Data Processing and Simulation in PandaX-4T

谌勋

上海交通大学 粒子物理与核物理研究所

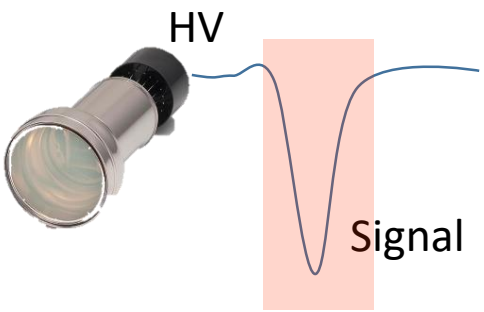
第十三届全国物理学术会议

Outline

- Data production in PandaX-4T
- Data pipeline
- Data processing details
- Data quality monitoring
- Simulation with BambooMC
- Summary and outlook

Signals in PandaX-4T

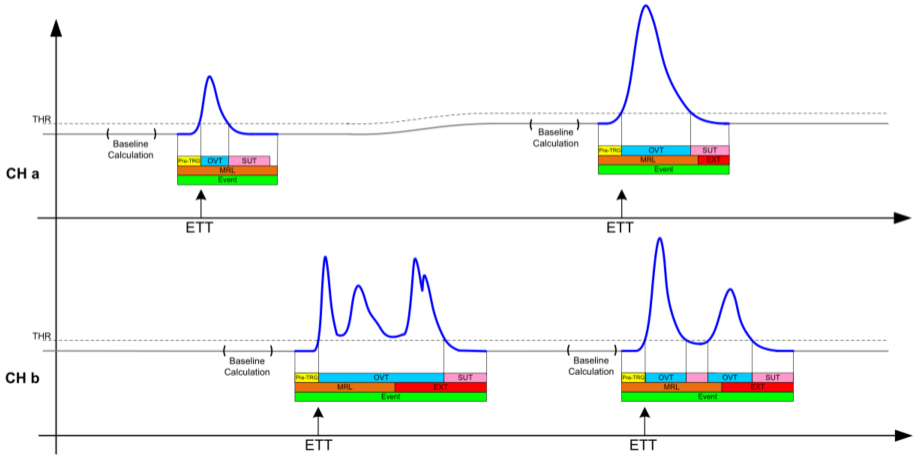
- Search light in the dark
  - Primary scintillation S1
  - Proportional scintillation S2
- Detect photons by PMTs
  - Photoelectric effect
  - Readout voltage



Signals to Data

For each PMT

- Digitization waveform at 250MHz (4ns/sample)
- Self-trigger with DAW firmware (triggerless!)

