

Journey of the data in PandaX dark matter experiments

Dan Zhang¹, Pengwei Xie², Xun Chen²

University of Maryland, College Park¹

Shanghai Jiaotong University²

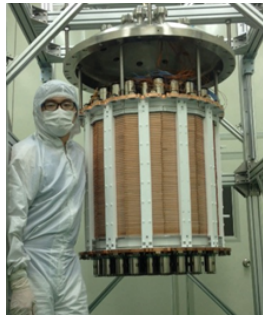
On the behalf of PandaX Collaboration

PandaX Experiment

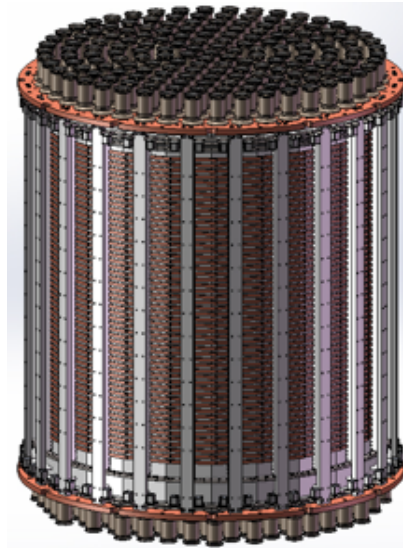
PandaX = Particle and Astrophysical Xenon Experiments



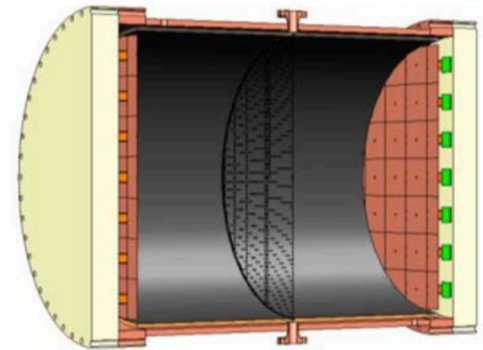
PandaX-I: 120 kg
DM experiment
2009-2014



PandaX-II: 580 kg
DM experiment
2014-2019



PandaX-4T: multi-ton
DM experiment
2019-



PandaX-III: 200kg-1ton
Gaseous Xenon detector
(future)

