DUNE Computing Model in a nutshell

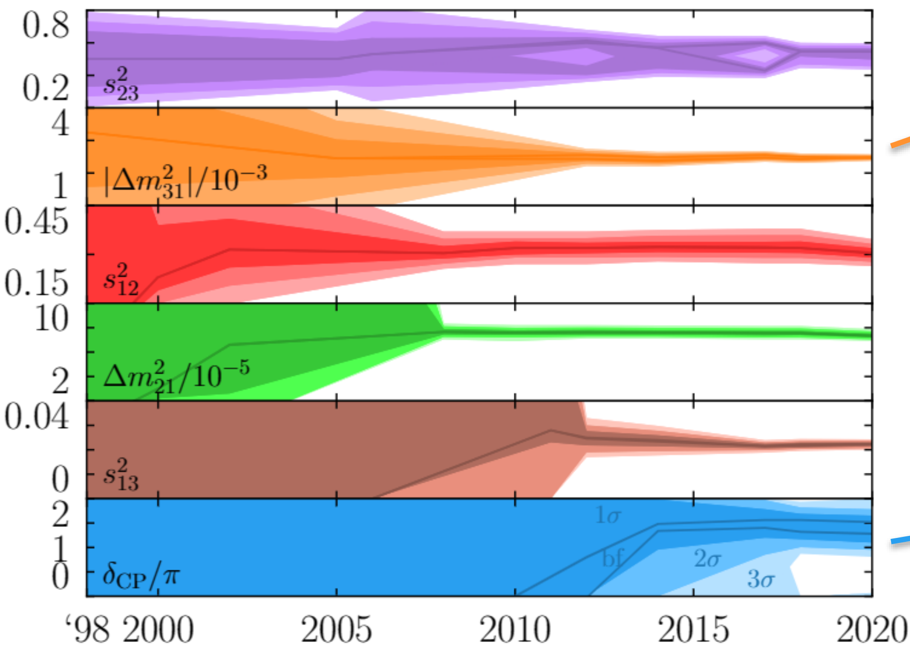
M. Tenti – INFN Bologna

15/09/2023

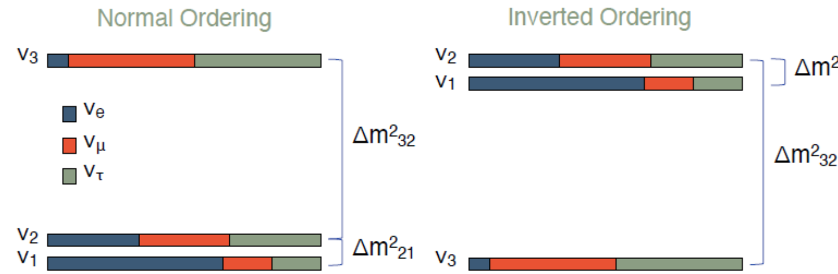
Istituto Nazionale di Fisica Nucleare

Motivations

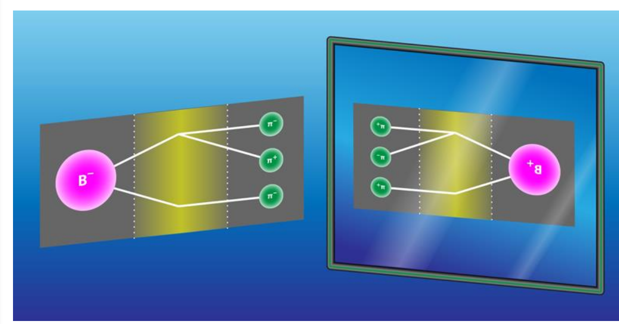
Impressive progress on the knowledge of neutrino oscillations since 1998. Nonetheless, there are **still open questions**.



Which Mass Ordering?

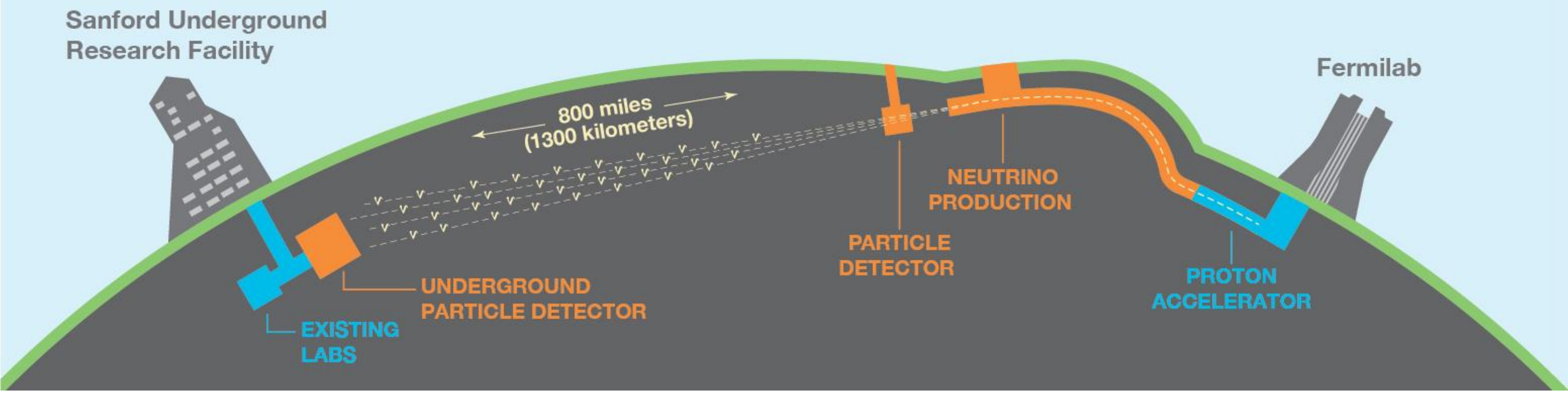


Is there CP symmetry violation in leptonic sector?



