SoLID HGC Update

Zhiwen Zhao
for HGC group

2022/05

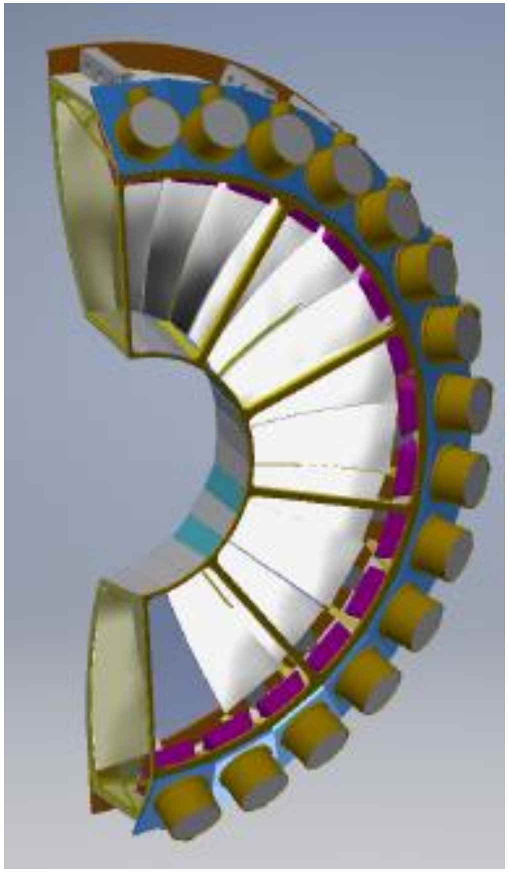
Outline

- HGC PID with AI/ML
- Regina funding application
- SBU mirror coating

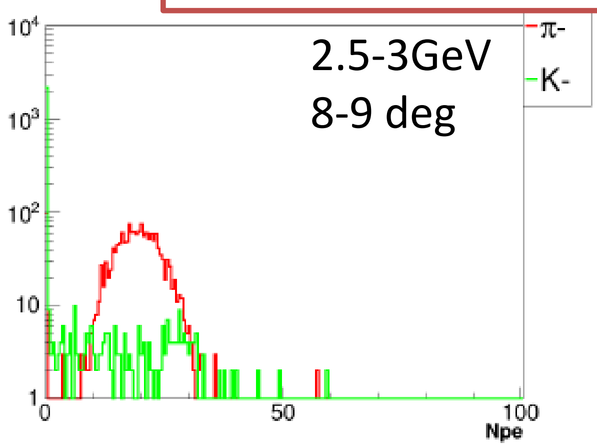
Cherenkov

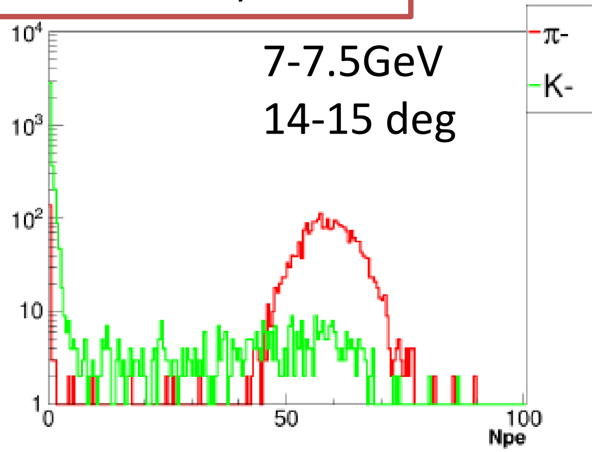
HGC:

- 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 pion
  - Higher background (within 50ns, each sector has 3Npe from background and minimum 10Npe from signal)



Sim of pi and K (no background added)





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

HGC Npe (Number of photoelectron)

- N of pe determined by z,p,theta at vertex
- distribution of pe determined by z,p,theta and phi at vertex

