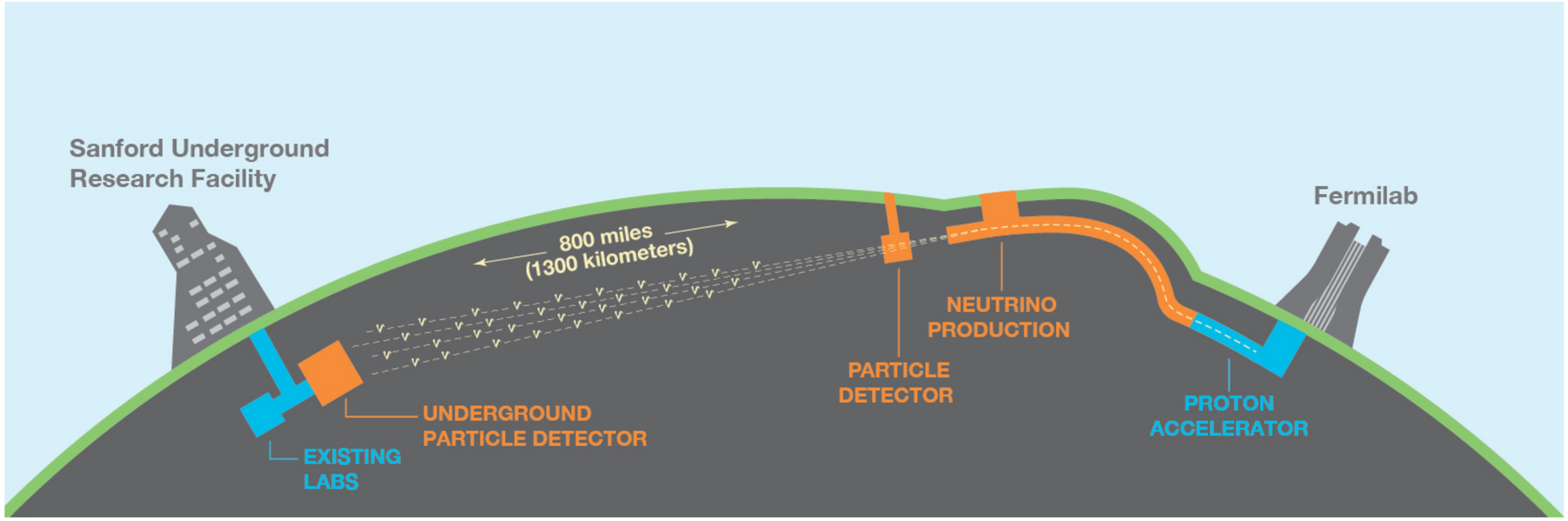
Imperial College London

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Kubernetes for DUNE DAQ

Pierre Lasorak
for the DUNE collaboration

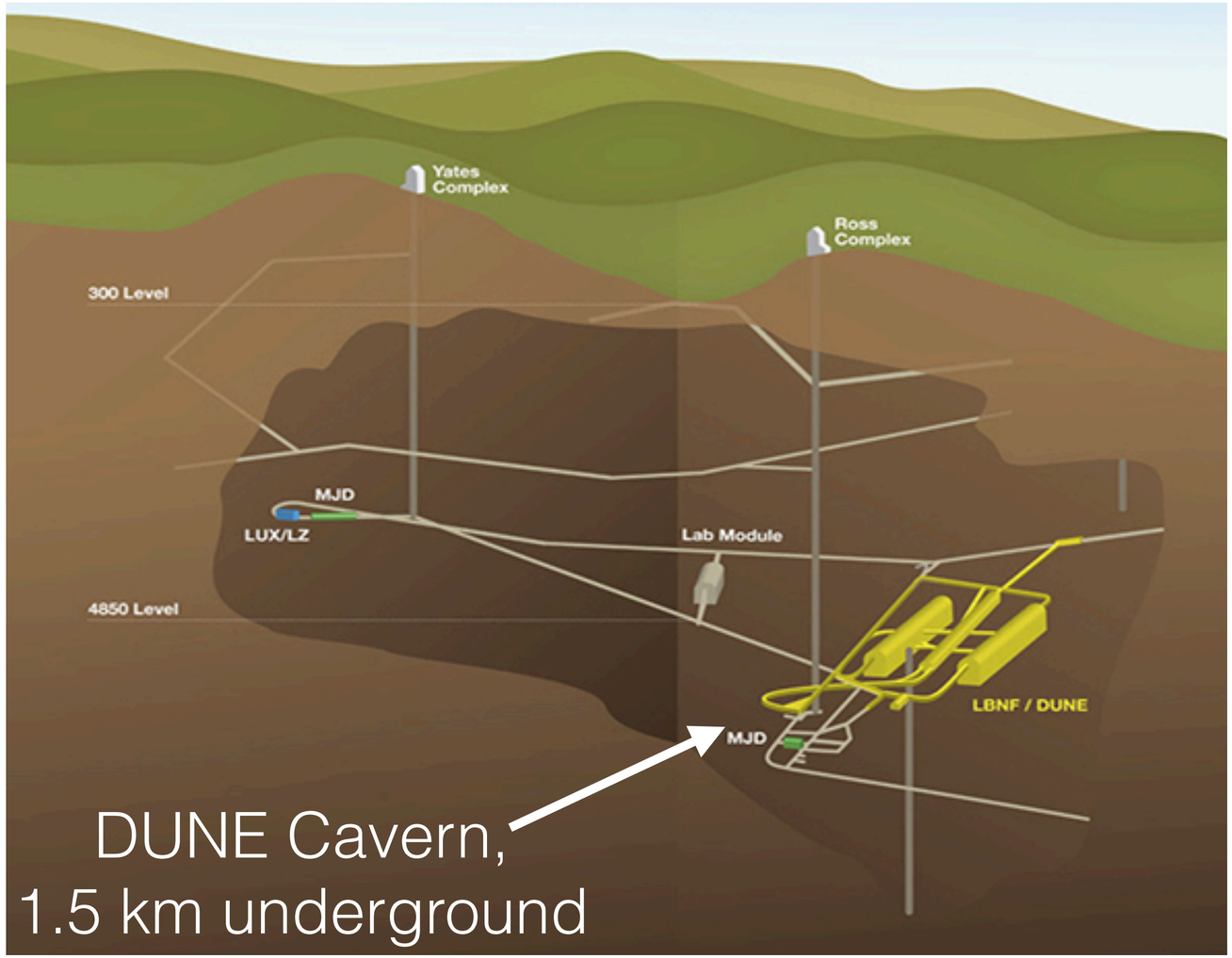
- Deep Underground Neutrino Experiment (DUNE): next-generation long-baseline neutrino oscillation experiment based in the US
  - High-intensity neutrino beam, produced at Fermilab (2 MW)
  - Neutrinos are measured at the near detector (0.5 km away from the source)
  - Neutrinos travel to SURF (Sanford Underground Research Facility) where oscillations are measured