INFN Istituto Nazionale di Fisica Nucleare
DUNE

Deep Underground Neutrino Experiment



A new generation **Long Baseline** (1300 km) neutrino oscillation experiment based on:

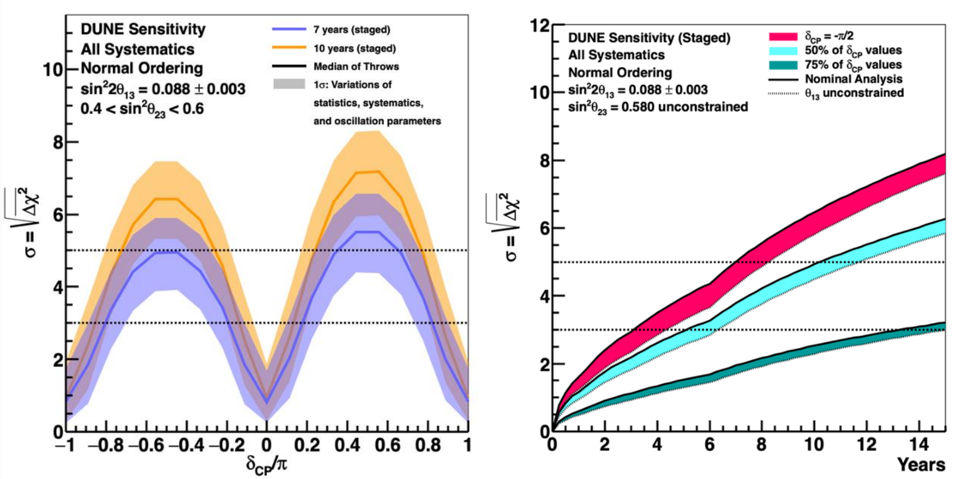
1. a wide band high intensity (1.2 MW upgradable to 2.4 MW) $\nu/\bar{\nu}$ neutrino **beam** produced at Fermilab

2. a large mass (~70 kton) **Far Detector** at the Sanford Underground Neutrino Facility (SURF) 1.5 km underground exploiting the Liquid Argon Time Projection Chamber (LArTPC) technology

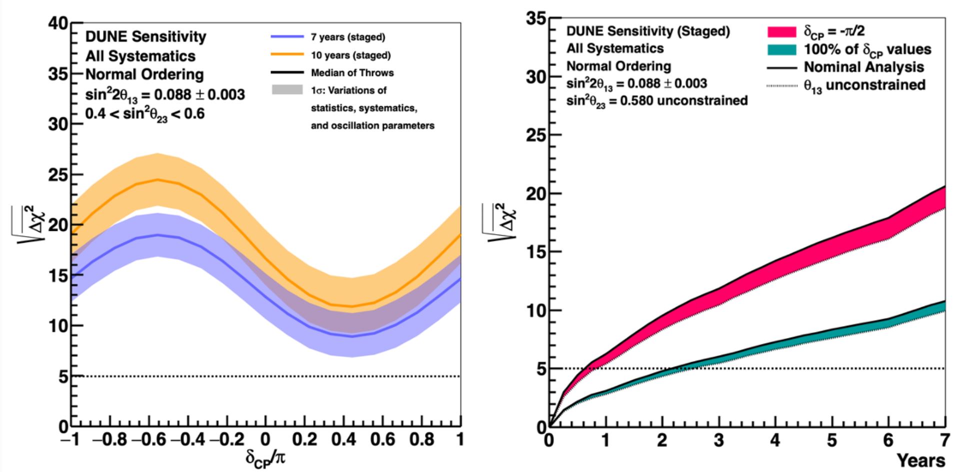
3. a **Near Detector** complex (ND) at Fermilab providing control of systematic uncertainties

