Computing Framework and MC Production

Francesca Di Lodovico, QMUL

5th Open Hyper-Kamiokande Meeting
19-23 July 2014, Vancouver

Outline

- Agreement w/ Super-Kamiokande and T2K
  - Description of the Agreement
  - Current Status
- Release
  - First Hyper-K release
- Production
  - Overview
  - Data Retrieval
- Web Pages (documentation)
  - Wiki
- Computing Model Strategy

Y.Suzuki@Super-K meeting on May 10th, 2013

Report

- Request from HK working group for an use of SK softaware and SK data in the HK software delopement work.
  - They asked to make a similar agreement between SK and T2K

Answer to the request

1) The SK Executive Committee feels that Hyper-Kamiokande is a natural extension of Super-Kamiokande, therefore, we should cooperate as much as possible.

2) We think that a HK working group will develop their own simulation code based on the 'real' HK geometry and other parameters since the HK detector size and shape are quite different from SK. As HK develops its own simulation code, then HK software becomes less dependent on SK software. We believe that HK should start their code development now.

3) Certainly, we agree that HK may adopt concepts of the SK software and if necessary to copy SK codes into HK codes freely. Once done, HK may modify the codes as they wish without reporting all the modifications back to SK. We would like to be informed of any modifications that could also improve the SK code.

4) SK software should not be used for any other purpose than for HK simulation and analysis software development.

5) You mentioned the T2K-SK agreement as an example, but the situation of T2K is quite different from HK. T2K uses SK detector as a far detector, therefore, T2K definitely needs SK software and SK data, which is the reason we have made such an agreement. HK will not use the SK detector, so we do not need the same kind of agreement, as you have proposed.

6) SK will not form a group to have responsibility for maintaining analysis software of the SK detector for the HK group.

7) SK does not see any reason that the HK working group needs to use SK data for the HK software development. SK believes that HK needs only the MC generator, detector simulator and data analysis codes, which we are willing to provide, but not data.

Hyper-K code vs Super-K vs T2K

Vector Generation

Beam/ND: modified T2K code
Atmospherics: Honda flux
Nucleon Decays: SK code (vec_gen_b8)
Solar: SK code (ndecay_vc)
SN: new code (being worked on)

Simulation

WCSim

Reconstruction

fiTQun, BONSAI

HighE

LowE

Analysis

Atmospherics ν: OSC++

Beam ν: PROB++

Solar ν: SOLFIT, spallation cuts and related libraries

Recent progress on the migration from the SK to the HK code:

- Working on a new SN vector production
- All the vector files will then be added to the Generalized Neutrino Vector Generator when ready, now only for atmospherics.
- BONSAI will be added to the HK repository when interfaced to WCSim.Ongoing work.

Hyper-K Software Release

- All needed documentation is in https://wiki.hyperk.org/Software

- Use git (https://github.com/) as code-repository
  - Git is a modern distributed code management system, being used by several particle physics experiments (eg. SNO+)
  - Hosted at QMUL. Aim to mirror to other sites.
    - github.com can host packages, but they are publicly accessible unless paid for.
  - To get an account email your public key to both:
    - Francesca Di Lodovico f.di.lodovico@qmul.ac.uk
    - Alex Owen r.a.owen@qmul.ac.uk

- Added all the relevant packages to the repository.
  - All open-source packages (e.g. ROOT, GEANT4, etc.) are treated as “third party” packages (no developement in git).
  - Production scripts for HK are added to git.

WCSim: SVN to GitHub Migration

- WCSim will be migrating to GitHub from svn. This is to make it easier for many people to contribute to the code.
- WCSim users and developers will soon receive an email asking them to let us know if there are any changes which are not in the svn so we can commit them before the migration. You will receive another email when the migration is complete with instructions on how to use GitHub.
- More info on GitHub: https://github.com/
- A dictionary for git commands for svn users: http://git.or.cz/course/svn.html

By Erin O'Sullivan

Hyper-K Repository

- A total of 9 external open source packages:

CLHEP fiTQun GEANT4 iRODS NEUT Prob3++ ROOT WChSandBox WCSim

are in the git repository as:

hk-clhep hk-fitqun hk-geant4 hk-irods hk-neut hk-prob3++ hk-root hk-wchsandbox hk-WCsim

- Also a total of 6 Hyper-K packages (and increasing) are in the repository:

hk-eventdisplay hk-hyperk hk-simplefitter hk-TITUSanalysis hk-TITUSeventdisplay hk-utils

- More packages (eg for vector file production) will be added as needed.