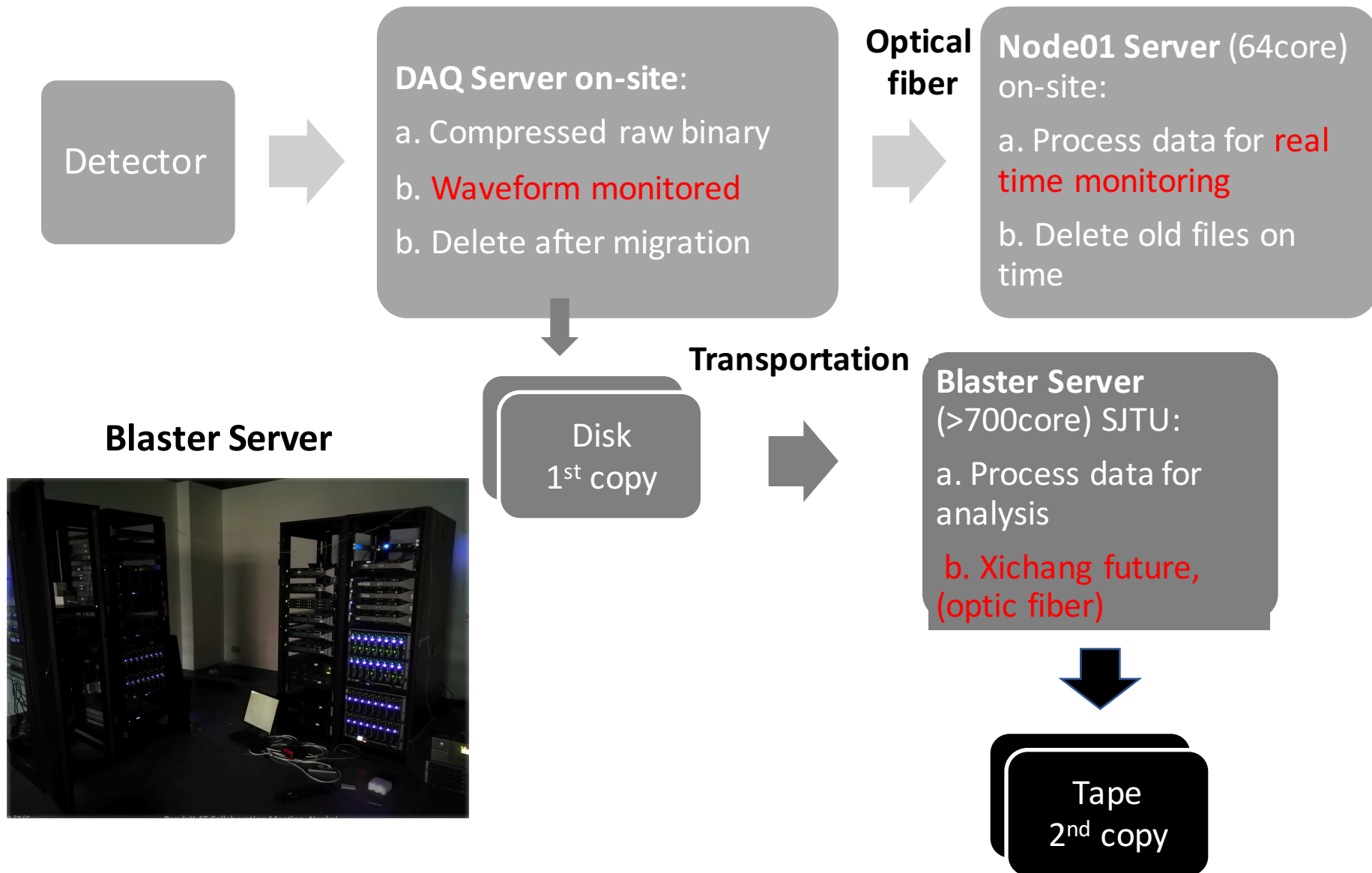
Outline

1. Data load
2. Journey of the data
 1. Storage and Transportation in PandaX-II and future
 2. Processing chain in PandaX-II and preparation for PandaX-4T
3. Software accessories

1. Data load

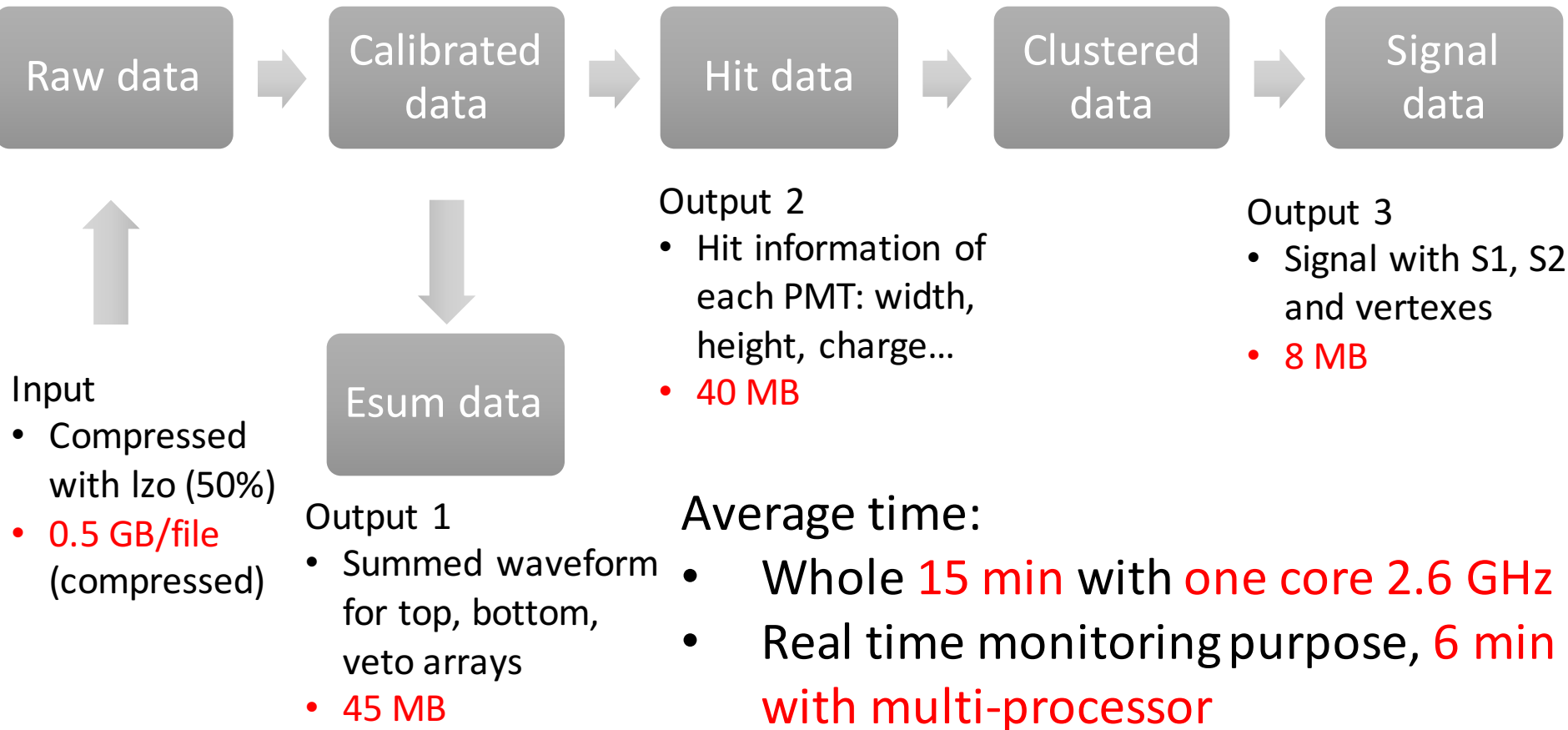
- PandaX-II (800 TB raw data)
 - 1.2 MB/s, $\sim 2.2\text{Hz}$ (dark matter run)
 - 42 MB/s, $\sim 100\text{Hz}$ (random trigger ^{220}Rn calibration)
- PandaX-4T
 - Estimation: 300 MB/s (100 Hz in calibration)
 - DAQ design limitation: 800 MB/s
- PandaX future experiments (30T, 100T)
 - $\sim\text{GB/s}$