Goals and Sensitivities

CP Violation Sensitivity



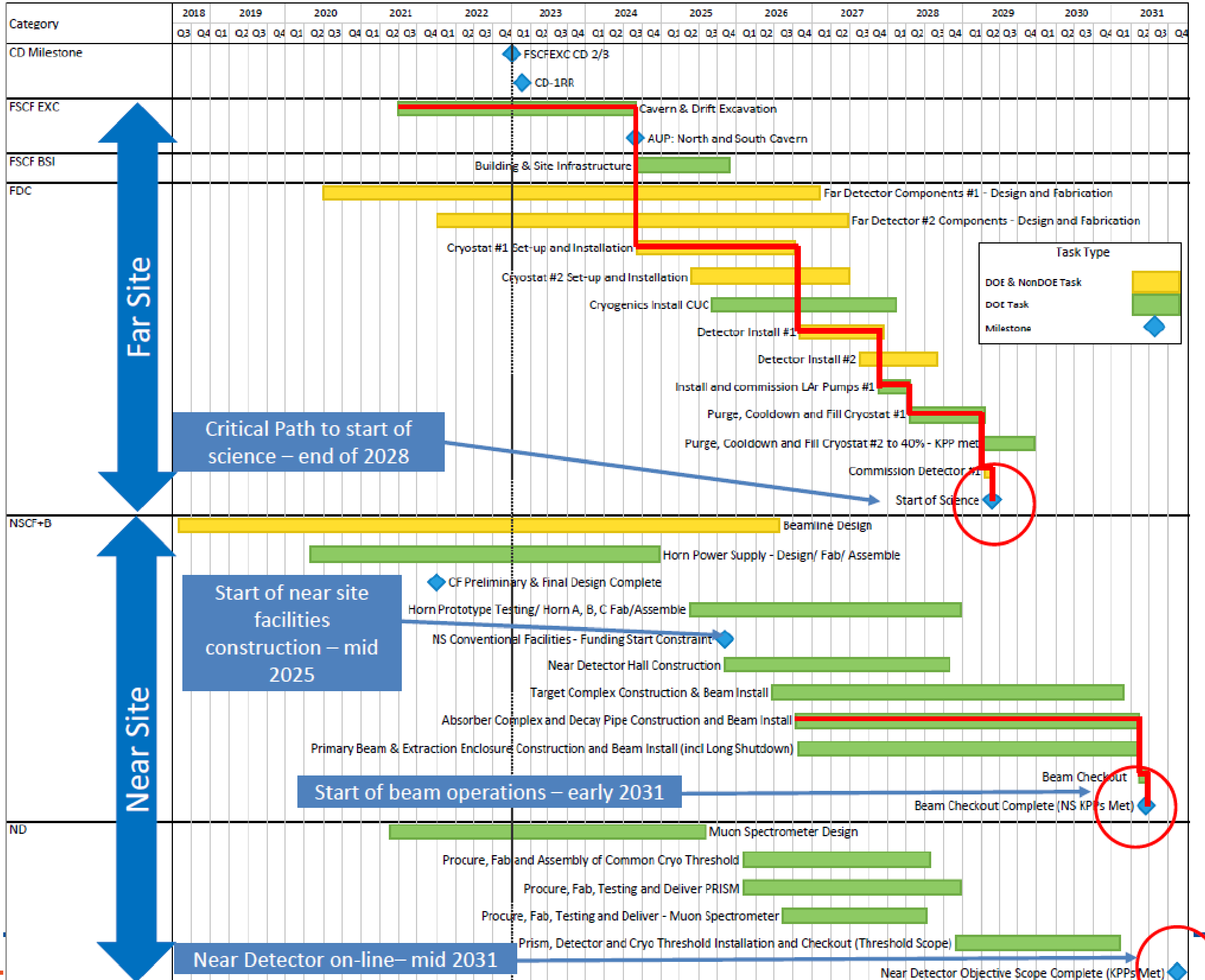
Mass Ordering Sensitivity



- Determination of all neutrino oscillations parameters in a single experiment

- Large mass, low energy threshold and low background environment:
  - Supernova neutrinos
  - Solar neutrinos
  - Proton decay search
  - BSM physics

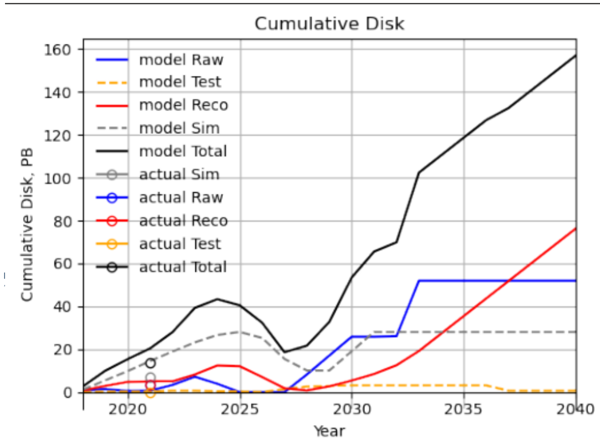
LBNF Timeline

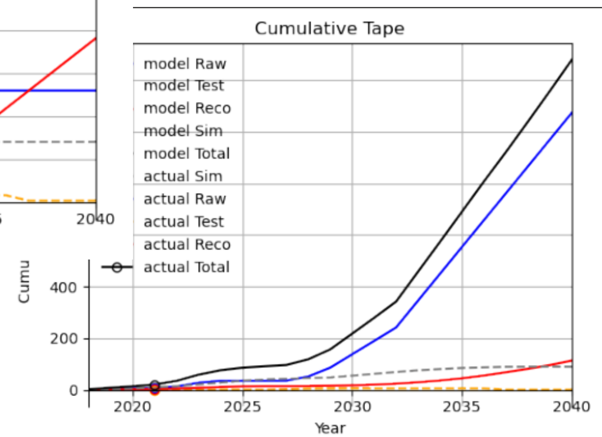


Data Production

Process	Rate/module	size/instance	size/module/year
Beam event	41/day	3.8 GB	30 TB/year
Cosmic rays	4,500/day	3.8 GB	6.2 PB/year
Supernova trigger	1/month	140 TB	1.7 PB/year
Solar neutrinos	10,000/year	≤3.8 GB	35 TB/year
Calibrations	2/year	750 TB	1.5 PB/year
Total			9.4 PB/year