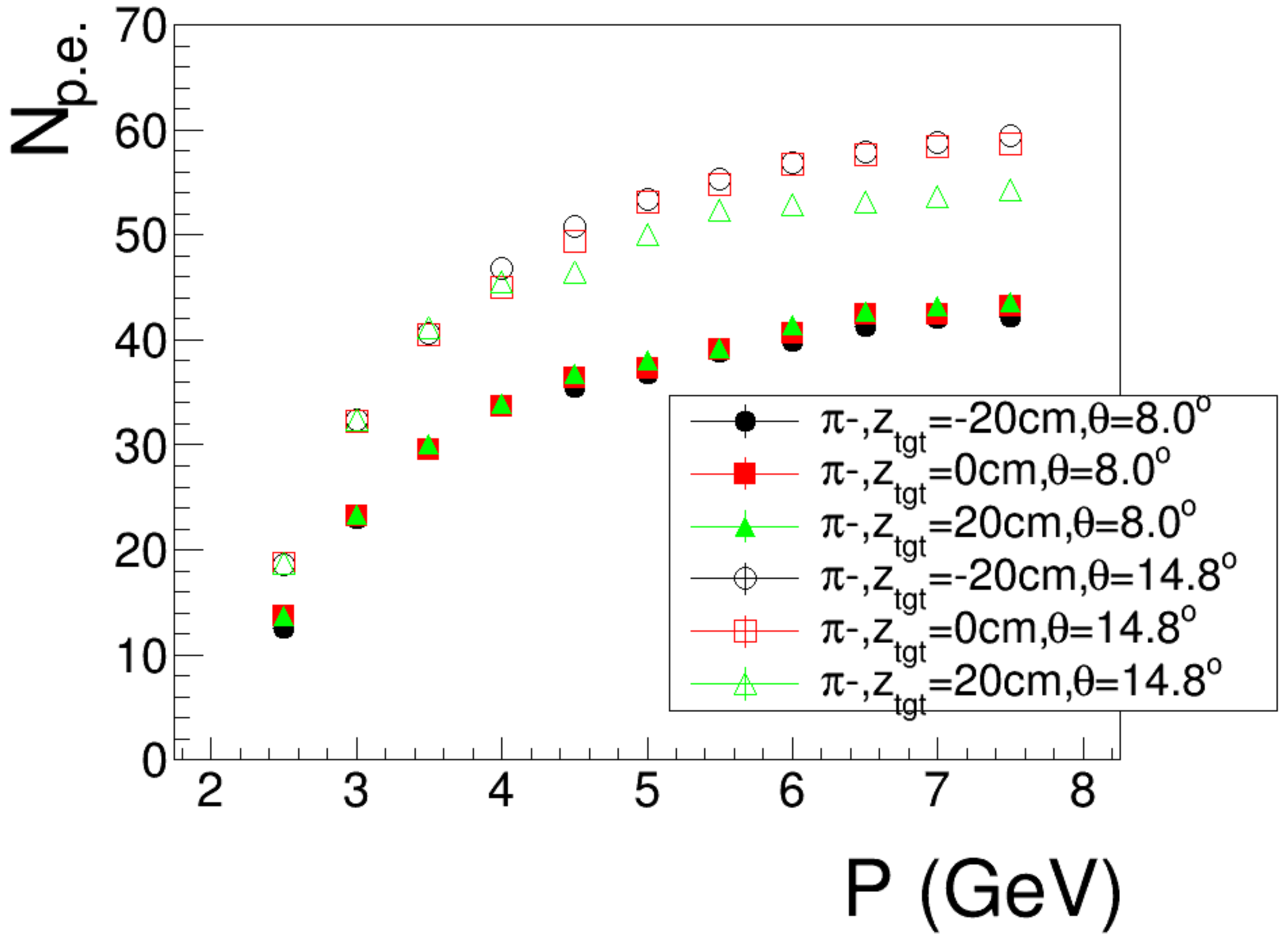
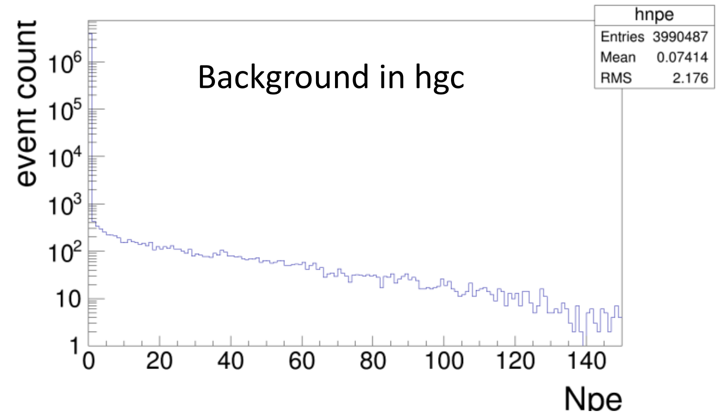


Figure 83: Simulated number of photoelectrons of negative pions as a function of momentum at various polar angles and target vertex positions. A very similar output is obtained for positive pions.

