DUNE Computing Requests for 2025

Computing Contributions Board Meeting

Heidi Schellman and Michael Kirby

Basics of the resource model

- Keep raw data on disk for 2-3 year, on tape to end of expt.
  - 2 tape copies, one disk copy (For 2025 Increased lifetime for raw data from 1 to 3 years, reduced copies to 1 to allow both NP02 and NP04 on disk through 2026)
  - protoDUNE tape – 1 copy each at CERN/FNAL
  - DUNE tape - 1 copy at FNAL, 1 copy at other institutions
- Reconstruct full sample every year (protoDUNE for 3 years, DUNE to end of expt.)
- Do new simulation campaigns each year
- Keep simulation and reconstructed data on disk for 2 years (always have 2 versions)
  - One copy in Americas, one in Europe where possible (model assumes 1.5 copies)
  - No need to stage from tape until it ages out
- One copy of reconstruction/simulation -> tape as it can be redone if necessary.

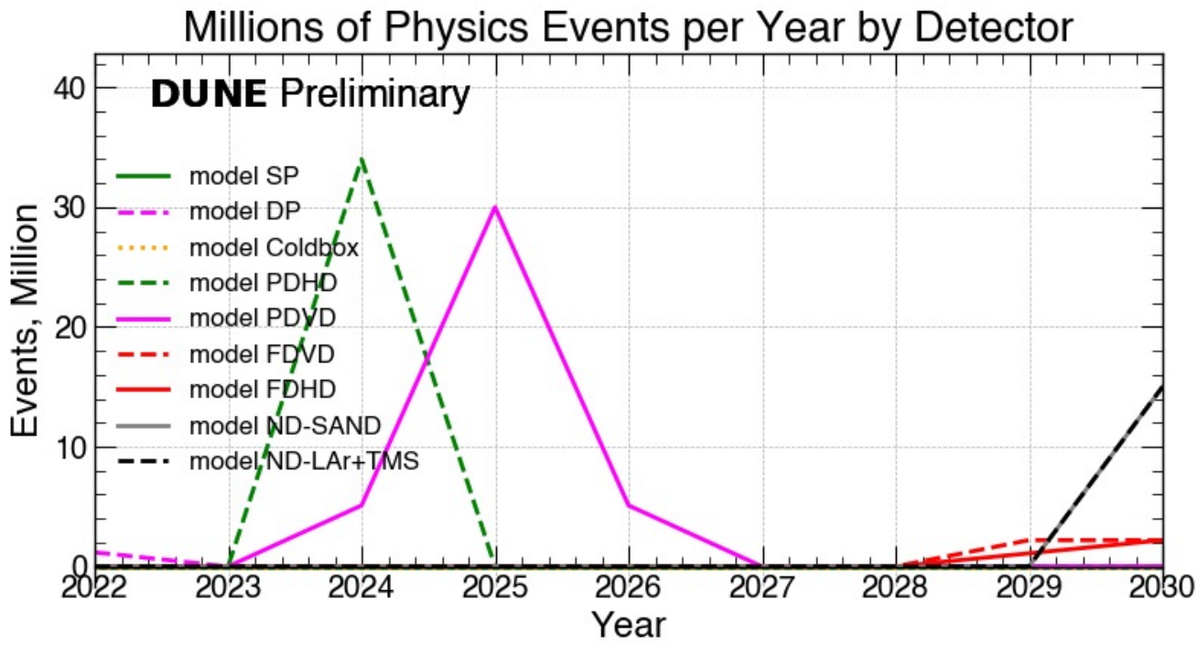
- **CPU estimates** are based on measurements from ProtoDUNE data and existing simulations and **for the FD/ND have large uncertainies**.

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Updates to the model

- Summer 2024: Complete rewrite of the underlying code for clarity and flexibility.
- Site splits now based on detector (PD, FD, ND), not year.
- Delayed ProtoDUNE II operations until Spring 2024, but more data than expected
- With the successful large-scale FD simulation campaigns, we have considerably better understanding of both our processing time expectations and our simulation needs
- not all of that additional information has been included in the model just yet (minor tweaks)
- Fixed a bug in the retention and reprocessing of events and lifetime of output on disk
- new estimates for FD processing time based upon multithreaded processing and subsequent smaller memory footprint
- updated files sizes for reconstruction and simulation output – no longer estimate based upon retention of the raw waveforms in data or rawdigits in the simulation
- still working on understanding the GPU requirements for 2x2 and ND-LAr and how those estimates can be translated
- We are transitioning from MWC to HEPScore23 (HS23) as the metric for CPU resources