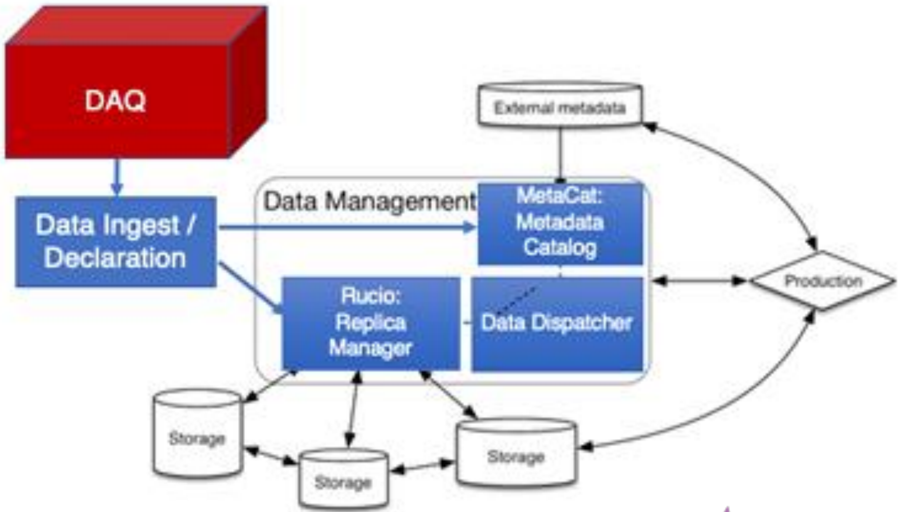
- FD: Set a cap at 30 PB/year: allows 2 modules + commissioning/test
- Data taking rate is ~ 10Gb/s with 100Gb/s bursts for SNB and calibration
- ND: O(300) TB/year overall



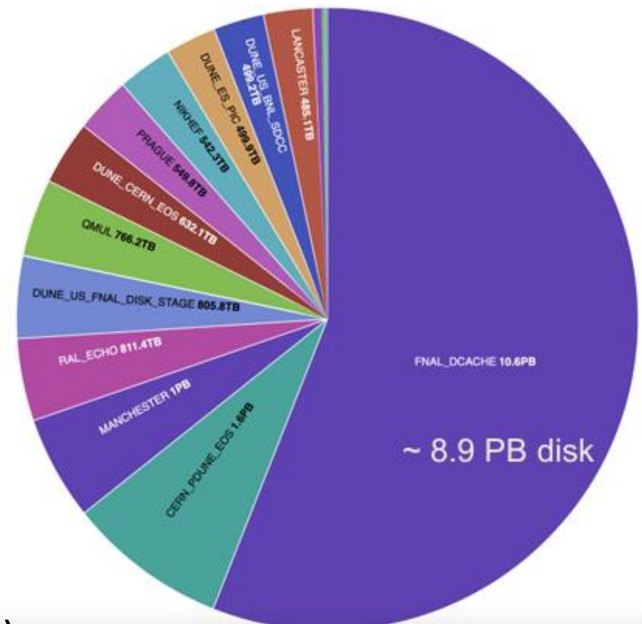


Tier	Description	Tape copies	Lifetime	Disk Copies	Disk Lifetime
Raw	Physics data	2	indefinitely	1	1 year
Test	test and commissioning	1	6 months	1	6 months
Waveforms	processed waveforms	1	10 years	1	1 month
Reco	pattern recognition	1	10 years	2	2 years

Data Management System



Rucio managed
FNAL ~ 8.9/11 PB
Other ~ 7.6/12.5 PB

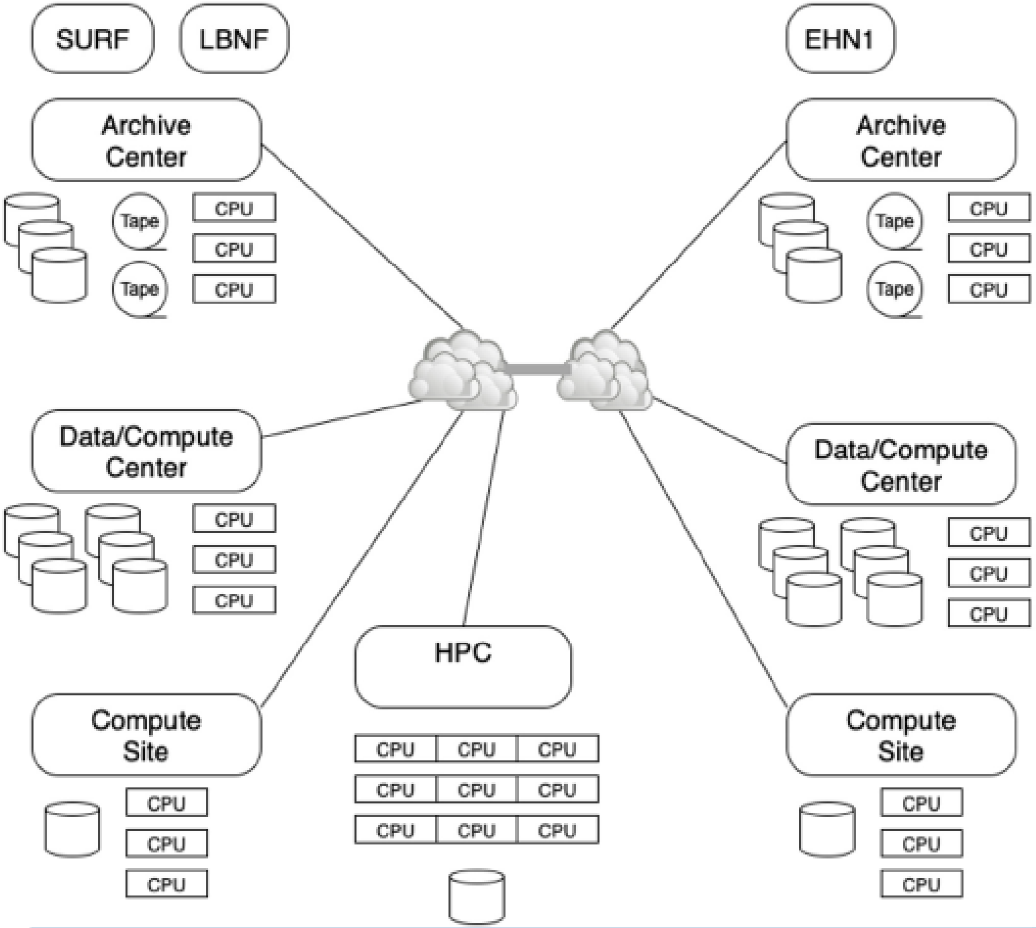


- **RUCIO**: Replica Manager
- **MetaCat**: Metadata Catalog
- **Data Dispatcher**:
  - Server keeps track of files in a project and when they are available for use
  - Client calls server to get next file to process
  - Server returns URL of next file (streamed or copied)
  - Client notifies server when file is either complete or has failed