HGC background mixing

- Pion from target center z=-350cm
  at fixed angle and mom with 0.5 sim safety factor
- kaon from target center z=-350cm
  at fixed angle and mom without sim safety factor



- Background from “beamontarget” (without sim safety factor)
  - File “/cache/halla/solid/sim/solid_gemc/SIDIS_He3_JLAB_VERSION_1.3/pass8/farm_solid_SIDIS_He3_moved_BeamOnTarget_0.561e10_skim_HGCwinCF1.root”
  - SoLID SIDIS He3 run use 15uA beam, so there 15e-6/1.6e-19*50e-9=4.7e6 e- within 50ns time window
  - this skim file has 0.561e10 beam e-. It is 0.561e10/4.7e6=1194 of 50ns time window
  - This file has 3990487 not-empty-anydetector events and ~9000 not-empty-hgc event. So each 50ns time window, there are 9000/1194=7.5 events in hgc
  - If only mixing Npe and considering background is symmetric for 30 sectors
    - In each sector, 50ns time window has 7.5/30=0.25 events in hgc
    - 1 HGC signal events should be in 1 or 2 sectors, but to know which 2 sector, we need to tracking info. So we can simply consider 3 neighboring sectors around the sector with highest Npe
    - In 1/2/3 sectors, 50ns time window has 0.25/0.5/0.75 events in hgc

HGC only

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

Decay from target to HGC
- 2.5GeV pion 4.6% kaon 30%
- 7.5GeV pion 1.6% kaon 11.5%
- Evenly kinematics pion 2.7% kaon 18%
- at most 1% decay within target and hope tracking can exclude those

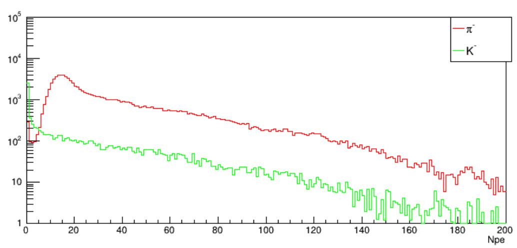
FOM	P=2.5GeV, Theta=8deg alltrack	P=2.5GeV, Theta=8deg Nodecaytrack	P=7.5GeV, Theta=14.5deg alltrack	P=7.5GeV, Theta=14.5deg nodecaytrack
No background	0.93	0.99	0.92	0.98
“3 sector” background	0.57	0.60	0.78	0.80
“3 sector double” background	0.52	0.53	0.64	0.66

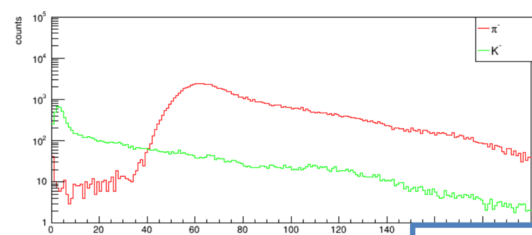
P=2.5GeV, Theta=8deg alltrack

P=7.5GeV, Theta=14.5deg alltrack

“3 sector” background

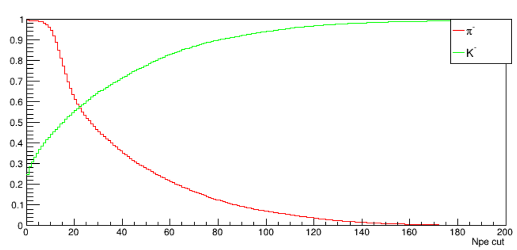
count

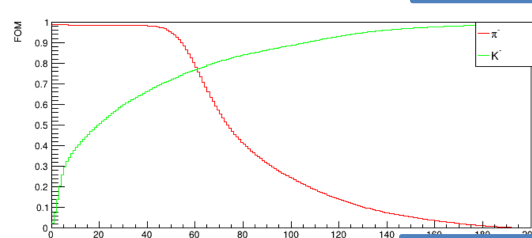




Npe

FOM





Npe cut

alltrack and nodecaytrack seem having similar FOM for Npe cut with background

Go beyond Npe cut to use sensor location info with traditional ray tracing or AI/ML

