SBS GEP Experiment Event/Track Reconstruction

Anuruddha Rathnayake

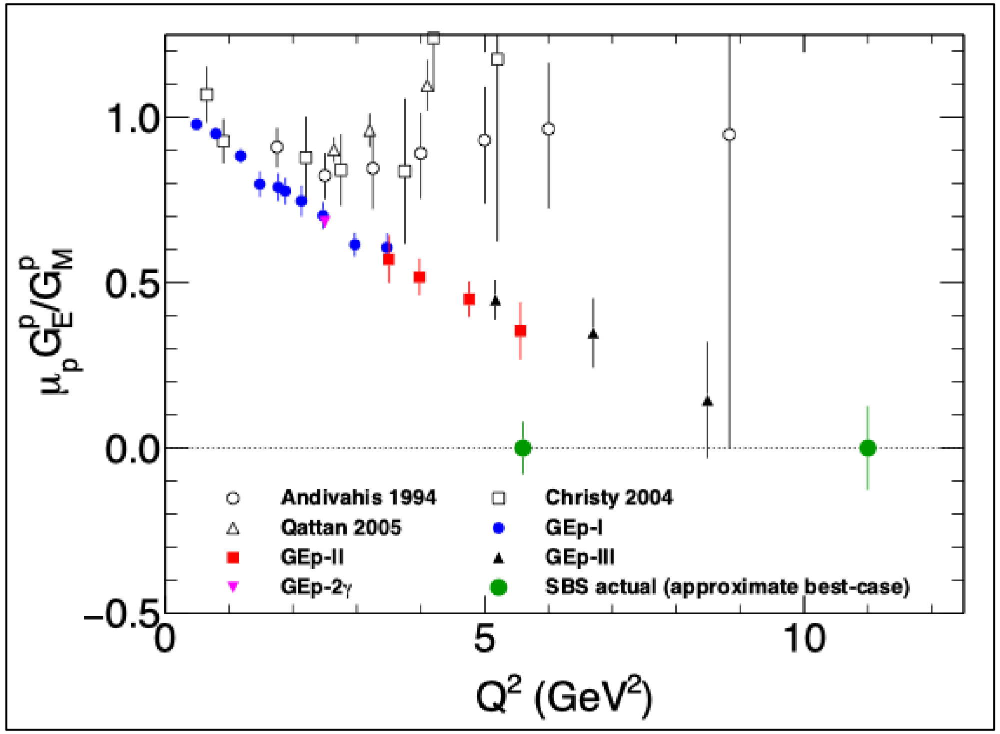
DNP Fall Meeting 2025, Chicago

Acknowledgements

- UConn group members: Andrew, Kip, and Sarah
- ALL the SBS collaborators
- JLab scientific and technical staff
- This work supported in part by DOE Office of Science, Office of Nuclear Physics, award DE-SC0021200

Motivation to Improve GEP Event-Reconstruction

- We have collected approx. 3.8 C and 94 C of data for kin1($5.7\ \mathrm{GeV}^2$) and kin3($11.1\ \mathrm{GeV}^2$), respectively
- Assuming overall trigger/detection/reconstruction efficiency of 70% (w/o radiative losses), we will have following error bounds as a best-case scenario
- But currently we are well below that, the <u>first major analysis task for GEP is to improve reconstruction</u>



Plot: A. Puckett

Analysis Machinery for GEP and other SBS Experiments

".evio" Raw data files from CODA

Event Reconstruction

SBS_OFFLINE
- Built "on" Hall-A analyzer / Podd
- Inherits from a lot of base classes in Podd but includes extensive additions
- Decoding
- Detector level reconstruction
- Spectrometer level reconstruction
- Physics variables for further analysis

SBS_REPLAY
- Data Bases to store detector / spectrometer specific information like channel maps
- Set replay parameters/thresholds
- Definitions of outputs from replay
- Replay scripts that use SBS_OFFLINE machinery to steer analysis