DUNE Grid Philosophy

- Beyond being connected to the wide area network, a site should provide at least one service among:
  - DUNE Computing Element (either HTC or HPC)
  - Data Cache (i.e. local temporary disk)
  - DUNE Storage Element (i.e. disk)
  - DUNE Data Archive (i.e. tape)
  - Interactive Analysis and Build Facility (i.e. user interfaces)
  - DUNE Analysis Facility (e.g. Jupyter notebooks)
- The DUNE model builds on the emergence of faster networks to move to a service-oriented model, where sites provide services – disk, CPU, real memory/core and archival tape – and projects are distributed to them based on their capabilities and available networking

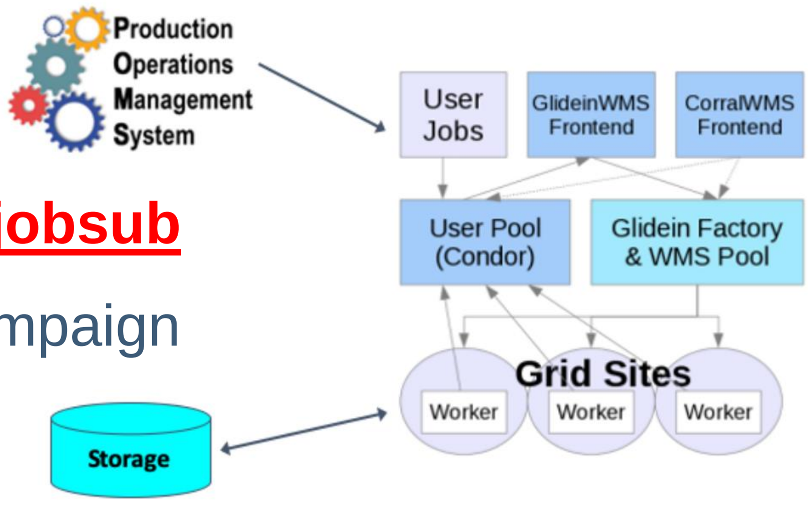
DUNE's global computing model

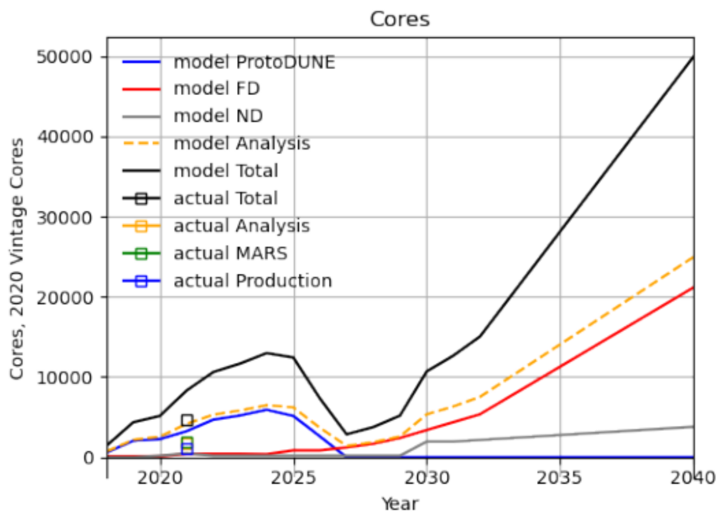


- Less hierarchical than original WLCG model
- Able to **use opportunistic resources**
- Sites offer services
  - Archival **(4)**
  - Disk (Rucio Storage Element **~10**)
  - CPU (**~40**)
  - HPC (**growing**)
- HTC ~50% US, 50% non-US
- HPC mainly US

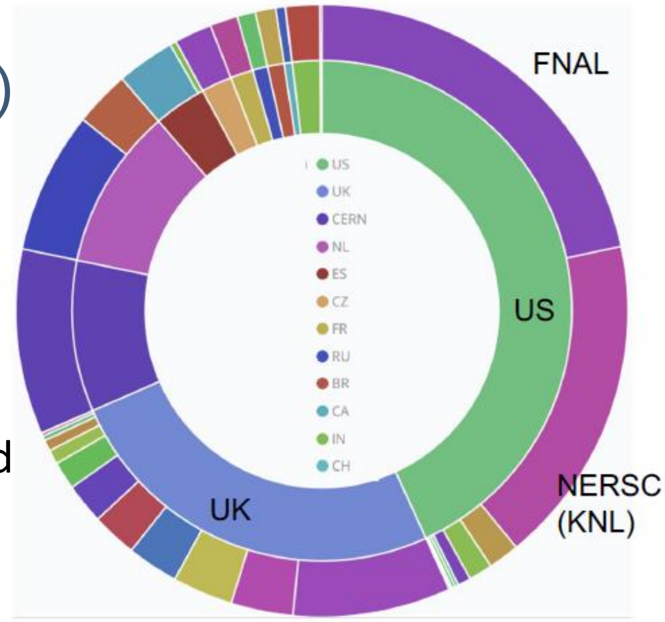
Computing

- Grid job submission with **FIFE-jobsub**
- **POMS** manages production campaign and provide a webUI
- Dedicated tools: **JustIN**
- Resource/slot provisioning is with **GlideinWMS** (on top of **HTCondorCE**)