HGC data format

/group/solid/www/solid/html/files/AIML/solid_hgc_sim/

https://solid.jlab.org/files/AIML/solid_hgc_sim/

background3sector/solid_SIDIS_He3_hgc_moved_pim_1e5_row_pixel.csv.zip

	0	1	2	3	4	...	3074	3075	3076	3077	3078
0	0	0	0	0	0	...	4934.06	-957.444	773.560	3130.01	1
1	0	0	0	0	0	...	6330.81	858.745	1258.640	3130.01	1
2	0	0	0	0	0	...	5621.80	-1550.810	110.266	3130.01	1
3	0	0	0	0	0	...	5706.73	-1519.000	345.001	3130.01	1
4	0	0	0	0	0	...	4574.93	1088.980	-1216.480	3130.01	1
...	...	...	...	...	...	...	...	...	...	...	...
99995	0	0	0	0	0	...	7399.62	-909.056	532.744	3130.01	1
99996	0	0	0	0	0	...	5497.90	1309.600	-455.812	3130.01	1
99997	0	0	0	0	0	...	6662.41	351.042	-814.176	3130.01	1
99998	0	0	0	0	0	...	5970.38	-725.987	724.503	3130.01	1
99999	0	0	0	0	0	...	5189.04	-350.627	-1545.720	3130.01	1

[100000 rows x 3

background3sector/solid_SIDIS_He3_hgc_moved_km_1e5_row_pixel.csv.zip

	0	1	2	3	4	...	3074	3075	3076	3077	3078
0	0	0	0	0	0	...	4106.00	277.924	-846.833	3130.01	0
1	0	0	0	0	0	...	0.00	0.000	0.000	0.00	0
2	0	0	0	0	0	...	3281.10	-1390.420	794.323	3130.01	0
3	0	0	0	0	0	...	4942.21	1325.020	-283.395	3130.01	0
4	0	0	0	0	0	...	5308.49	1451.740	-904.008	3130.01	0
...	...	...	...	...	...	...	...	...	...	...	...
99995	0	0	0	0	0	...	4665.86	-559.912	-1062.220	3130.01	0
99996	0	0	0	0	0	...	0.00	0.000	0.000	0.00	0
99997	0	0	0	0	0	...	4329.50	-805.390	-996.263	3130.01	0
99998	0	0	0	0	0	...	5739.35	-673.397	751.146	3130.01	0
99999	0	0	0	0	0	...	3088.49	1377.370	-653.491	3130.01	0

[100000 rows x 3079 columns]

ML code on google colab

1. “train_pid_solid.ipynb” row+track

https://colab.research.google.com/drive/13Y18LYnazxFZfu_nABrsn3gDZS6mF8Ga?usp=sharing

