Hyper-K / T2K Comp

Overview

Sophie King

2024/02/20

KING'S College LONDON

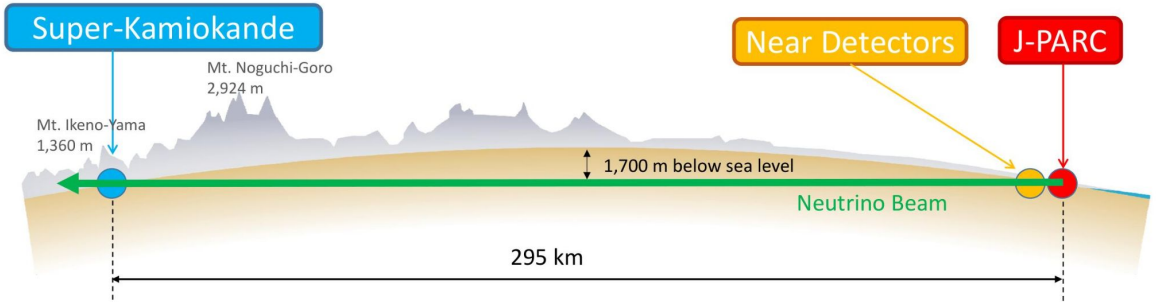
Hyper-K and T2K

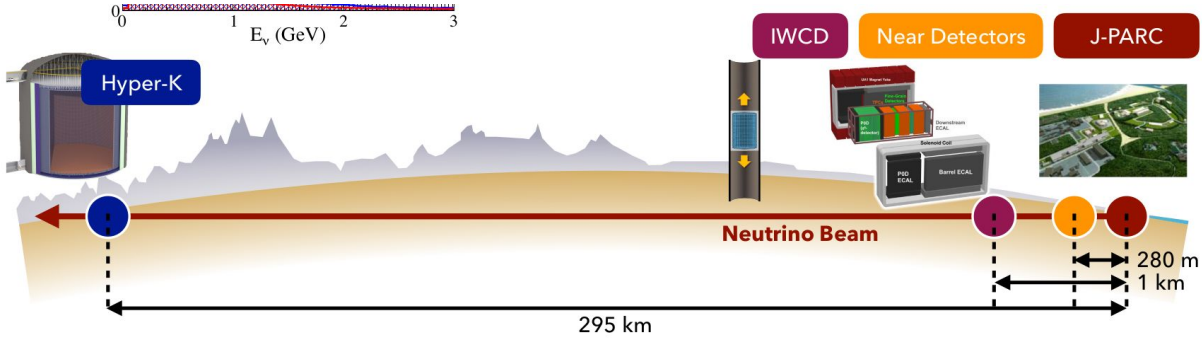
T2K consists of a far detector, SK, and near detectors including ND280.

When we refer to T2K computing this really means ND280 computing, as we do not cover the far detector.

Hyper-K computing will encompass not only ND280, but a new intermediate detector and a far with more than 20,000 PMTs.

→ **Big increase in the computing requirements going into Hyper-K era**





I won't distinguish discussion of T2K and Hyper-K Computing since use the same tools