Hyper-K Repository

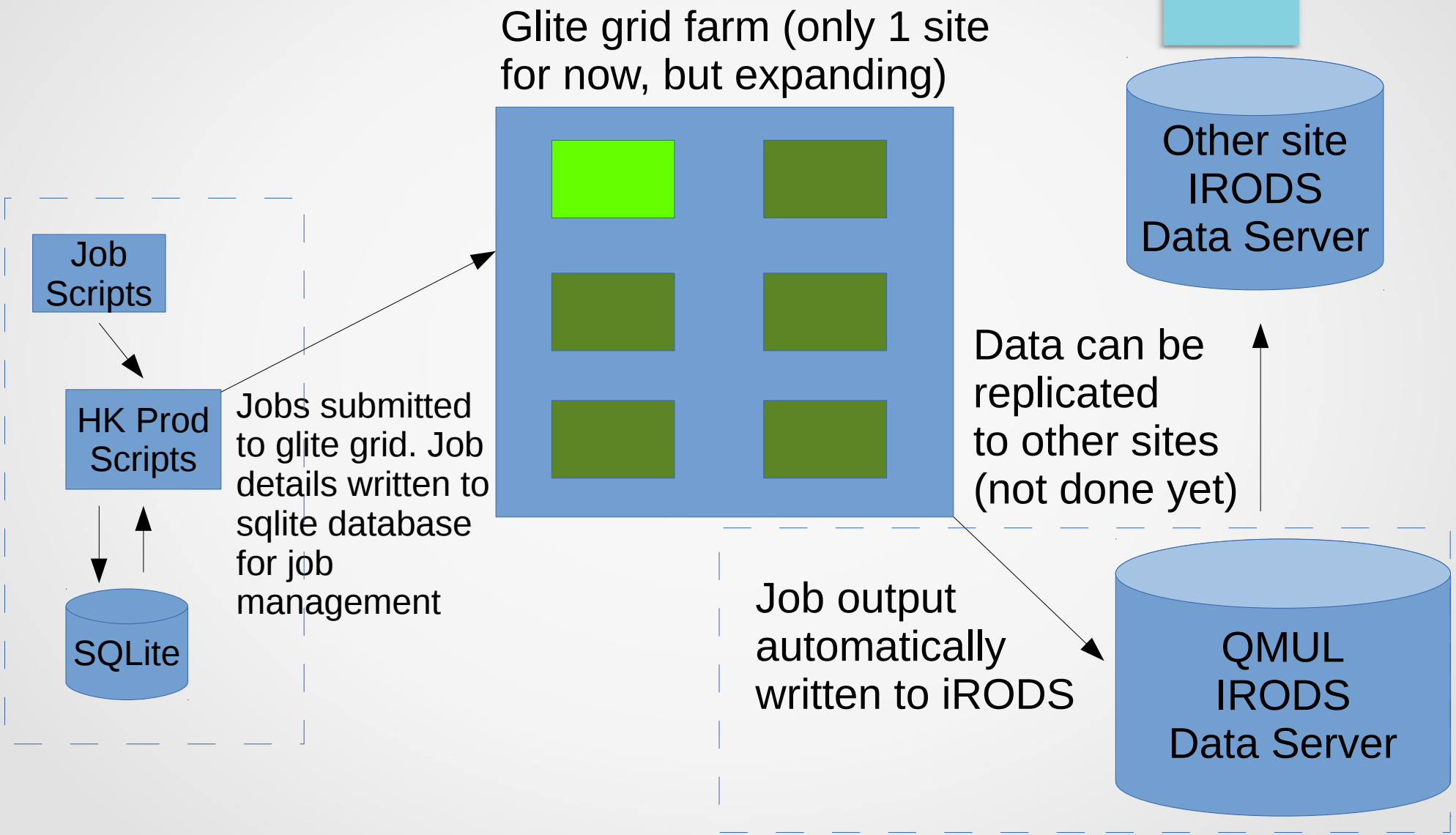
- Instructions to download the files are on the Hyper-K Wiki https://wiki.hyperk.org/

  > git clone ssh://git@poset.ph.qmul.ac.uk/hk-hyperk hk-hyperk
  > cd hk-hyperk
  > ./hk_get_release.sh <option>
  <option> : ALL, PROD, RECO, SIMU, HKANA, TITUSANA
- The above command will retrieve only basic files for given releases.
- build_hk.sh used to build production (simu+reco) packages.
  - Added build_reco.sh, build_simu.sh for simu and reco, separately.