2. Journey of the data in PandaX-II and future



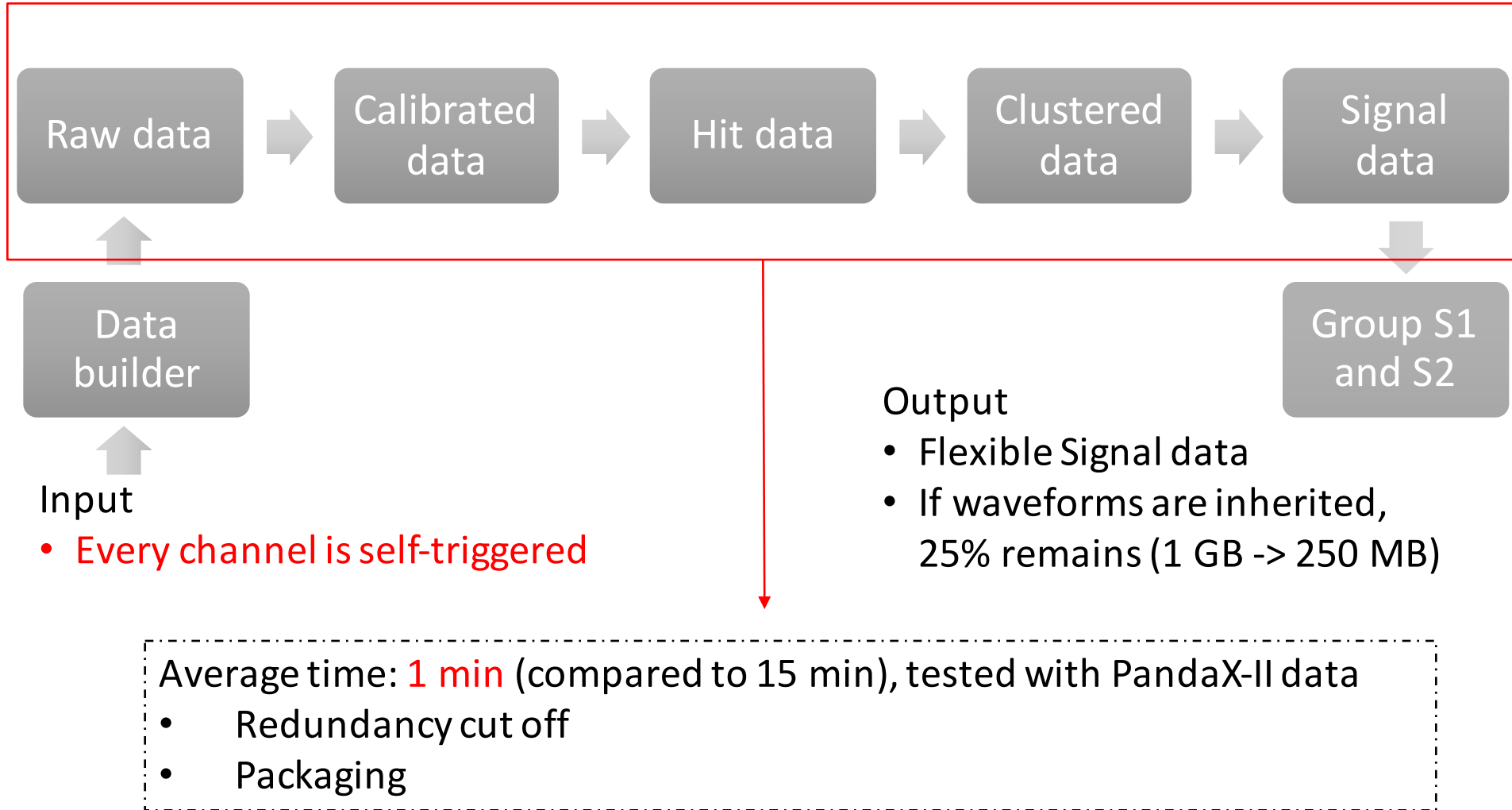
Data processing chain in PandaX-II

C++98, ROOT5



Data processing chain prepared for PandaX-4T

C++14, ROOT6



3. Software accessories

- Improvement in persistent layer
- Online real-time monitor
- GitLab software management

- Machine learning/deep learning applications

Improvement in persistent layer

- Bamboo-shoot2

(PandaX-I, PandaX-II)

- Based on ROOT TTree, ROOT serves as the persistency layer
- Mapping Customized TObject to the Branch

- Problem

- **Compatibility (root 5.34/20)**

- Bamboo-shoot3

(PandaX-4T)

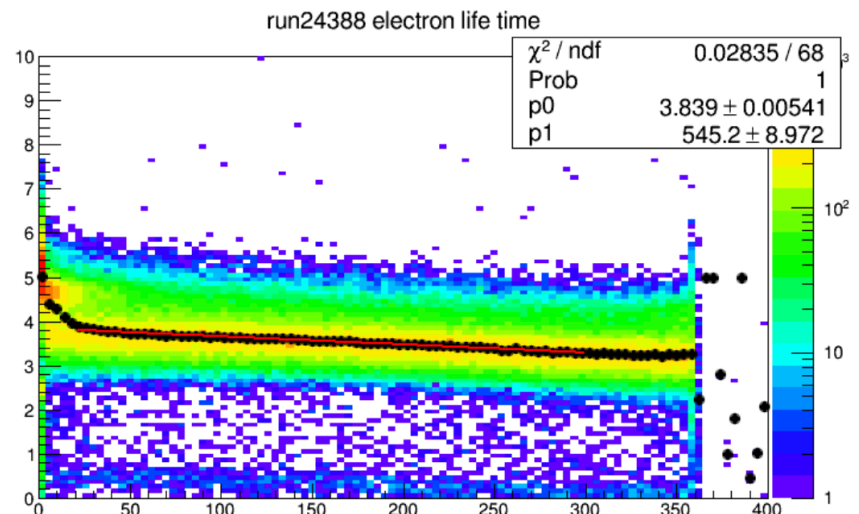
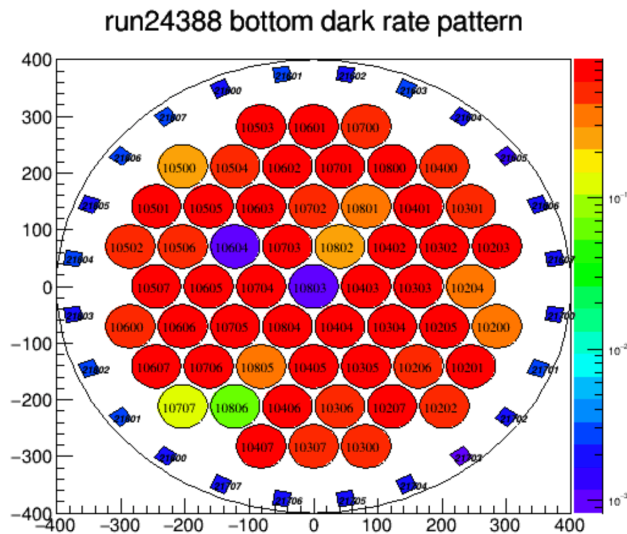
- No dependency on ROOT
 - Object serialization library
 - File I/O library
 - Other utilities
- Data format
 - Bamboo-shoot data (bsd)
 - ZSTD compressed file (support lzo and no compression)

