Event index based correlation analysis for the JUNO experiment

Weidong Li, **Tao Lin**, Yan Liu, Jiaheng Zou
(on behalf of the JUNO collaboration)

lintao@ihep.ac.cn
IHEP

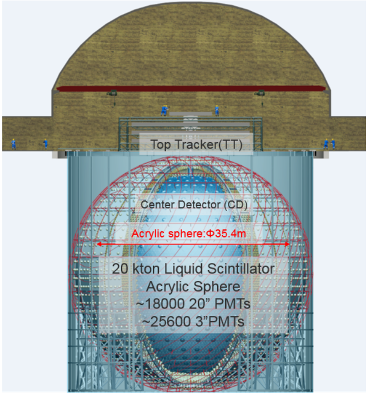
GRID 2021
July 6, 2021

Outline

1. Introduction
2. Design and Implementation
3. Performance
4. Summary and Plan

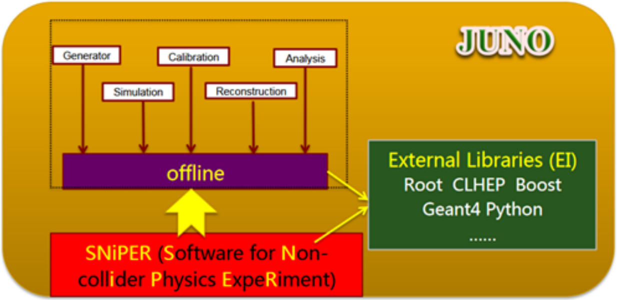
JUNO experiment

- The JUNO experiment under construction in southern China, will run for more than 20 years.
- It will have a rich physics programme;
  - Neutrino Mass Ordering and Precise measurement of 3 oscillation parameters.
  - Reactor neutrinos, Supernova burst neutrinos, Geo neutrinos, Atmospheric neutrinos, Solar neutrinos
- The JUNO detector is 700 m deep underground:
  - Central Detector, 20 kton LS with $3\%/\sqrt{E(MeV)}$ of energy resolution
  - Water Cherenkov detector
  - Top Tracker



Offline software

- The JUNO offline software consists of Generator, Simulation, Calibration, Reconstruction and Analysis tools.
- SNiPER is adopted as the underlying data processing framework.
    - SNiPER provides the ability to manage multiple events in the same time window, which is one of important features in the analysis.



Challenges in the analysis

Rare signals:

- The total event rate is about 1 kHz.
- The event rate of reactor antineutrinos is 60 per day.
- Most of the events are background.

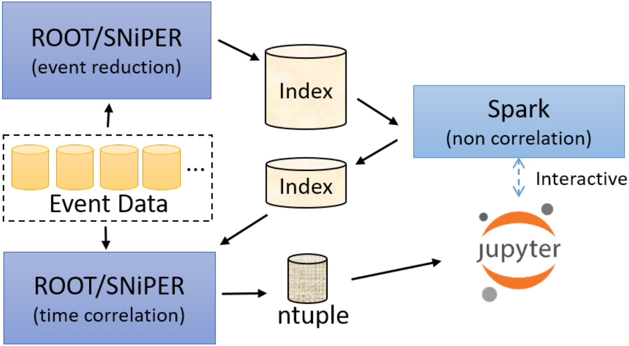
Time correlation:

- A neutrino is detected via the IBD process, coincidence in time, space and energy:
  - IBD: $\bar{\nu}_e + p \rightarrow e^+ + n$
  - A prompt signal e^+.
  - A delayed signal neutron with an average capture time of $\sim$200 us.
- Need the muon vetos to reject the possible backgrounds.

Due to the time correlation, difficult to use the big data technologies.

Design of the event index based method

- In order to speed up the time correlation analysis, an event index based method is designed.
  - Generation of index data: generate the index data with necessary information by processing the full event data.
  - Single event analysis: use Spark to do the event selection by analyzing the index data.
  - Time correlation analysis: analyze the events within the same time window.