Hyper-K Production

- Release deployed to CVMFS. Binaries read from there.
- HK Prod Scripts – generate grid job scripts and manage bookkeeping and manage submission to the grid
- SQLite database – stores bookkeeping information
- Job Scripts – shell scripts containing workflow: download files from iRODS, run job, store files In iRODS
- Package with scripts in HK repository: hk-utils

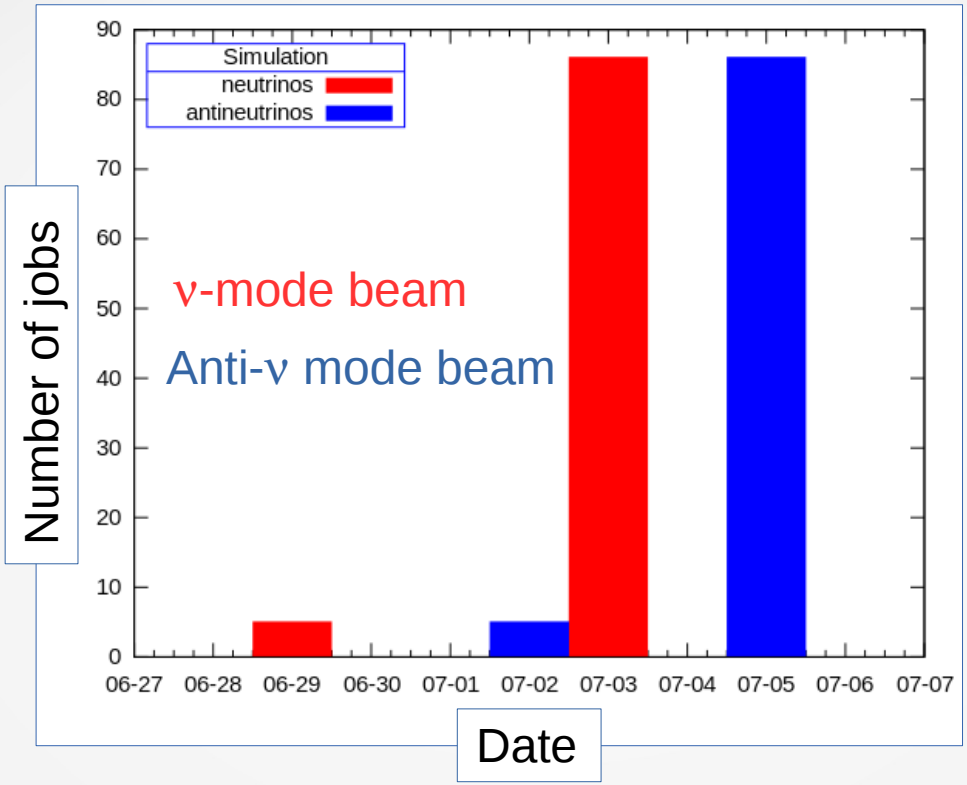
Hyper-K Production



Results

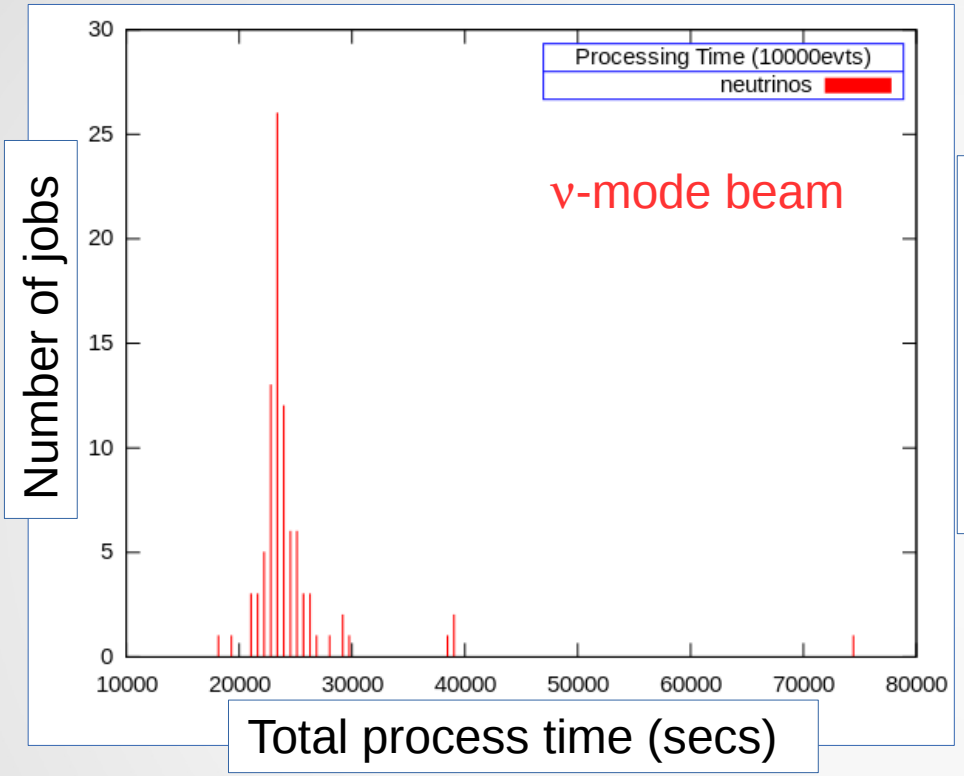
- Run 900K simulation (10K events/job) ν-mode and 900K simulation (10K events/job) $\bar{\nu}$-mode events
- Run 900K reconstruction (200 events/job) ν-mode and 900K reconstruction (200 events/job) $\bar{\nu}$-mode events
- Data stored and available in iRODS.
- Files can be easily downloaded and studied!
- All jobs run on the Grid – competing with processing from existing experiments.