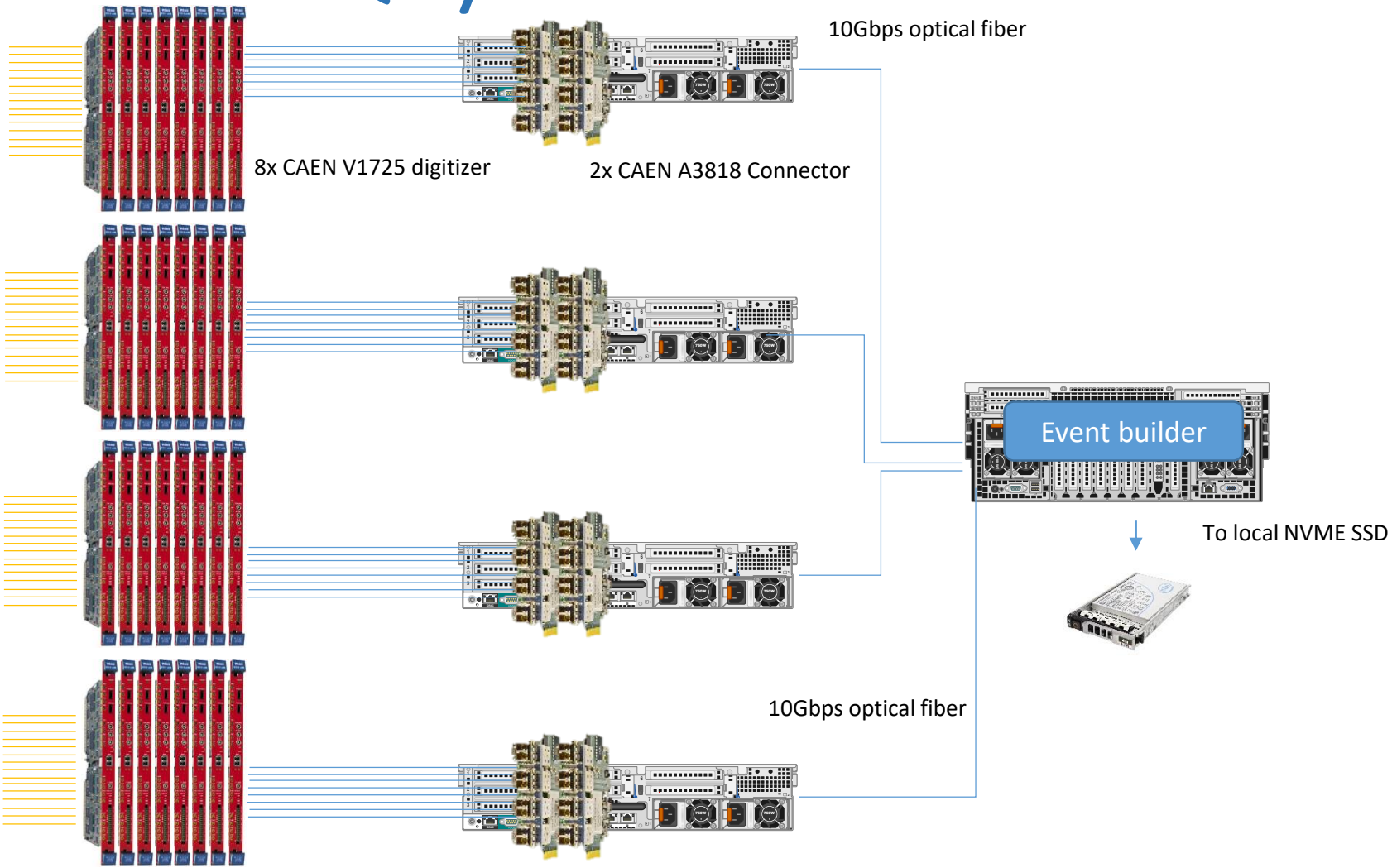


~ 487 readout channels

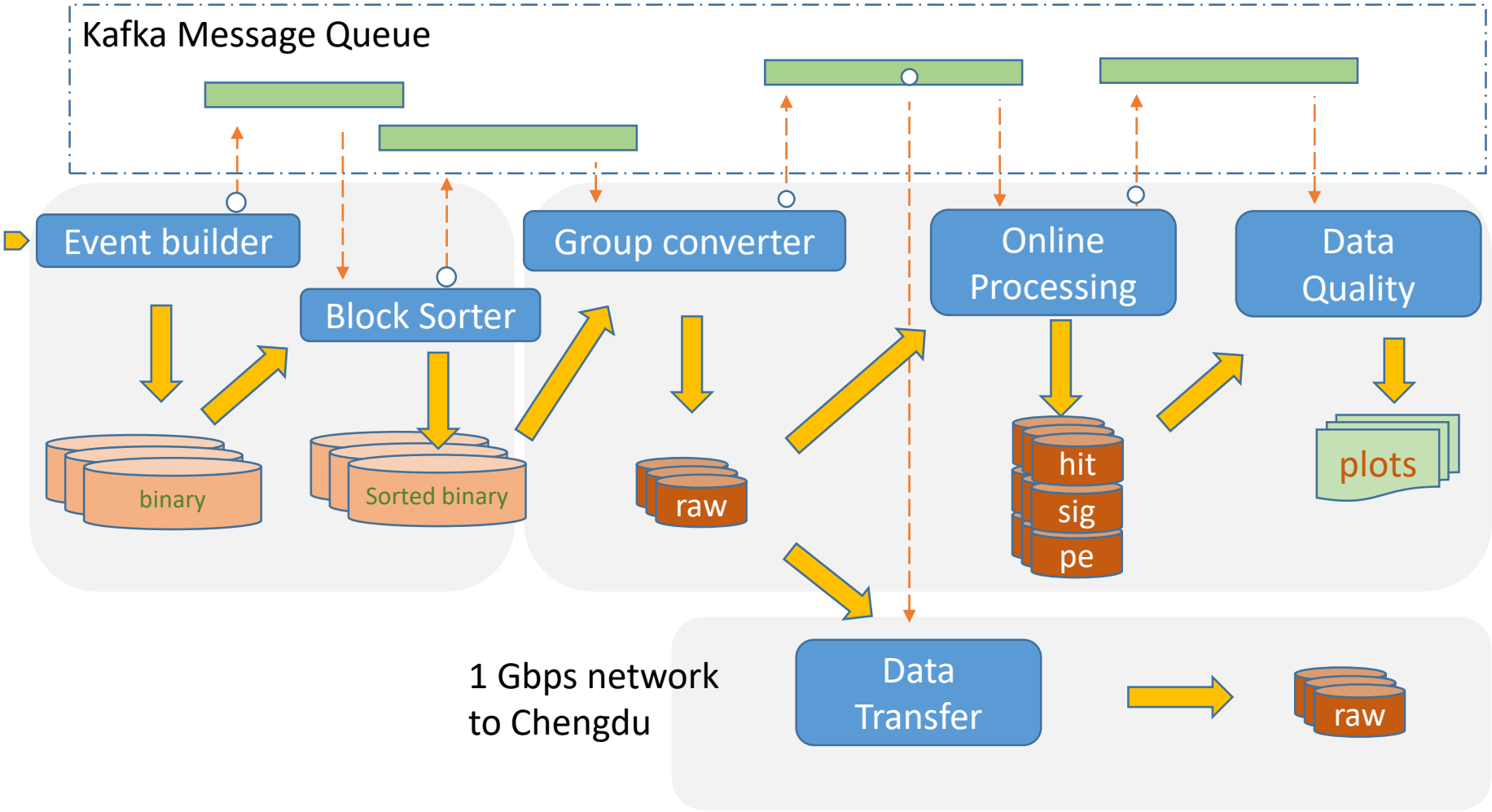
- 199 top PMTs (with 14 low gain channels)
- 169 bottom PMTs
- 105 veto PMTs