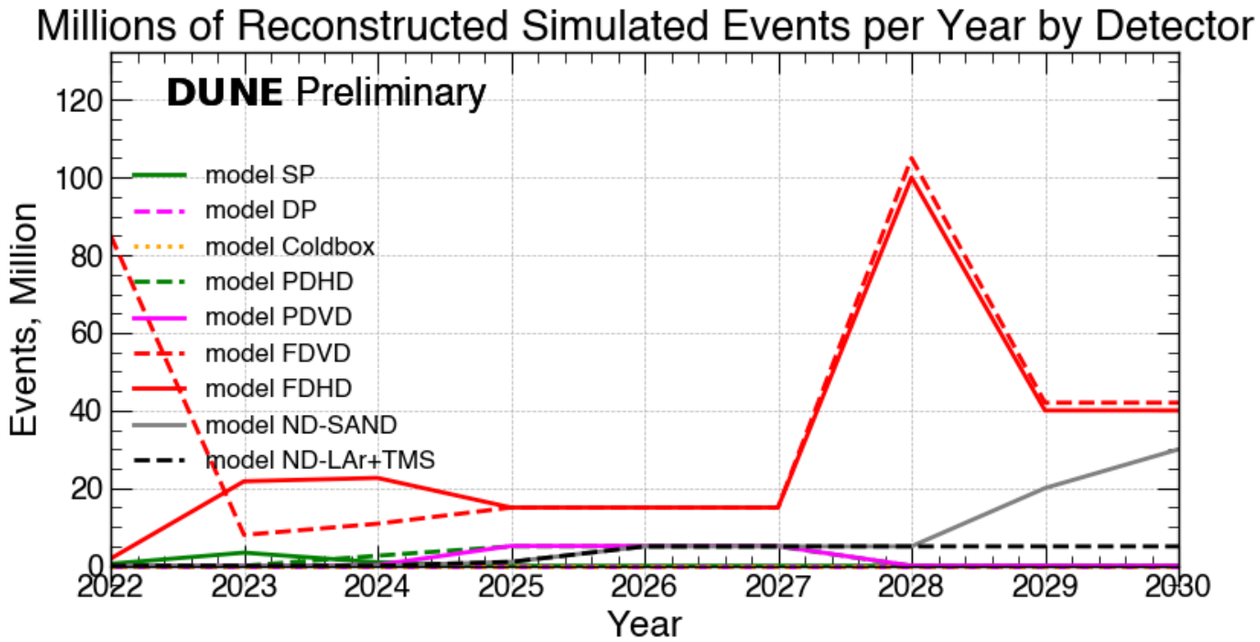
Spill/trigger records per year input into the model



Millions-of-Physics-Events-per-Year-by-Detector-Events

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DP	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	0.0	0.0	34.0	0.0	0.0	0.0	0.0	0.0	0.0
PDVD	0.0	0.0	5.1	30.0	5.1	0.0	0.0	0.0	0.0
FDVD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	2.2
FDHD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	2.2
ND-SAND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0
ND-LAr+TMS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0

Simulated Spills/trigger records per year



Millions-of-Reconstructed-Simulated-Events-per-Year-by-

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	0.5	3.4	0.9	0.0	0.0	0.0	0.0	0.0	0.0
DP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	0.0	0.0	2.6	5.0	5.0	5.0	0.0	0.0	0.0
PDVD	0.0	0.0	0.2	5.1	5.1	5.1	0.0	0.0	0.0
FDVD	85.1	8.0	10.8	15.0	15.0	15.0	105.0	42.0	42.0
FDHD	1.8	21.8	22.6	15.0	15.0	15.0	100.0	40.0	40.0
ND-SAND	0.0	0.0	0.0	1.0	5.0	5.0	5.0	20.0	30.0
ND-LAr+TMS	0.0	0.0	0.0	1.0	5.0	5.0	5.0	5.0	5.0

Lifetimes and copies

Parameters	DiskCopies	DiskLifetimes	TapeCopies	TapeLifetimes
Raw-Data	1	3	2	100
Test	1	0.5	1	1.0
TP	1	0.5	1	100
Reco-Data	2	2	1	15
Reco-Sim	1.5	2	1	15
Analysis-Data	2	5	1	15
Analysis-Sim	2	5	1	15