<https://github.com/revive/bamboo-shoot3>

Online real-time monitor

- Waveform display (DAQ server)
 - Pipe stream, not stored but plotted when the webpage is visited
- Data quality page (64core server, **6 min right after getting the data**)
 - Not deleted in the database
 - Detector and photomultiplier condition for each run and file

Run Nr.	Started	Duration	Rate	Type	Description
24388					
24388	Sep 11 21:10	10h 31m	8.47	PHYSICS_KR_ZEO	Zeolite Kr83 injection, 52+5SLPM



GitLab – software management

The screenshot shows the GitLab interface with a dark blue header. The main content area is titled "Projects" and includes a "New project" button. Below the title, there are navigation links for "Your projects 38", "Stared projects 1", and "Explore projects". A search bar labeled "Filter by name..." and a dropdown menu for "Last updated" are also present. The project list is filtered to "All Personal" projects. Four projects are listed, each with a letter icon, name, description, developer badge, and statistics (stars, forks, issues, merge requests).

Project Name	Description	Developer	Stars	Forks	Issues	Merge Requests	Last Updated
P PandaX / p4_tools		Developer	0	0	0	0	Updated 2 days ago
N PandaX / nest-lite	pandax version of nest	Developer	1	0	0	0	Updated 1 week ago
P PandaX / p4sc	slow control related code for pandax-xt	Developer	0	0	0	0	Updated 2 weeks ago
B software / BambooMC	BambooMC is a modularized simulation program based on Geant4.	Developer	3	3	0	0	Updated 1 month ago

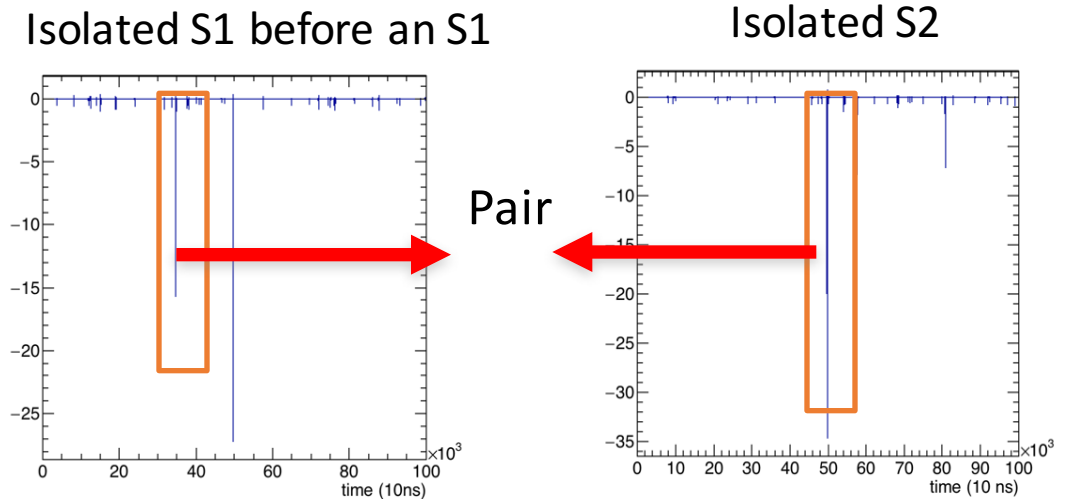
Machine learning /deep learning applications

- Accidental background rejection in dark matter experiment

- A multivariate approach (Boost decision tree)

- Signal identifying in neutrinoless double beta decay (NLDBD) experiment

- Conventional neuron network
- Qiao Hao et. al., Signal-background discrimination in PandaX-III