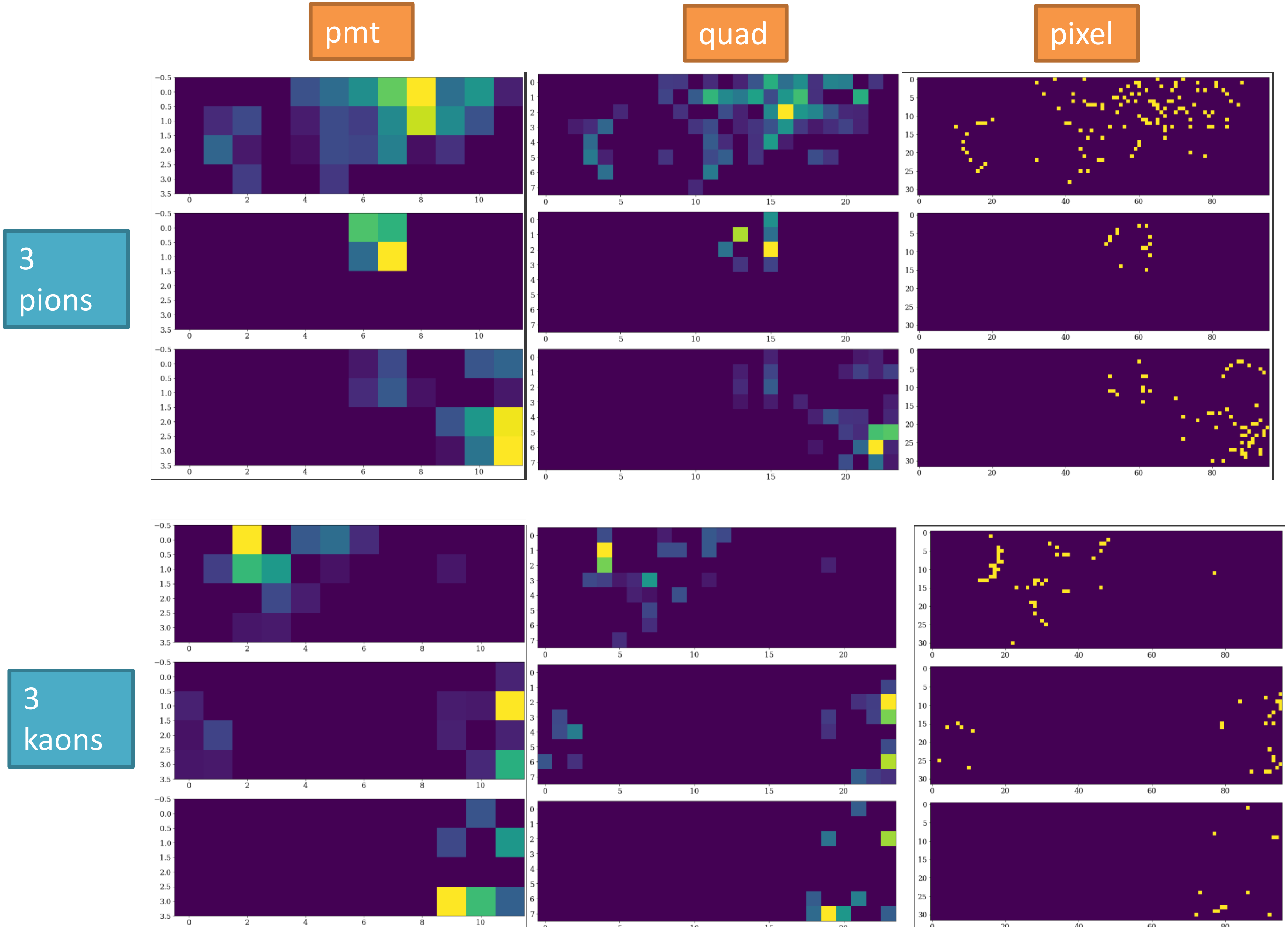
2. “TrainSoLID_PID.ipynb” image+track

https://colab.research.google.com/drive/1AIBIrOgJloSpwV2v3qcGtbKvnC5z2ZlM?usp=sharing

