB, …
  - —ESNet, Gloriad, CSTNet, DayaNet, DTNs
    - Much of our data traverse the WAN at some point.
- IHEP (Beijing) Cluster and Castor provide equivalent data and CPU resources.

Scientific Objectives (2017)

- Daya Bay primary scientific goals:
  - Precise measurement of θ_{13} without ambiguities
    - $\sin^2 2\theta_{13}$=0.089±0.010±0.005 ➔Reduce total error to ±0.003 after 4 years
  - Measure Δm_{ee}^2 ➔as good as MINOs for $\Delta m_{\mu\mu}^2$
- Further scientific goals:
  - Detailed comparison of the predicted and observed antineutrino spectra
  - Multi-year measurement of the absolute reactor flux throughout the Daya Bay reactor fuel cycles.
  - Measure the production of neutrons and other spallation products over a wide range of muon energies at different modest depths.
  - Search for signatures of new antineutrino interactions in the observed reactor flux and spectra.
  - Look for supernovae neutrinos.

HPC Requirements for 2017

- ~550 TB raw data: end of 4 years of 8-AD running
  - —~6-8 M CPU hours/yr plus Simulation
  - —Take more advantage of Carver-like resources
  - —~2-3 PB production data (HPSS)
  - —~1 PB disk space
- Curation of Daya Bay's irreplaceable data
  - —Much more than mere data storage.
  - —Long-term archiving & integrity of data
    - Raw, Processed, Simulated, Databases
  - —Virtualized production releases
  - —Human knowledge

Strategies for New Architectures

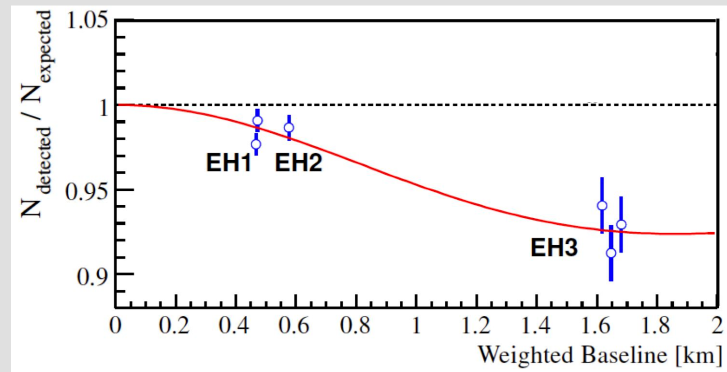
- Daya Bay is largely uninterested in new architectures except as mandatory.
  - —Common tools and techniques for migration
- ATLAS has done research on many-core architectures:
  - —Memory sharing
  - —Inter-procedural optimization based upon code & data “closeness”
  - —Automated parallelization through dataflow analysis
  - —ATLAS (Athena) and Daya Bay (NuWa) share framework architecture (Gaudi) so adoption is straight-forward.

Summary

- Daya Bay is typical of the preponderance of HEP experiments.
- Daya Bay established NERSC/ASCR partnership from the very beginning of project.
- PDSF + NERSC has been an extraordinarily productive facility for data-intensive science.
  - —We still regularly see I/O and throughput limits
- ASCR's science impact payoff is very high for these mid-range experiments.
- Data Curation will be a near/mid-term challenge.
- Simulation is now data-intensive science along with experiment and observation.
  - —Other offices are following in HEP's footsteps.

March 2012 Discovery of θ_{13}

- **NERSC and ESNet were vital in the successful measurement of the theta$_{13}$ neutrino transformation.**
  - — **The last and most elusive piece of a longstanding puzzle: how neutrinos can appear to vanish as they travel**
  - — **The answer – a new, surprisingly large type of neutrino oscillation – affords new understanding of fundamental physics; may eventually help solve the riddle of matter-antimatter asymmetry in the universe.**
- **Experiment could not have been done without NERSC: PDSF for simulation and analysis; HPSS and data transfer capabilities; and NGF and Science Gateways for distributing results**
  - **NERSC is the *only US* site where all raw, simulated, and derived data are analyzed and archived.**
  - **Investment in experimental physics requires investment in HPC infrastructure.**



The Daya Bay experiment counts antineutrinos at detectors in three experimental halls (EH) near the Daya Bay nuclear reactor and calculates how many would reach the detectors if there were no oscillation. The plot shows measured disappearance of antineutrinos at the halls as a function of distance from the reactor. The 6.0% rate deficit at EH3 provides clear evidence of the new transformation.

NERSC's PDSF cluster

HEP

PI: Kam-Biu Luk (LBNL)

THANK YOU
(XIE-XIE 谢谢)

URLs & References

- http://nms.dyb.ihep.ac.cn/nms/index-dyb.html
- http://dayabay.lbl.gov/dybspade/daily-transfers/index.html?ihep
- http://dayabay.lbl.gov/dybspade/daily-placements/index.html?pdsf
- http://dayabay.lbl.gov/dybspade/counts/index.html?ihep
- http://dayabay.lbl.gov/dybspade/daqdm/summary/
- http://dayabay.lbl.gov/dybspade/sentry/dbay/closedOnDisk/
- http://dayabay.lbl.gov/dybspade/delivery/index.html
- https://portal-auth.nersc.gov/dayabay/odm/
- http://www.nersc.gov/users/computational-systems/pdsf/
- http://portal.nersc.gov/project/pdsf/ganglia/
- http://www.gloriad.org/maps/cactigraphs/nagios/status.html