ROOT files with output *TTree* branches for interactive physics analysis

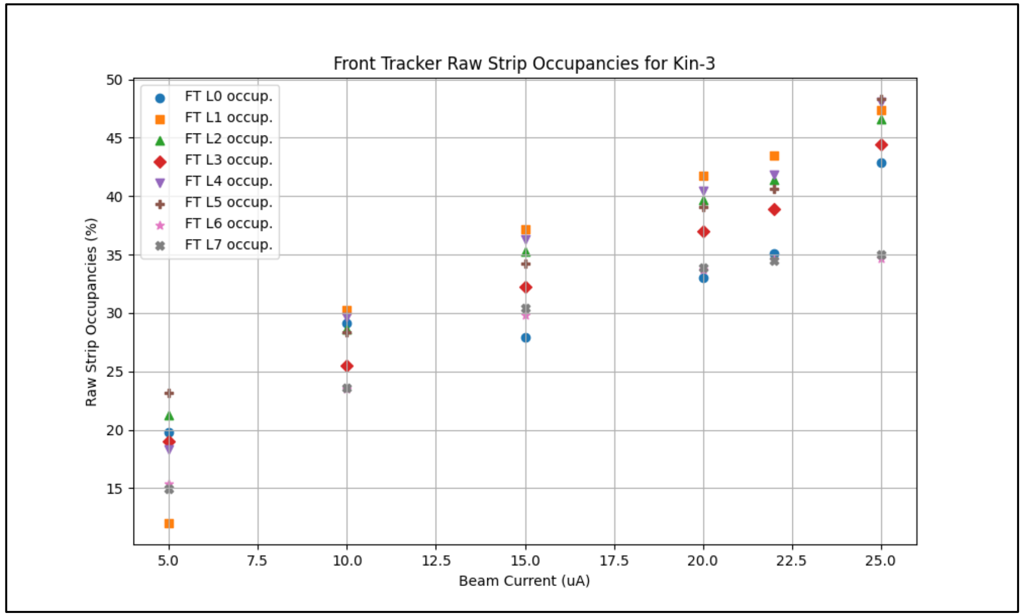
G4SBS
- *Geant4* based simulation for SBS experiments
- Elastic/inelastic/gun event generation
- Detector/spectrometer response simulation
- Returns energy depositions

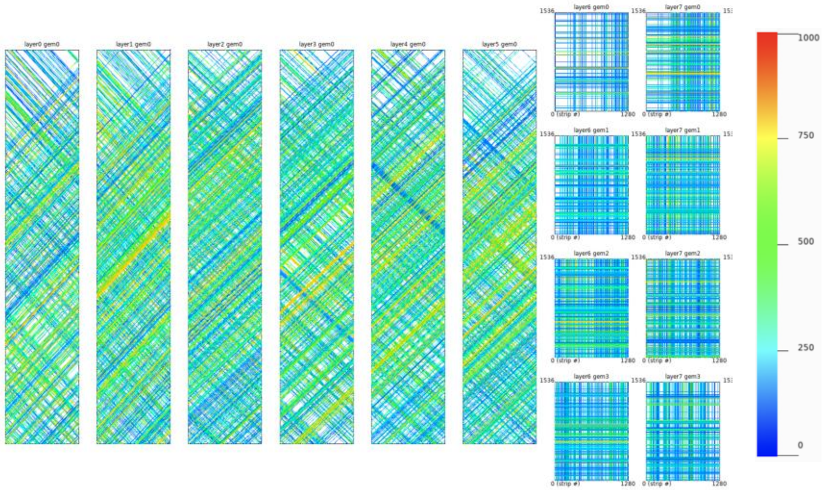
LIBSBSDIG
- Library for digitization of simulation output
- Converts g4sbs time, position, energy deposit etc., into ADC, TDC, crate, slot, channel, etc.

UCONN

GEP Experiment Event Reconstruction Challenges

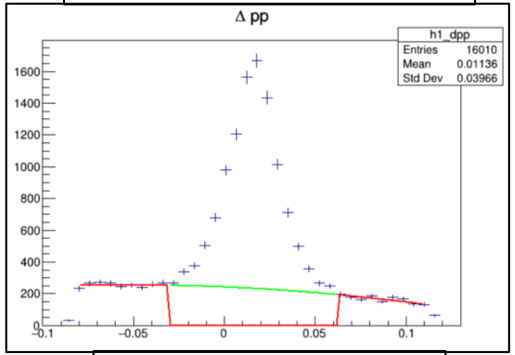
- Main challenge by far is *GEM tracking* under high occupancy (space and time signal pileup) and low signal to background ratios
- The hit comes from the large number of combinatorics → high computational times + (fake hits + fake tracks)
- 1D cluster formation and 2D hit reconstruction is especially affected by high occupancy → lots of *fake hit* reconstruction
- Tracking portion itself is done in field-free regions → straight lines; hence not the most complicated part *if* the hits fed in are mostly *true/real* proton hits





