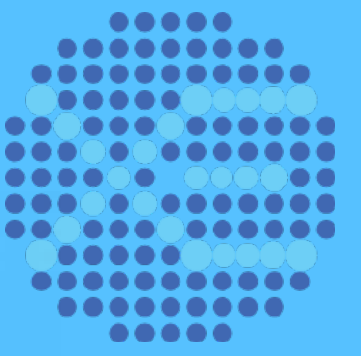# Data processing in XENONnT dark matter project

Dacheng Xu ⟵ a PhD student
Columbia University
HTC25 @ UW–Madison
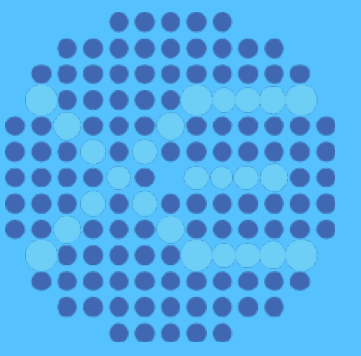**June 4th, 2025**

XENON Collaboration

- 200+ members
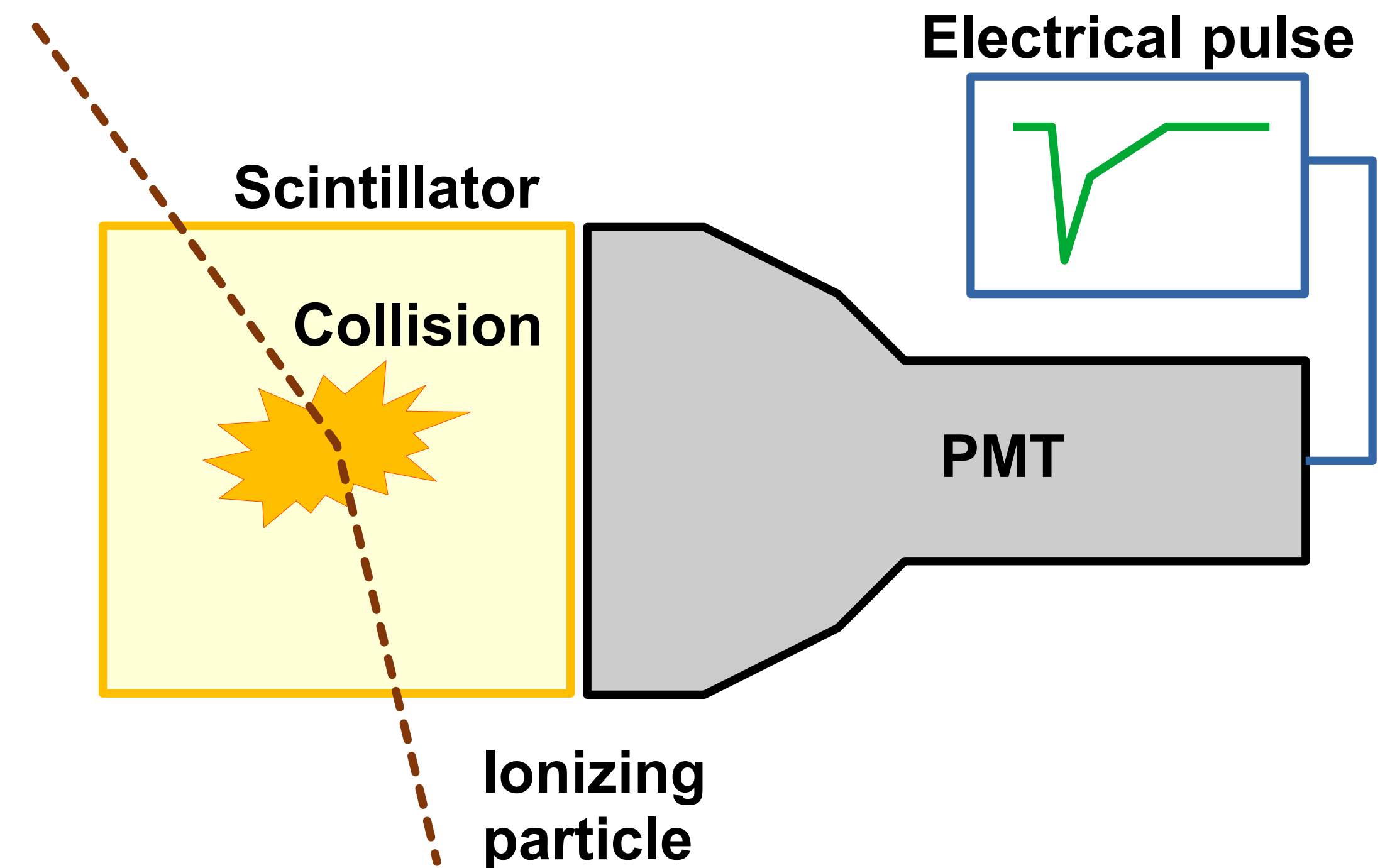
- 30 institutes

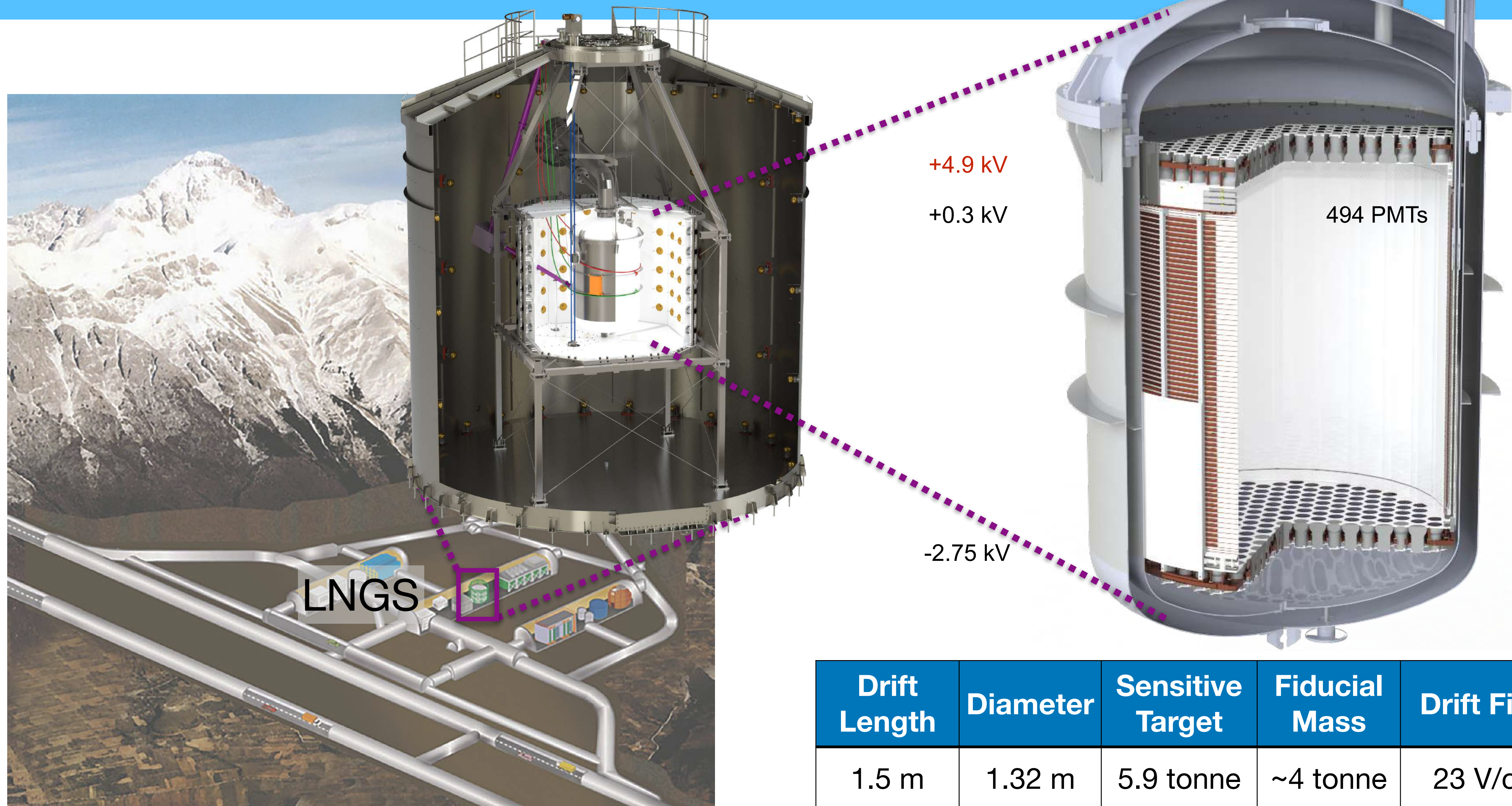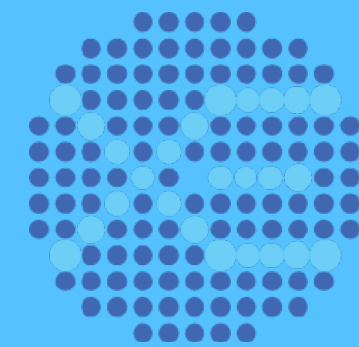- 12 countries

Dark matter direct detection

- We search for a dark matter candidate called weakly interacting massive particle (WIMP).

- The WIMP (ionizing particle) will collide the xenon (Xe) nucleus and deposit energy in liquid Xe (scintillator).

- The energy deposition will be converted into photons that photo-sensors can see. In XENONnT, we use photomultiplier tubes (PMTs) as sensor.