See Ren Xiangxiang's talk on Monday

The DAQ System



Message controlled data pipeline



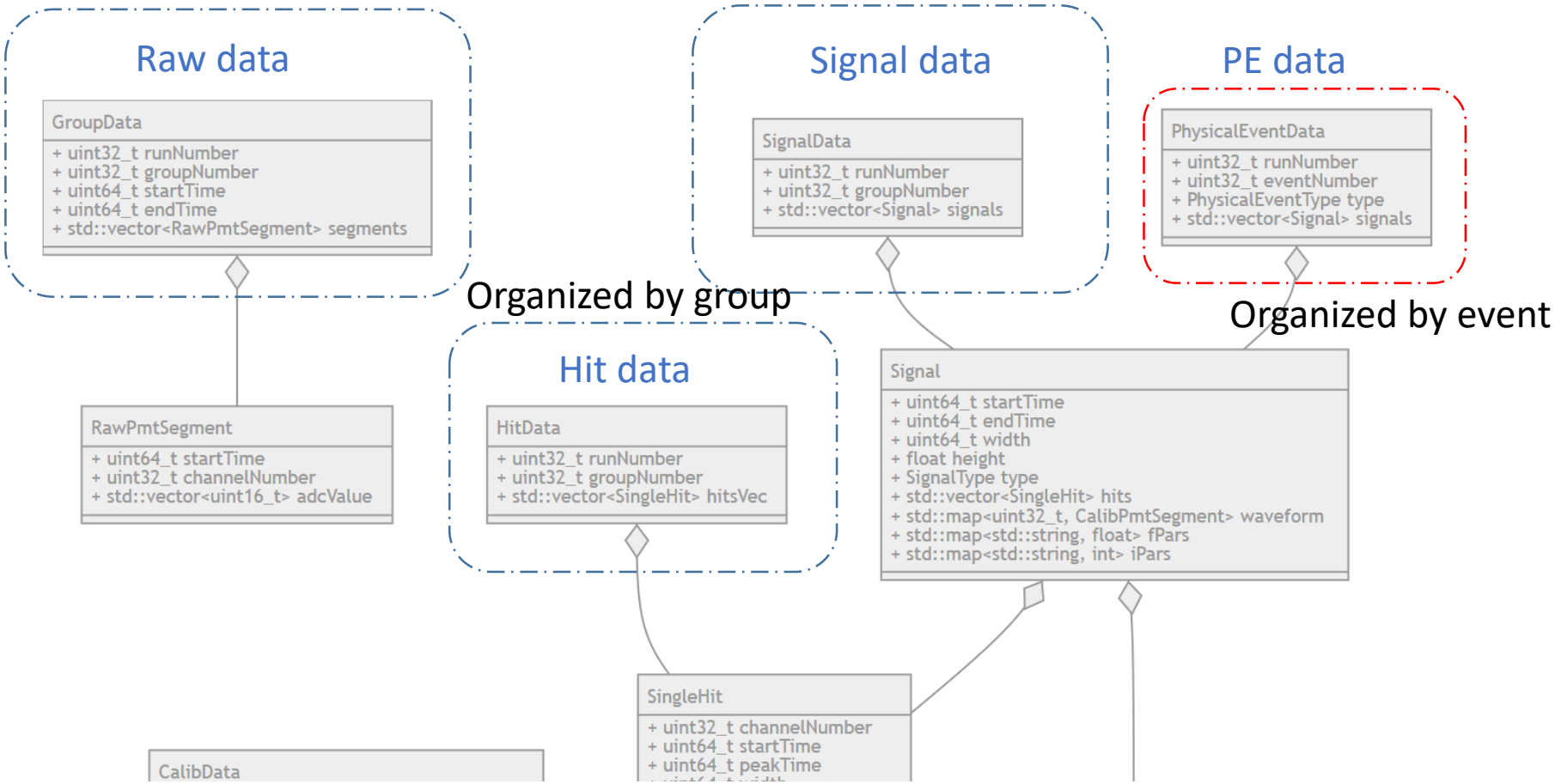
Data file Format

- Based on fast I/O library of bamboo-shoot3
  - https://github.com/pandax-experiments/bamboo-shoot3
  - Requires C++14
  - Improved I/O speed
  - Internal compression with ZSTD by default, ~50% compression ratio for PandaX raw data.
- Agile data structure defined by pseudo code with code generation

```
tuple RawPmtSegment {
  uint64 startTime;
  uint32 channelNumber;
  [uint16] adcValue;
};
```