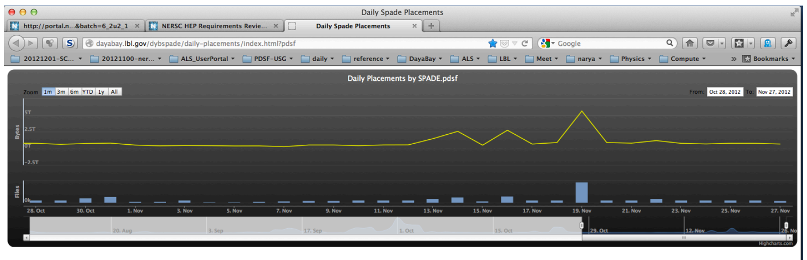
Outline

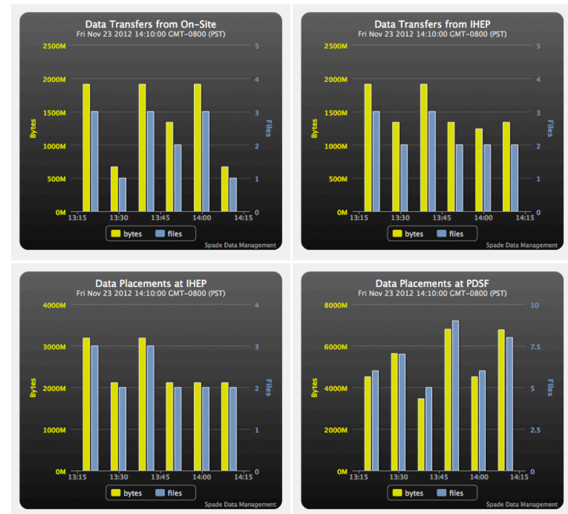
- Project Description
- Computational Strategies
- Current HPC Usage
- HPC Requirements for 2017
- Strategies for New Architectures
- Summary

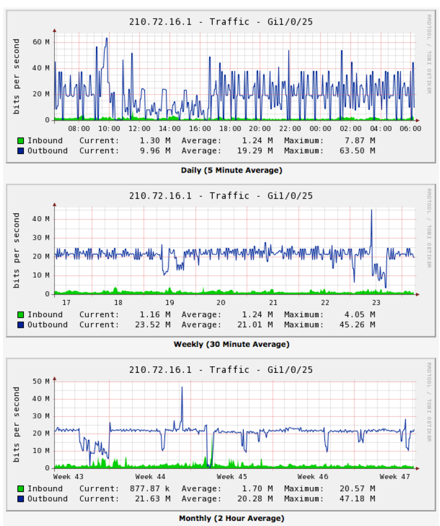
Long Term Science Goals

- **Daya Bay primary scientific goals:**
  - — **Precise measurement of θ13 without ambiguities**
    - Important measurement of a fundamental parameter.
    - Improve the measurement of other mixing parameters by accelerator experiments.
    - Overconstrain the νSM interpretation of oscillations with accelerator measurements.
    - Improve the ultimate precision on JCPν and unitarity of the PMNS matrix.
  - — **Measure Δm_{ee}^2**
    - Measurement of new ν mixing parameter
    - Complementary and with similar precision to accelerator-based measurements
    - Constrains 3-ν model
- **Further scientific goals:**
  - — Daya Bay collects reactor antineutrino data at a tremendous rate, enabling a precision measurement of the reactor antineutrino spectrum. A detailed comparison of the predicted and observed spectra will validate and improve reactor flux calculations and possibly resolve ambiguities in reactor antineutrino predictions.
  - — Make a multi-year measurement of the absolute reactor flux throughout the Daya Bay reactor fuel cycles.
  - — Measure the production of neutrons and other spallation products over a wide range of muon energies at different modest depths.
  - — Search for signatures of new antineutrino interactions in the observed reactor flux and spectra.
  - — Look for supernovae neutrinos.