- The output (electrical pulse) of PMT will be stored for further offline analysis.
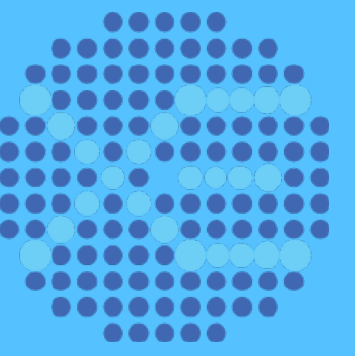
Symmetry 2024, 16(2), 201

Electrical pulse

Scintillator

Collision

PMT

Ionizing particle

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

+4.9 kV

+0.3 kV

494 PMTs

-2.75 kV

LNGS

Drift Length	Diameter	Sensitive Target	Fiducial Mass	Drift Field
1.5 m	1.32 m	5.9 tonne	~4 tonne	23 V/cm

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

dacheng.xu@columbia.edu

Eur. Phys. J. C 84, 784 (2024)   4

Incarnations of XENON detectors

XENON10	XENON100	XENON1T	XENONnT
2005 - 2007	2008 - 2016	2015 - 2018	2020 - now
25 kg - 15 cm	160 kg - 30 cm	3.2 T - 1 m	8.6 T - 1.5 m
89 PMTs	178 PMTs	248 PMTs	494 PMTs
Triggered	Triggered	Triggered	Trigger-less

For more information, please go to website https://xenonexperiment.org/history

dacheng.xu@columbia.edu

COLUMBIA UNIVERSITY
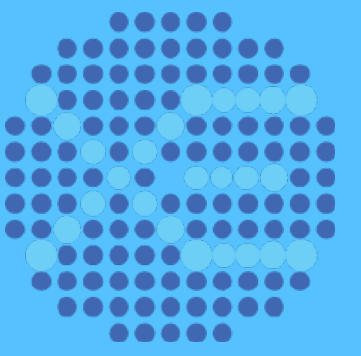IN THE CITY OF NEW YORK
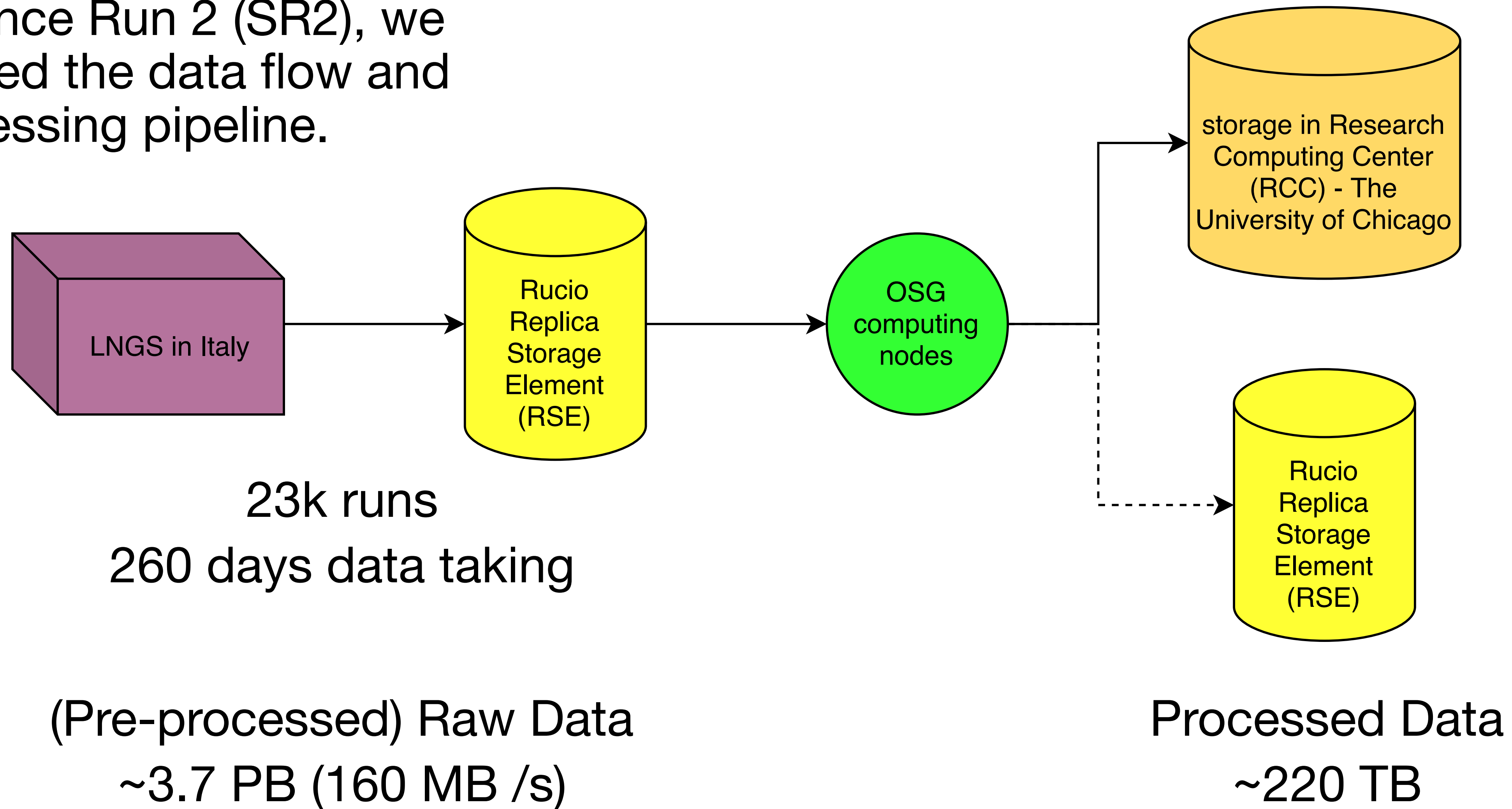
The goals of data reprocessing

- Extract the features of signal from the raw data and build meaningful observables.

