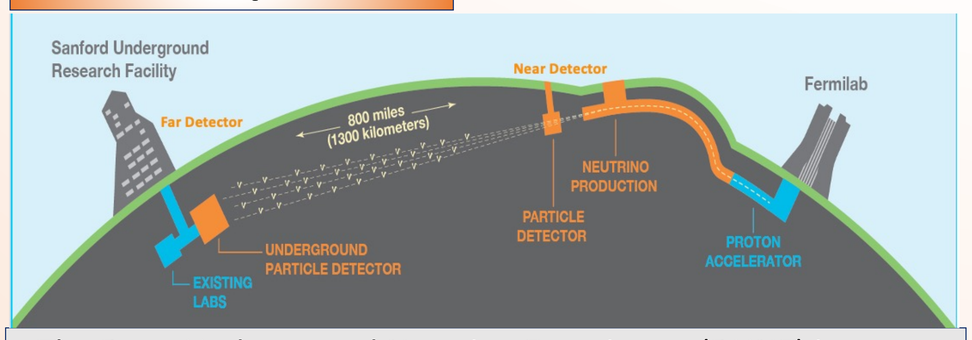
# FORM, a Fine-grained Object Reading/Writing Model for DUNE

Barnali Chowdhury, Peter van Gemmeren, Michael Kirby
on behalf of the DUNE Collaboration

INTERNATIONAL CONFERENCE ON COMPUTING IN HIGH ENERGY & NUCLEAR PHYSICS

## 1. DUNE Experiment

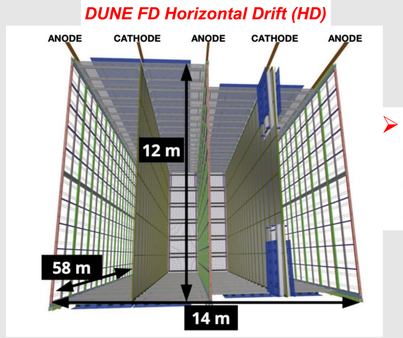



The Deep Underground Neutrino Experiment (DUNE) is a cutting-edge experiment for neutrino science. Major components:

- Neutrino beam from Fermilab to SURF Lab (South Dakota)
- Large 70kt LArTPC far detector (FD) and capable near detectors
- Search for neutrino CP violation, measure neutrino properties, look for supernova neutrinos.

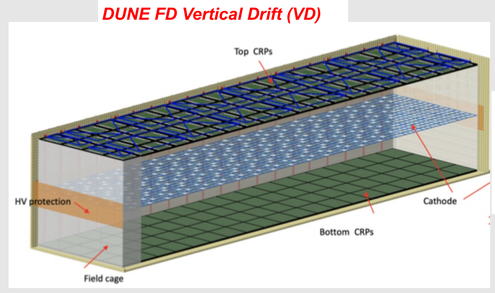
## 2. DUNE's Existing Framework Challenge

*DUNE FD Horizontal Drift (HD)*




- Raw data volume of DUNE far detector HD is 3.8 GB /event

*DUNE FD Vertical Drift (VD)*




- Raw data volume of DUNE far detector VD is 8 GB /event

- DUNE data-processing framework, *art,* is essential to process raw data for precise physics measurements
- Existing computing frameworks are based on "collider-physics" concepts
  - Data products are based on run, subrun, and event
  - An ATLAS/CMS event is significantly smaller, less than a few MB
  - But a DUNE event is more than a factor of 1000 times larger in size
- With large event size DUNE's main challenge is efficient "memory management" for data processing

## 3. Development of new Fine-grained Framework

- DUNE is expanding significant effort on developing new Fine Grained Data Processing Framework to overcome current framework limitations
- The new framework will SUPPORT flexibly defined, context-aware processing units to address the varying granularity necessary for processing different kinds of data.
- The framework MUST be able to read and write very large data collections (10s of GB in size) whilst maintaining a low (maximum few GB) memory footprint
- DUNE is working on several software framework requirements to ensure the future needs of the experiment and to integrate it with modern technology, for example, HPC.