Help from data science group:
Kishansingh Rajput
Malachi Schram

3 sector event view

Nodecaytrack, full, Bg_3s*2
z350_p2.5_theta8.0

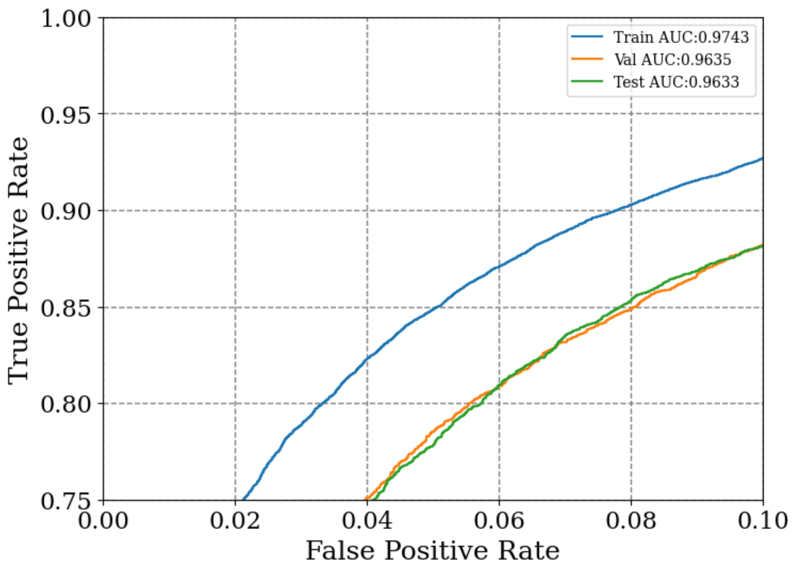


Nodecaytrack
image+track

ROC (pion)

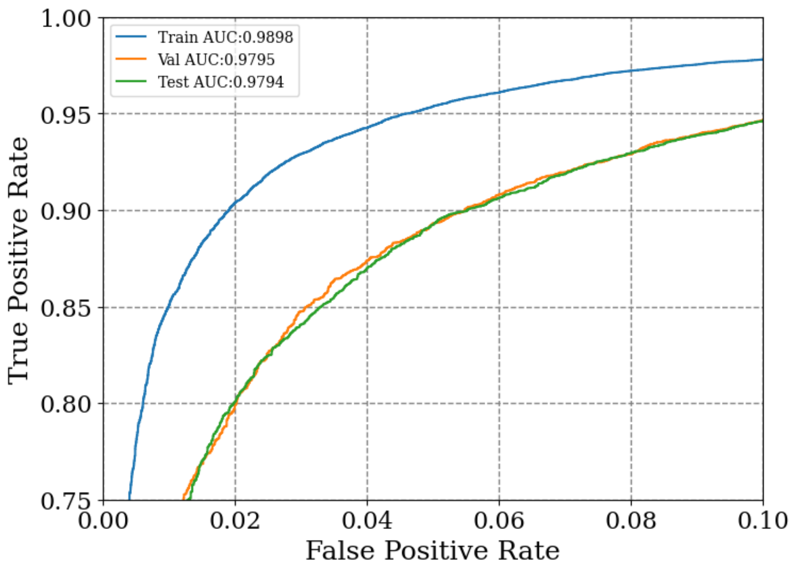
- An **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds
- ROC error is at 0.01 level from data science group initial study

pmt



ROC_full_z350_p2.5_theta8.0_background3sectordouble_48_6_pion.png

quad



ROC_full_z350_p2.5_theta8.0_background3sectordouble_192_6_pion

FOM results

Nodecaytrack
image+track

		Bg_no		Bg_3s		Bg_3s*2	
z350_p2.5_theta8.0	pmt	0.996	0.996	0.965	0.920	0.900	0.770
	quad	0.996	0.996	0.975	0.960	0.950	0.880
z350_p7.5_theta14.5	pmt	0.999	0.998	0.996	0.995	0.994	0.991
	quad	0.999	0.998	0.998	0.996	0.996	0.994

- FOM is chosen as true pion eff at false kaon rate = 0.05
- Red is hgc only simulation, blue is full simulation
- Each model is individually trained and FOM obtained from test data
- More background can reduce FOM, smaller sensor size can increase FOM
- small angle low P is more difficult than large angle high P
- Pixel result is not shown as it require more data and cpu and mem to train, but the improvement is expected to be relatively small

Q&A with data science group

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

More AI/ML study

- LGC can do it similar to HGC
- EC may take a different model
- Develop a model to train LGC&EC for PID
- Do GEM tracking
- Combine PID and tracking?
- Detector optimization

Potential Canadian Funds for SoLID HGC

- **CFI Innovation Fund (IF)** funds research infrastructure in Canada. There is a ~C$400 million competition every two years, covering all disciplines.
- SoLID HGC vessel cost is ~C$1 million, based on pCDR budget. U.Regina VP-Research has agreed to support an application to CFI-IF for SoLID HGC Vessel for C$509.5k.
- If approved, U.Regina would ask the Province of Saskatchewan to match this amount, to fully cover HGC vessel cost estimate in pCDR.
- This could reduce pressure on funds provided by US-DOE, and may allow some other de-scoped SoLID component to go forward.
- **Updated 2023 CFI-IF Competition Deadlines:**
  - Notice of intent (submitted): February 2022
  - Proposal due: July 2022
  - Decisions announced: June 2023
  - Decision for SK match: Dec 2023
- **Funds will be contingent on US-DOE Critical Decisions, same as MOLLER CFI-IF funds that were awarded in 2021 competition**