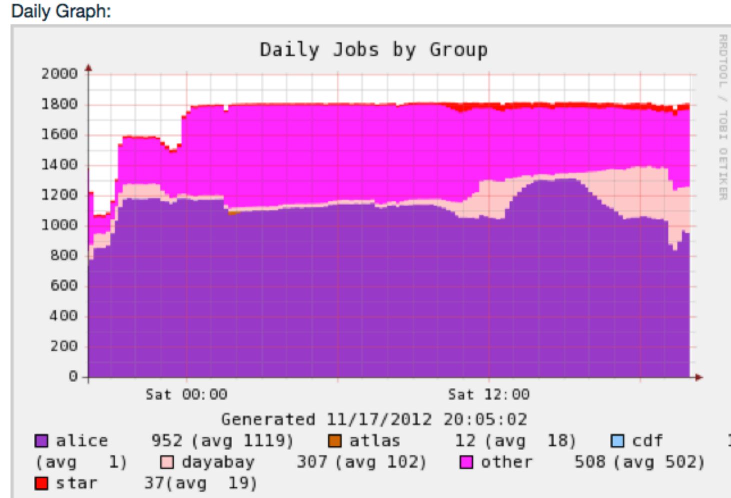
Monitoring



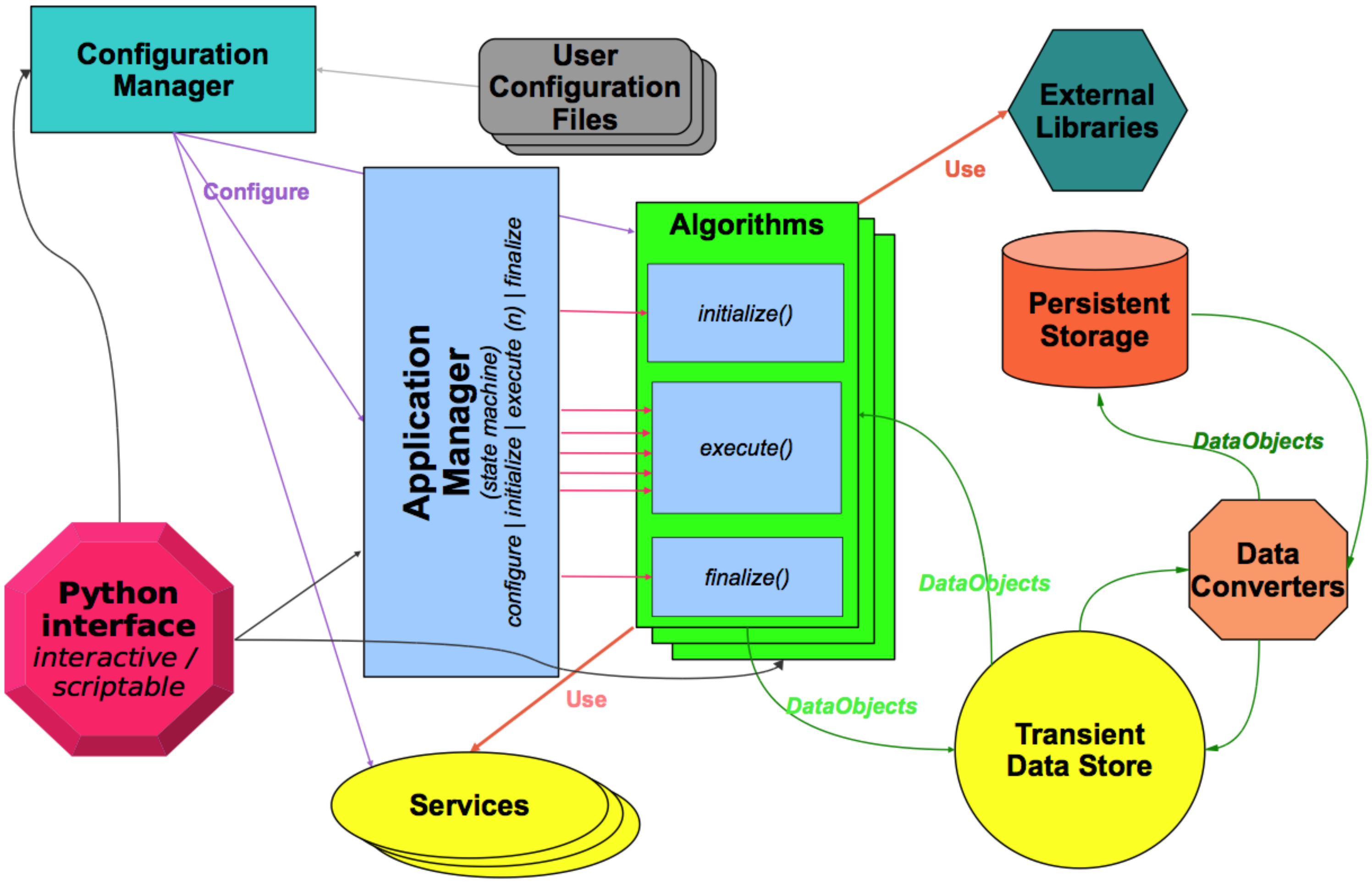




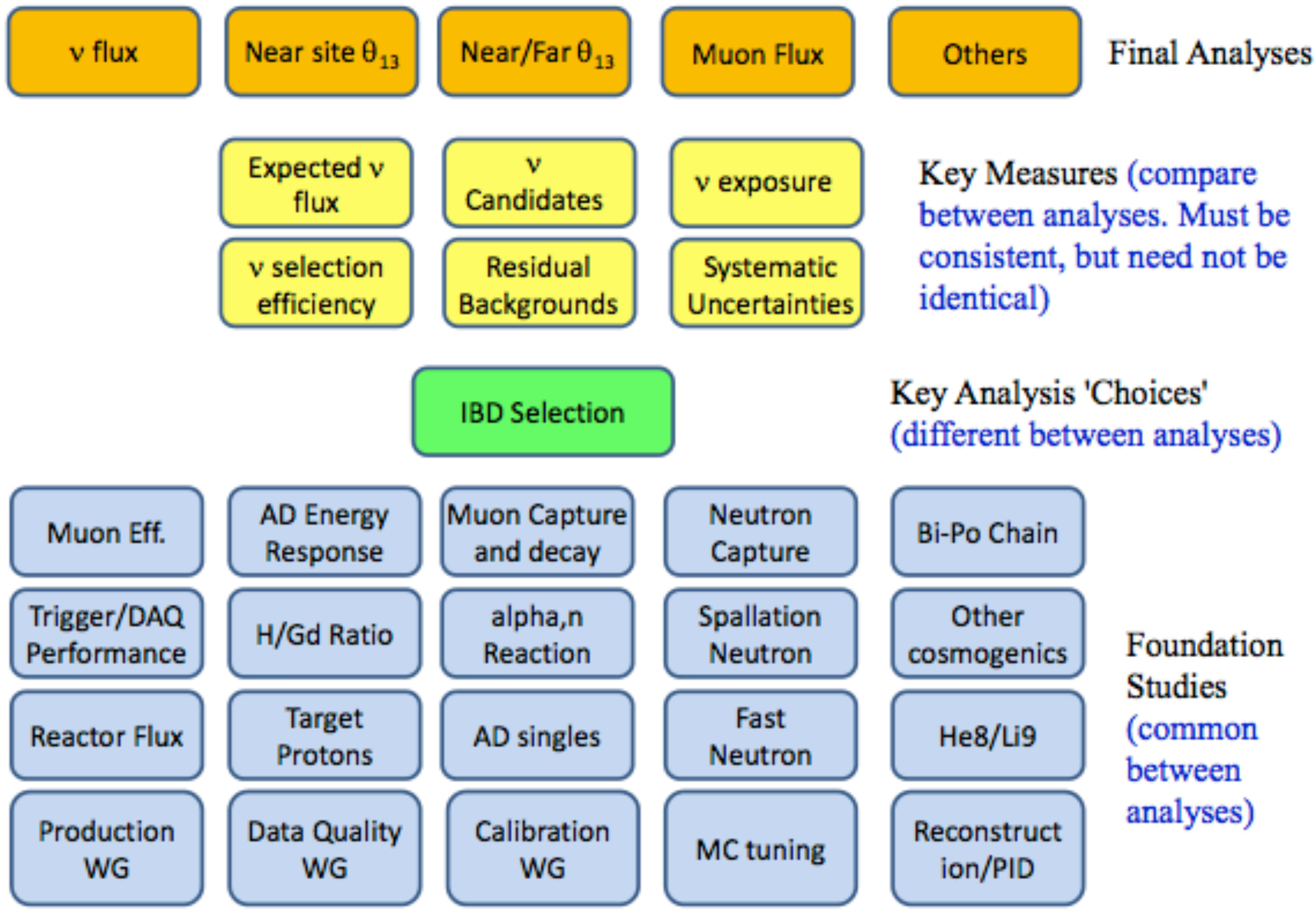
Daily Graph:



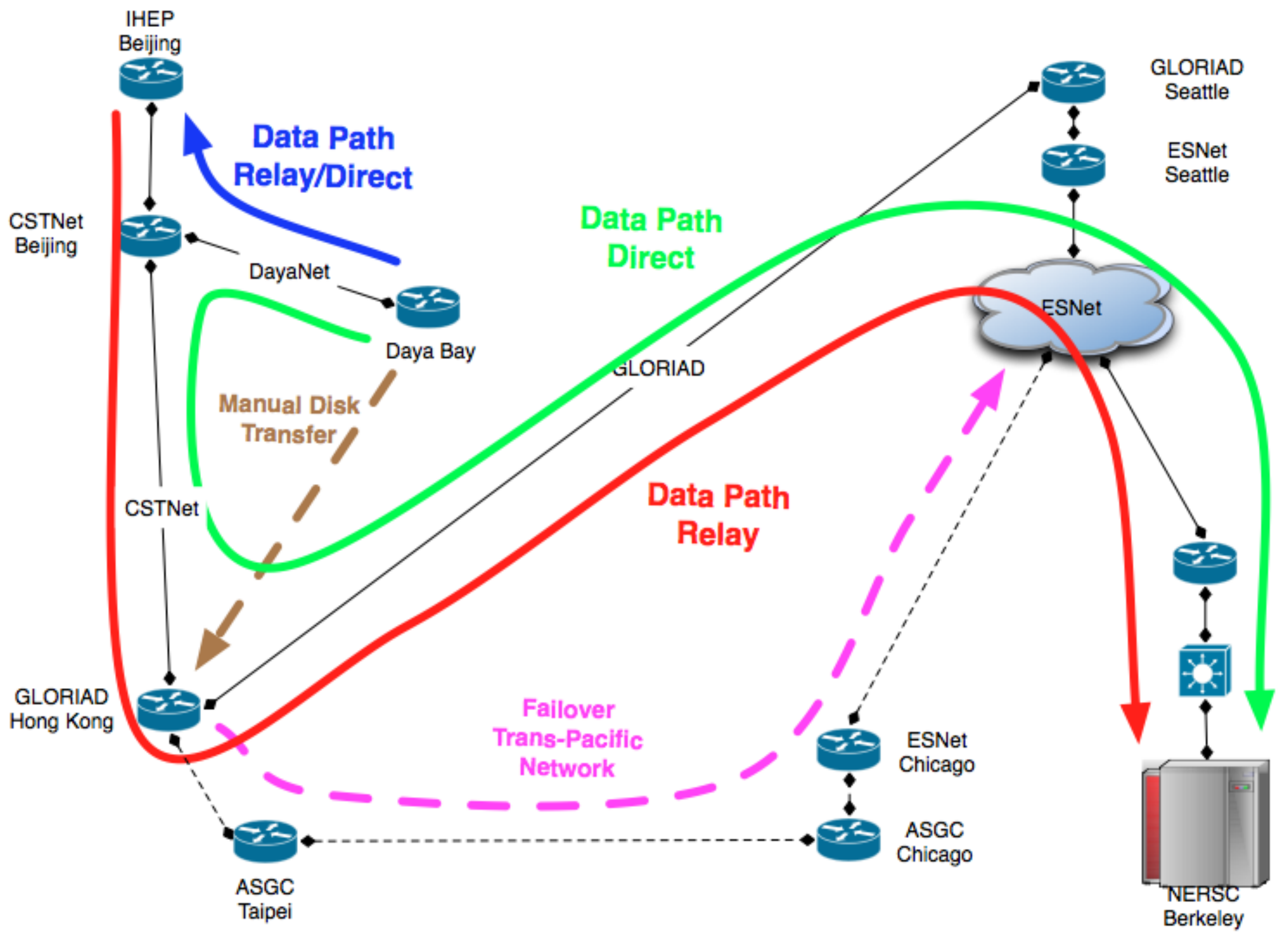
NuWa - Gaudi-based Framework



Foundation Analysis

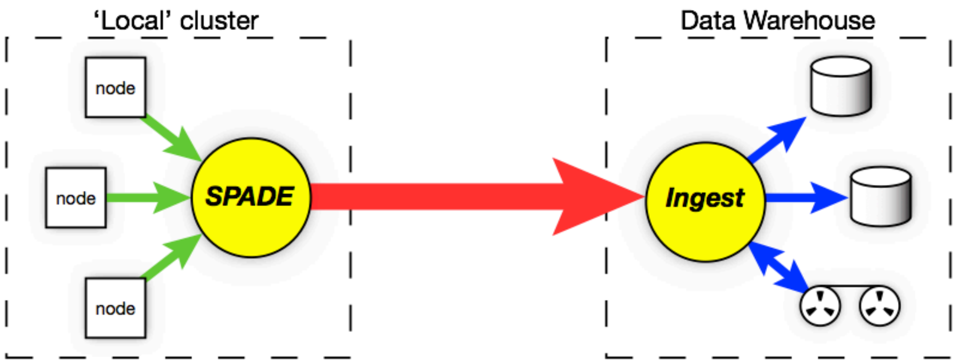


Daya Bay Networking



- Relay Path requires daisy-chaining 2 data transfers (default)
- Direct Path requires 2 data transfers out of site.
- DayaNet: Dedicated 150 Mbps optical link
- CSTNet: Chinese national network
- GLORIAD: Trans-Pacific scientific network (NSF)
- ASGC: Trans-Pacific eScience network (fallback for GLORIAD)
- ESNet: US national Energy Science Network
- Hot-swappable disk transport between Daya Bay and Hong Kong in case of long-term network failure of either DayaNet or CSTNet