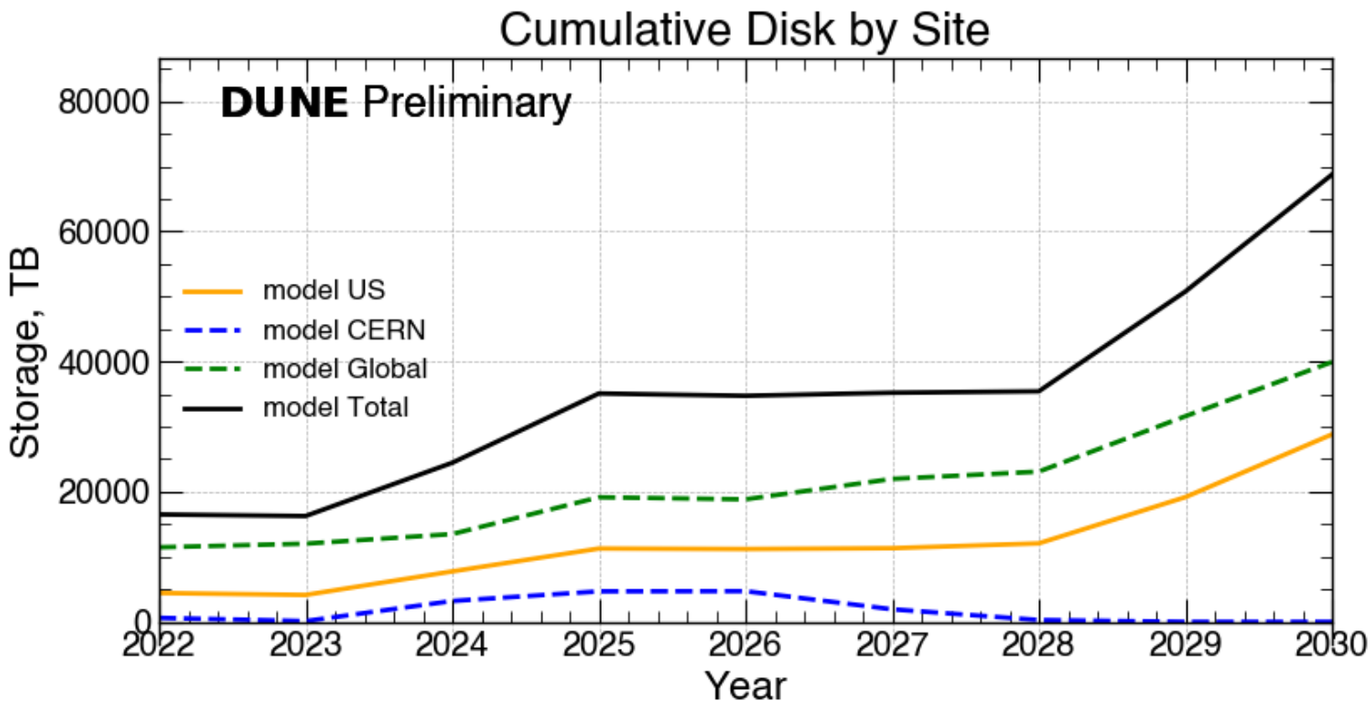
Table 3: Lifetimes and number of copies for different kinds of data. An exception, we assume protoDUNE raw radata will stay on disk for up to three years for reprocessing. Far detector data are assumed to stay for 2 years.

Division between National Labs and Countries

Detector Class	Data Type	US	CERN	Global
PD	Raw-Data	0.5	0.5	0.0
PD	Reco-Sim	0.25	0.0	0.75
PD	Reco-Data	0.25	0.0	0.75
PD	Test	0.5	0.5	0.0
PD	TP	0.5	0.5	0.0
FD	Raw-Data	0.5	0.0	0.5
FD	Reco-Sim	0.25	0.0	0.75
FD	Reco-Data	0.25	0.0	0.75
FD	Test	0.5	0.0	0.5
FD	TP	0.5	0.0	0.5
ND	Raw-Data	0.5	0.0	0.5
ND	Reco-Sim	0.25	0.0	0.75
ND	Reco-Data	0.25	0.0	0.75
ND	Test	0.5	0.0	0.5
ND	TP	0.5	0.0	0.5

Table 4: Assumptions about splits of Disk resources between the US, CERN and Global.