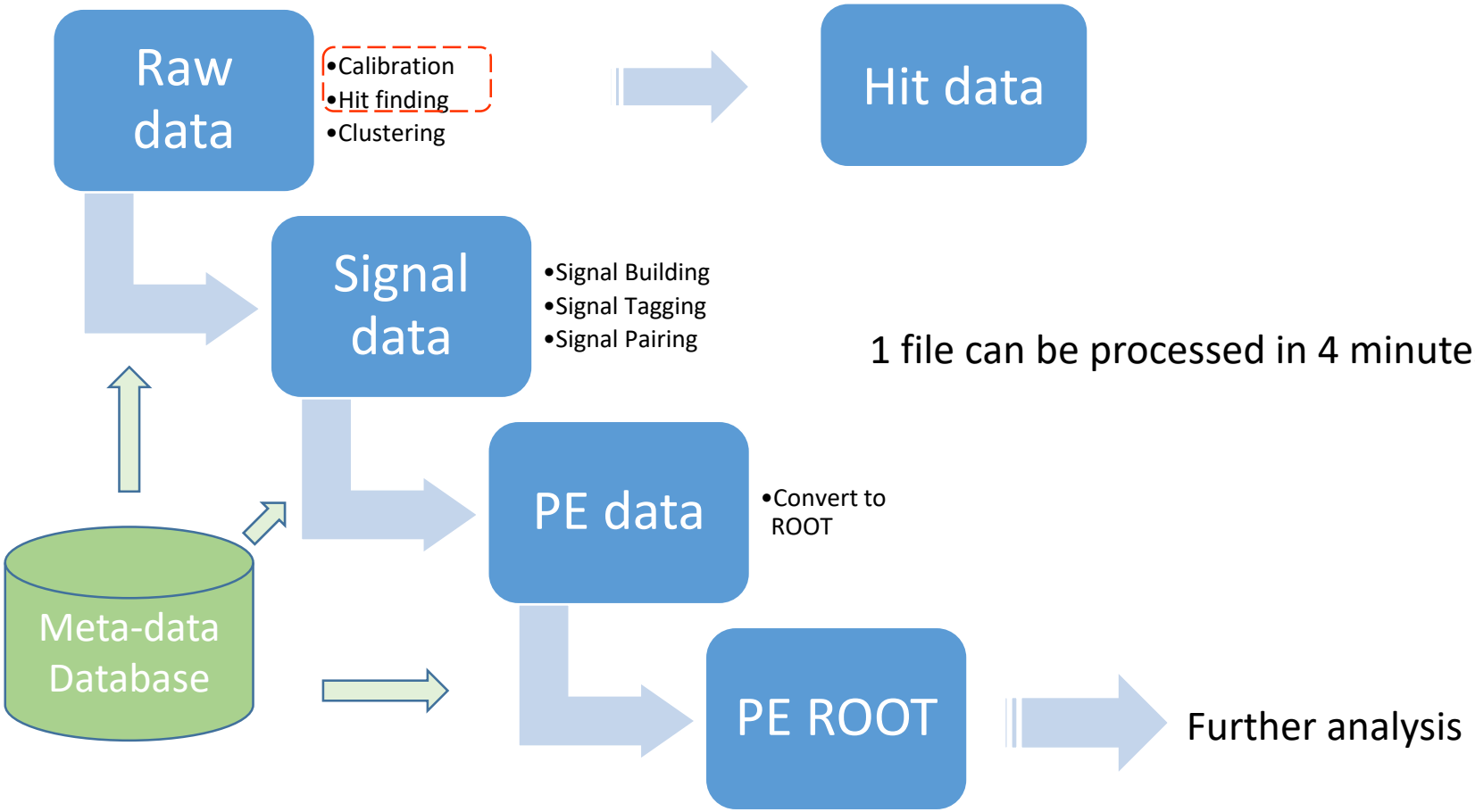
```
struct GroupData {
  uint32 runNumber = 1;
  uint32 groupNumber = 2;
  uint64 startTime = 3;
  uint64 endTime = 4;
  [RawPmtSegment] segments = 5;
};
```

Data files in PandaX-4T



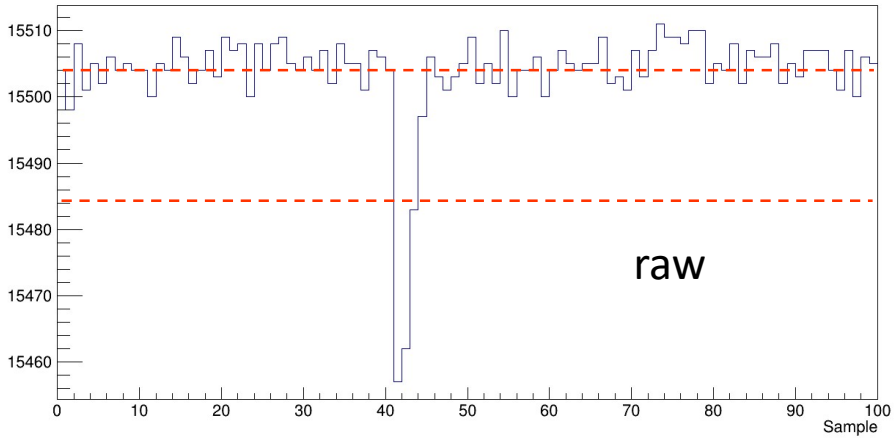
Visualization of data can be found in Abdusalam's poster.