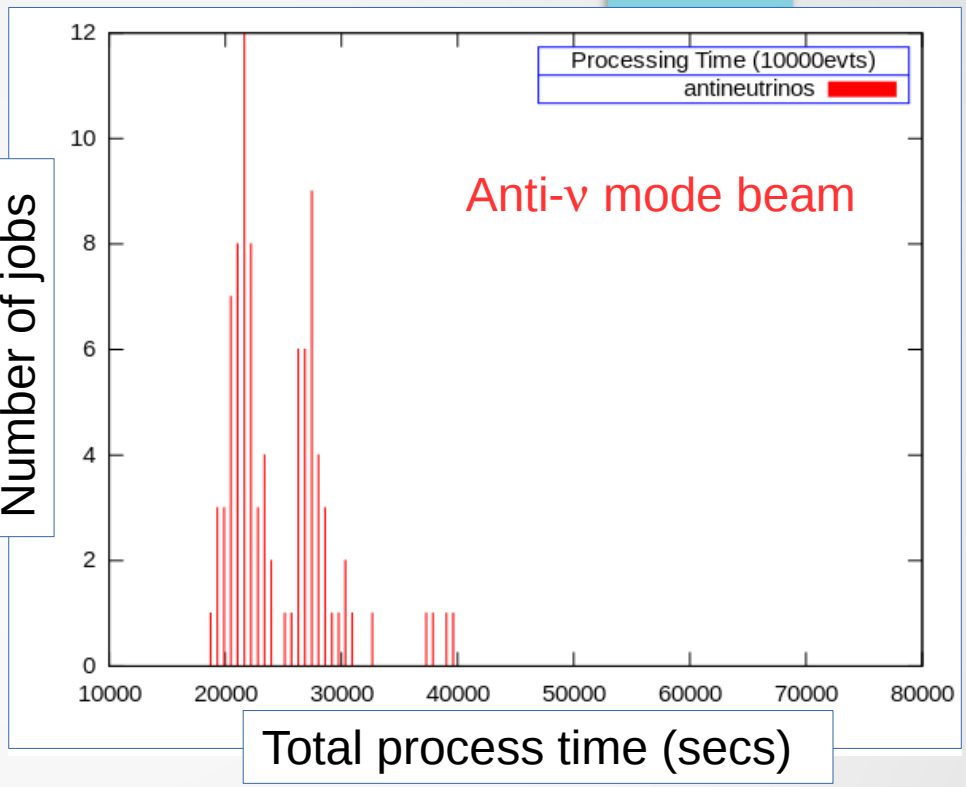
Simulation



- Simulation in 2 stages: small sample for initial validation (checks on the ntuple content done by Ryan Terri) then full sample

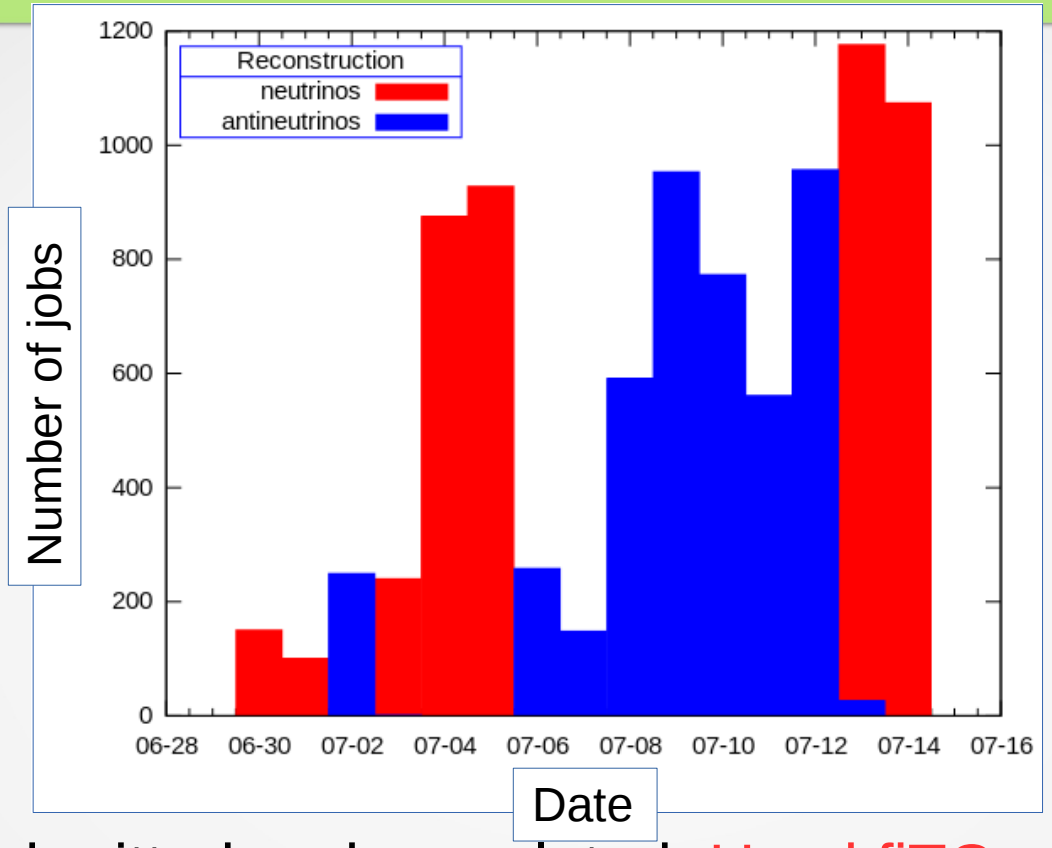
Simulation





- ~90 entries (jobs) per plot. Each jobs is 10k events
- Job duration for anti-neutrinos much wider than for neutrinos
- “Double peak” should be checked with more stat (it is expected not to be there – see eg reconstructed jobs)
- Average job time for simulation around ~2.5s/evt in both cases