Projected Disk Needs

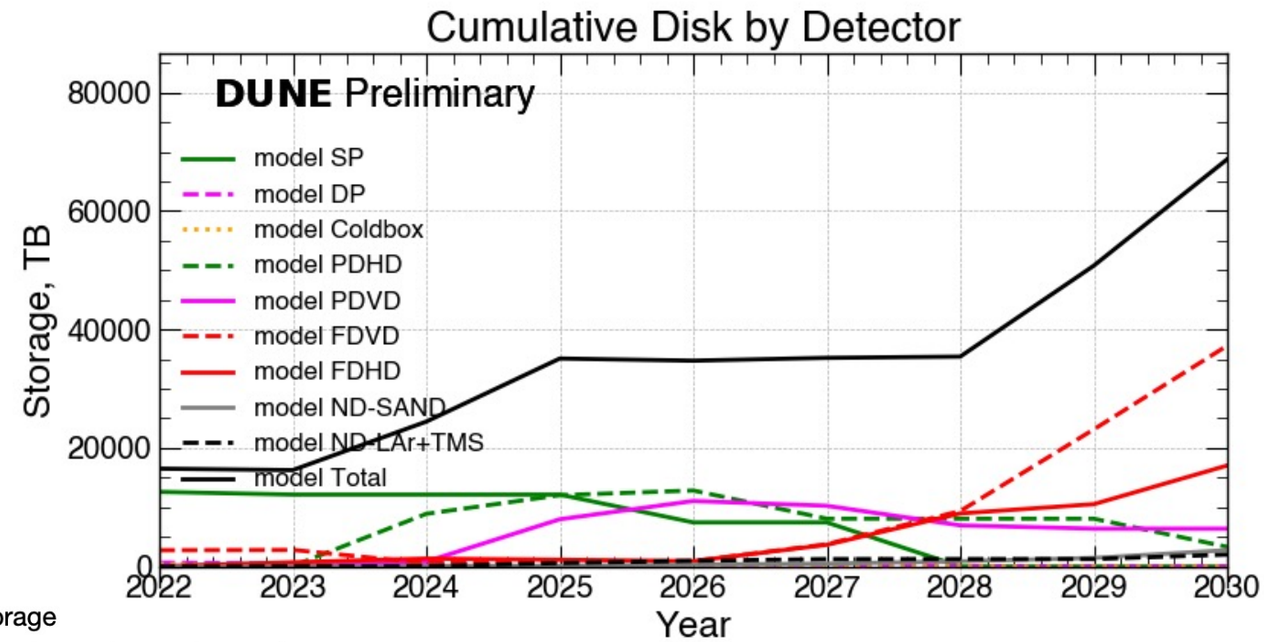


Cumulative-Disk-by-Site-Storage

Locations	2022	2023	2024	2025	2026	2027	2028	2029	2030
US	4432	4134	7752	11266	11191	11318	12058	19165	28806
CERN	612	112	3205	4685	4716	1930	280	0	0
Global	11459	12015	13489	19143	18827	21965	23084	31599	39932
Total	16504	16263	24447	35095	34735	35214	35424	50764	68738

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Projected Disk Needs by Detector