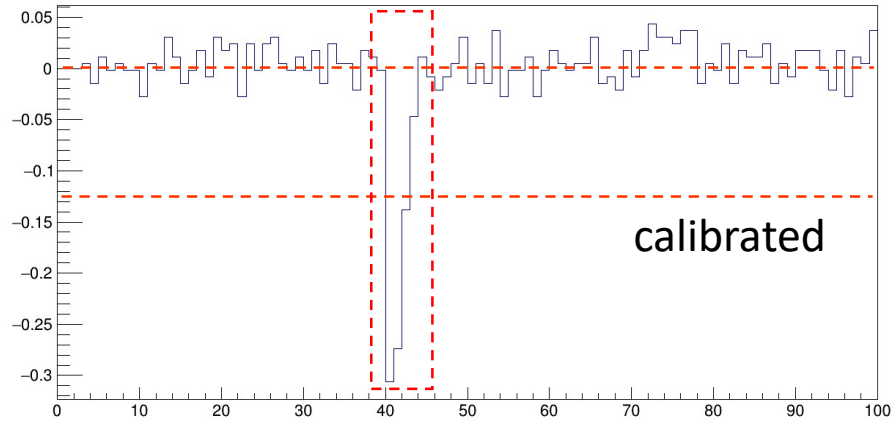
Data Processing



1 file can be processed in 4 minutes.