Reconstruction

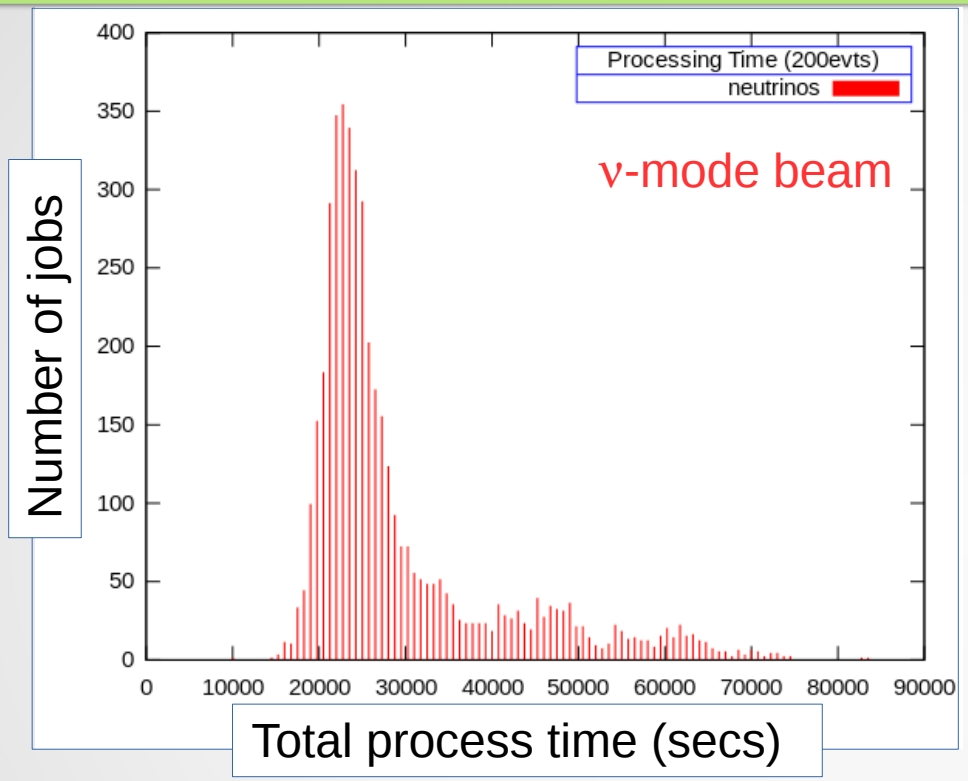


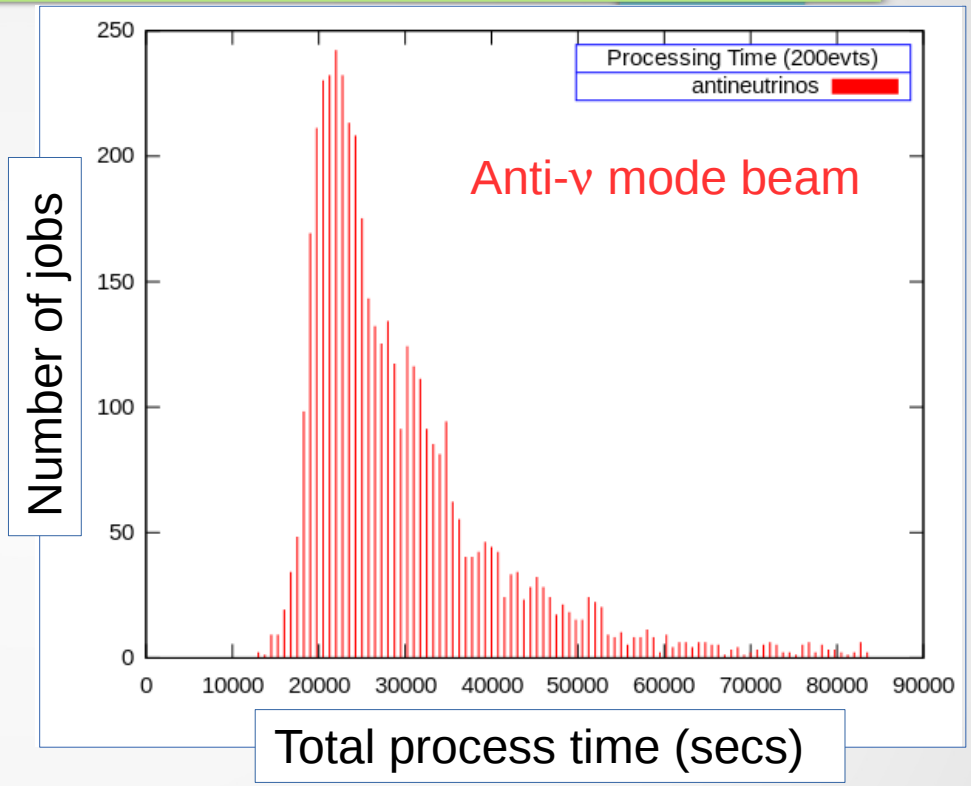
ν-mode beam

Anti-ν mode beam

- 9000 jobs submitted and completed. Used fiTQun v4r1, no MR.
- Dip in processing during 6/Jul-8/Jul due to certificate renewal problem (identified workaround for the future)
- Spanned ICHEP when queues relatively empty – processing at other times (30 June – July 3) will see a smaller throughput
- Aiming to increase number of sites providing compute resources

Reconstruction





- ~4500 jobs/plot, 200 events per job
- Peak centred about ~2min/evt for reconstruction jobs
- But, long tail with ~40% of jobs
- Means we can only safely process 200 events/job within certificate limit (~1 day)
- Profile for neutrinos and anti-neutrinos follows that of simulation

Hyper-K Data Storage

- Data available in iRODS (no Grid certificate needed):
  - /QMULZone2/home/hyperk/production/simu/
  - /QMULZone2/home/hyperk/production/reco/
- Data can being replicated to sites
- Production version maps to frozen versions of WCSim and fiTQun. File names have form
  - neutvect-<run#>-<sim#>.root
  - fitqun-<sim#>-<rec#>.root
  - In this case 1 run = 1 sim file
  - Log files written to 'success', 'failed' collections.

Data Location

- Instructions on the wiki for iRODS:
  https://wiki.hyperk.org/Software/iRODSHK
- Simu neutrinos:
  /QMULZone2/home/hyperk/production/simu/v0r0p3
- Simu antineutrinos:
  /QMULZone2/home/hyperk/production/simu/v0r0p4
- Reco neutrinos:
  /QMULZone2/home/hyperk/production/reco/v0r0p4
- Reco antineutrinos:
- /QMULZone2/home/hyperk/production/reco/v0r0p3

It's crucial to check the files. It's easy to download them. Please aim to look at the files and report at the SW meeting or to me.

Production Summary

- Production scripts worked well
  - Capable of handling peak throughput of close to 1200 jobs/day with little management overhead
  - Fully automated workflow for simulation: run job and store data in iRODS
  - Almost fully automated workflow for reconstruction: download simu from irods, run job and store outout in iRODS
    - Only manual part is merging of small ROOT files which is necessary due to processing time/event allowing much fewer events to be processed
- Further improvements to scripts planned to avoid ~15% failure rate in reco due to 2 jobs running on same node:
  - Second job cannot download the simu as it already exists. Simply resubmitting later works
  - Solution is to have each job download the simu with a unique filename
- Will need to modify collection names to include event type
- **Available 1.8M reconstructed ν and anti-ν beam events to check.**