- Physics goals
  - Neutrino oscillations (δ_{CP}, mass ordering)
  - Supernova neutrinos burst detection
  - Beyond the standard model

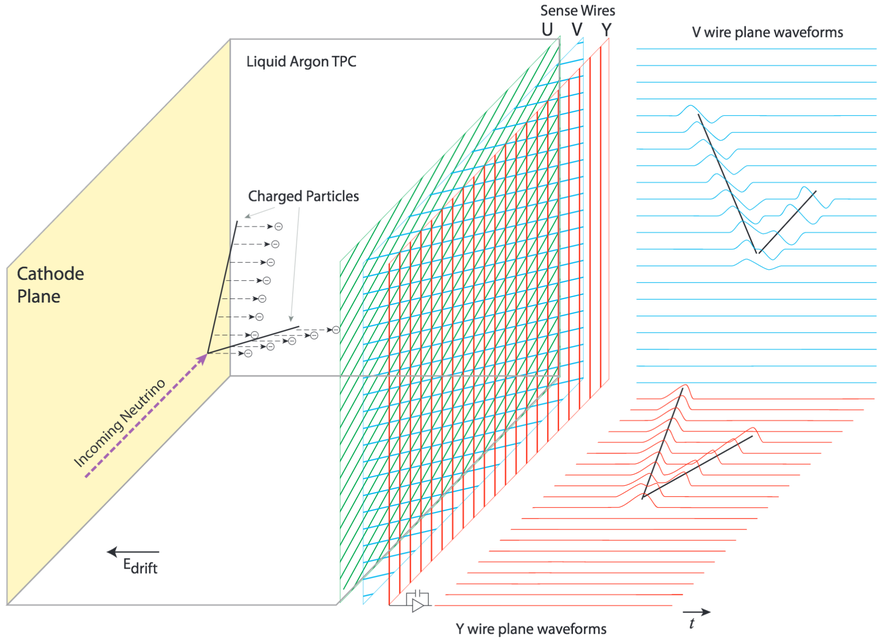
DUNE FD

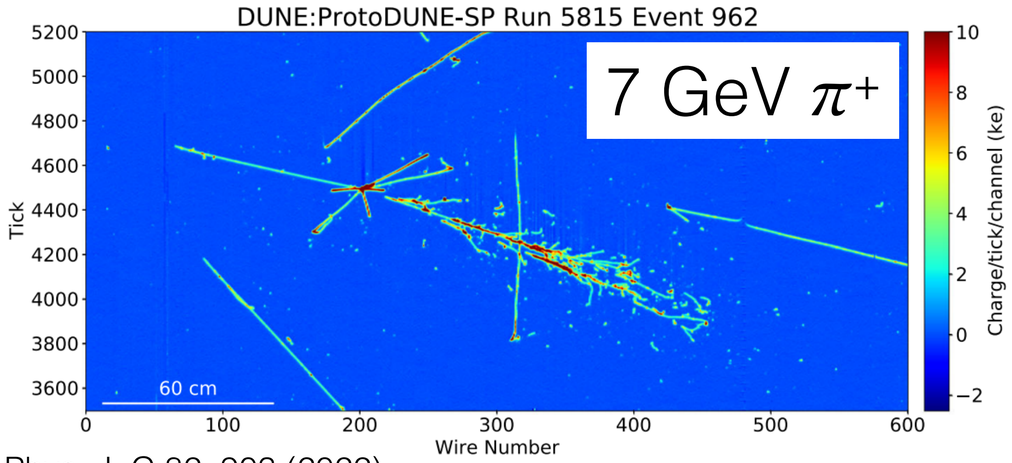
- DUNE Far Detectors (FDs)
  - 4x17 kt liquid argon module
  - 1.5 km underground
  - Remote area in South Dakota
- Implications
  - Difficult access
    - Remote operation of the experiment
    - Immediate intervention/support is impossible
  - Limited power & cooling



DAQ-Physics

- LArTPC detectors (Liquid Argon Time Projection Chamber)
  - 1st FD module: 165 readout units (TPC + photon detectors)
  - 2560 channels / TPC unit
  - 14 bits @ sampling rate of 1.95 MHz / wire
  - Total throughput: 1.2 TB / sec / FD module
    - Real-time processing in CPU
- A range of events types need to be handled by the DAQ
  - Couple of 100's MB to > 100 TB
- We cannot miss the next supernova!
  - Uptime requirement 99%