Calibration & Hit Finding





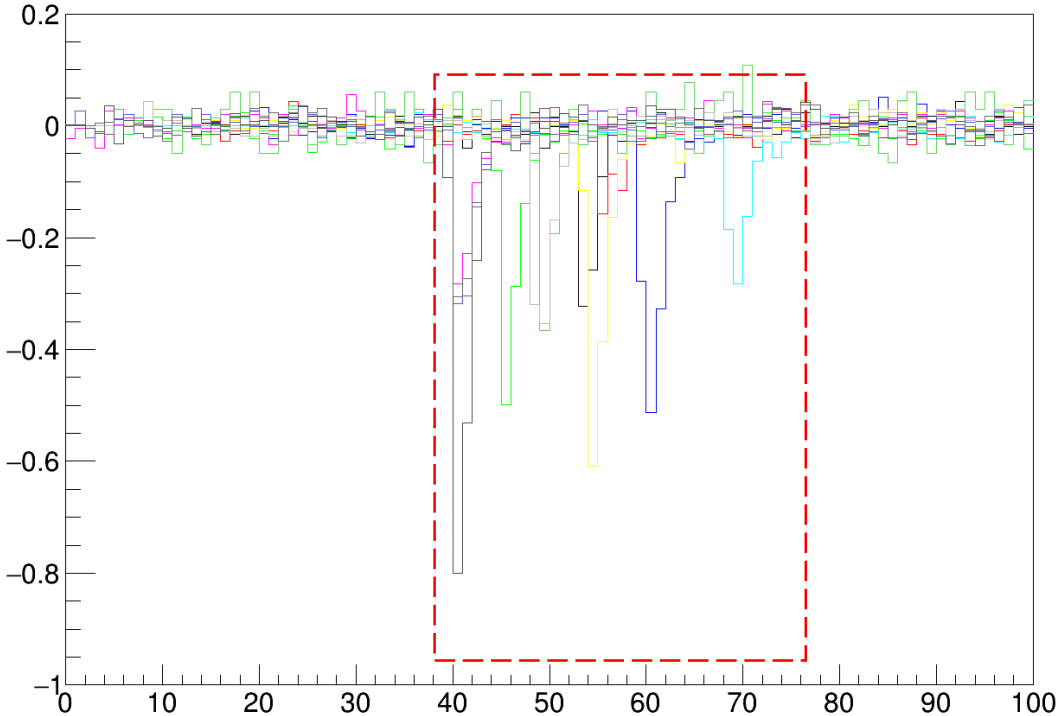
Calibration

- Baseline calculation & subtraction
- Gain modification
- Time modification

Hit finding

- Threshold
- Calculate pre- and post-pedestal, rms, area and height

Signal clustering, building, tagging



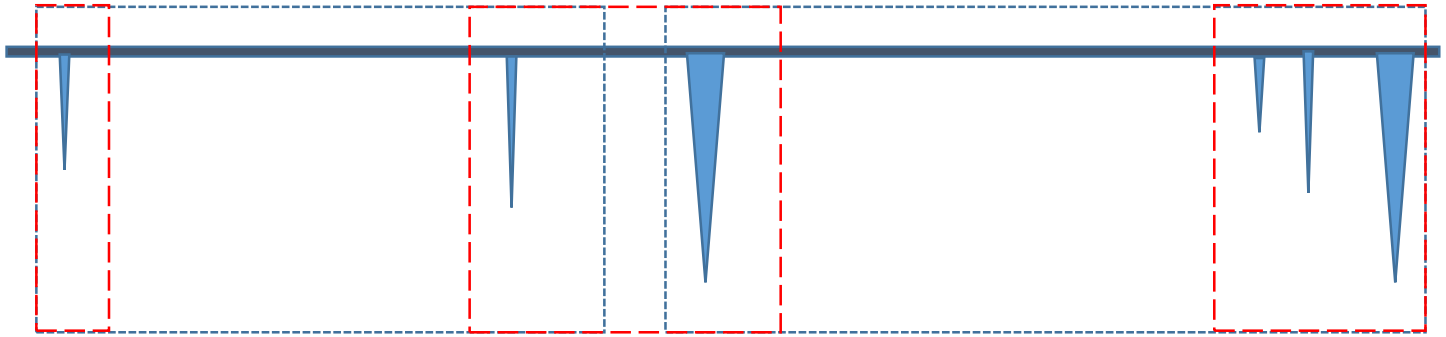
Clustering: groups hits together

Building -
calculate properties of signal:
Charge, height, width,
position…

Tagging -
Recognition of the signal types:
S1, S2, Sparking …

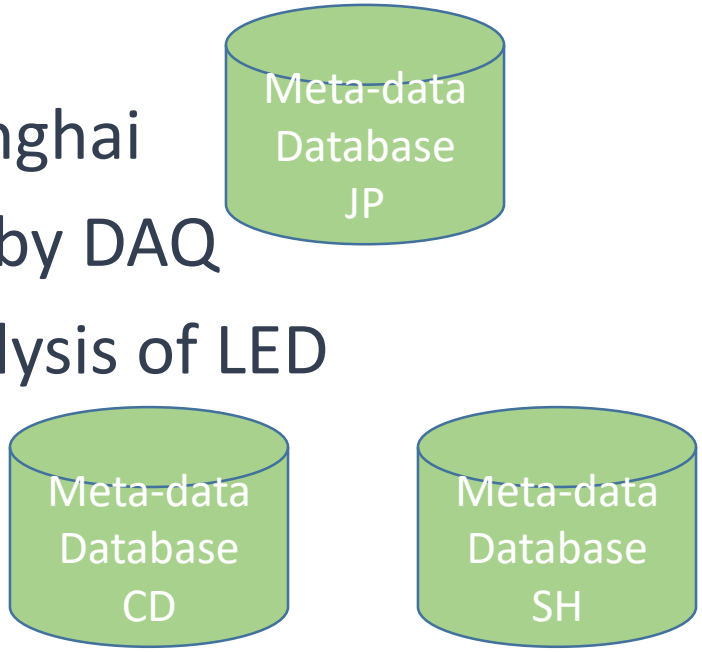
Signals to Events

- Need to pair signals to form physical events
- One group may contain many events
- One event may across two groups
- Analysis start from events



Database for meta data storage

- Based on PostgreSQL
- Main server in Jinping
- Hot standby in Chengdu and Shanghai
- Run and file information created by DAQ
- Gain information updated by analysis of LED calibration data
- Mapping of PMTs



Computing infrastructure in Chengdu

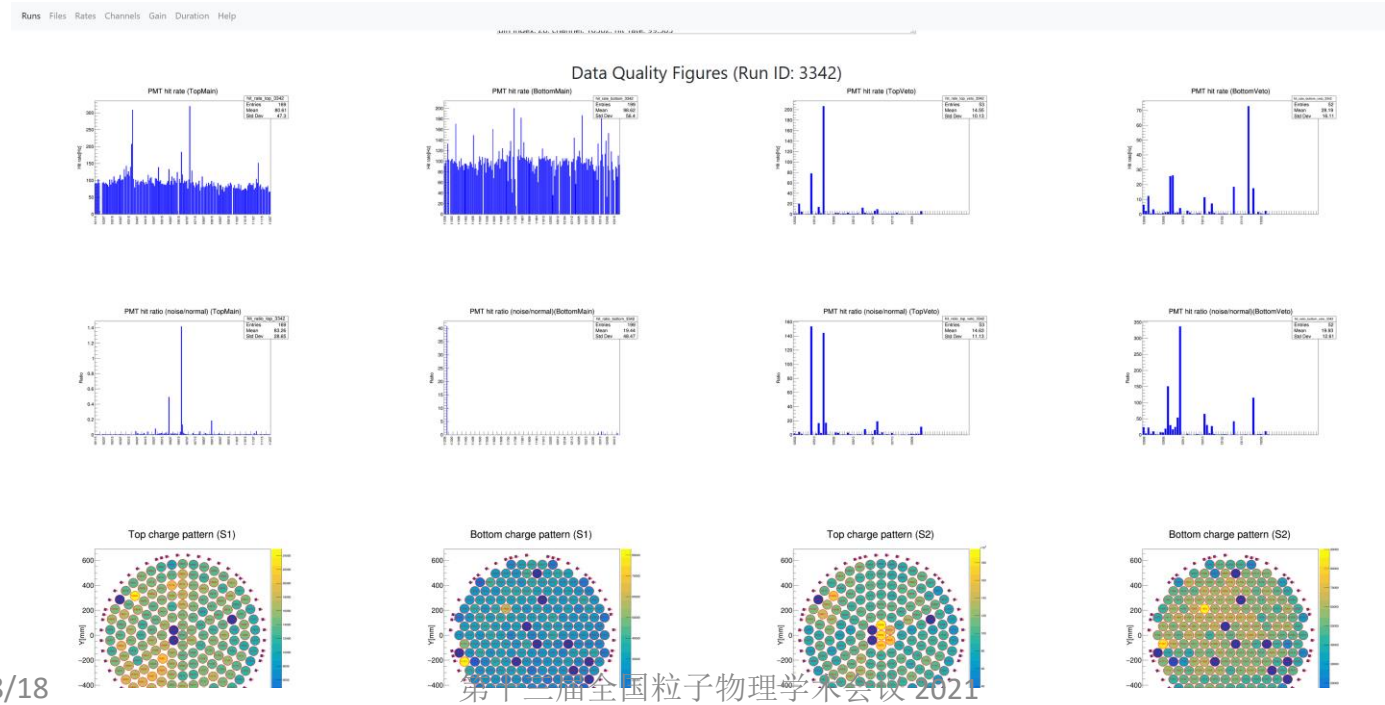
- Computing:
  - 320 cores (AMD EPYC 7452)
  - 4G mem/core
  - Job managed by HTcondor
- Storage
  - 1.7 PB high speed parallel storage
  - Backup data in LTO-8 tapes
- Network:
  - 100Gbps IB (internal)
  - 1Gbps to JP
  - 500Mbps (external)
- Over 480 TB data collected since the start of the commissioning run