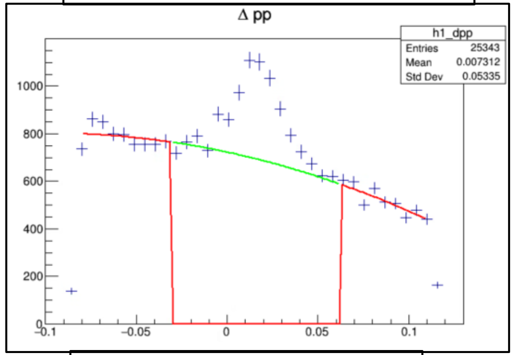
Front Tracker strips fired in an event color coded in ADC strength. 22 uA on LH2 target; one of the highest background cases in SBS.

15 uA run elastic yield



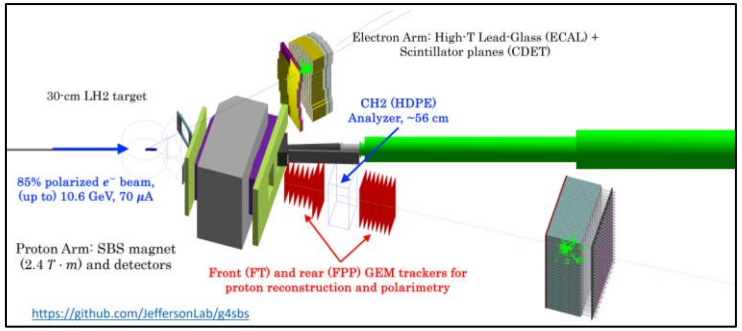
Heep yield/C = **24K**

22 uA run elastic yield



Heep yield/C = **11K**

GEP Proton Arm GEM Trackers



GEP in Monte Carlo – A. Puckett

Front Tracker (FT)

- Hee'p Scattering vertex recon.
- Momentum recon
- Incident proton tr. on analyzer
- 8 GEM layers
  - 150 cm x 40 cm X-W type readout layers x **2**
  - 150 cm x 40 cm U-V type readout layers x 4
  - 200 cm x 60 cm X-Y type readout layer x **2**

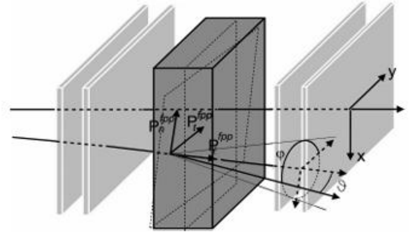


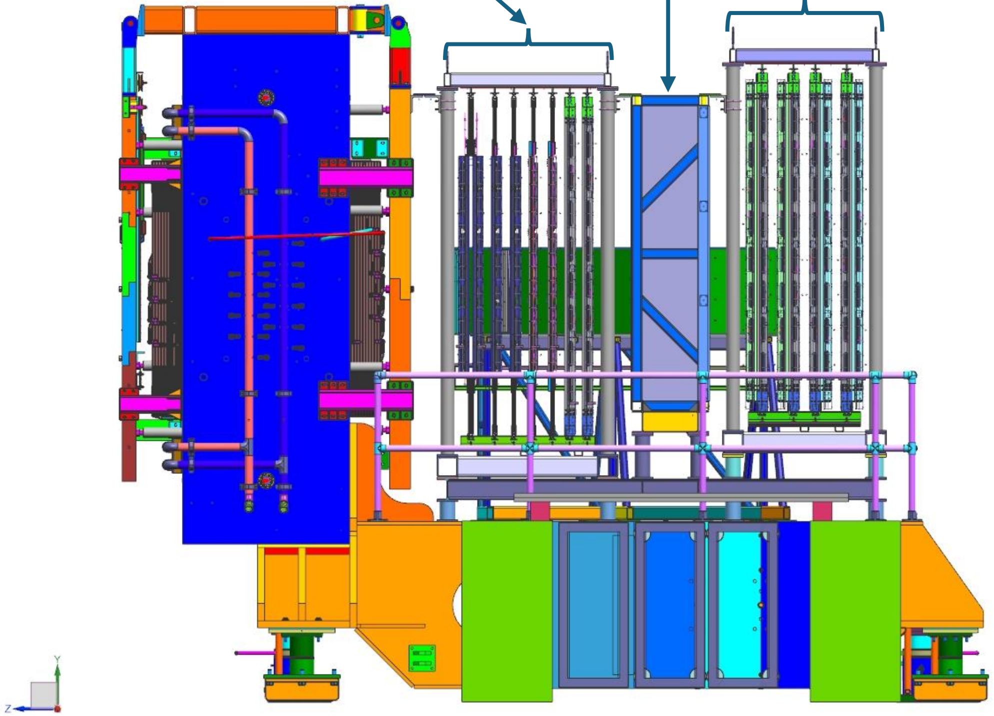
FIG. 9. Principle of the polarimeter, showing a noncentral trajectory through the front chambers, scattering in the analyzer, and a track through the back chambers; ϑ is the polar angle, and φ is the azimuthal angle from the y direction counterclockwise.