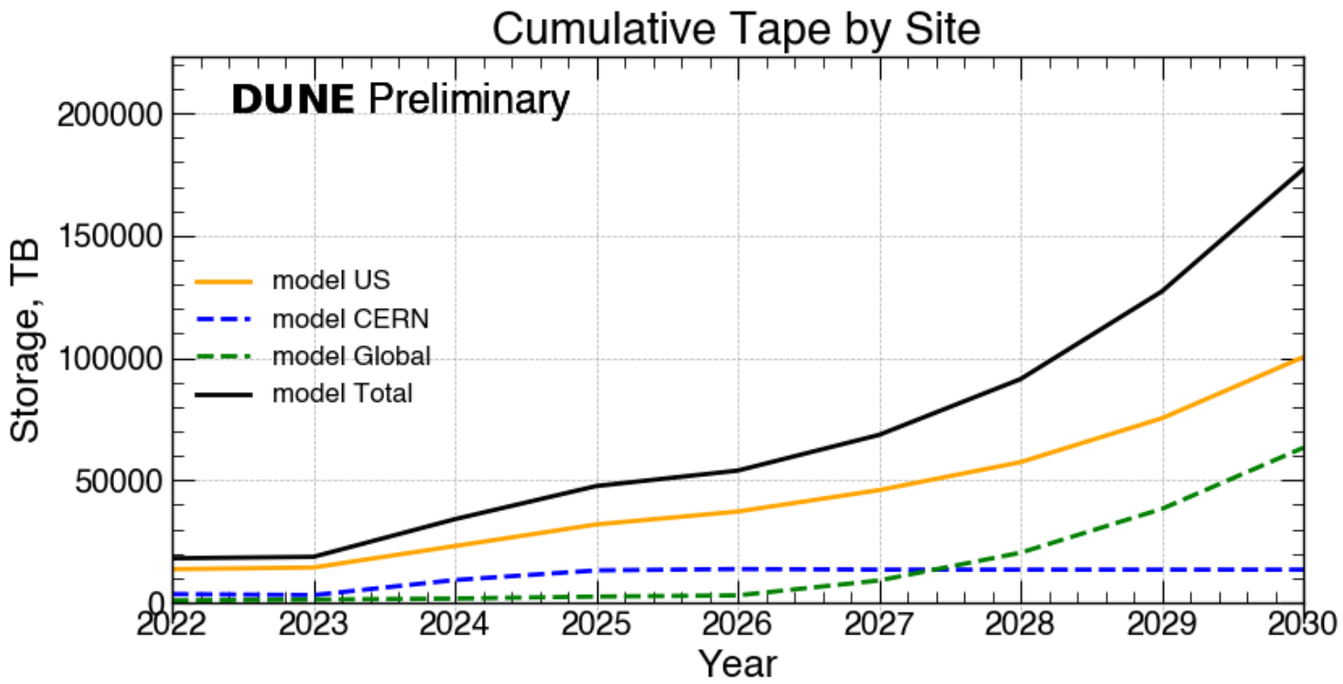


Cumulative-Disk-by-Detector-Storage

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	12587	12132	12132	12132	7435	7435	0	0	0
DP	652	485	485	355	355	355	0	0	0
Coldbox	468	20	75	0	0	0	0	0	0
PDHD	4	71	8882	12028	12820	8060	8060	8060	3300
PDVD	0	2	826	7963	11050	10239	6939	6378	6378
FDVD	2736	2792	563	775	900	3750	9350	23131	37269
FDHD	54	707	1331	1128	900	3650	8950	10508	17041
ND-SAND	0	0	100	137	325	475	875	1437	2750
ND-LAr+TMS	0	50	50	575	950	1250	1250	1250	2000
Total	16504	16263	24447	35095	34735	35214	35424	50764	68738