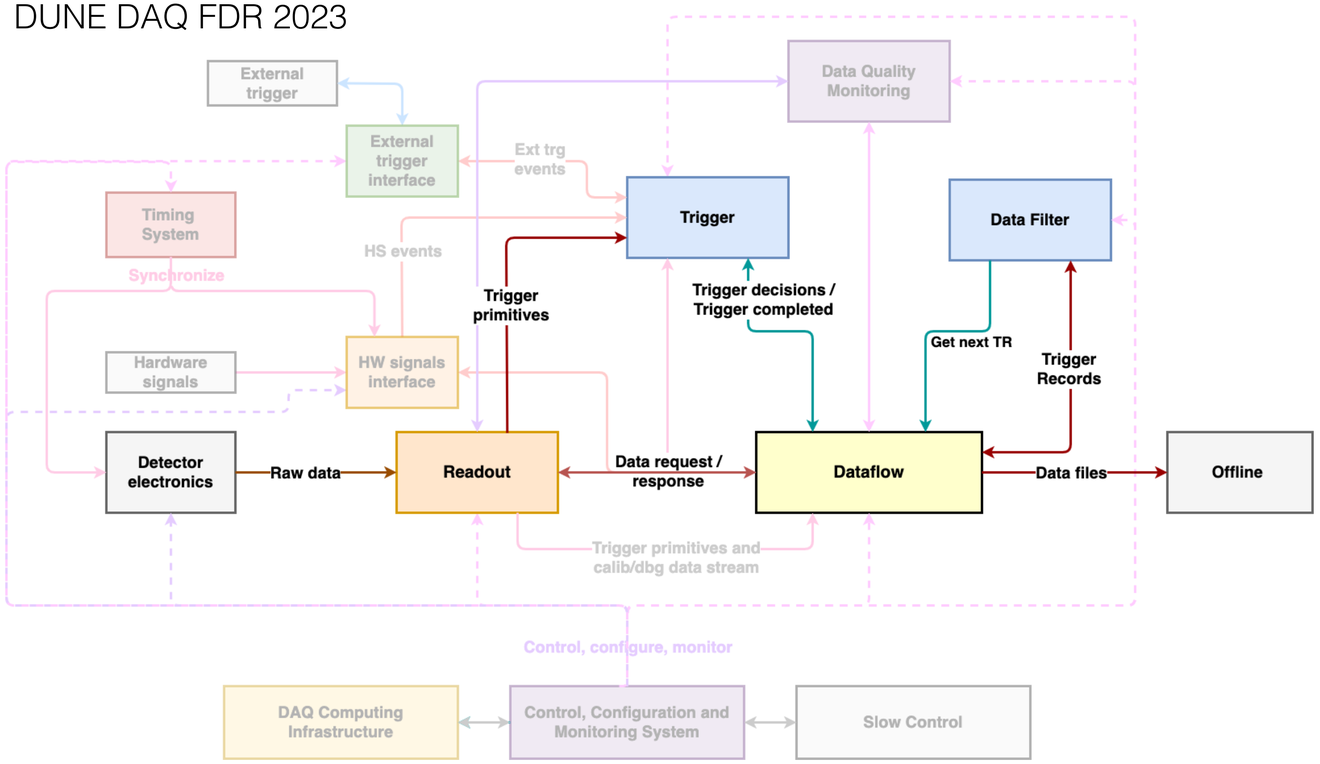




Eur. Phys. J. C 82, 903 (2022)

DAQ overview/infrastructure

DUNE DAQ FDR 2023

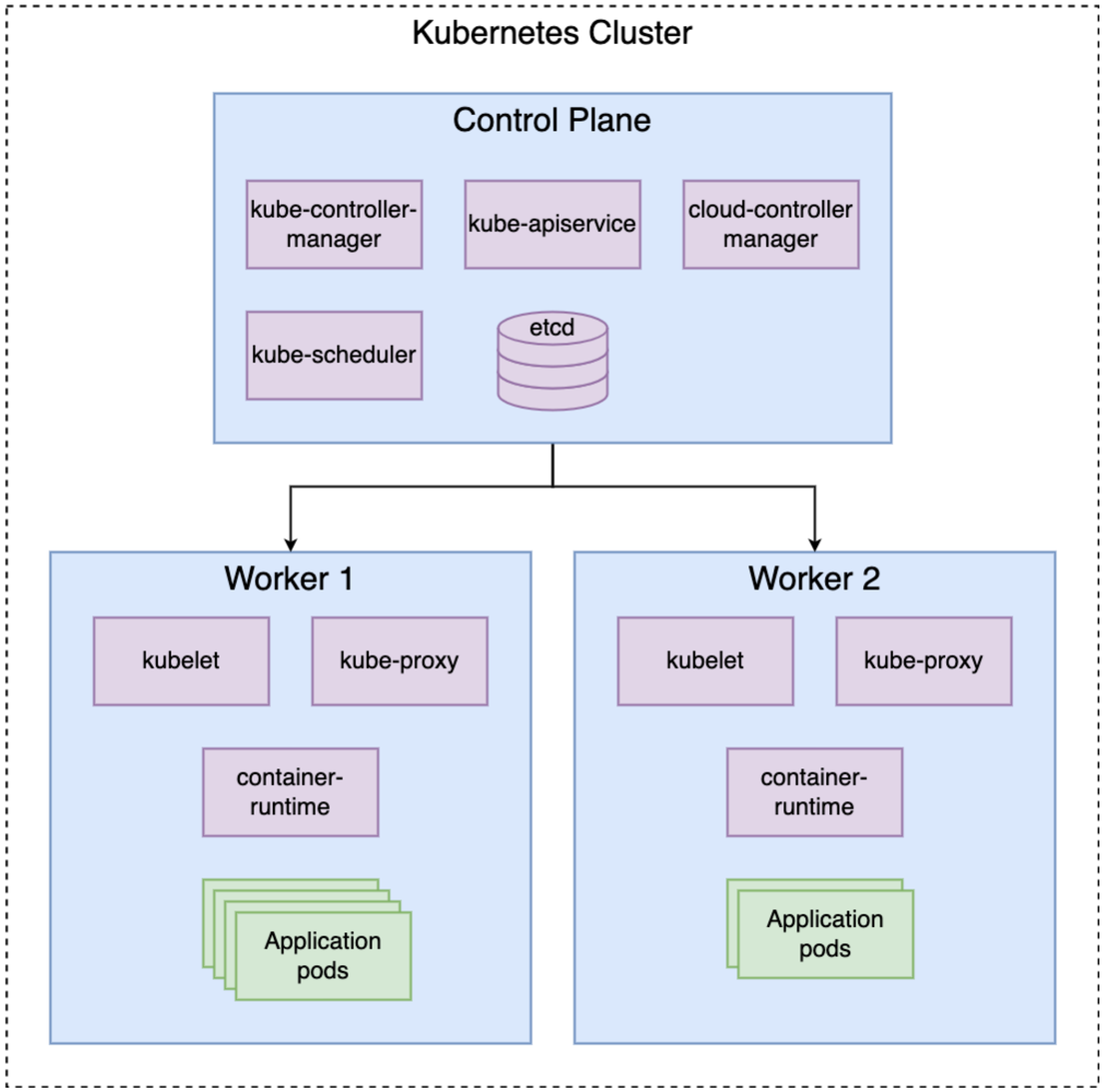


- Timing
  - Custom hardware
- Infrastructure, control and monitoring: ~30 servers underground
  - 20 (micro-)services and databases