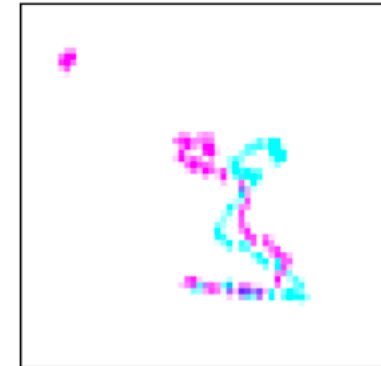


PandaX-III MC Simulation analysis

NLDBD signal



^{214}Bi decay gamma



Challenges for discussion

Discussion 1: self-trigger

Group S1
and S2

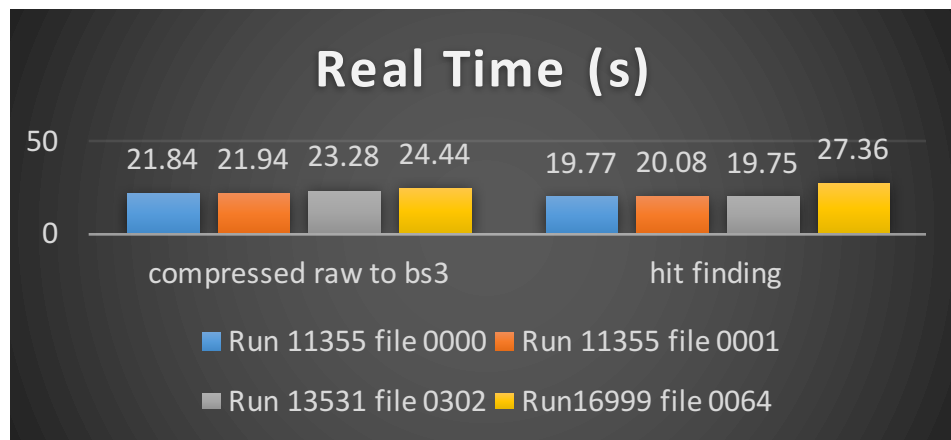
Black box for PandaX!

Every channel is self-triggered after passing the threshold.

No event number is recorded at the first place.

Discussion 2: decompression

- Trade-off between compressing ratio and speed
- ZSTD is used in Bamboo-shoot3, mainly selected for compressing ratio, 1GB \leftrightarrow 0.5GB



Discussion 3: algorithms

- Position reconstruction: bad PMTs (horizontal), low drifting electric field (vertical)
- A better energy resolution.

Thank you!

Backup

```
auto res = pu::map(
    [&](const RawData &raw) {
        auto cd = cal(raw);
        auto hd = hit_finder(cd);
        auto sd = hit_clustering(hd);
        auto ssd = sig_builder(sd);
        auto tsd = sig_tagging(ssd);
        auto psd = sig_pos_rec(tsd);
        return std::make_tuple(std::move(hd), std::move(psd));
    },
    raw_data);
```


New data reconstruction with Bamboo-shoot3

Test with PandaX-II data

```
tuple RawPmtSegment {
  uint32 startTime;
  uint32 channelNumber;
  [uint16] adcValue;
};

struct RawData {
  uint32 eventNumber = 1;
  uint64 triggerTime = 2;
  [RawPmtSegment] segments = 3;
};
```



Decompress

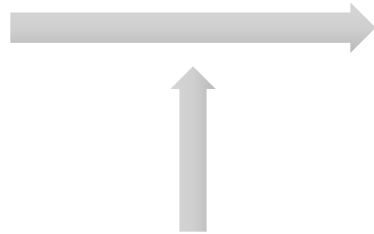


Calibrated
data

PMT gain in the
database, done by
LED calibration daily

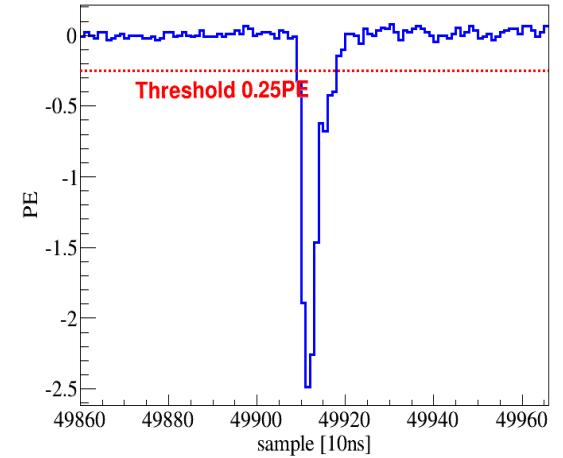
- Input 1GB per file
- PandaX-II
 - 1.2 MB/s, ~ 2.2Hz (dark matter run)
 - 42 MB/s, ~ 100Hz (random trigger ^{220}Rn calibration)
- PandaX-4T
 - Estimation: **300 MB/s** (**100 Hz** in 4T sensitive volume)