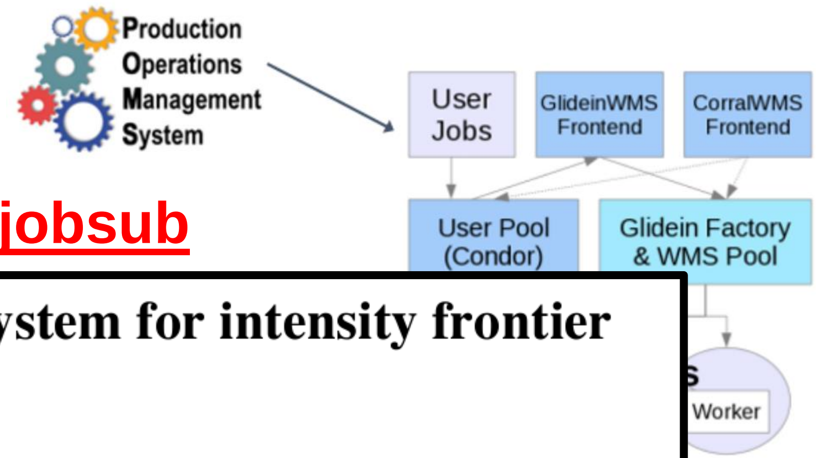




DUNE Production CPU for 2022 used 24M wall hours



Computing



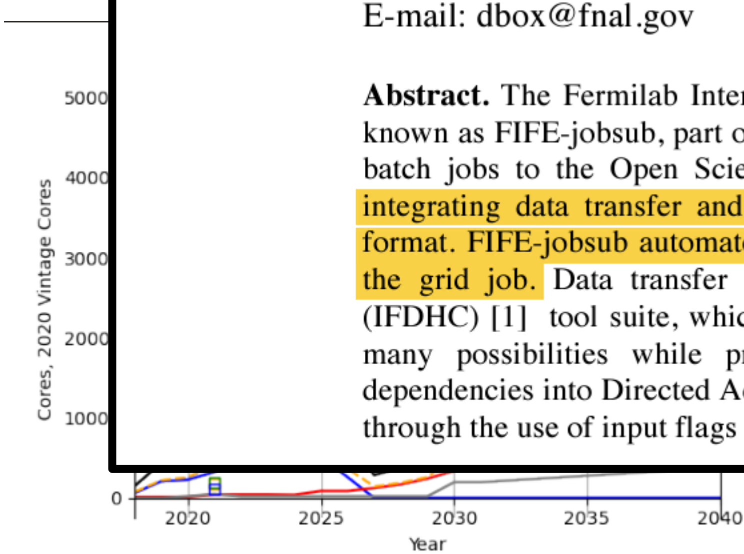
- Grid job submission with **FIFE-jobsub**

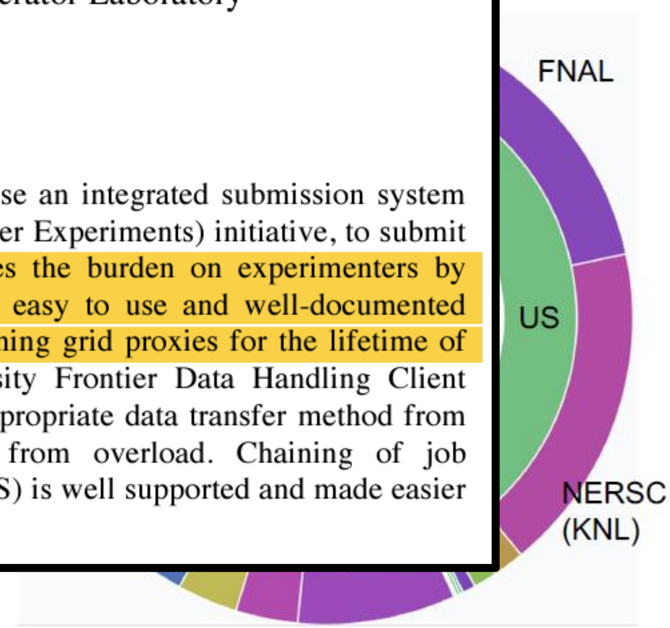
FIFE-Jobsub: a grid submission system for intensity frontier experiments at Fermilab

Dennis Box[1]
Scientific Computing Division, Fermi National Accelerator Laboratory
PO Box 500, Batavia, IL USA 60510

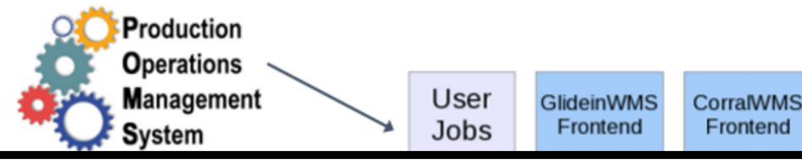
E-mail: dbox@fnal.gov

Abstract. The Fermilab Intensity Frontier Experiments use an integrated submission system known as FIFE-jobsub, part of the FIFE (Fabric for Frontier Experiments) initiative, to submit batch jobs to the Open Science Grid. FIFE-jobsub eases the burden on experimenters by integrating data transfer and site selection details in an easy to use and well-documented format. FIFE-jobsub automates tedious details of maintaining grid proxies for the lifetime of the grid job. Data transfer is handled using the Intensity Frontier Data Handling Client (IFDHC) [1] tool suite, which facilitates selecting the appropriate data transfer method from many possibilities while protecting shared resources from overload. Chaining of job dependencies into Directed Acyclic Graphs (Condor DAGS) is well supported and made easier through the use of input flags and parameters.