- Remove background in higher level data.

- Compress the needed size of data. Eventually we have O(100) events (or even less) in the final dataset after years of data-taking. The reduction is O(PB) -> O(MB). We have to store some intermediate results, so the processed data is O(100)TB.

WIMP (?) events

higher level

Upper

Lower

Output of PMTs

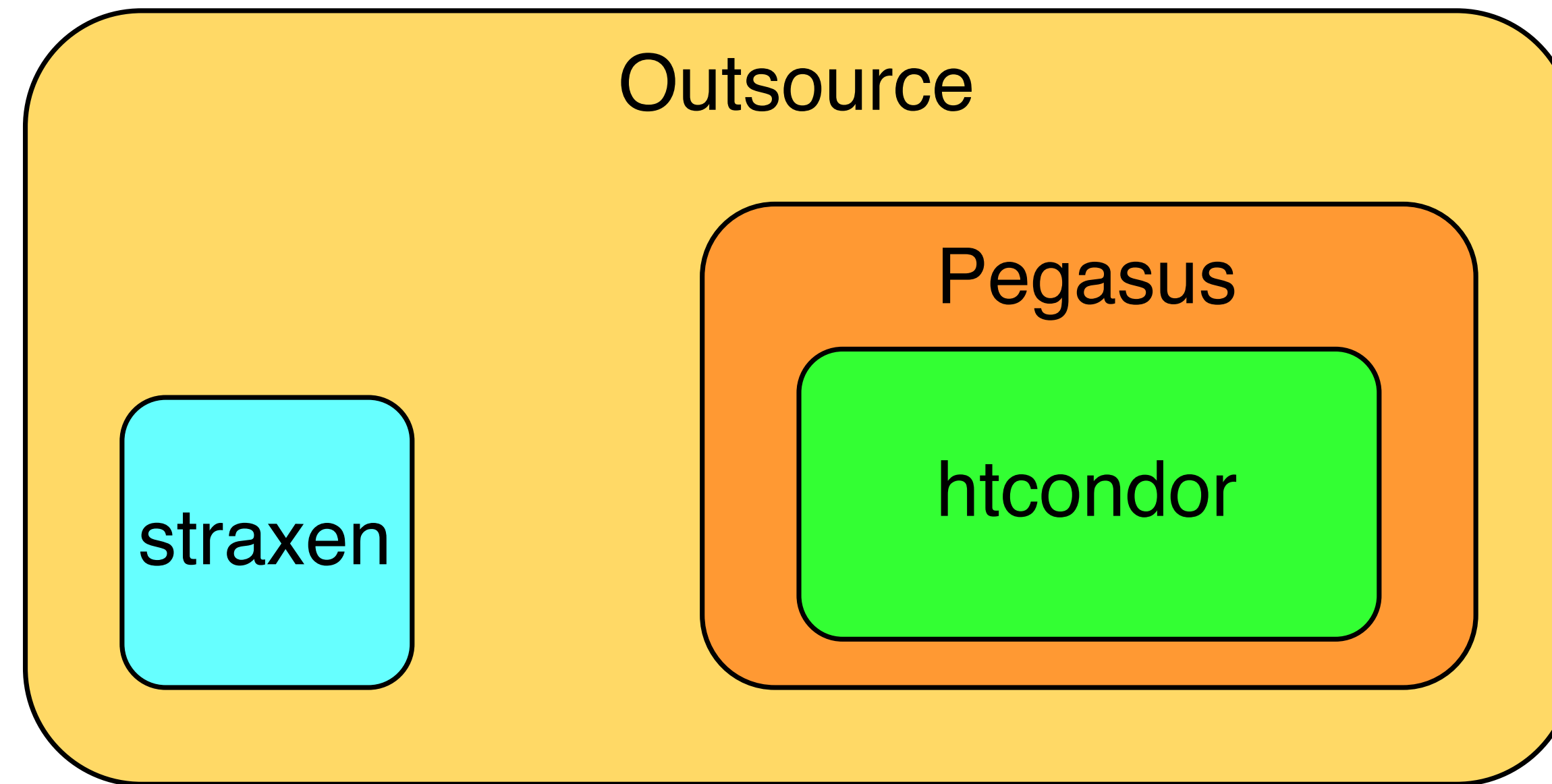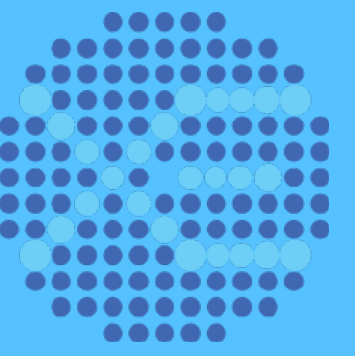COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK
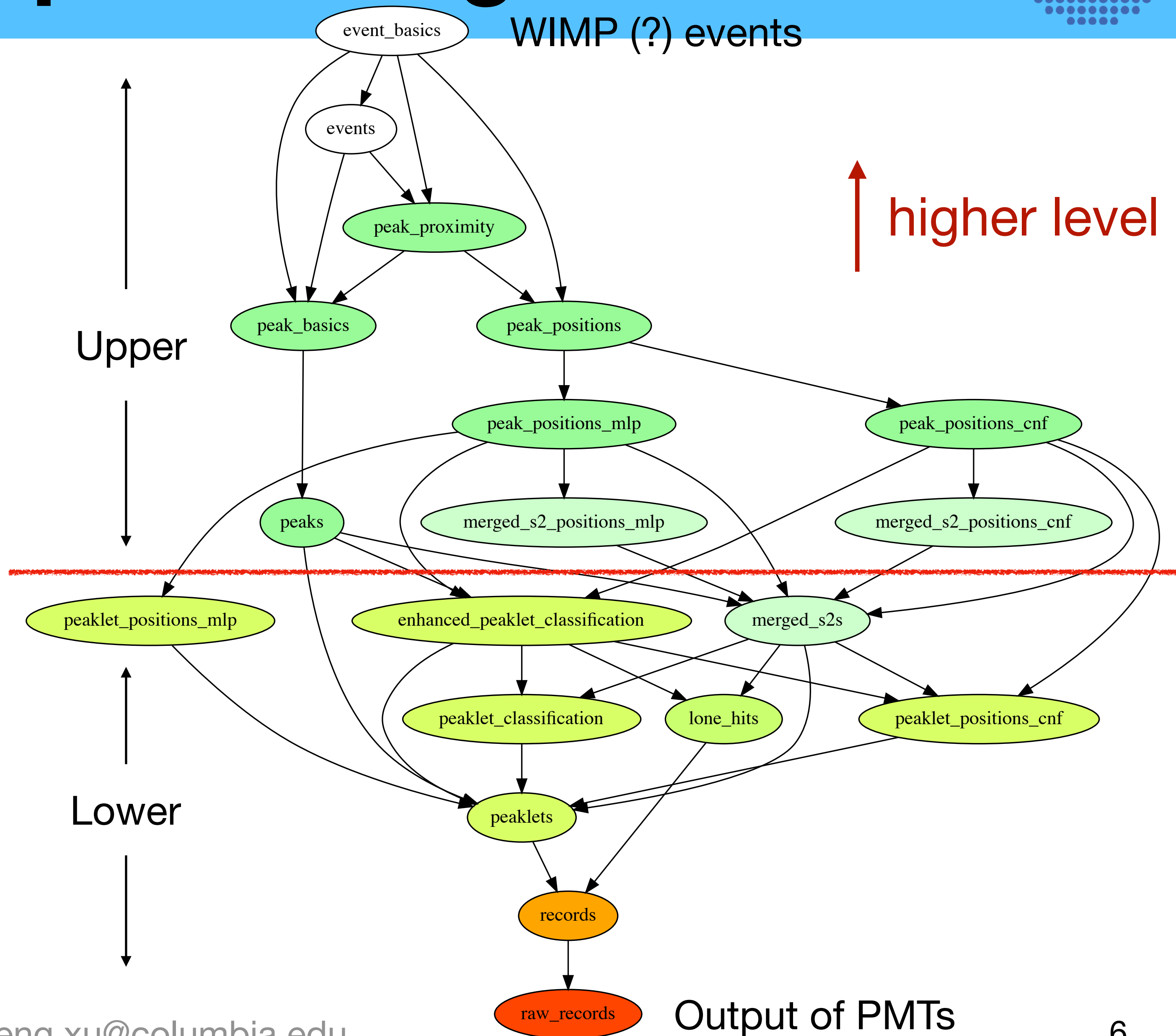
XENONnT SR2 data flow

In Science Run 2 (SR2), we improved the data flow and reprocessing pipeline.