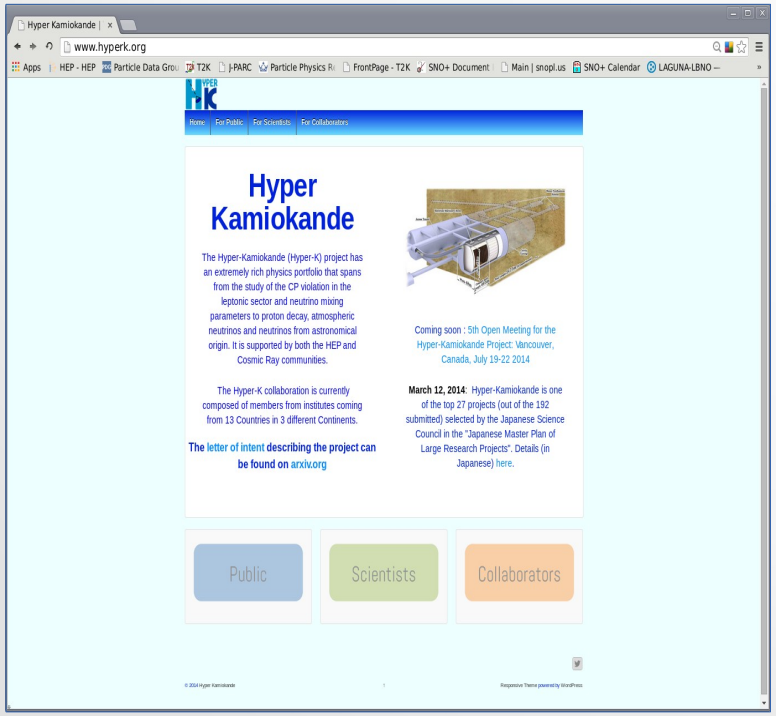
Web Sites

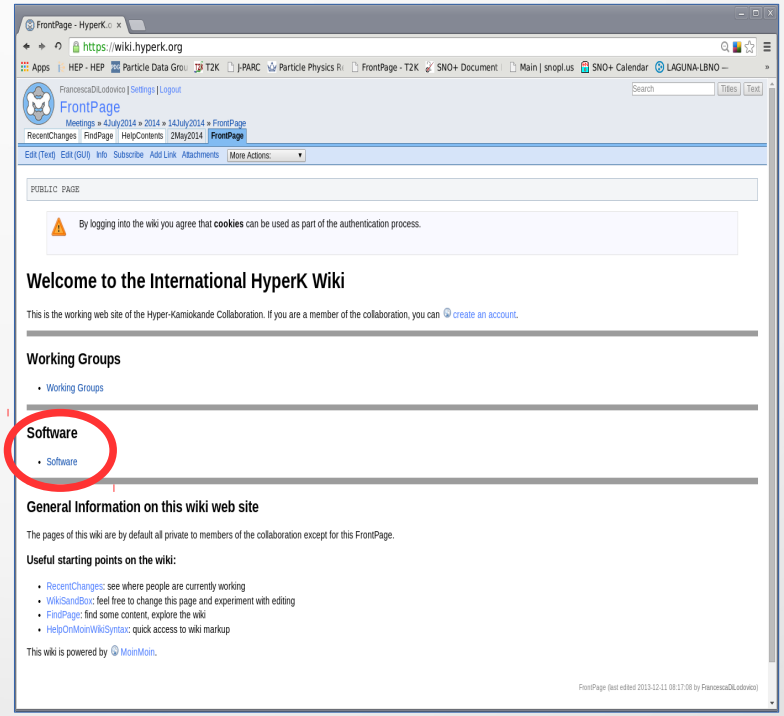
As a by-product of the Grid VO (Virtual Organization), we have a domain hyperk.org.

Two web sites:

Public web site: http://www.hyperk.org/

Working wiki web site: http://wiki.hyperk.org





Hyper-K Wiki

Working wiki web site: http://wiki.hyperk.org

The Wiki is meant to be the working web site for the experiment.

Currently contains all the needed software documentation:

- how to install a release

- how to retrieve the production files

If you are a T2K member and use the same email address as for t2k.org you will be automatically registered, otherwise I will get an email requesting to approve the account.

For questions on the registration, email Alex Owen (r.a.owen@qmul.ac.uk) and myself (f.di.lodovico@qmul.ac.uk)

Computing Model

Working on two major timelines:

Short Term future:

•Need to have a system automated and up and running for the current simulation production: use the Grid (VO: hyperk.org) - ongoing

Middle/Longer Term Future:

•Keep up to date with the current developments towards a Computing Model for Hyper-Kamiokande

•Current proposal presented @ the Second Open HK Meeting: http://indico.ipmu.jp/indico/contributionDisplay.py?contribId=5&confId=10

•Main aspects:

- Cloud
- Virtualization
- Digital Preservation

Conclusions

- Developed a fully comprehensive system for
  - Release code management
  - Processing on the Grid
  - Data Storage
  - Documentation
- The system is complete, but being extended.
- First MC beam simulated events produced. 1.8M reconstructed events available. Checks are needed before extending the production.
- Future plans:
  - Add more sites to production (started)
  - Produce atmospherics and other events
  - Implement comments from checks