Data Transfer & Management



daq.Neutrino.0021508.Physics.EH1-Merged.SFO-1._0001

Tracking of daq.Neutrino.0021508.Physics.EH1-Merged.SFO-1._0001	
Closed by DAQ	Sat, Jan 7 2012, 16:34
IHEP warehouse	Sat, Jan 7 2012, 16:45
PDSF warehouse	Sat, Jan 7 2012, 16:51
HPSS at PDSF	Sat, Jan 7 2012, 16:51
KUP at PDSF	Sat, Jan 7 2012, 18:28

- **Spade/Ingest suite is designed to reliably transfer data from an experiment to it's data warehouse.**
- **We have 2 production data warehouses (IHEP & LBNL) containing all raw data and all production processed data.**
  - **Web query interface and accessible by all collaborators**
- **Nominal data transfer latencies from Daya Bay are ~10-15 minutes (IHEP) and ~15-20 minutes (LBNL) after file is closed by DAQ.**
- **Data is buffered onsite and at IHEP to accommodate network interruptions (>30 days buffer at Daya Bay).**
- **Keep Up Production (KUP), data archiving and conversion to ROOT are triggered automatically by Ingest.**

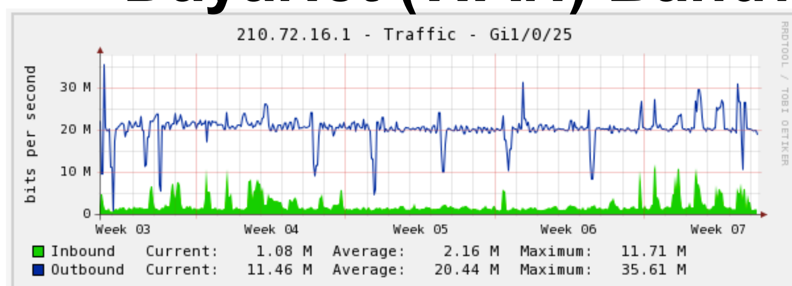
Data Volumes

- **Feb 17, 2012 data volume on disk at PDSF**
  - **105 TB raw**
  - **88 TB derived, MC, user**
- **3 Hall / 6 AD data rate of 310 GB/day**
- **Automated data transfer (Spade) can buffer >30 days raw data for network interruptions.**
- **All critical data are stored on tape redundantly (LBNL HPSS & IHEP Castor)**
  - **Raw data, calibrations, publication production, database**

Raw Data Rates

	Average (GB/d)	Sum (GB)	Maximum (GB/d)	Minimum (GB/d)
Jul-11	71	1,994	489	0
Aug-11	260	8,051	475	76
Sep-11	279	8,366	485	113
Oct-11	314	9,748	469	177
Nov-11	418	12,534	757	191
Dec-11	366	11,358	576	227
Jan-12	310	9,619	328	252
Feb-12 *	310	4,964	318	272