LNGS in Italy

Rucio Replica Storage Element (RSE)

OSG computing nodes

storage in Research Computing Center (RCC) - The University of Chicago

Rucio Replica Storage Element (RSE)

23k runs
260 days data taking

(Pre-processed) Raw Data
~3.7 PB (160 MB /s)
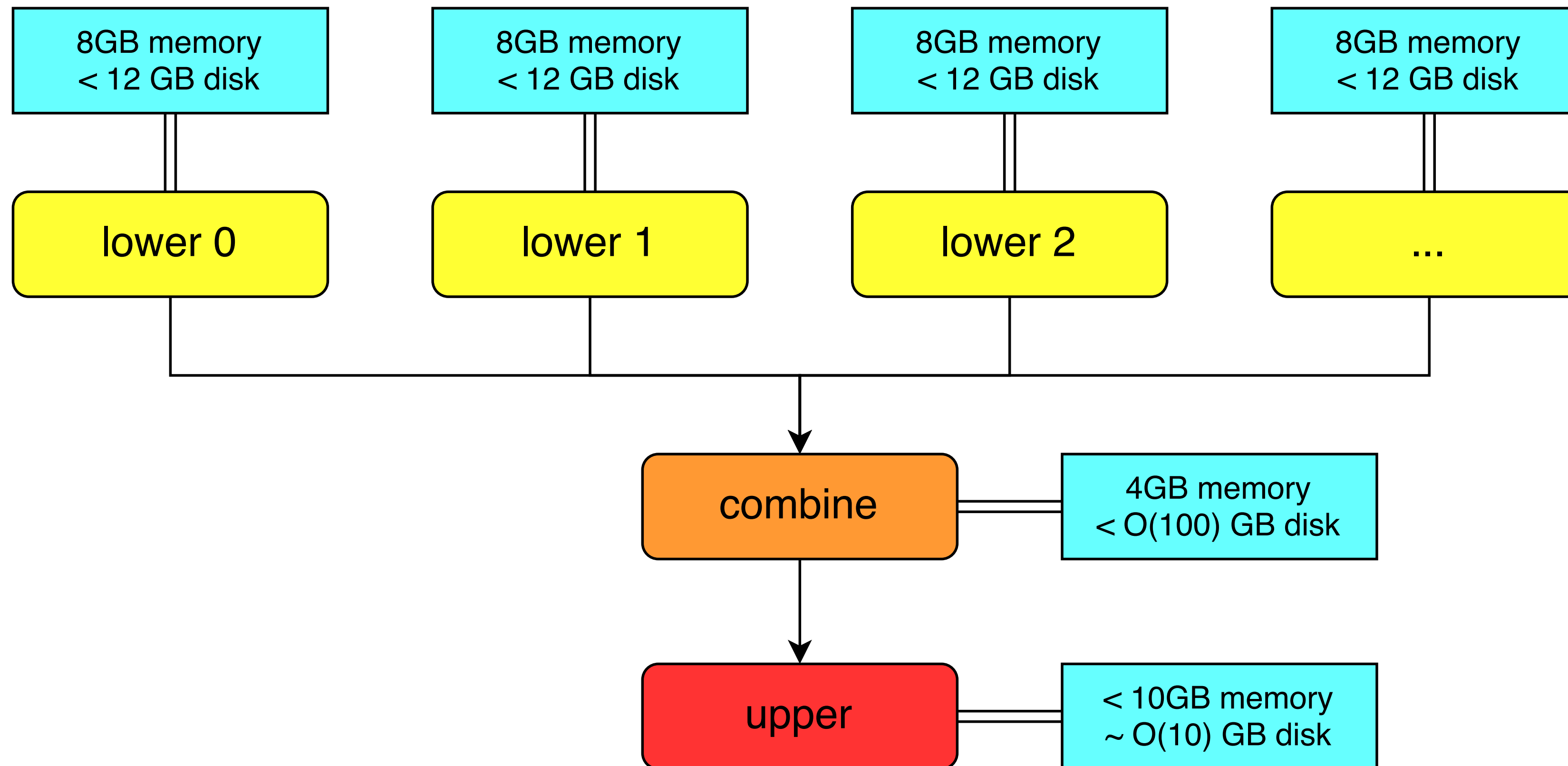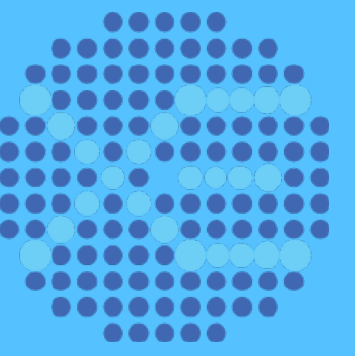
Processed Data
~220 TB
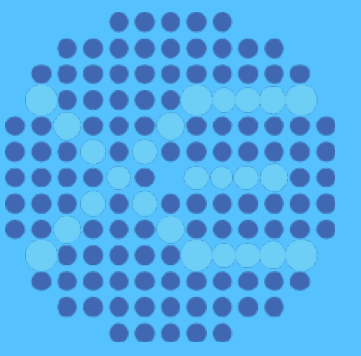
Outsource as a wrapper

