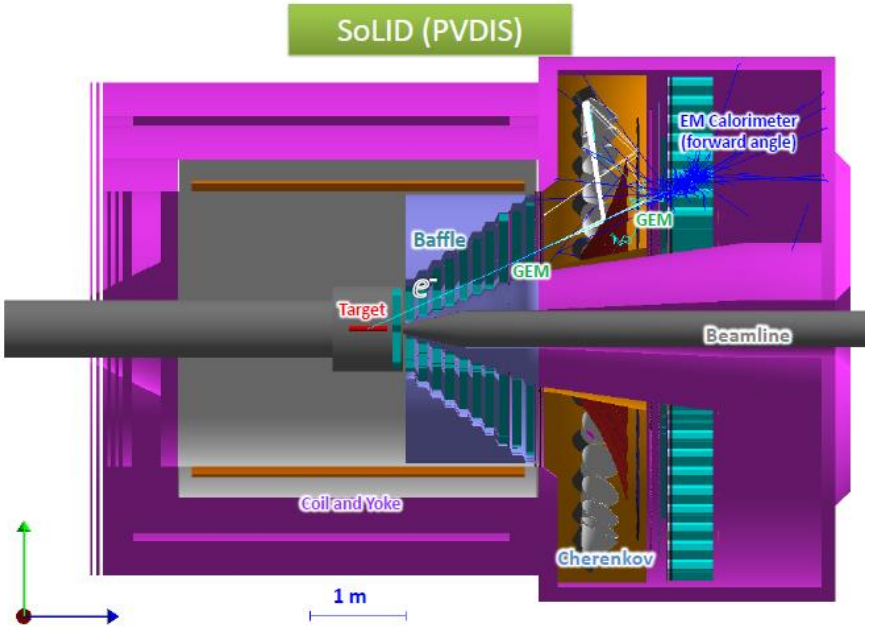
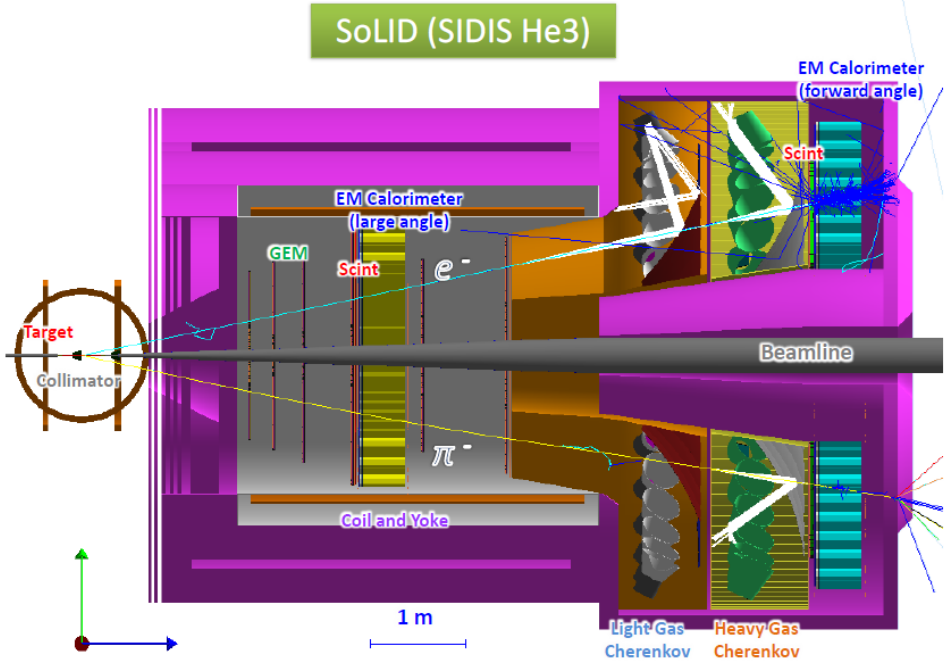
SoLID detector overview

- GEM for tracking (offline in analysis)
- PID detectors (both offline in analysis and online in trigger)
  - Cherenkov: LGC to identify e and reject pion, HGC to identify pion and reject kaon
  - EM Calorimeter: to identify e and reject pion and other hadrons
  - Scintillator Pad detector: photon rejection and time of flight with tracking
  - MRPC: time of flight with tracking
- Particle and 4momentmum from combining all detectors
- Common challenge
  - High rate and high background