Massive data re-production

- Required when
  - database updated (mostly gain)
  - Algorithm updated
    - Tagging
    - Re-clustering
    - Position reconstruction
- Reprocess all “golden” data (> 220TB) in the commissioning run with all the computation cores in Chengdu in 3-4 days.

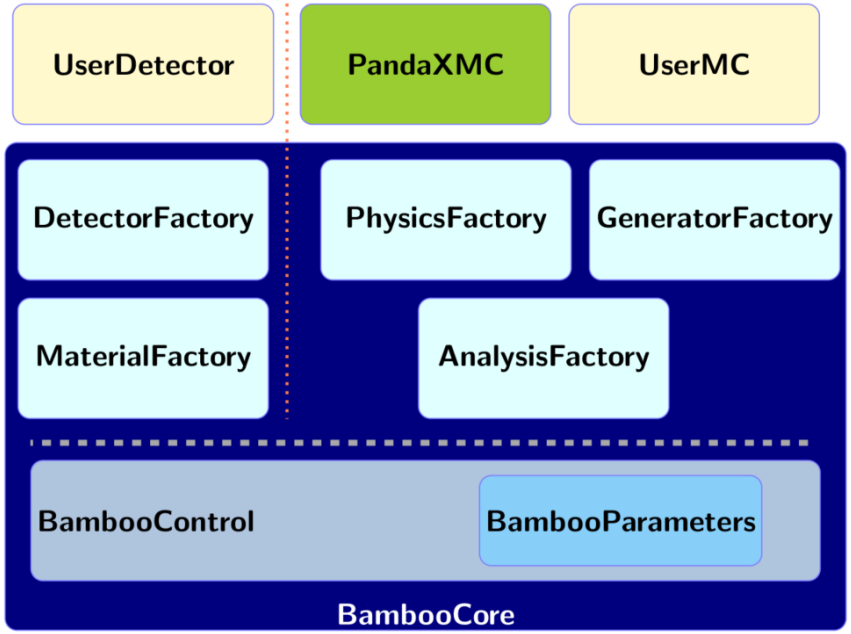
Data quality monitoring

- Process selected data in JP with limited resources
- Optimized some plots with Javascript
- Integrated functions of file querying and tagging, bad channels marking…
- Sequentially read available data to generate plots
  - Limited throughput



Simulation with BambooMC

- A modular MC simulation program base on Geant4 used in PandaX experiments
- Designed to be used as a simple framework for different experiments.
- Modular - It is possible to combine different geometry physics, generator, analysis in one program.