- Pegasus: a workflow management software (not by XENON)
  - Great help from Mats Rynge
- straxen: XENONnT data reprocessing software
- Outsource: the wrapper combining the functionality of Pegasus and straxen (tell Pegasus the resources straxen needs)
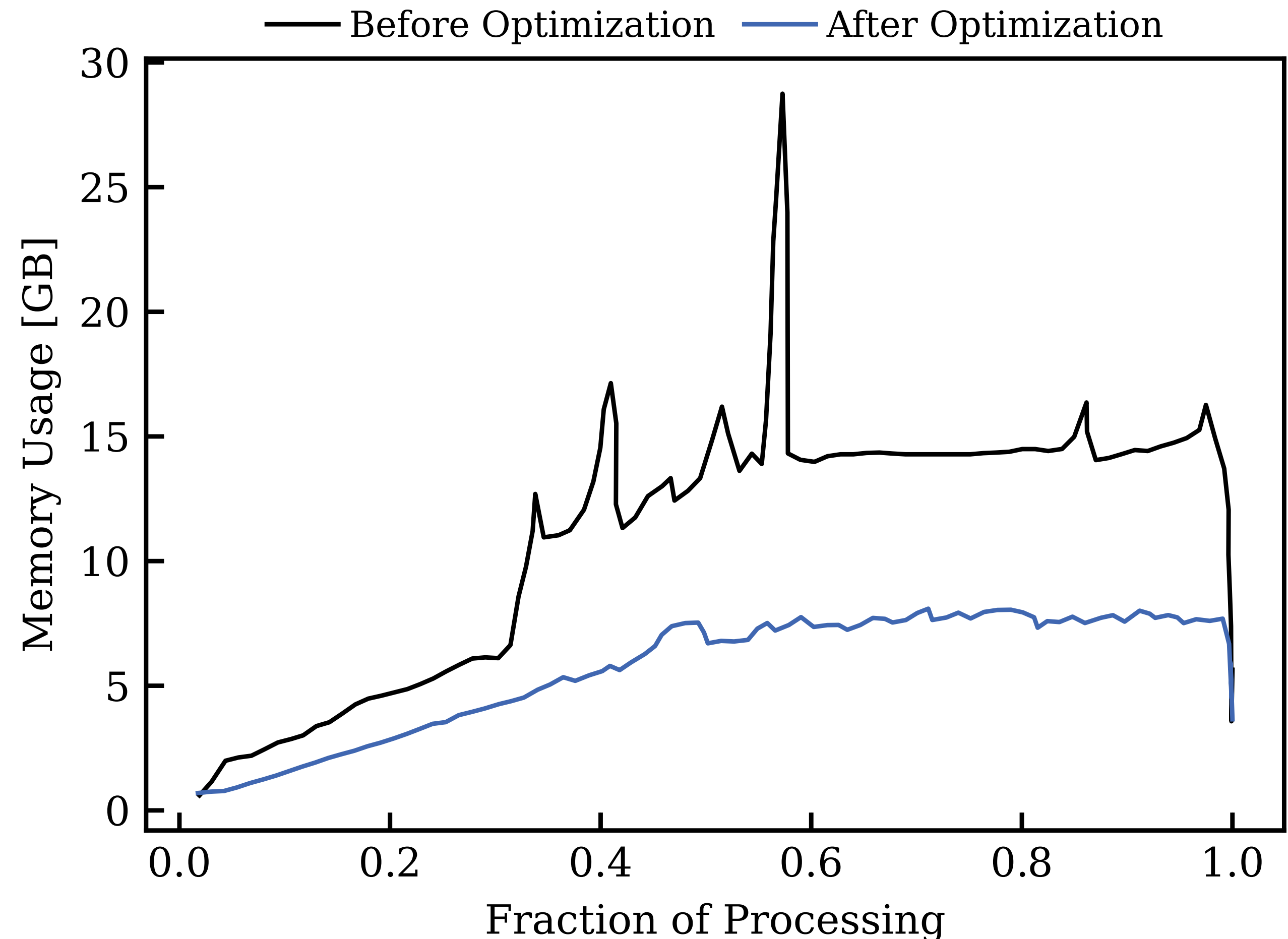
Outsource (Pegasus) workflow