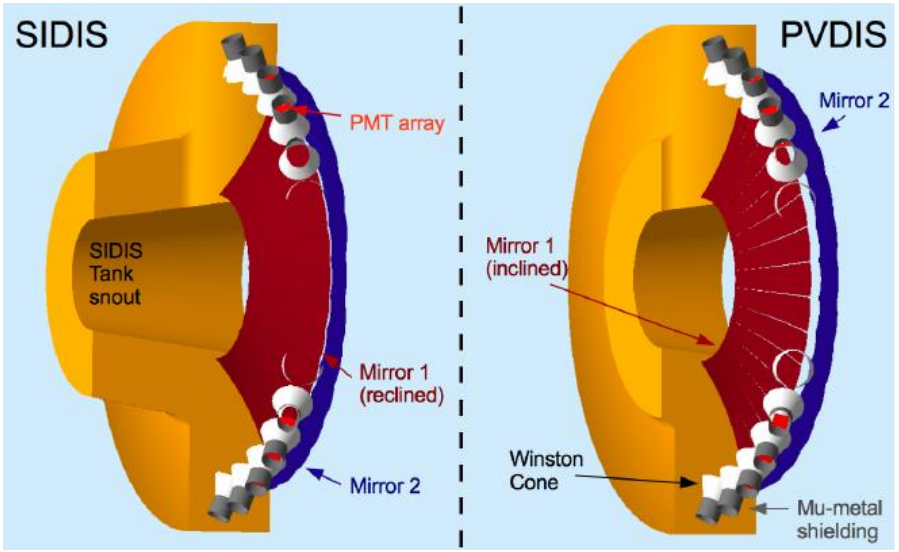


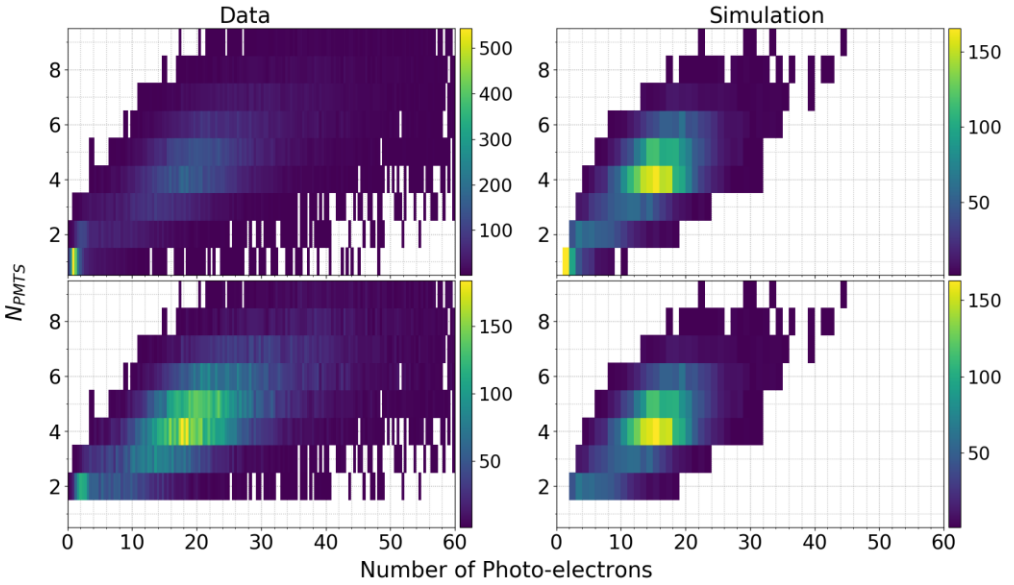


Cherenkov

LGC:

- Threshold detector: identify e and reject pion
- 30 sectors of 3x3 MAPMT array
- Background rate 2MHz/MAPMT
- Online Trigger:
  - Combine with EC
  - Maintain high e efficiency while suppress background
  - Tentative condition: 2Npe of 2MAPMT (as Cherenkov ring of high energy e should distribute more in space)