- Readout: ~80 servers underground
  - High-end NIC
  - Multi socket CPU
  - High-end SSD storage
- Trigger: ~20 servers underground
  - Dynamically reconfigurable
- Dataflow: ~10 servers on the surface
  - Disk access
- Data filter: ~20 servers on the surface
  - Disk access

Kubernetes for DAQ

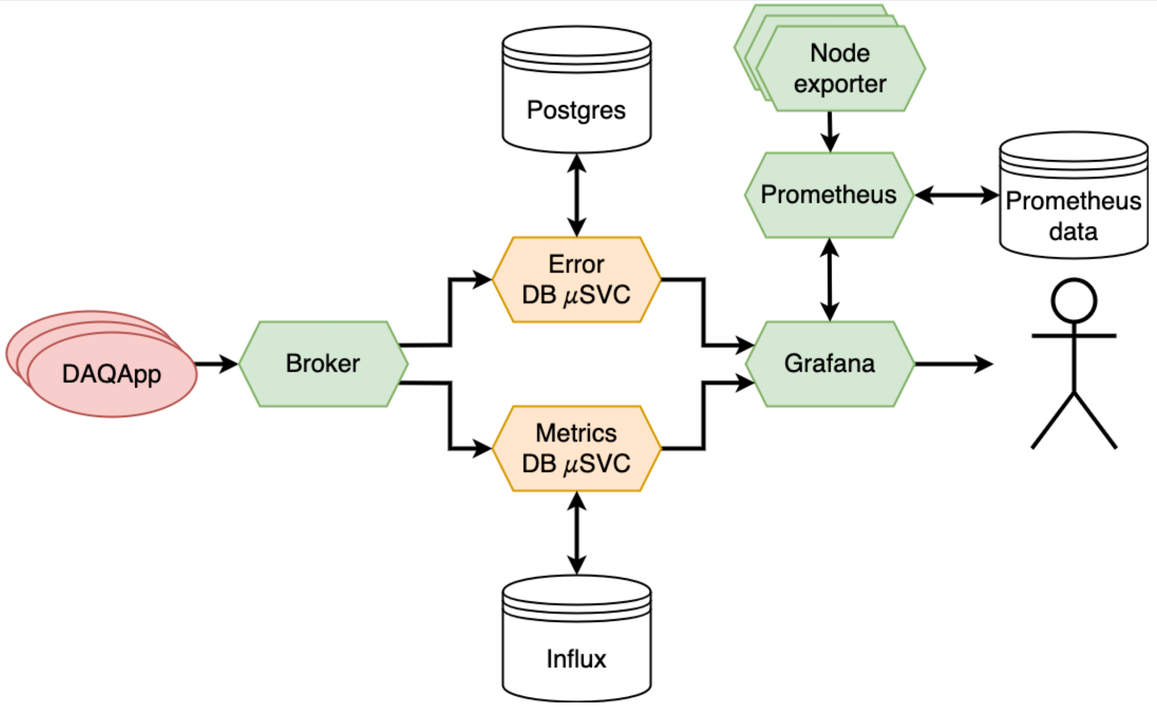
- Kubernetes (K8s) is a container orchestration tool
  - Created by Google in 2014 to manage data centres
  - Open source
- Allows distribution of containerised workloads on any host in the K8s cluster
  - High flexibility, resilience
  - Designed to maximise uptime of data centre
  - Container images used for portability
  - Allow versioning of the system
  - Multiple and expendable resource definitions
  - Many tools provide K8s-ready images and YAML
- Many features align with the high uptime, high-reliability requirements of the DAQ



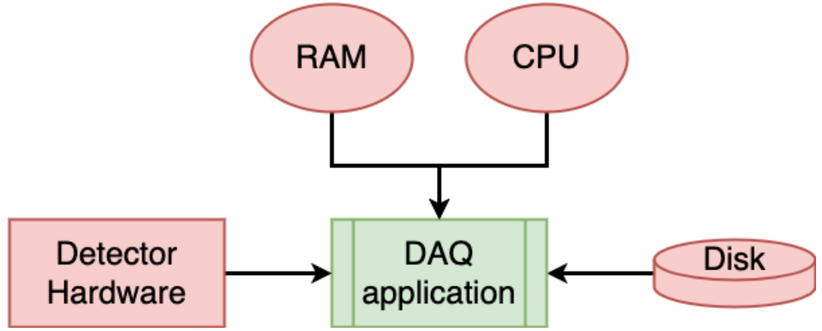
Kubernetes for DAQ

- DAQ system ≠ data centre!
- Naturally fitting in K8s
  - Services (flask,…)
  - Web UI
  - Databases
- Potentially challenging
  - DAQ readout processes
    - Hardware interaction
    - Pinning to Host, CPU, RAM
    - Networking
  - Data flow, data filtering