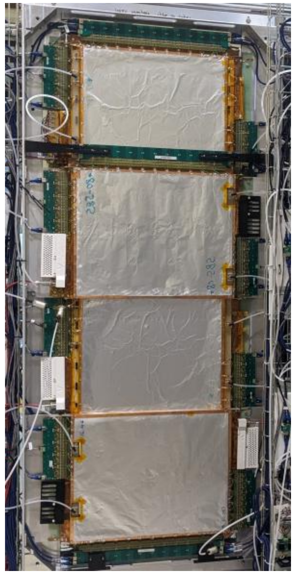
Proton-nucleon spin-orbit interaction causes azimuthal asymmetry

Focal Plane Polarimeter Tracker (FPP)

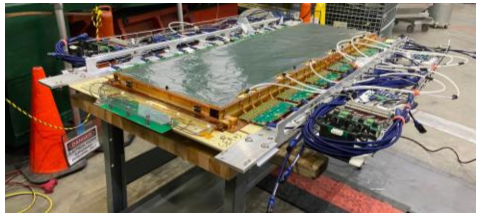
- Reconstruct scattered proton track from CH_2 analyzer for proton polarimetry
- 8 GEM layers
  - 200 cm x 60 cm X-Y type readout layers x **6**

CH_2 analyzer

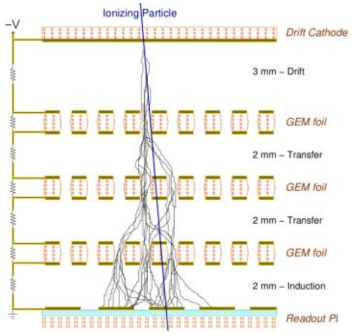