Calibrated data

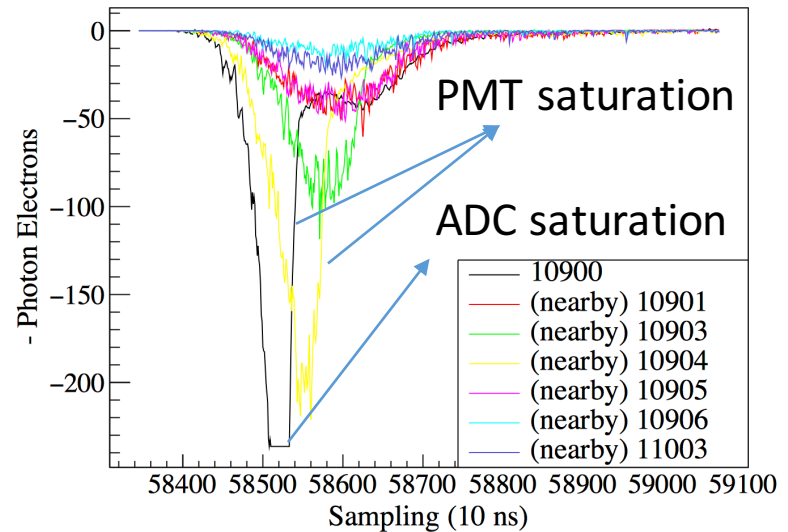
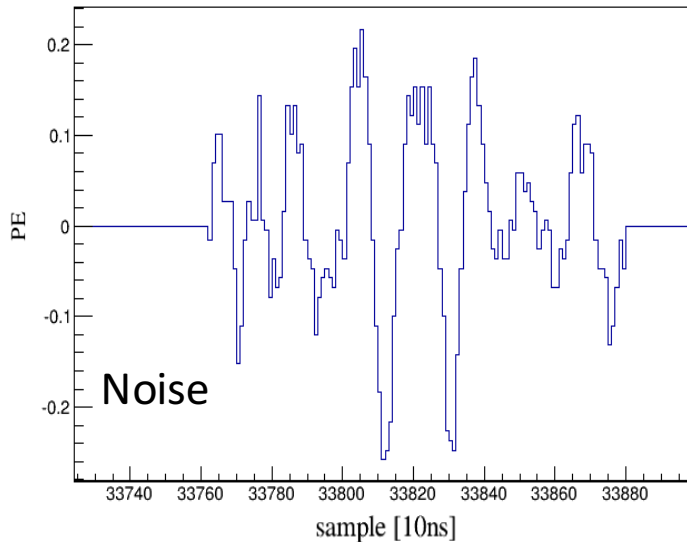


Hit data

- Hit finding with a threshold (0.25 PE in PandaX-II)
- Hit tagging (noise, saturated, normal)
- Hit properties (area, peak, width, etc.)