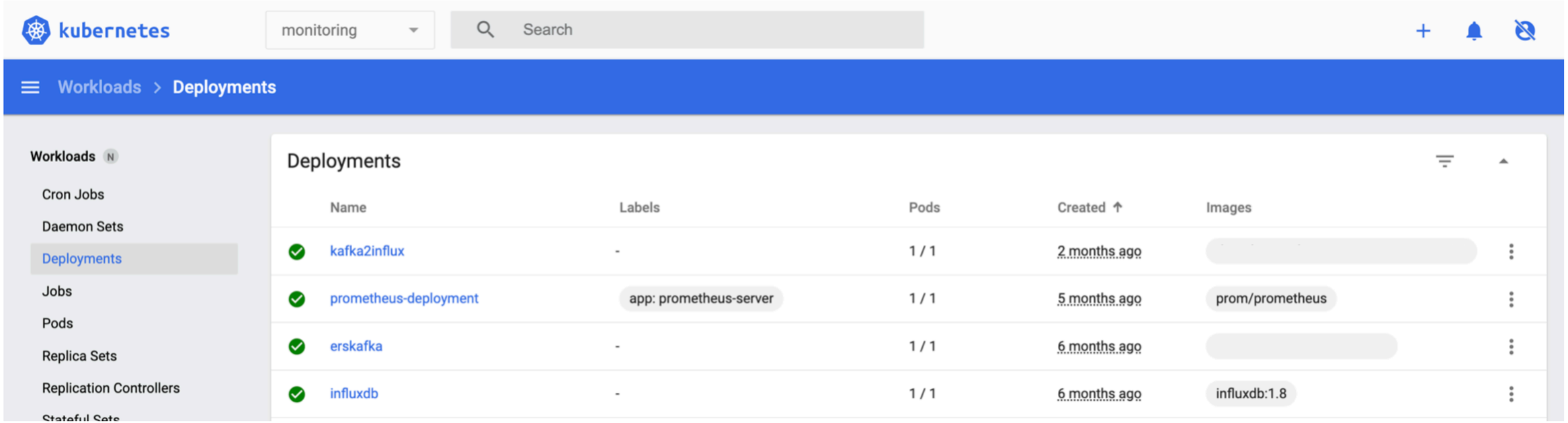
Monitoring



Generic DAQ application



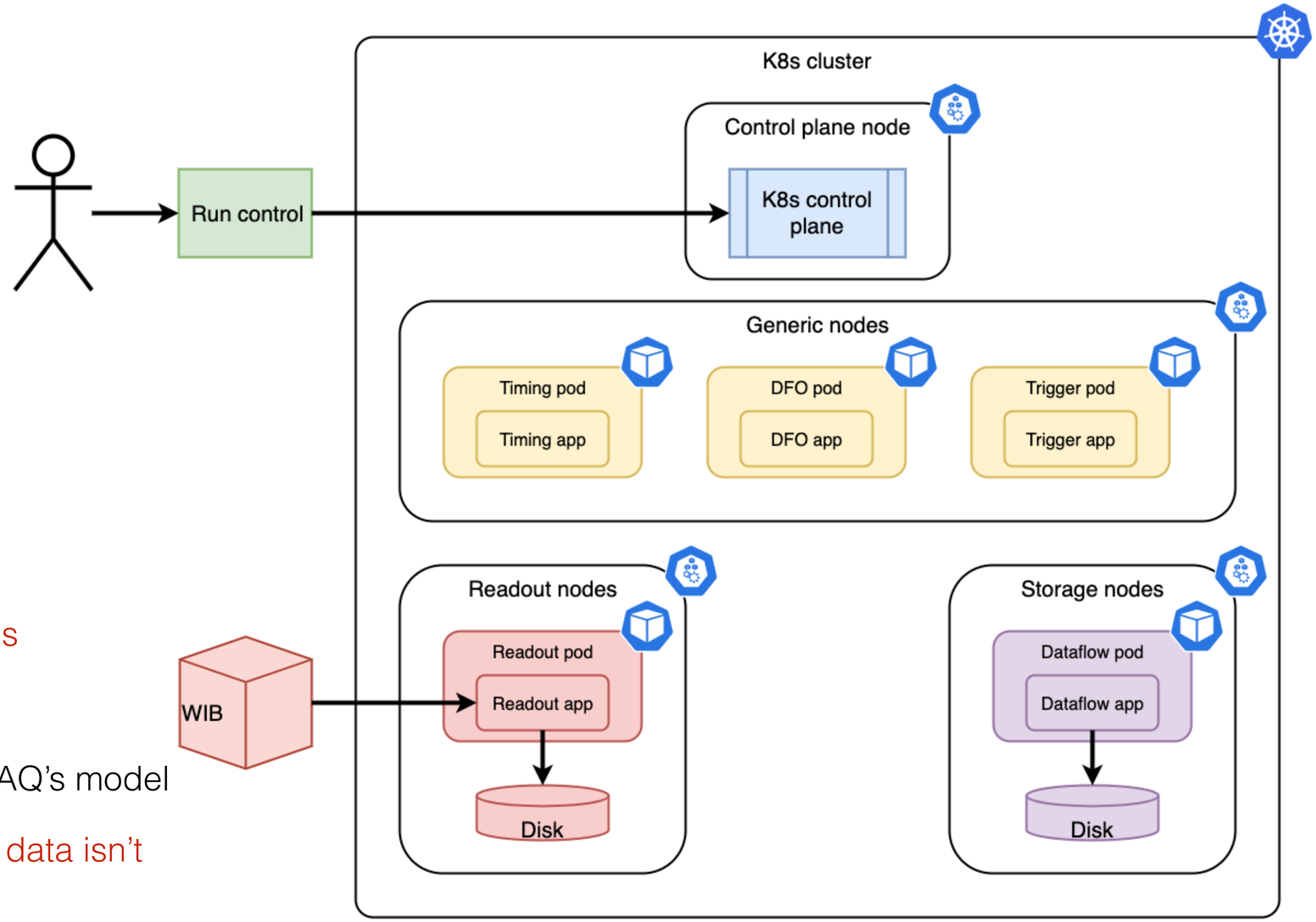
DAQ Services



- A subset of custom DAQ services deployed on our ProtoDUNE K8s server @CERN
  - InfluxDB holds DAQ metrics (trigger rate etc.)
  - ERS (ATLAS' Error Reporting System) used to handle error/warning/info messages
  - Prometheus for node monitoring (RAM etc.)
- Already proven to be an efficient way to deploy & monitor services (ours and from others)!