DayaNet (WAN) Bandwidth

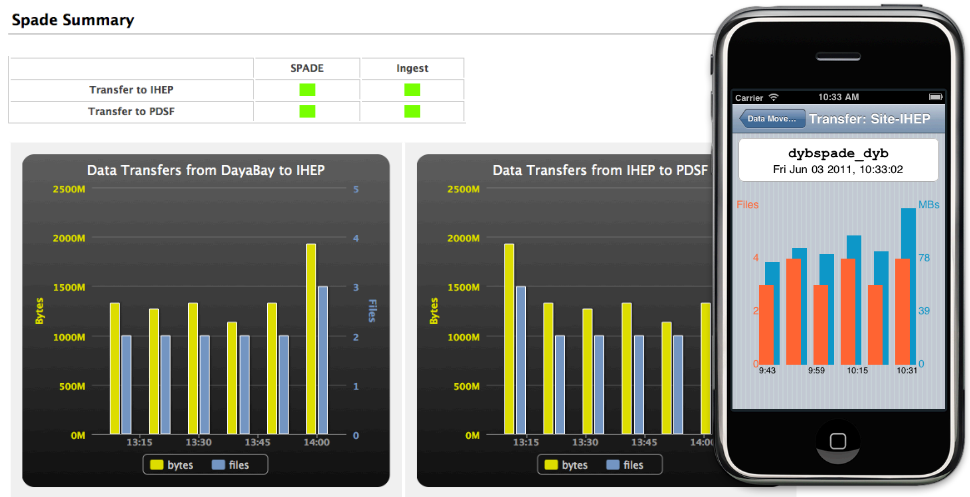


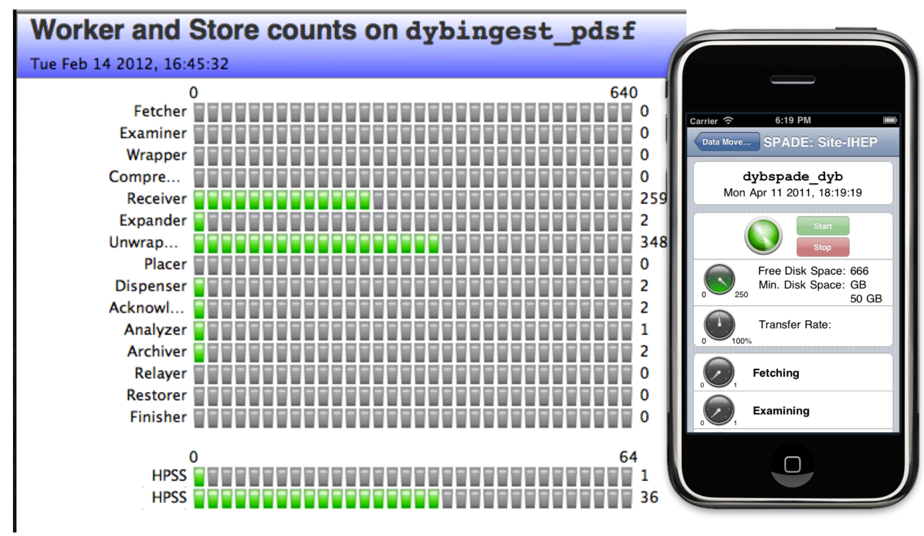
Monthly (2 Hour Average)

N.B. Data files are compress
over network and on HPSS.

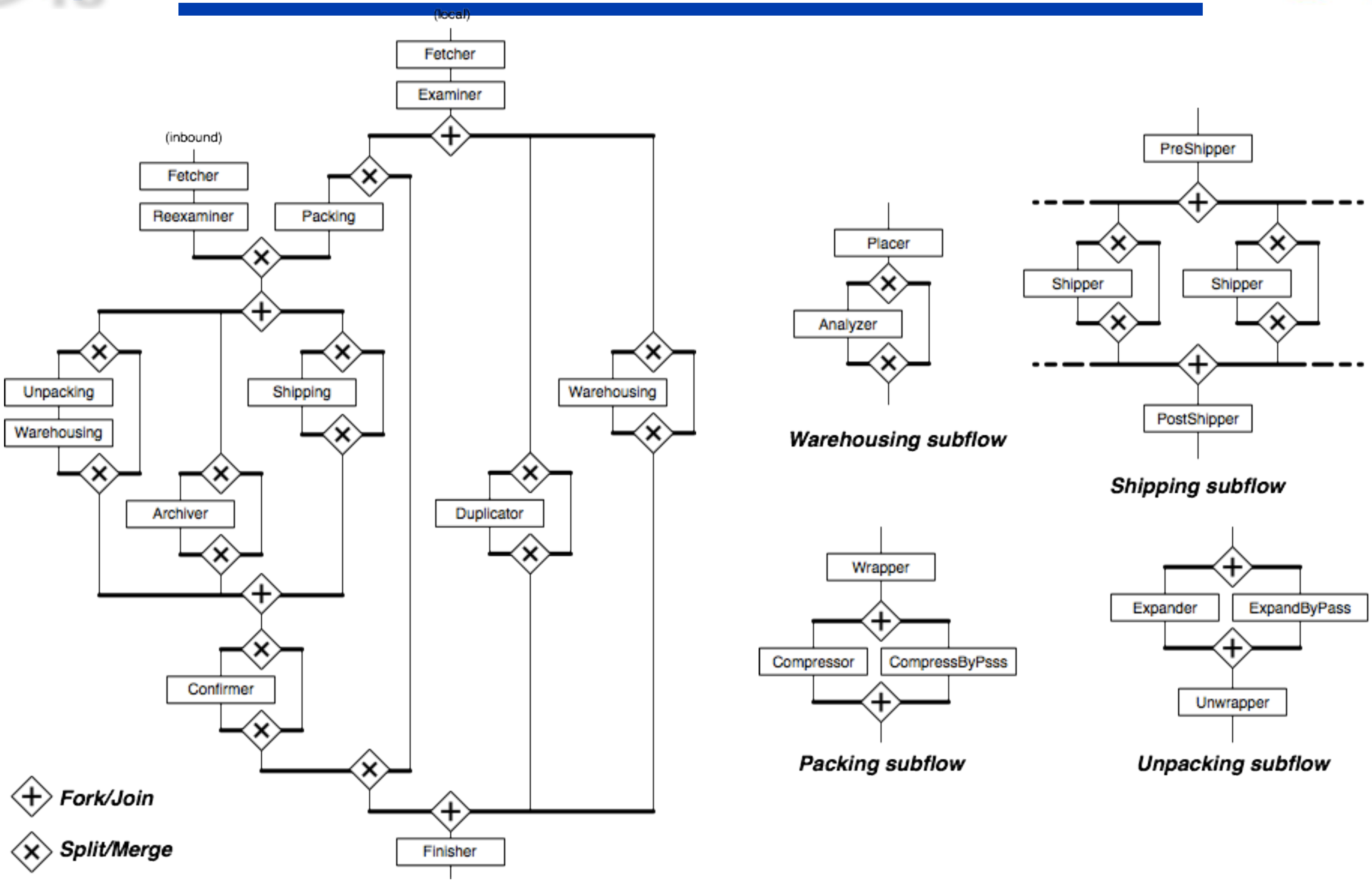
Data Transfer Monitoring

- Spade/Ingest is fully automated and normally requires no manual operation/intervention.
- Most failure modes are recovered automatically.
- Each Spade/Ingest process is fully monitored.
- Each Spade/Ingest data flow is fully monitored.
- Dedicated DayaNet link is fully monitored.
- Asynchronous validation of data integrity done by data file checksum.