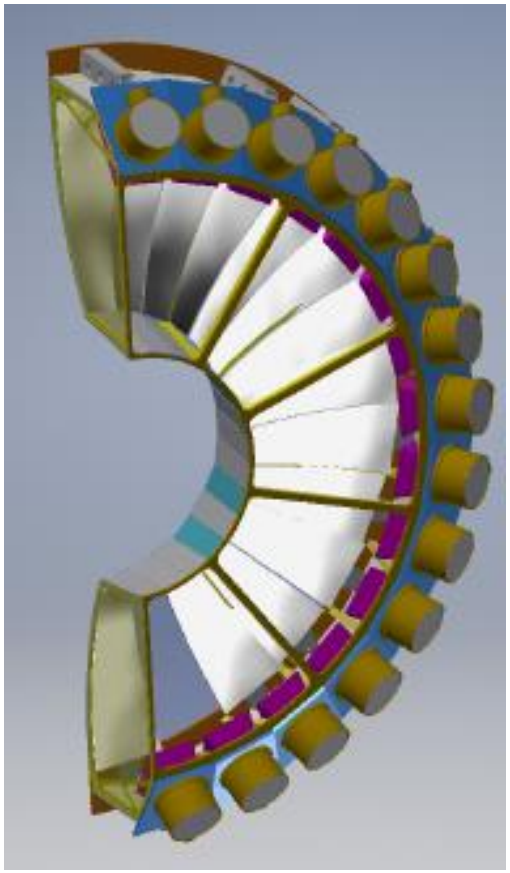
For online trigger, can AI/ML help with better trigger condition for higher e efficiency while suppress more background?
Limitation from running on trigger hardware in DAQ at combined trigger rate ~100kHz?

High rate beam test of trigger cut on 2MAPMT

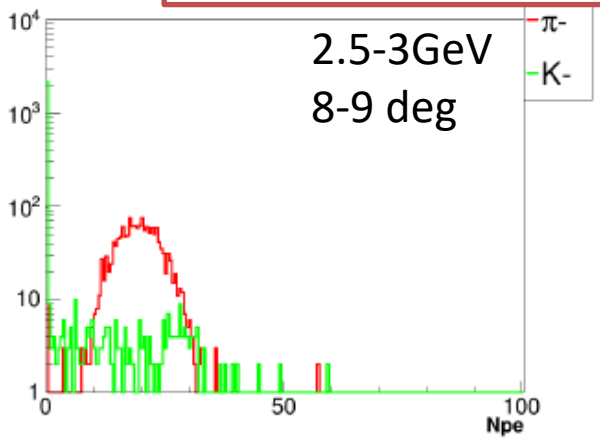
Cherenkov

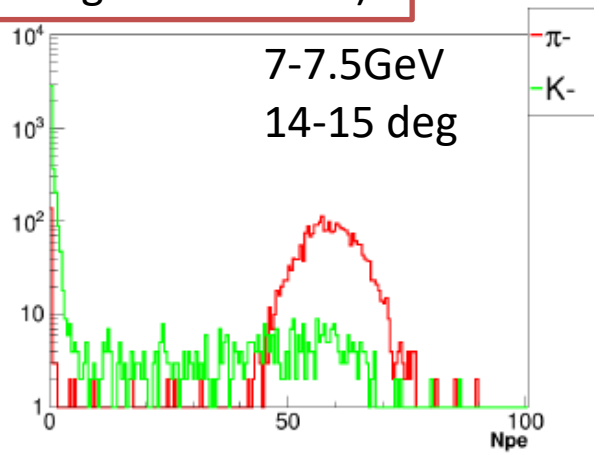
HGC:

- Threshold detector: identify pi and reject kaon
- 30 sectors of 4x4 MAPMT array
- Background rate 4MHz/MAPMT
- Not in trigger
- More difficult than LGC in offline analysis
  - Npe and ring size have strong angle and momentum dependance (combine with tracking info)
  - Kaon decay 10-30% into pi and muon which will have Cherenkov light like pi
  - Higher background (within 50ns, each sector has 3Npe from background and minimum 10Npe from signal)



Sim of pi and K (no other background added)