Hyper-K Computing

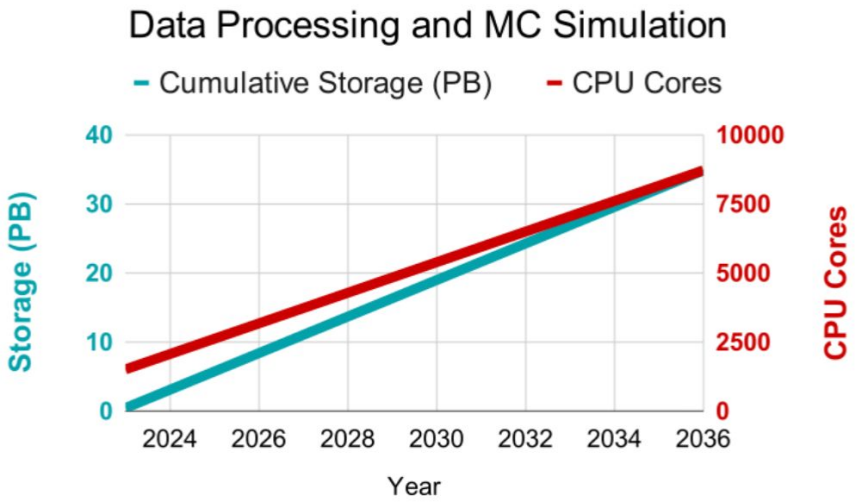


Figure 1: The total computing forecasts, covering all the Hyper-K detectors, considering both data and MC needs, and for all signal, background and control/calibration samples. Note that the storage estimate is per replica.

Storage requirements of Hyper-K are driven by the raw data of the far detector.

CPU requirements are driven by the reconstruction of the water Cherenkov detectors (far detector and intermediate detector)

Also plan to cover largest high-level analysis tasks (generally GPU based fitters), which is not yet estimated.

T2K and Hyper-K Computing

Convention: ‘T’ for ‘Tier’ - Data and MC storage attached to DIRAC
‘S’ for standalone (i.e. not centrally managed by DIRAC)

T0 - Primary raw data storage. KEKCC and Kamioka.
Not registered in DFC, but synced to T1s via Grid SE

T1 - DIRAC storage and/or CPU/GPU
Enough to cover data and MC of 1 or more detectors

T2 - DIRAC storage and/or CPU/GPU
Any resources connected through DIRAC that are smaller than T1

T2K and Hyper-K Computing

Convention: ‘T’ for ‘Tier’ - Data and MC storage attached to DIRAC
‘S’ for standalone (i.e. not centrally managed by DIRAC)

SC1 - Standalone Compute 1.
e.g. HTC/HPC clusters, cloud resources
- CPU/GPU at a similar level to a T1.

SC1 - Standalone Compute 2.
Smaller sites (clusters,cloud etc) where a given service task is handled,

SS3 - Standalone Storage 3
Any non-DIRAC stroage accessible to all Hyper-K collaborators.
Not used for data or MC, more commonly analysis level files.
e.g nextcloud, irods

GridPP DIRAC - WMS

GridPP (UK, Imperial) provide multi-VO service that both T2K and Hyper-K use

GridPP DIRAC WMS

- connect various Grid resources (UK, France, Italy)
- also tested with cloud resources in JENNIFER2 (INFN, IN2p3)

T2K/HK Comp tools use the DIRAC API for job submission

Advanced features such as production transformation not available on multi-VO, but is something T2K and Hyper-K are interested in.

→ Longterm Hyper-K would like to contribute person power to help realise features

- currently in the process of building a computing group
- do not yet have enough resources

GridPP DIRAC - DFC

GridPP (UK, Imperial) provide multi-VO service that both T2K and Hyper-K use

GridPP DIRAC DFC

- Hyper-K and T2K use a mix of
  - DIRAC API (comp scripts)
  - DIRAC RMS for FTS (comp scripts)
  - dirac-dms command line (users)
  - gfal command line (digging into issues, checking for dark files etc)

- Tested use of DFC metadata
  - plan to roll this out into standard production and comp tools

- Would be interested in use of transformation system when available to multi-VO

T2K and Hyper-K Computing

T2K and Hyper-K largely follow the same computing model and use the same tools (seeking to further consolidate these into a single computing package and model).

→ Based around **DIRAC, hosted by GridPP/Imperial.**

Hyper-K Computing will need to cover everything T2K/ND280 Computing covers, but on top of this we need to consider

- **larger data rate for Hyper-K** (4-5TB/day to be stored off-site)
- much **more demanding CPU and Storage requirements** for production across multiple detectors and different physics samples
- evolve to **incorporate machine learning into comp model and workflows**
- expand to cover demanding **high level analysis tasks.**

Hyper-K will be running into 2040s, so we need to look ahead and keep Hyper-K up-to-date with community tools and technology.

Backup

Title