DEEP UNDERGROUND NEUTRINO EXPERIMENT

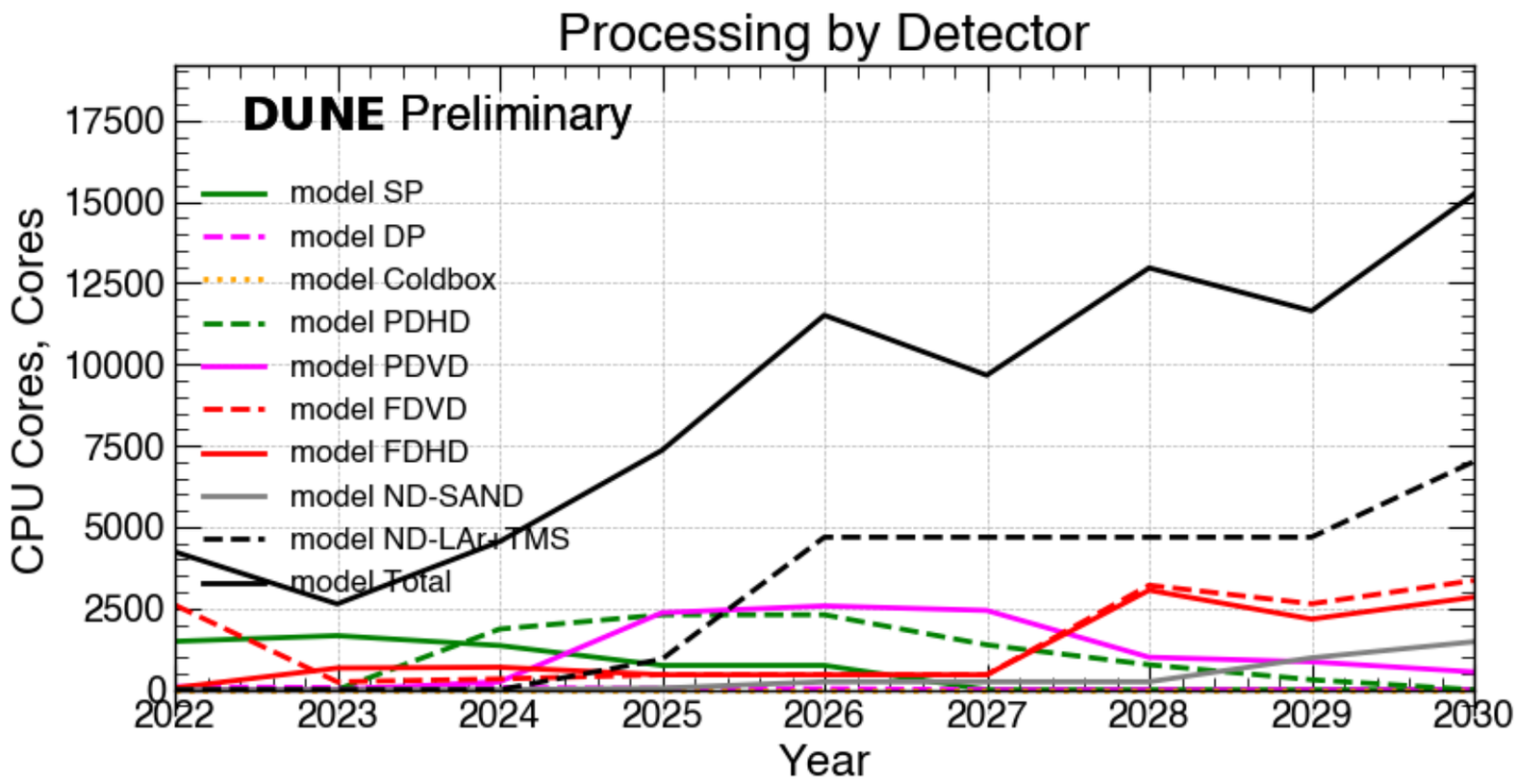
Projected Tape by Site



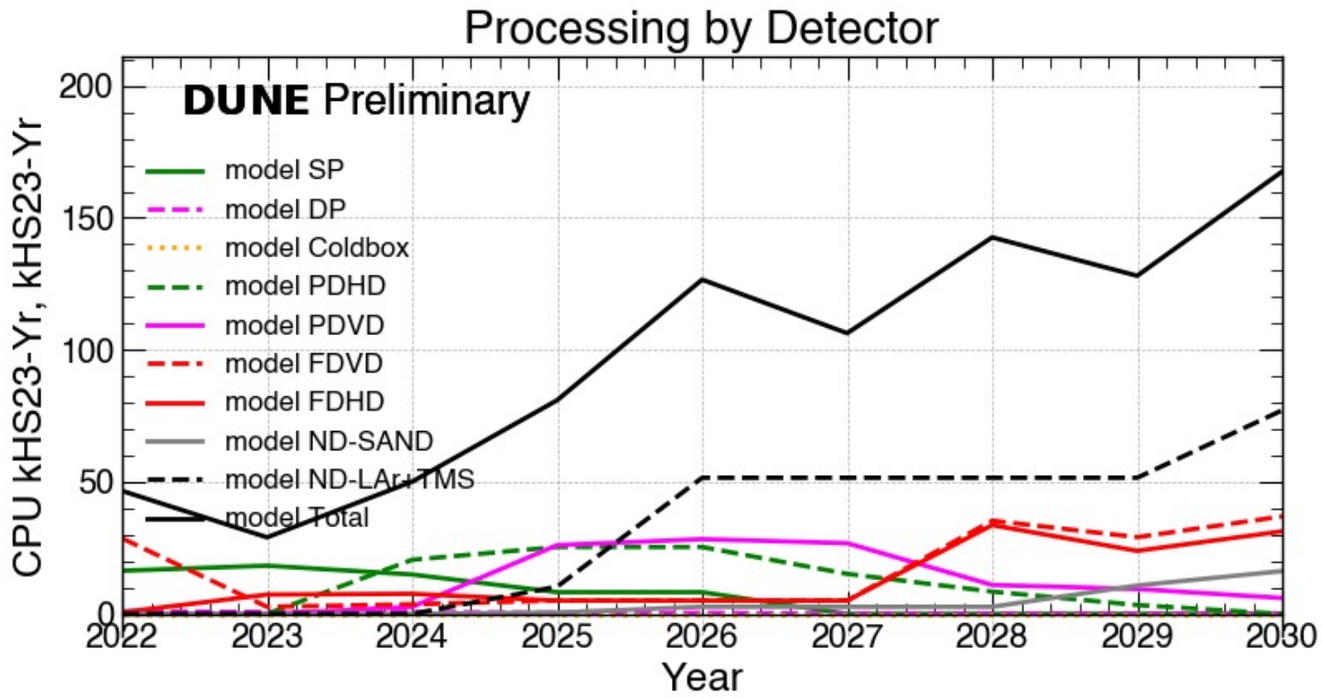
Cumulative-Tape-by-Site-Storage

Locations	2022	2023	2024	2025	2026	2027	2028	2029	2030
US	13692	14381	23143	31974	37239	46000	57440	75365	100354
CERN	3522	3080	9266	13195	13756	13506	13506	13506	13506
Global	930	1277	1712	2499	2987	9074	20412	38337	63326
Total	18145	18739	34122	47670	53983	68582	91359	127210	177188

Projected CPU needs - Cores



Processing CPU Needs – kHS23



Processing-by-Detector-CPU-kHS23-Yr

Detectors	2022	2023	2024	2025	2026	2027	2028	20	
SP	16.3	18.3	14.9	8.2	8.2	0.0	0.0	(	
DP	1.0	0.7	0.7	0.4	0.4	0.0	0.0	(	
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	0.0	0.0	20.5	25.3	25.3	15.2	8.4	3.4	0.0
PDVD	0.0	0.0	2.7	26.0	28.3	26.8	11.0	9.4	6.0
FDVD	28.6	2.7	3.6	5.0	5.0	5.0	35.3	29.1	36.9
FDHD	0.6	7.3	7.6	5.0	5.0	5.0	33.6	23.9	31.3
ND-SAND	0.0	0.0	0.0	0.5	2.7	2.7	2.7	10.8	16.2
ND-LAr+TMS	0.0	0.0	0.0	10.3	51.6	51.6	51.6	51.6	77.1
Total	46.6	29.0	50.1	80.9	126.6	106.3	142.6	128.1	167.6

Processing CPU Needs – split based upon author fraction

Country	DUNE members	# Fraction	Filtered	CPU FY25 HS23	# CPU FY25 Cores	# CPU FY26 H	# CPU FY26 Core
USA	669	46.9	46.9	37927	3448	59352	5396