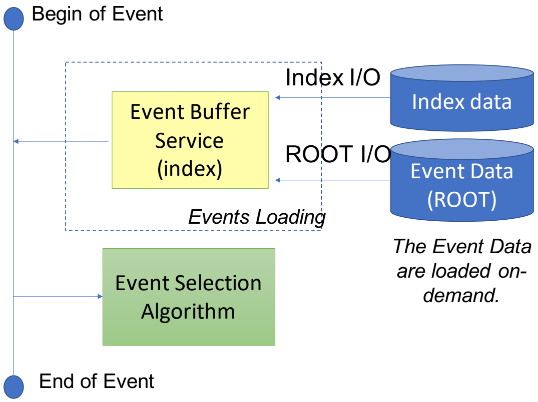


Design of the index data

- The index data contains two parts.
- The first part is the reference to the event data:
  - `fileID`: the reference to the event data file.
  - `entryID`: the reference to the entry in the data file.
- The second part is the necessary event level information used for single event selection, which could be extended by the users:
  - `Energy`: the reconstructed energy.
  - `Vertex`: the reconstructed vertex.
  - `Time`: the event time.
- The index event could be analyzed by the big data technologies, such as Spark.
- The file formats of the index data could be in the plain text file, the ROOT format file, HDF5 etc.

Design of the correlation analysis framework (1)

- This analysis framework is developed based on the SNiPER framework.
- The event loop is driven by the index data, instead of the event data.



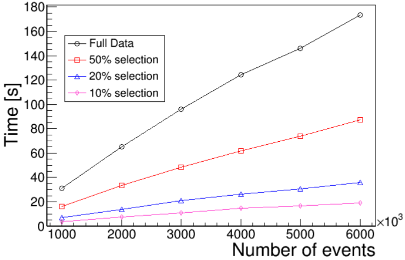
Design of the correlation analysis framework (2)

Detail flow:

- An "index" is loaded via the Index I/O.
- The `fileID` and `entryID` is retrieved from the "index".
- According to the `fileID`, check whether the file is loaded or not. If not loaded, need to open the corresponding file.
- According to the `entryID`, load the corresponding event and get the timestamp of the current event.
- Load the other events within the same time window and pack the events into an object called `EventWindow`.
  - The events at the boundary of the files need to be handled properly.
- The analysis algorithms process the events in the time window.

Performance of the event loading

- Measure the event loading with different ratios.
  - Randomize the event selection according to the ratios.
  - Only the selected events are loaded, the events in the time window are not loaded.
  - Repeat 30 times to reduce the errors.
- Result: the index data could speed up the event loading by reducing the I/O.
  - The overhead: from the index I/O



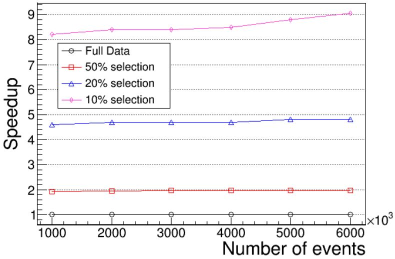


Figure: Left is the time to load the events with different ratios. Right is the speedup.

Performance of the correlation analysis (1)

Dataset and selection criteria

- Radioactivity background samples in liquid scintillator (LS) are generated to mimic the IBDs.
  - All the isotopes in U238/Th232/K40 decay chains are considered.
  - The intervals between two events are sampled according to the rates.
  - The positions are sampled in the LS uniformly.
- The selection of single events
  - Fiducial volume cut
  - Energy cut (delayed signals)
- The selection of correlated events
  - Energy cut (prompt signals)
  - Time interval cut
  - Distance cut

Performance of the correlation analysis (2)

- Compared to the normal analysis, there is about 14-fold speedup.
  - About 2.5% of events are selected in the index data.
  - There are about 5% of events loaded from the event data.

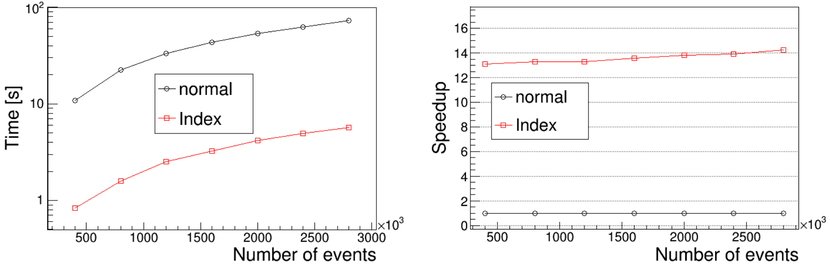


Figure: Left is the time. Right is the speedup.

Summary and Plan

Summary

- An index based method is implemented to speed up the correlation analysis of rare events in JUNO.
- By reducing the I/O, the index data could improve the analysis speed.
- The speedup of correlation analysis is about 14 when 5% of events are really loaded.

Plan

- In order to further speed up the analysis, the parallelized version is under development.

Thank you!