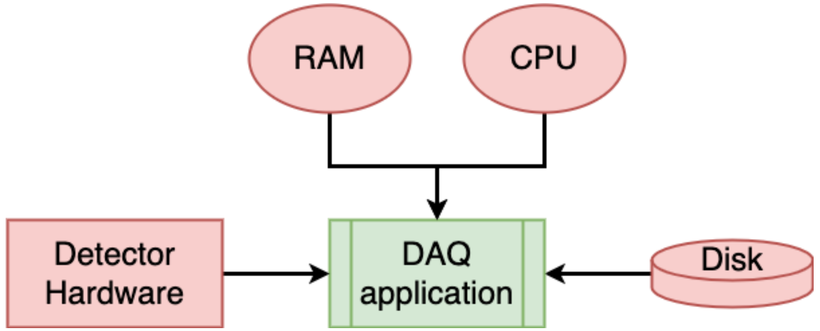
DAQ applications in Kubernetes

- Process management
  - K8s uses scheduling
    - Several solutions to start a manage processes
    - Not naturally aligned with DAQ
  - Investigating how to best use K8s scheduling, prototype so far:
    - DAQ applications in Deployments
    - Run Control starts and stops the processes
- Networking
  - Multiple sessions concurrently with namespaces
- Storage
  - K8s model for storage is not aligned with the DAQ's model
  - Continuously flushing output directories, so the data isn't persistent in the same way as implied by K8s



Readout application in Kubernetes

- Readout applications need to be locked on specific resources
  - I/O hardware connected to the detector
    - User will want to know exactly which part of the detector is read out
    - DeamonSet runs on all the nodes, discovers connected devices and creates resources
  - CPU and RAM in the same NUMA region
    - Hit finding is done on the readout host
  - Disk
    - Readout can stream the complete data stream on SSD (supernova)
      - Requires write optimisation (O_DIRECT, etc…)



Software development in Kubernetes environment

- Development needs to be integrated with K8s to enable deployment to the production system to be smooth
- Some of the issues encountered so far
  - Complete DAQ libraries + dependencies are large
    - Standalone image ~ 5 GB
    - Resort to mounting CVMFS? Not envisaged in the final system
  - Extra compilation time to build the image
    - Building an image for every build → Storage & distribution becomes complicated
    - Resort to NFS and mount the library directory? Not available everywhere
  - Application logs also tend to get lost…

Conclusion

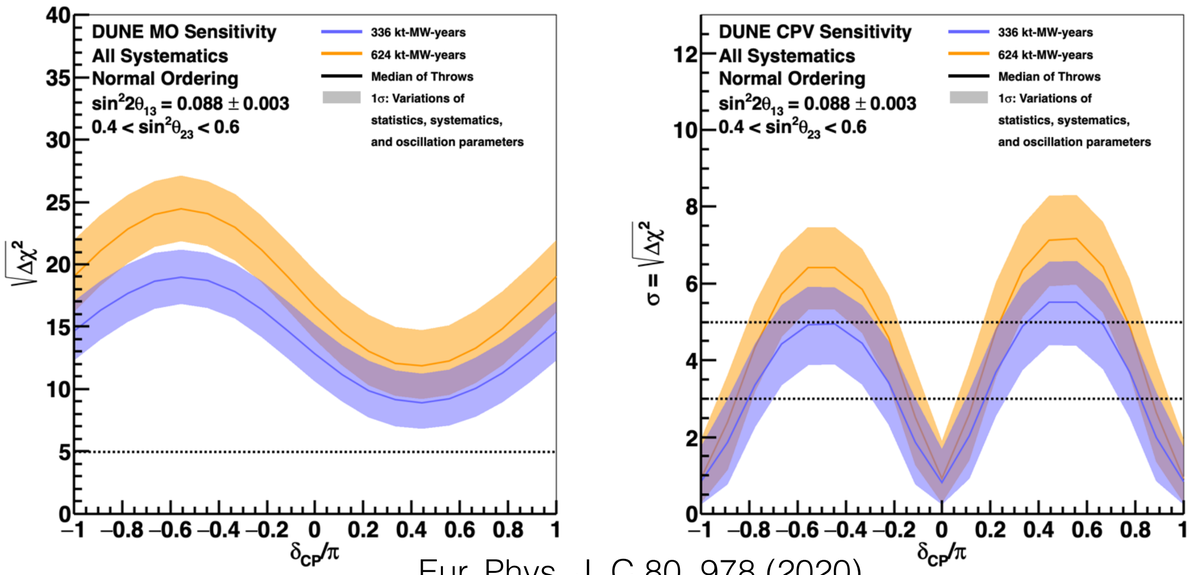
- We plan to use Kubernetes for the DUNE DAQ
  - Services will run in containers in the Kubernetes cluster
  - DAQ applications may run in the cluster
    - Solves some problems
      - Smart resources and process management
      - Application-level networking becomes simpler
    - Still many challenges
      - Networking, processing and IO overhead need to be understood
      - Development with containers
      - Details of hardware interactions and process pinning

Backup

DUNE Physics

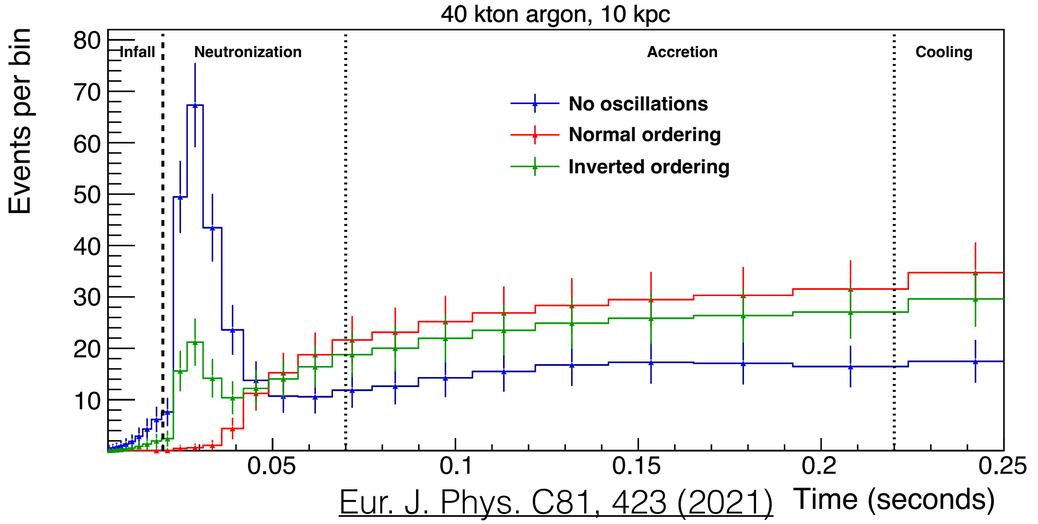
- Wide-ranging physics goals
  - Accelerator neutrino oscillation program + atmospheric neutrinos
    - Neutrino δ_{CP}
    - Neutrino mass hierarchy
    - Measure known parameters with increased precision
  - Core collapse supernova neutrinos detection
  - Solar neutrinos
    - HEP neutrinos
  - Beyond the standard model
    - Proton decay
    - Non-standard neutrino oscillations
    - Dark matter detection