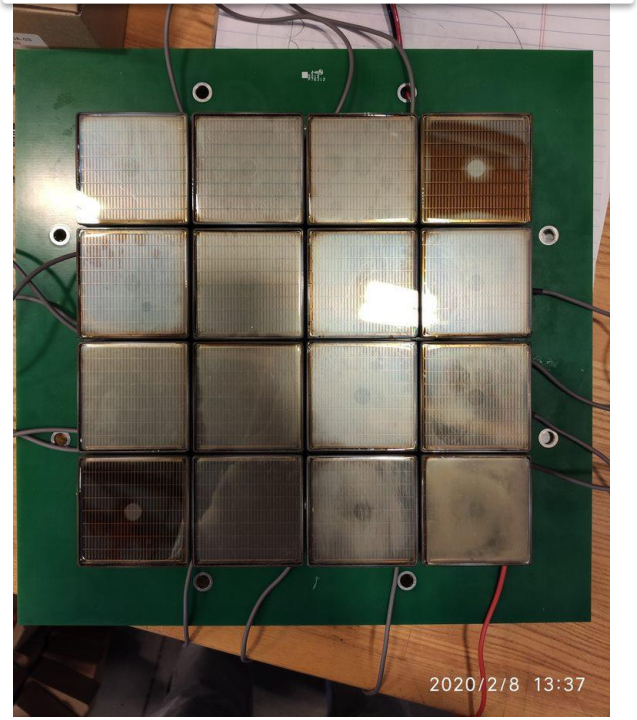




For offline analysis, can AI/ML help with better signal particle identification while suppress more background by using spatial information?

Cherenkov

4x4 MAPMT array



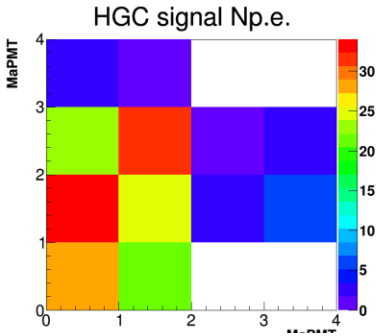
Readout:

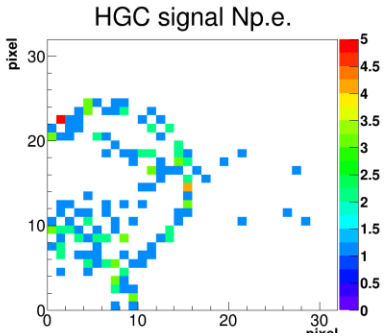
each 5cmx5cm MAPMT can have

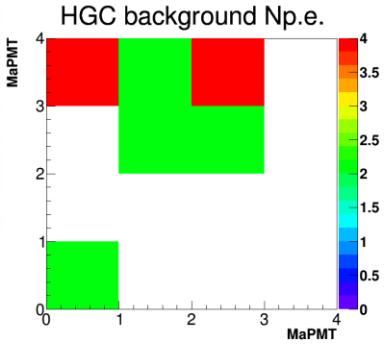
- 1 pmt sum output with Npe
- 4 2.5cmx2.5cm quad sum output with Npe
- 64 6mmx6mm pixels with 0/1

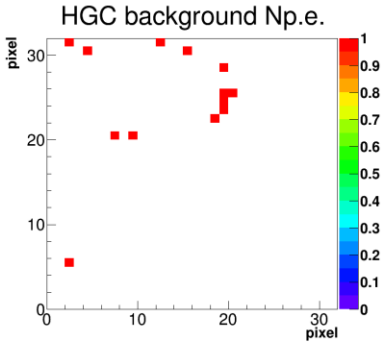
Cost increase with more readout channels and pixel readout is expensive

Use AI/ML to understand what level of readout is really needed

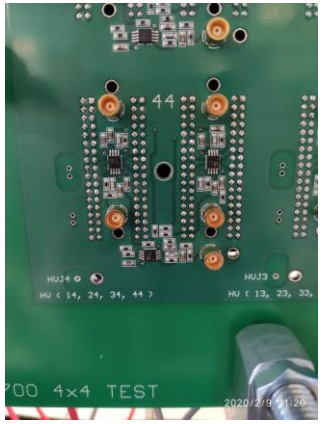




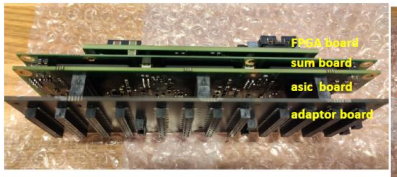




Simple sum readout (total+quad)



MAROC sum readout (total+quad+pixel)



MAROC sum boards

General thought

- Cherenkov may use AI/ML algorithm for image processing
- EC may use a different algorithm
- How to combine those two?
- Is it even possible to use one algorithm combine all detectors including tracking and PID?