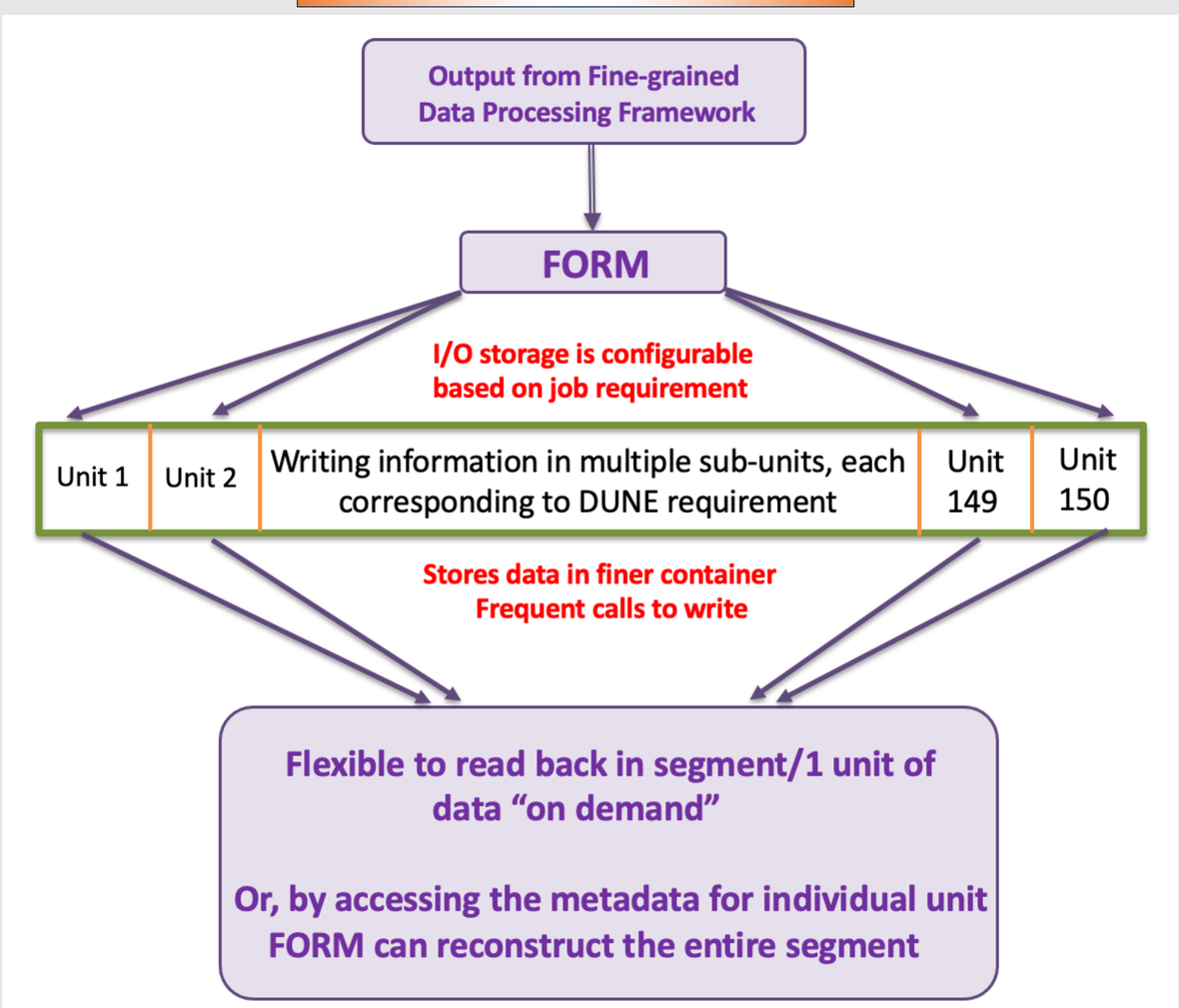
## 4. Development of FORM, Fine-grained I/O and Storage Infrastructure

- Memory Optimization will not work unless we integrate the evolving "Data Processing Framework" with Fine Grained I/O infrastructure
- DUNE is currently developing Fine-grained Object Reading/Writing Model (FORM)
  - to facilitate writing in the same configuration as data-unit processing
  - to store data in finer container and read back in segments
- The I/O infrastructure will SUPPORT writing multiple entries from a large DUNE event worth of several GBs in size
- The I/O infrastructure will also SUPPORT reading and accessing an individual entry written by FORM and storing metadata to retrieve information later
- With FORM, DUNE framework will ADAPT to eager writing and delayed reading resulting in memory optimization

## 6. FORM Infrastructure

- FORM will be designed to reading and writing in multiple persistent data formats
- The infrastructure will be transparent to I/O technology layer allowing integration with HDF5, TTree, RNTuple and others
- The infrastructure should be compatible with widely-used scientific computing systems to fully utilize DUNE computing resources.
- Parallel execution of reading/writing by the infrastructure
- The infrastructure can run framework-provided input, hence support reading/writing the entire DUNE Far Detector/Near Detector spills, or a subset of hits in that spill and then return the full spill

## 5. FORM Design




## 7. FORM Demonstrator

- DUNE has successfully developed FORM demonstrator as a stand-alone code in ROOT, based on TTree
- This demonstrator shows that data can be stored separately by APA, DUNE's far detector processing unit, into a dedicated TTree
- Processing data on APA level can be done by reading and writing data in smaller grains, resulting in significant memory reduction
- Memory for writing and reading was reduced sufficiently enough to allow DUNE to stay within memory limits
- Using persistence references allows reading complete events as well