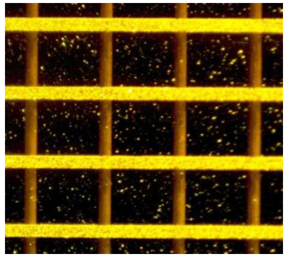
200 cm x 60 cm XY GEM layer



150cm x 40 cm UV GEM layer

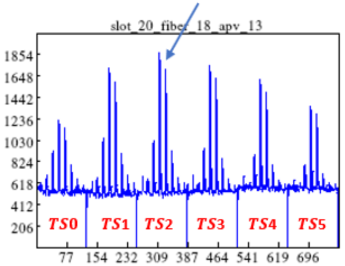


Tripp le GEM



X-Y readout

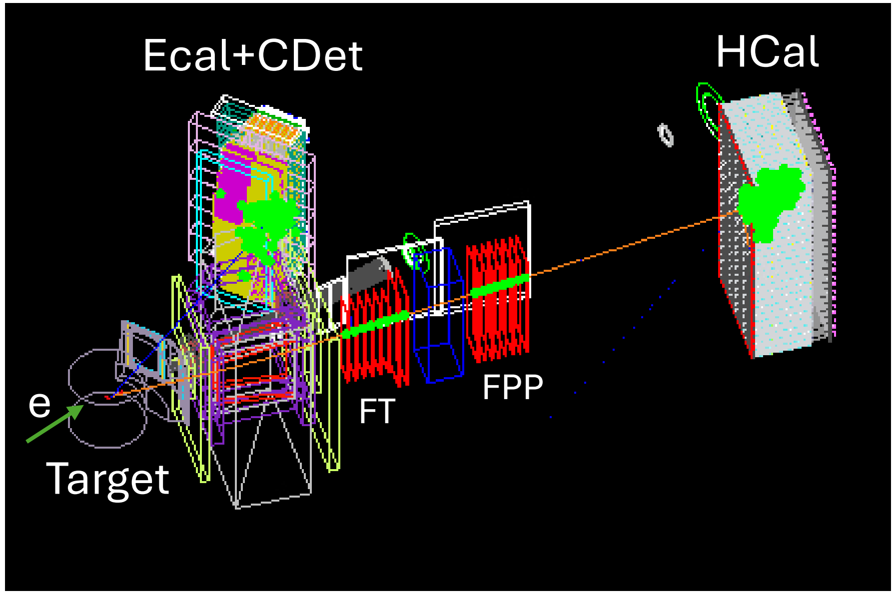
Good Signals



Raw ADC signals from 128 strips

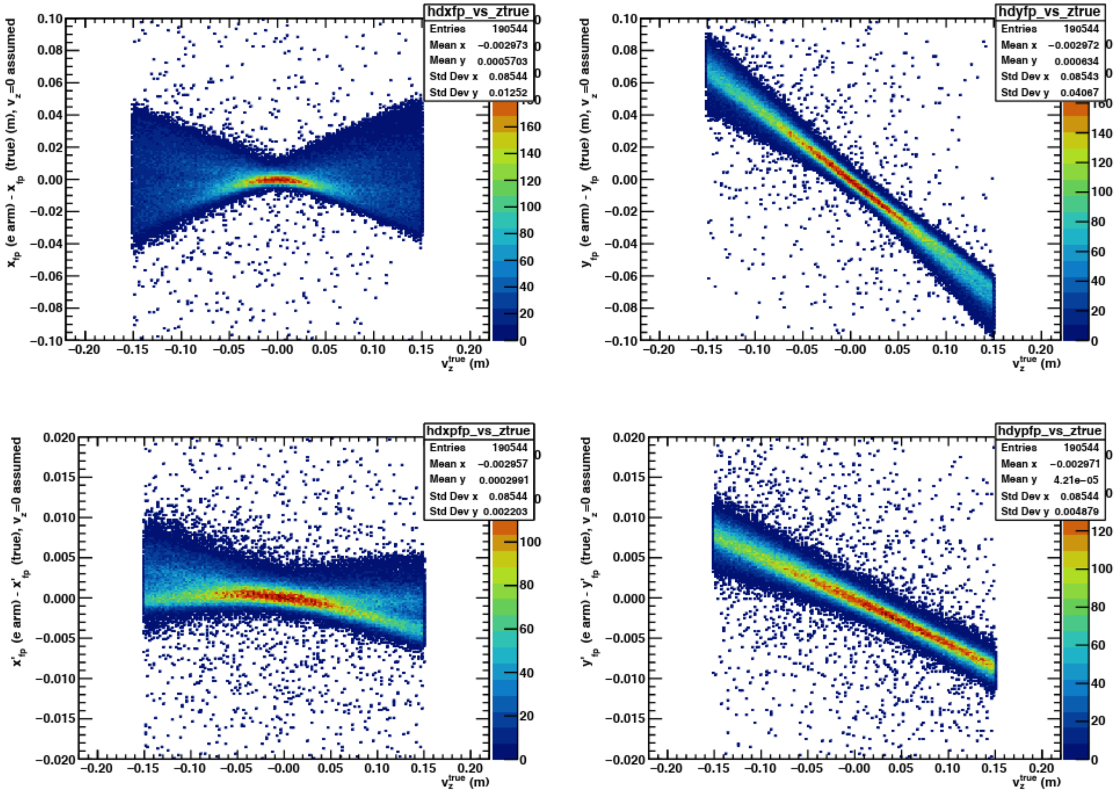
High-Level Overview of Current GEM Track Reconstruction