CERN	36	2.5	2.5	2041	186	3194	290
Brazil	42	2.9	2.9	2381	216	3726	339
Colombia	17	1.2	1.2	964	88	1508	137
United King	173	12.1	12.1	9808	892	15348	1395
Italy	164	11.5	11.5	9298	845	14550	1323
Spain	52	3.6	3.6	2948	268	4613	419
Madagasca	9	0.6	0.6	510	46	798	73
Portugal	11	0.8	0.8	624	57	976	89
Mexico	10	0.7	0.7	567	52	887	81
India	48	3.4	3.4	2721	247	4258	387
France	59	4.1	4.1	3345	304	5234	476
Switzerland	9	0.6	0.6	510	46	798	73
Canada	11	0.8	0.8	624	57	976	89
Netherlands	10	0.7	0.7	567	52	887	81
Czech Repu	14	1.0	1.0	794	72	1242	113
total	1427	100.0	93.5	80900	6875	126600	11509

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Long term projections

- The model can extend to 2038-40
- Storage needs will be dominated by the Far Detector
  - FD raw data capped at 30 PB/year but still dominates
- CPU/GPU dominated by the Near Detectors
  - 60 interactions/spill
  - More channels
  - Need to 2x2 data/sim to get better estimates
  - Likely to use HPC instead of grid resources
- Analysis processing is currently estimated as a % of reconstruction/simulation CPU
- Updated to have separation of the disk needs by site
- updated to include GPU projections
- last week received updates to workflows, and confirm the accuracy as those workflows are processed through the year