SoLID	Tracking	Cherenkov	EC
1. What are we trying to do? Articulate the objectives of the 3 efforts. * Including the figure of merit	improve the performance of GEM clustering improve the performance of tracking reconstruction	Improve Cherenkov PID beyond simple Npe cut. For HGC with background , efficiency (> 90%) and rejection (>10) Improve LGC with trigger design	Improve EC PID performance with background. We want to keep pion rejection > (50:1) with electron efficiency>90%.
2. Explain what is done today, and what are the limits of current practice? (baseline)	Not much	Not much, start to explore AI simple Npe cut performance degrade with high background	Not much. the traditional cuts couldn't keep the pion rejection as high due to energy leak at edge
3. If we are successful, what difference will it make?	a few times improvement on the speed and around 10% improvement on the tracking reconstruction efficiency and accuracy. GEM clustering will benefit SBS also	Improve Cherekov performance baseline at high background Help with readout choice to determine if pixel/quad/sum are needed	significantly improve the ECAL PID performance at the edges of EC
4. Data available (raw and simulated) * File format (root?) * Data format and variable summary (tabular?) * Data size (number of samples?) * Where is the data located? When can we have access?	Unlimited simulation data in root or text format available on ifarm as soon as we agree on a format	Unlimited simulation data in root or text format Both low rate and high rate data from HallC test (~10 thousands events) Cosmic with background data from bench (~thousands events) available on ifarm as soon as we agree on a format	Unlimited simulation data in root or text format Some low rate real data from Fermi lab test (~thousands events) available on ifarm as soon as we agree on a format
5. Timeline? * Publications/conferences?	Not sure	Working on note/short paper about readout aiming for next year. AI would be a nice part of it or a separated paper	Not sure
6. Who is available to work on this with the data science dept.?	Weizhi Xiong until Feb, someone else afterwards	Zhiwen Zhao, Bo Yu, Michael Paolone	Ye Tian, Zhenyu Ye

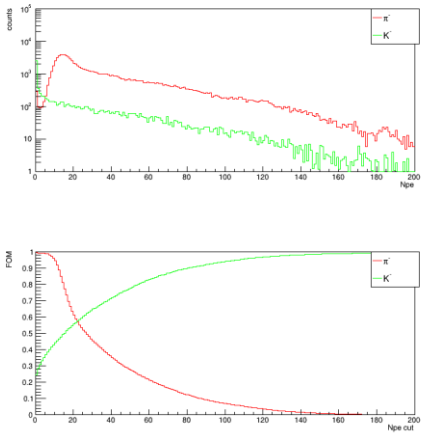
HGC FOM

- HGC performance can be judged by the following figure of merit:

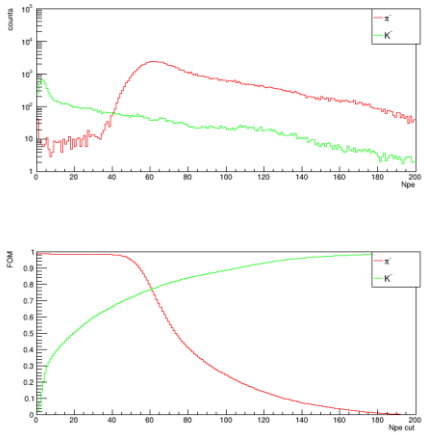
1. FOM pion:
   efficiency =(Nevent of >Npe)/Ntotal
2. FOM kaon:
   1-1/rejection = (Nevent of <Npe)/Ntotal

- Hope we have both FOM reach 0.9 at the same time

P=2.5GeV, Theta=8deg



P=7.5GeV, Theta=14.5deg



FOM	P=2.5GeV, Theta=8deg	P=7.5GeV, Theta=14.8deg
No background	0.93	0.92
“1 sector” background	0.77	0.86
“2 sector” background	0.64	0.80
“3 sector” background	0.57	0.78

Go beyond Npe cut to use location info with AI/ML

HGC data format

- Basic info
  - Event recorded in 3 sectors (for solid sim), each sector has 16 pmt in 4x4 array, each pmt can have readout by pmt/quad/pixel
  - 1 event is made of N=48/192/3072 numbers and each number is the Npe in this pmt/quad/pixel for this event
  - Files are listed for pion and kaon, for different momentum and angle, for with/without background, and for different readout
- File format
  - Text row: 1 event in 1 row
    - row has columns ordered in with readout sensor ID
  - Text matrix: 1 event in 1 matrix
    - matrix has spatial info:
      - Fake spatial: 4*12/8*24/32*96
      - Real spatial: 3 sector in real physical space without many additional 0
  - Any other format without many 0 to reduce file size?