Potential Canadian Funds for SoLID HGC

Progress to Date:

- Draft of full CFI-IF proposal is in progress (25 pages)
- 1st draft planned to be circulated next ~week to applicants and JP Chen
- Lead Applicant: Garth Huber (Regina)
- Co-Applicants:
  - Klaus Dehmelt (Stony Brook)
  - Abhay Deshpande (Stony Brook)
  - Haiyan Gao (Duke)
  - Michael Paolone (NMSU)
  - Nikos Sparveris (Temple)
  - Aram Teymurazyan (Regina)
  - Zhiwen Zhao (Duke)

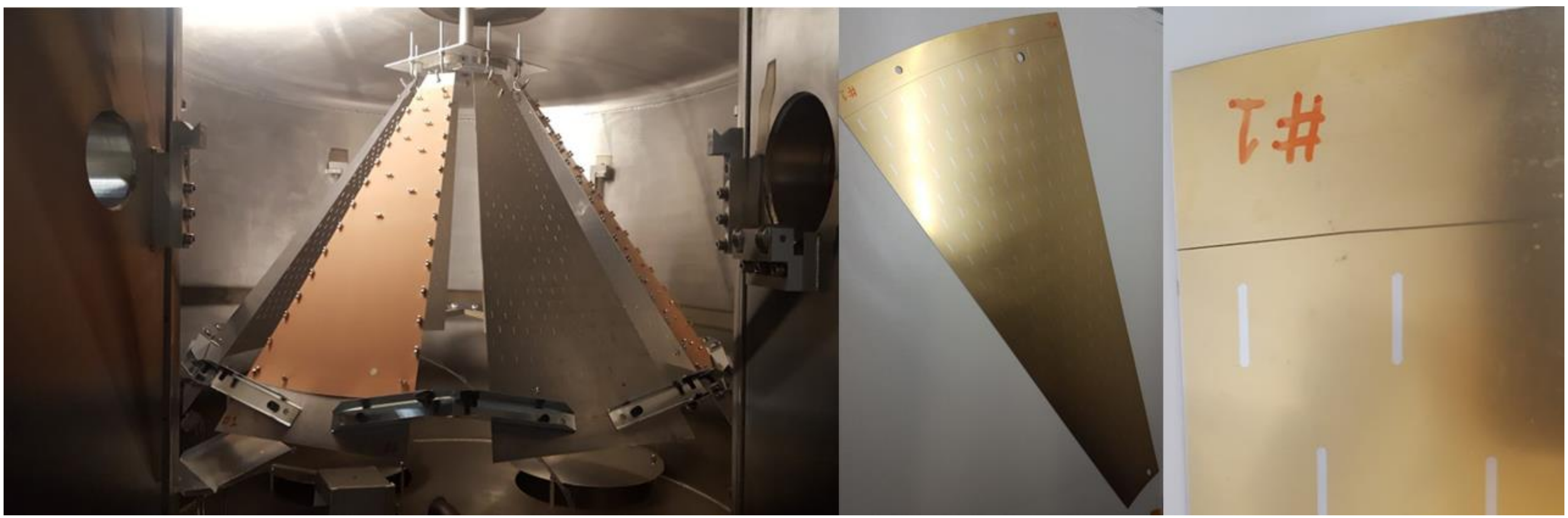
Equity, Diversity and Inclusion

- **An important part of the CFI-IF application is the EDI plan.** We have been working with Aurora Realin (JLab DEI Officer) and Pauline Streete (Regina EDI Officer). Once the SoLID Collaboration is fully constituted, it will be essential that we have a EDI Committee (as GlueX does)

Evaporator @ sbu

Klaus Dehmelt

- Evaporator w/ Physical Vapor Deposition device (PVD) commissioned
  - TPC related project → central membrane
- PVD operated → will be prepared for mirror coating



Hope to coat the small SoLID mirror samples around this fall, before coating larger pieces