- Lower jobs are light weighted

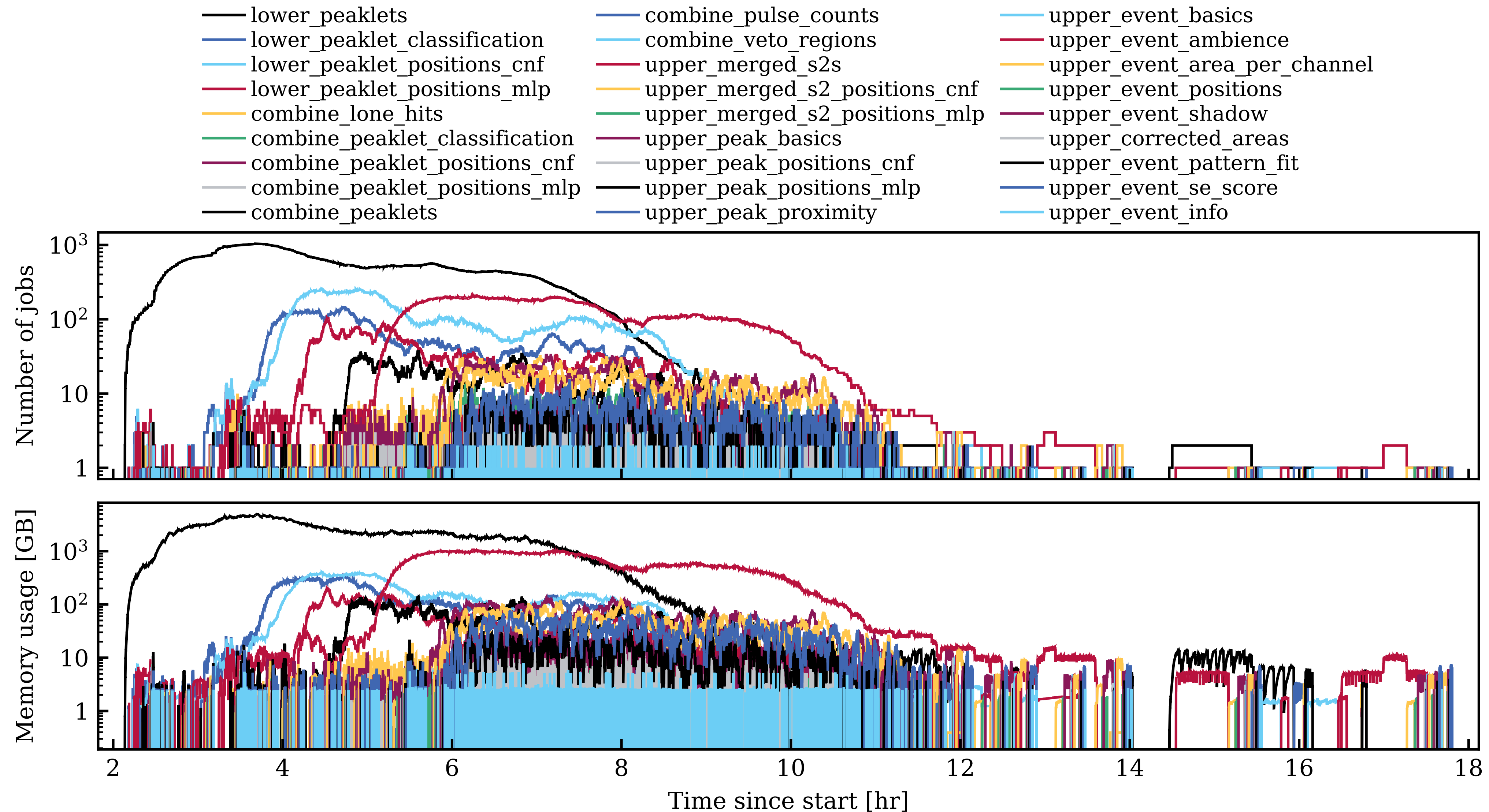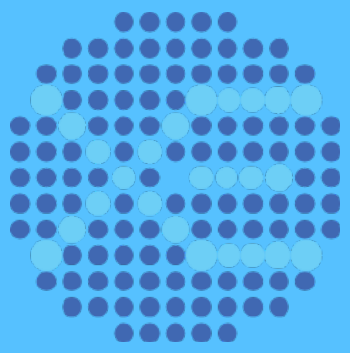- Combine and upper jobs are heavy in disk usage
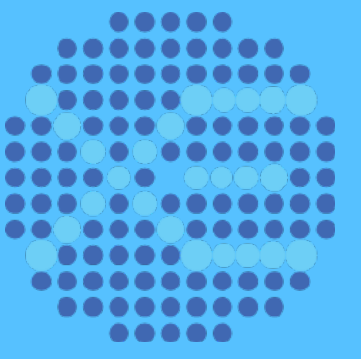
Efforts on reducing memory usage

- Memory consumption beasts are killed.

- The memory usage after the optimization is limited by the size of chunk in a single run.

- The reduced memory usage greatly shortened the reprocessing timeline.

Resources used in one example workflow
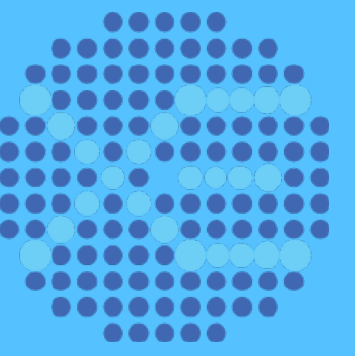
dacheng.xu@columbia.edu

Resources used in SR2 data reprocessing

- Wall time: ~52 days

- CPU time: ~17 years

- Number of (successful) workflows: 53

- Number of (successful) jobs:

  - Lower: ~51k

  - Combine: ~24k

  - Upper: ~24k

Conclusion and Outlook

- We had an overview of XENONnT SR2 reprocessing.

- The XENONnT is greatly benefited from the computing facilities from OSG.

- XENONnT has a reprocessing pipeline that uses self-aware resources.

- The next generation of LXe TPC dark matter project, XLZD (https://arxiv.org/abs/2410.17137), will have ~1200 PMTs. So the amount of data will continuously increase in the next decade.

Many thanks to Pascal Paschos, Judith Stephen,
Mats Rynge and all contributors of OSG!

Thank you!