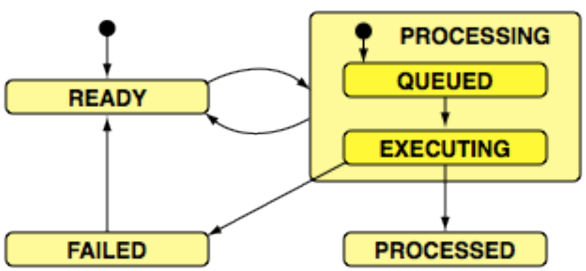


Spade/Ingest workflow

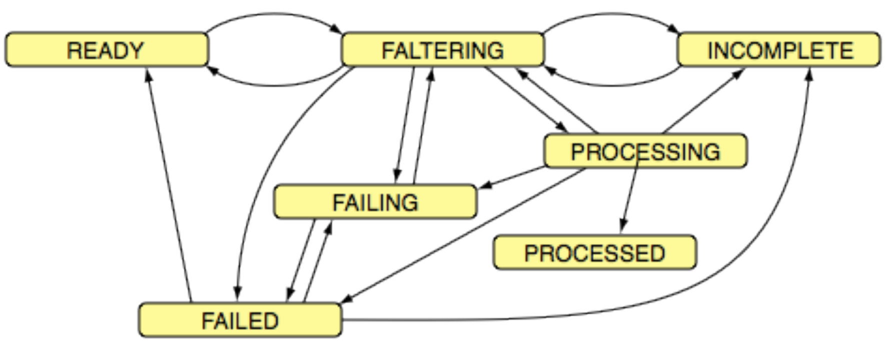


psquared

- Create a pair, binding a file with a processing configuration
- Moves the ‘pair’ through its state machine.
  - Only allows any ‘pair’ to be processed once unless explicitly reset.
- Interface is a web service.



State diagram from single file



State diagram from collection of files

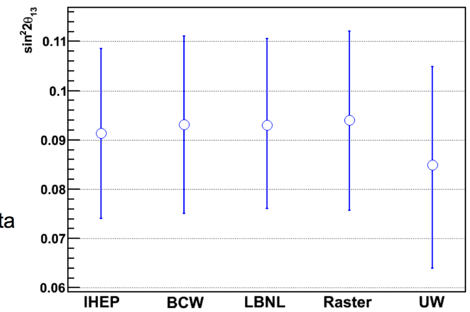
Internal Comparison & Validation of Results

Estimate of θ13

Very consistent results for multiple independent estimates.

IHEP slightly lower (0.0914 vs. 0.093): seems due to 0.8% core-to-core reactor uncertainty. Others assume 2%.

UW using smaller data set (Dec. 24-Feb.3).

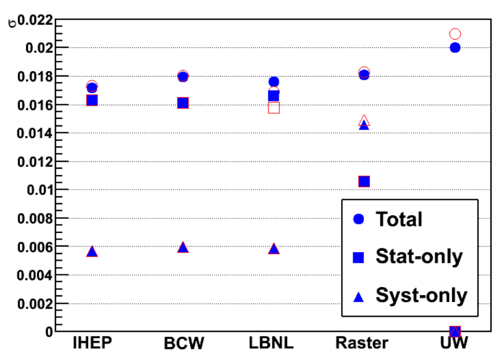


Mar. 1, 2012 Comparison of θ_{13} Oscillation Results 3/5

Statistics v.s. Systematics

Most analyses find similar stat/syst contributions.

Closed (open) symbols provide asymmetric +(-) uncertainties.

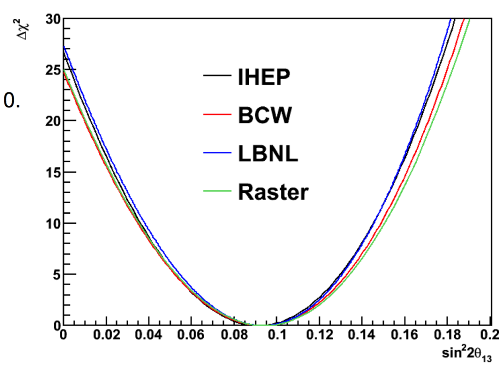


Mar. 1, 2012 Comparison of θ_{13} Oscillation Results 4/5

$\Delta\chi^2$ Distribution

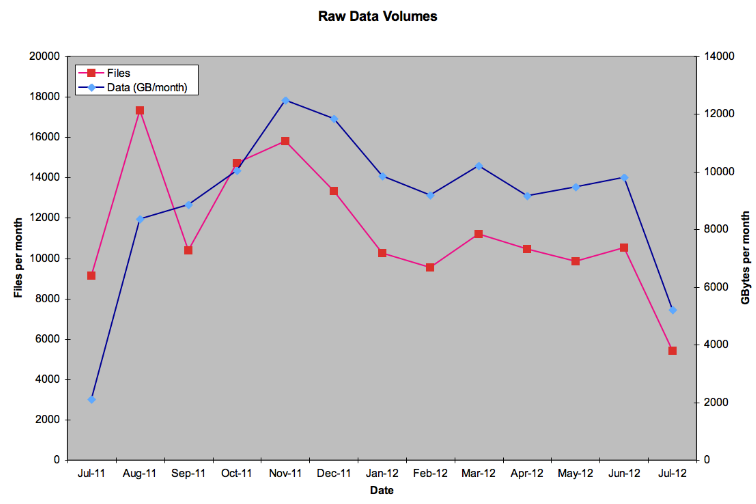
All report ~5σ exclusion of θ_{13} = 0.

LBNL and IHEP slightly narrower: due to smaller systematic uncertainties.