Computing



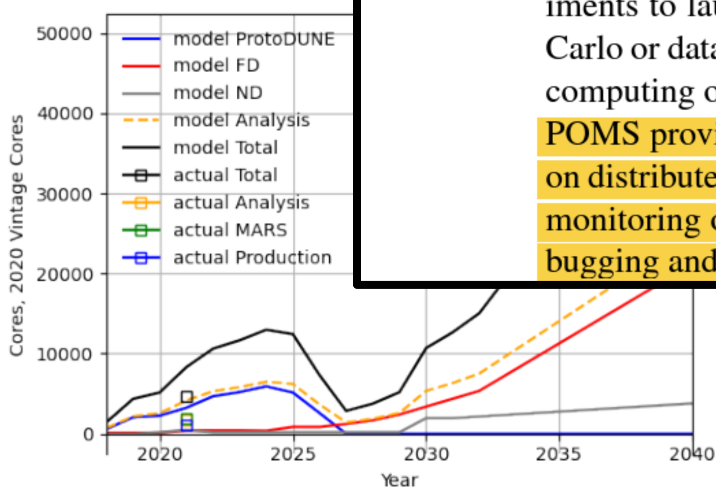
- Grid job s
- **POMS** ma and provi
- Dedicated
- Resource **GlideinW**

Production Operations Management System (POMS) for Fermilab Experiments

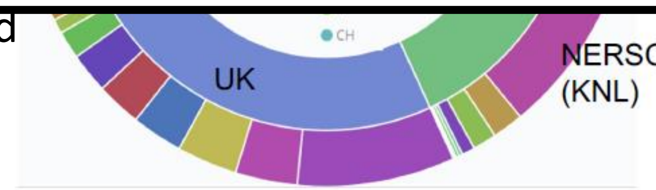
Marc Mengel[1,*], *Stephen* White[1,**], *Vladimir* Podstavkov[1,***], *Margherita* Wiersma[1,****], *Anna* Mazzacane[1,†], and *Kenneth* Herner[1,‡]

[1]Fermi National Accellerator Laboratory
P.O. Box 500
Batavia IL, 60532

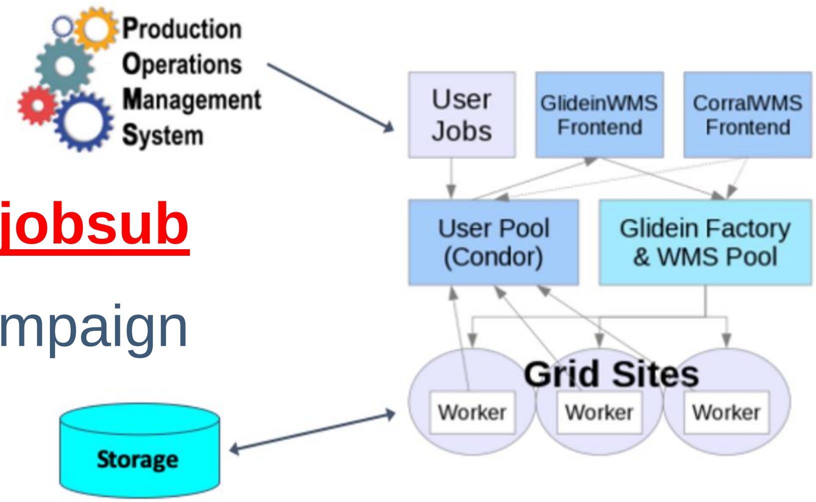
Abstract. The Production Operations Management System (POMS) software allows production teams and analysis groups across multiple Fermilab experiments to launch, modify and monitor large scale campaigns of related Monte Carlo or data processing jobs, and currently manages the majority of production computing of experiments at Fermilab. POMS provides a web service interface that enables automated job submission on distributed resources according to customers' requests, as well as subsequent monitoring of those submissions as well as recovery of failed submissions, debugging and record keeping.



CPU for 2022 used
24M wall hours



Computing



- Grid job submission with **FIFE-jobsub**
- **POMS** manages production campaign and provide a webUI
- Dedicated tools: **JustIN**

Overview of the justIN workflow system

justIN implements the workflow system design described in chapter 13 of the "DUNE Offline Computing Conceptual Design Report", A. Abed Abud et al (the DUNE Collaboration), FERMILAB-DESIGN-2022-01, 28 October 2022.