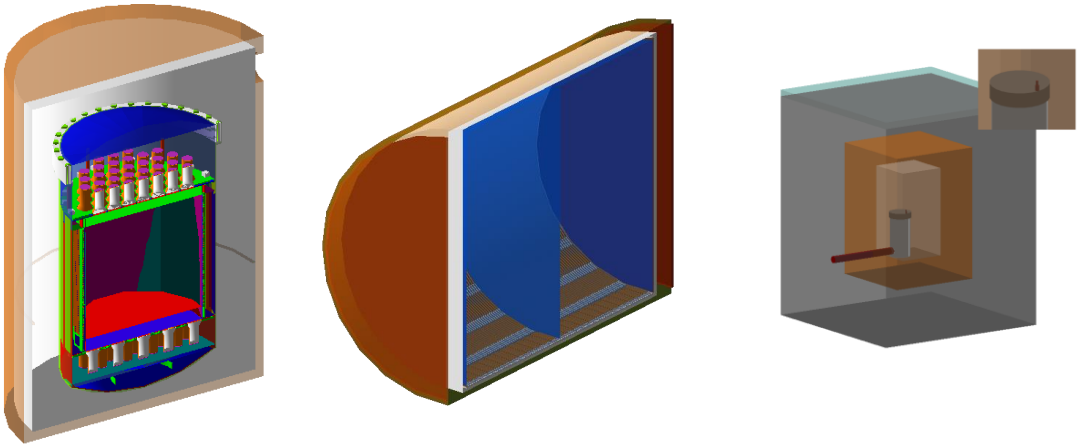


https://arxiv.org/abs/2107.05935

Open source in https://github.com/pandax-experiments/BambooMC

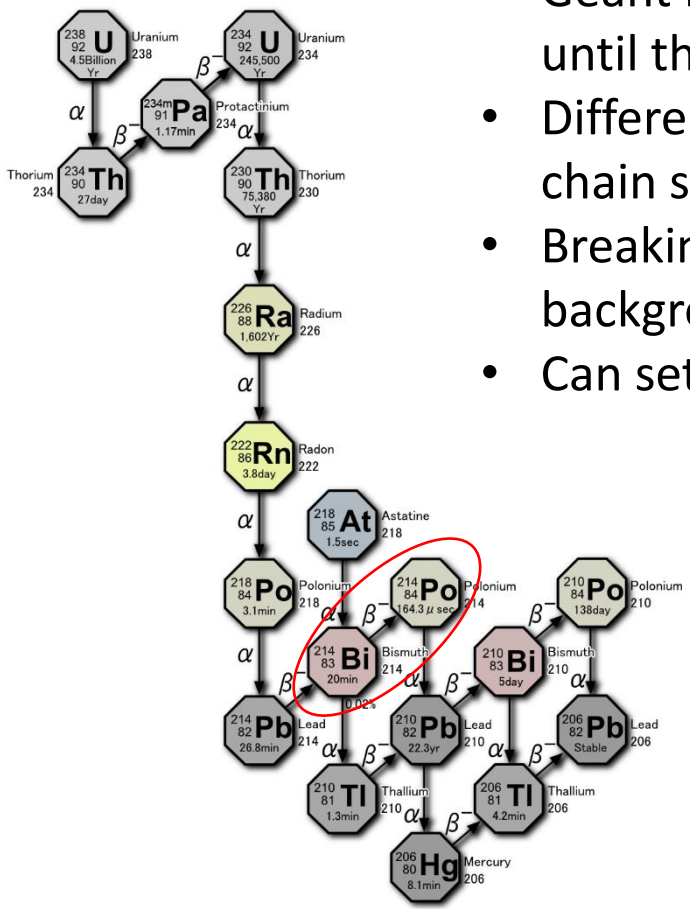
Modularized structure

- Free combination of detector geometry, physics, generator and analysis
  - Controlled by the config file
  - Detector hierarchy and reuse

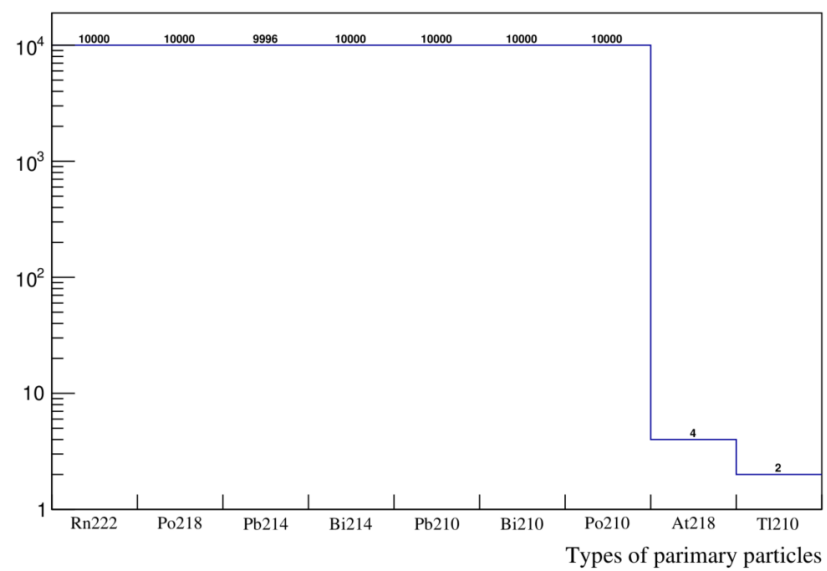


```yaml
run: 2312
geometry:
  material:
    name: SampleMaterial
  detectors:
  - type: SampleWorld
    name: World
    parameters:
      half_x: 2*m
      half_y: 2*m
      half_z: 2*m
  - type: BoxBlock
    name: VacuumBox
    parent: World
    parameters:
      half_x: 1*m
      half_y: 1*m
      half_z: 1*m
      material: vacuum
      sensitive: 2
      filter: neutron,gamma
  - type: BoxBlock
    name: PTFEBlock
    parent: VacuumBox
    parameters:
      half_x: 0.15*mm
      half_y: 0.15*mm
      half_z: 0.15*mm
      material: ptfe
      sensitive: 0
physics:
  name: PandaXPhysicsMod
  parameters:
    cutlength: 0.1*mm
generator:
  name: SimpleGPSGenerator
analysis:
  name: PandaXAnalysis
```

Breaking up of long decay chains



- Geant4 simulate the radioactive decay in a chain until the stable nuclei reached
- Different half-life of isotopes, the results from a chain should not be stored in the same event
- Breaking up long chain for the study of detector backgrounds
- Can set the threshold for the half life.



Summary and Outlook

- A high efficiency data process pipeline is built for PandaX-4T.
- A framework of simulation is created for various simulations in PandaX experiments.
- The data quality monitoring system need to be optimized.
- Another event builder system with higher efficiency is under development.