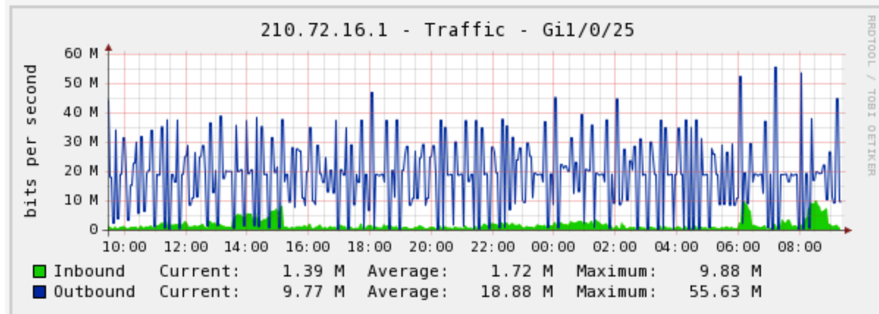


Mar. 1, 2012 Comparison of θ_{13} Oscillation Results 5/5

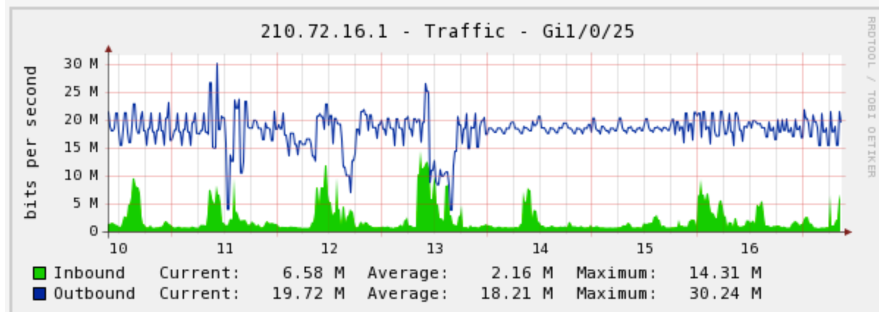
Data Volume & Transfer



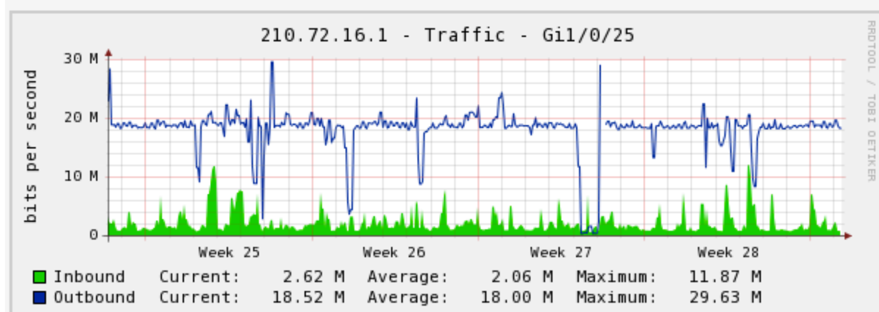
- **6-AD Raw data (19jul12):**
  - **— 70,000 files; 65 TB**
- **All raw data stored to tape.**
- **All processed data for major results stored to tape.**
- **KUP, redundant data, minor results kept on disk until superceeded.**



Daily (5 Minute Average)

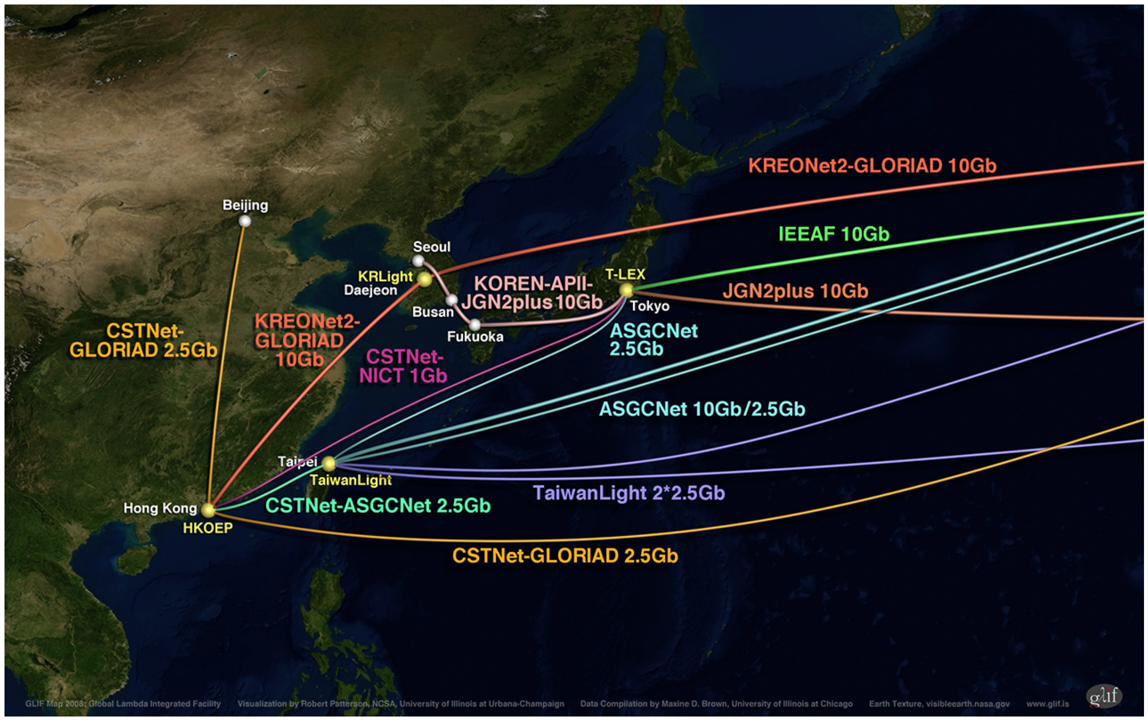


Weekly (30 Minute Average)



Monthly (2 Hour Average)

Network Outages



- **14Feb12: GLORIAD sub-oceanic cable damaged.**
  - **— Switch to ASCGNet (500 Mbps => 300 Mbps)**
  - **— 28 days to repair cable and return to service**
  - **— No impact on Daya Bay data transfer & communication**
  - **— Similar to ~6 week outage in Oct. 2011 (70 km offshore)**