Oscillation physics



Eur. Phys. J. C 80, 978 (2020)

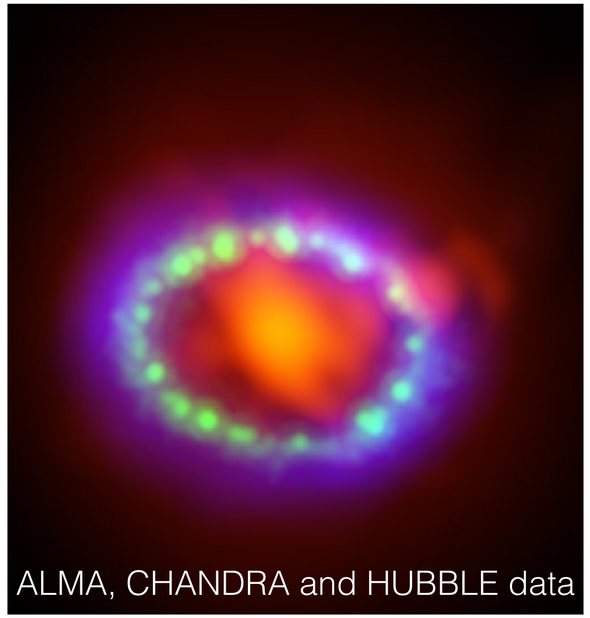
Supernova physics



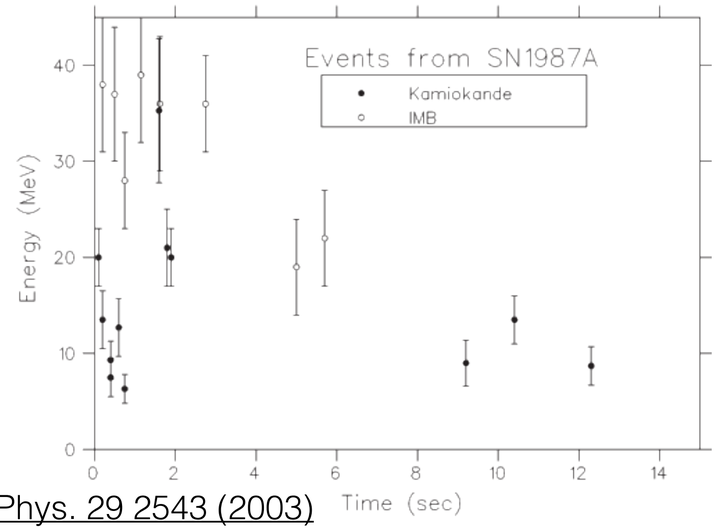
Eur. J. Phys. C81, 423 (2021)

Uptime requirement

- Detectable supernova neutrino bursts are expected to happen ~ every 100 years
- One recorded so far, SN1987A
  - ~20-25 recorded neutrinos + clear light signal
  - > 2000 citations! Huge implications for supernova physics!
- Cannot miss the next one
- Stringent up-time requirement on the Data AcQuisition (DAQ)
  - Ability to take out part of the detector during a run