Hit finding schematic

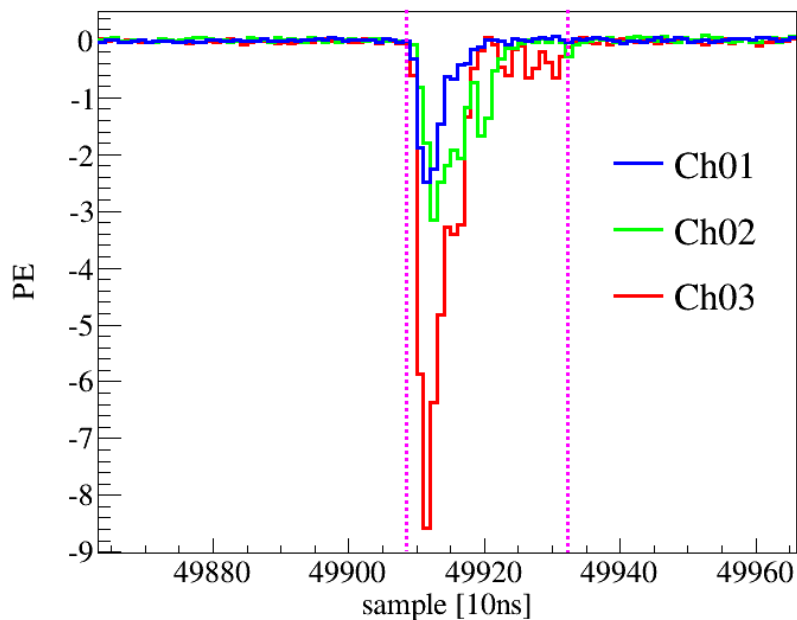


Hit data

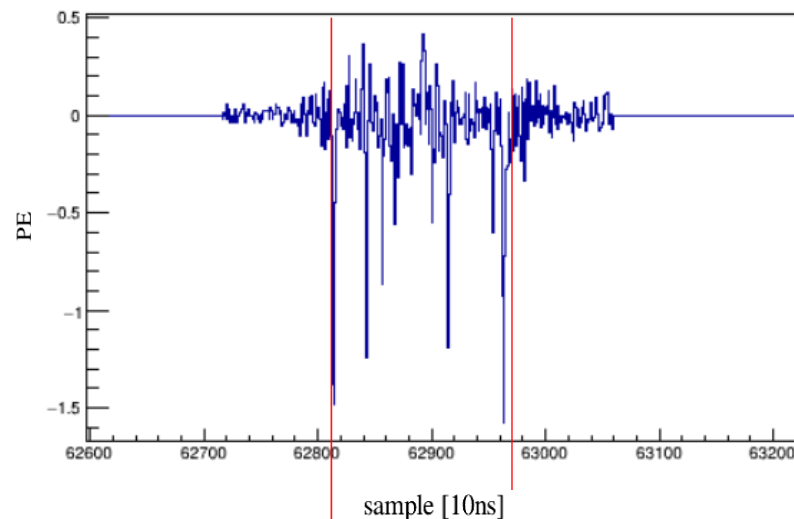


Clustered data

- Criterion: hits overlap in time will form a signal
- Special treatment: Small S2 (time separation of hits less than 500ns)



Typical clustering example



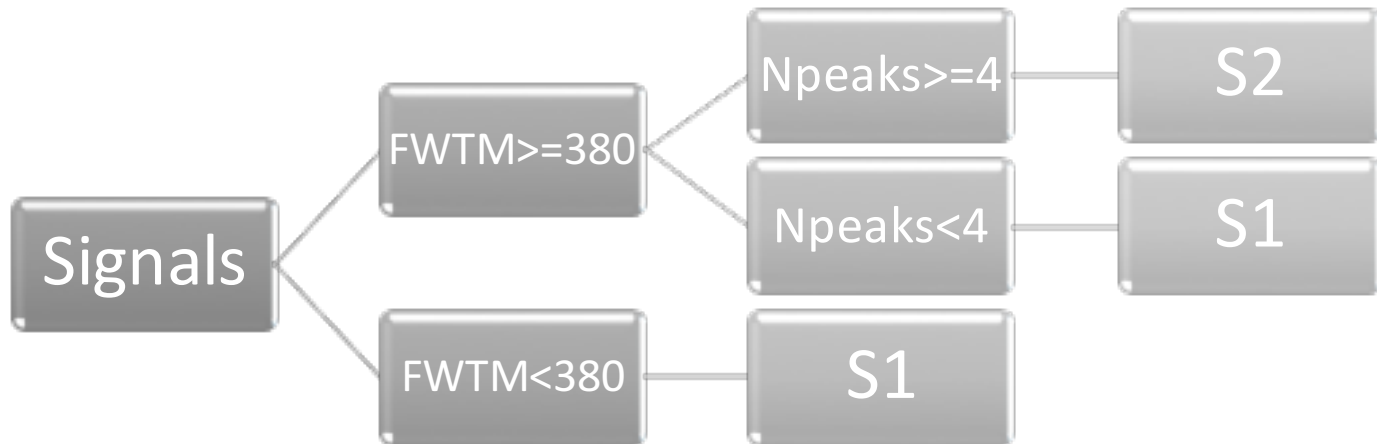
A top array summed waveform for single electron S2

Clustered
data



Signal
data

- S1 and S2 discrimination
- Horizontal position reconstruction



Horizontal position reconstruction

- Center of gravity (calculated for S1 or S2, top or bottom separately)
- Photon acceptance function (PAF) + statistic inference (maximum likelihood)

^{83m}Kr events
comparison

Turned off 4 out
of 55 top PMTs

Details will be
Published soon.