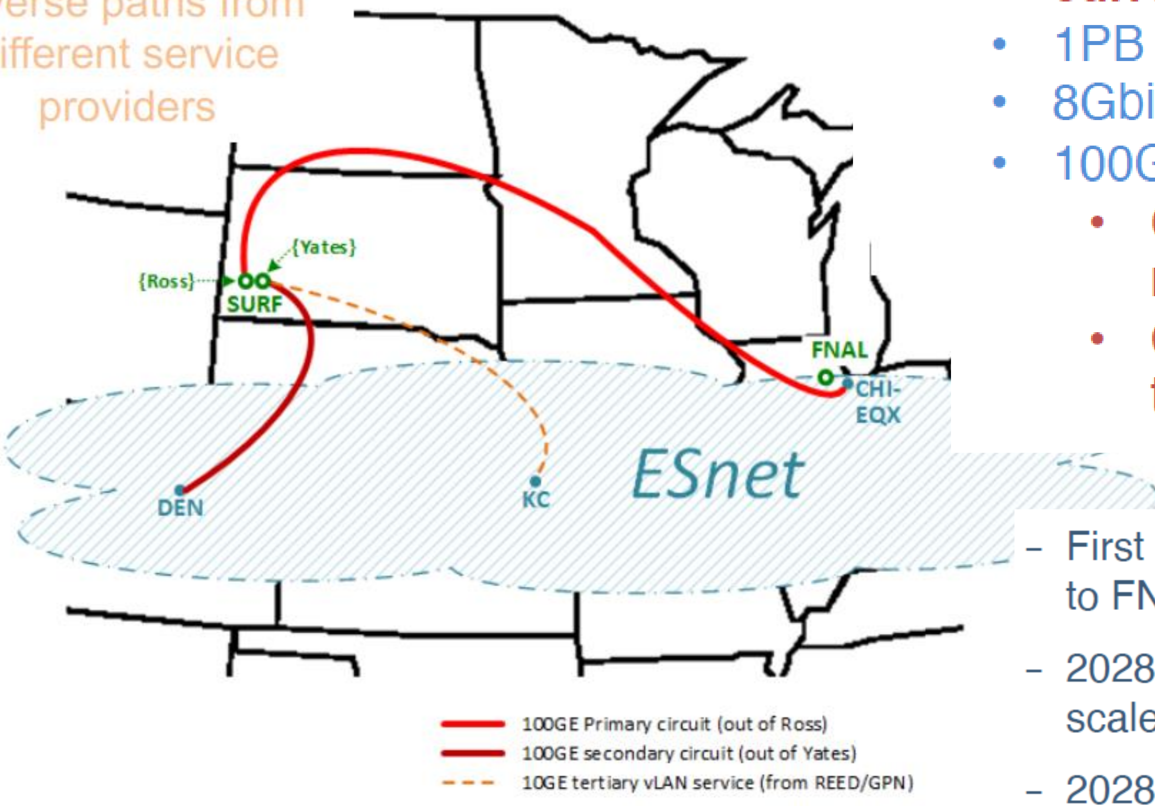
The justIN system includes all aspects of orchestrating the execution of code to generate simulated data and to process real or simulated data at computing sites around the world.

To make the most efficient use of the finite computing, network, and storage resources available to the experiment, the design of justIN was driven by the location and availability of data to be processed and its proximity to computing capacity as it becomes available.

Efficiently matching CPU and data is a long-standing problem in HEP computing. We have developed a relatively un-hierachical system that uses improved knowledge of the computing properties of applications (I/O rate, memory needs, data size) and the network connections between Rucio Storage Elements (RSEs) and CPU resources to optimally match processing and data.

Networking in US

Geographically-diverse paths from different service providers



- **Primary SURF-FNAL link currently live at 10Gbit/s**
- 1PB of disk cache available at FD
- 8Gbit/s steady state
- 100Gbit/s necessary to
  - Get supernova data out of mine
  - Catch up in timely manner if there's outage.

- First 100Gbit/s network path from SURF to FNAL 2027
- 2028 onwards—monthly "supernova-scale" tests of high volume transfers
- 2028 Second geographically distinct 100Gbit/s network path.

Global Networking

- DUNE has regular sequence of meetings with **ESNet**
  - Go over updated requirements particularly for fast transfer of supernova data
- DUNE participates in **LHCOne**-related meetings
  - Using LHCOne where convenient or necessary (some sites are configured to only use it).
  - Not planning our own DUNEOne at this time.
- ESNet has set up **OSCARS** dedicated virtual circuits from CERN<->FNAL in ProtoDUNE era, these have proved very useful in delivering bandwidth necessary.
- Important to test the “last mile” from border routers to the computing facilities, esp. at the HPC sites.

CNAF

- Current resources and 2024 request for nu_at_fnal (DUNE from 2024) INFN group:
  - 510 TB disk + **500 TB**
  - 510 TB tape
  - 3.5 kHS06 + **1.5 kHS06**
  - HPC resource (kindly provided) → **machine w/ GPU and > 300 GB RAM**
- Resource are used:
  - to optimize the design, develop reconstruction tools and evaluate physics potential for the SAND detector through MC simulation of neutrino interactions
  - to optimize (computationally intense) reconstruction algorithm for GRAIN

2023

		CPU (Cores)	CPU (MWC)	Disk (PB) [2]	Tape (PB)
Total DUNE Requrements [1]					
DUNE total			15,159	25.8	45.5
DUNE Requirements at FNAL					
FNAL		3,792	3,792	8.9	36.2
DUNE Requirments at CERN					
CERN		1,500	1,500	4.0	9.2
DUNE Requirements other sites					
Request to "others"			*9,867*	*12.9*	*0.1*
Pledges 2023					
BNL	BNL	1,000	2,000	1.0	
USA - other	(OSG opportunistic)	1,500	1,500		
UK	GridPP	1,000	2,000	4.0	0.5
FR	CC-IN2P3	320	640	0.5	1
ES	PIC Tier-1	512	512	0.7	
NL	NL/LHC Tier-1	514	1,028	1.5	
CZ	CZ-Prague-T2	2,400	3,150	1.2	
CH	Bern	200	200	0.2	
BR	CBPF	200	200		
IN	Tata	450	1,350	0.8	
CA	Canada	360	720	0.3	
Total pledge "others"		8,456	13,300	10.1	1.5
Overall Total pledge		13,748	18,592	23.0	46.9
Overall shortfall			-3,433	2.8	-1.4

Finally, we would like CNAF to contribute to **«official» DUNE computing**

Thank you