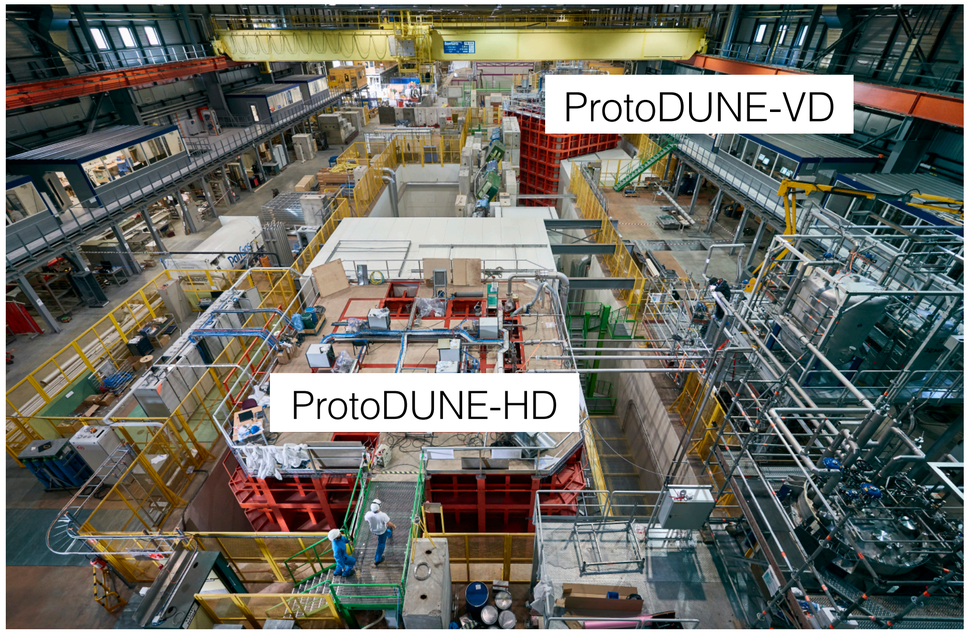
ALMA, CHANDRA and HUBBLE data

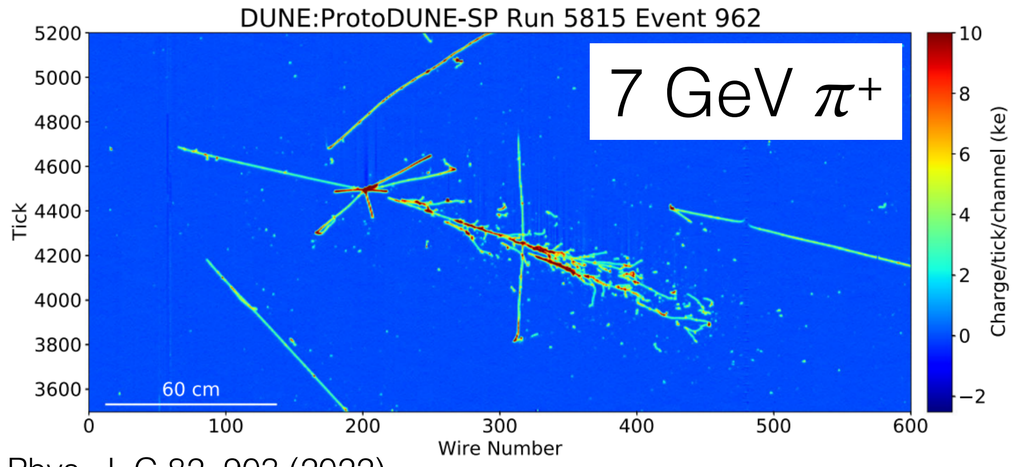


J. Phys. G: Nucl. Part. Phys. 29 2543 (2003)

ProtoDUNE

- At CERN Neutrino Platform
  - 2 test beam detectors using horizontal and vertical drift TPC (ProtoDUNE-HD and ProtoDUNE-VD)
  - 2 smaller coldboxes
- Used to exercise the detectors
  - APAs and VD readout modules
  - Readout technology
  - Trigger and DAQ
- Physics too!
  - Test beam measurements (electron-argon, hadron-argon cross-sections…)

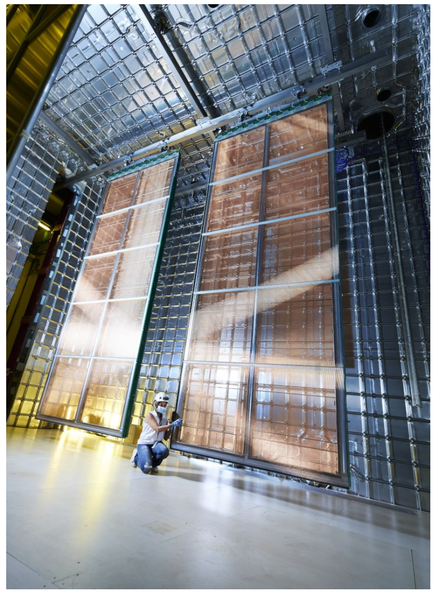


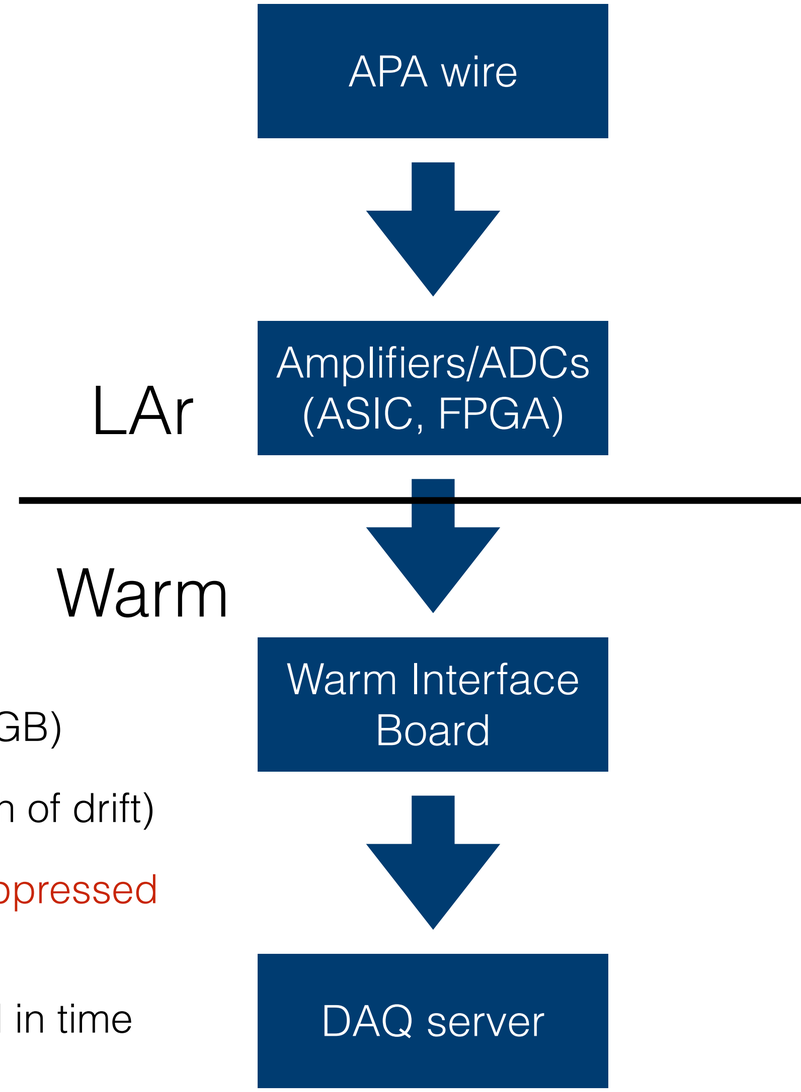


Eur. Phys. J. C 82, 903 (2022)

Readout at the HD

- Anode Plane Assembly (APA)
  - 2560 readout wires detect the drifting electrons
  - 150 APAs for the 1st FD module
- Signal amplified and digitised in the liquid argon
  - 14 bits @ sampling rate of 1.95 MHz / wire
- Warm interface boards (WIBs),
  - 4 WIBs / APA
  - Collect data from 4 ASICs
  - Throughput: 2.2 GB/s
- DAQ readout server
  - Connected via fibre-Ethernet to the WIB
  - One server / 2 APAs
  - No data reduction





- Events size
  - Beam / Atmospheric: 2.6 ms readout (3.8 GB)
    - Driven by the size of the detector (length of drift)
  - Supernova: 100 s continuous non-zero suppressed readout (140 TB)
    - Driven by supernova physics (extended in time and low energy)

DAQ control software

- Handle shifter/expert interactions
- Process management
  - Start and stop processes
- Send commands and keeping the DAQ system in a coherent state
- Resources management
  - Lock used resources and prevent oversubscription
- Executing automated recovery actions
  - Based on formatted error messages