- Region of Interest calculation on p-arm tracks from e-arm information
  - Attempting tracking using the entire GEM active areas impossible as combinatorics explode
  - Focus on “Regions of Interest” within the GEM modules(2D hit-finding)/layers(track-formation)
  - Use external detector constraints (calorimeter clusters) and heep (elastic) kinematics



Simulated heep event

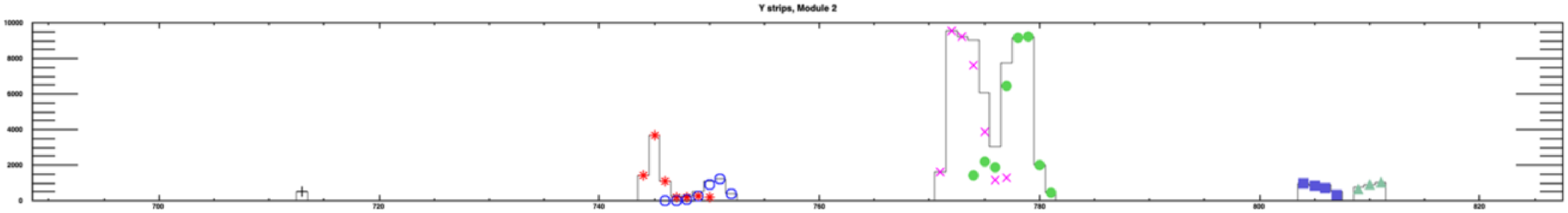
Simulation: electron angles reconstructed from ECal/CDet assuming $v_z = 0$



Plots: A. Puckett

1D clustering and 2D hit formation within GEM modules

- 1D clustering: combine "strip hits" caused by the *same MIP/ionization cloud* in each dimension (U/V)
  - Loop on all fired strips and make a list of strips with "local maxima" of ADC values → apply thresholds and timing cuts
  - For each local maxima, check for contiguous higher peaks either to the left or to the right and evaluate the current peak's "prominence" compared to the higher peak if found. If "not-prominent", remove the current strip from list of local maxima
  - Now using the above list of refined *local maxima* as "cluster seeds", grow the local maxima into individual clusters
  - However, there could be still overlapping clusters and therefore *cluster splitting* is needed
  - If there is another local maxima within a specified distance to the local maxima in question, consider that to be from a different cluster, and calculate the contribution from that cluster to the current cluster as a function of local maxima ADC strength and distance
  - Continue growing the cluster around the seed considering the ADC weighted time differences and Pearson correlation coefficient
  - Calculate cluster properties such as position, avg ADC, ADC asymmetry, avg time, and etc.

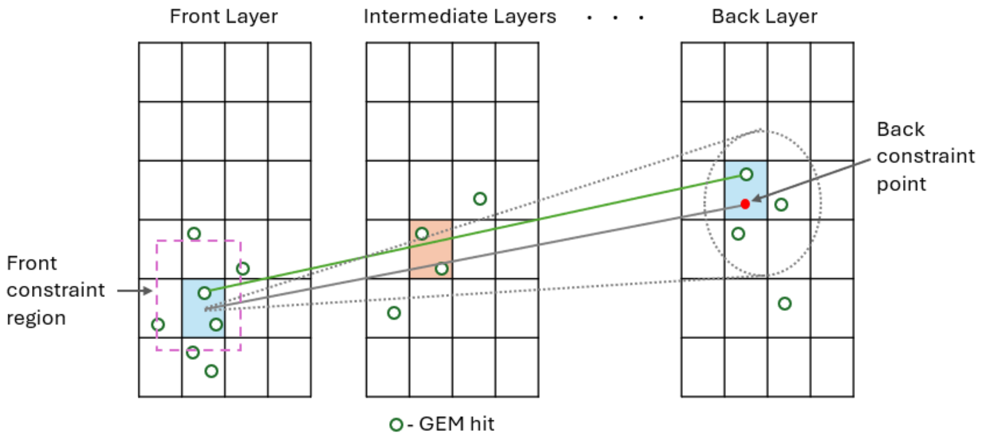


Plot from A.Puckett

- 2D hit formation: by comparing each 'U cluster' with 'V cluster'
  - First check whether the 2D hit is within the detector active area / constraint region
  - U-V cluster ADC correlation via Pearson coefficient test
  - U-V cluster time difference
  - ADC asymmetry
  - If the above criteria is met, the 2D hit is marked to be "kept"

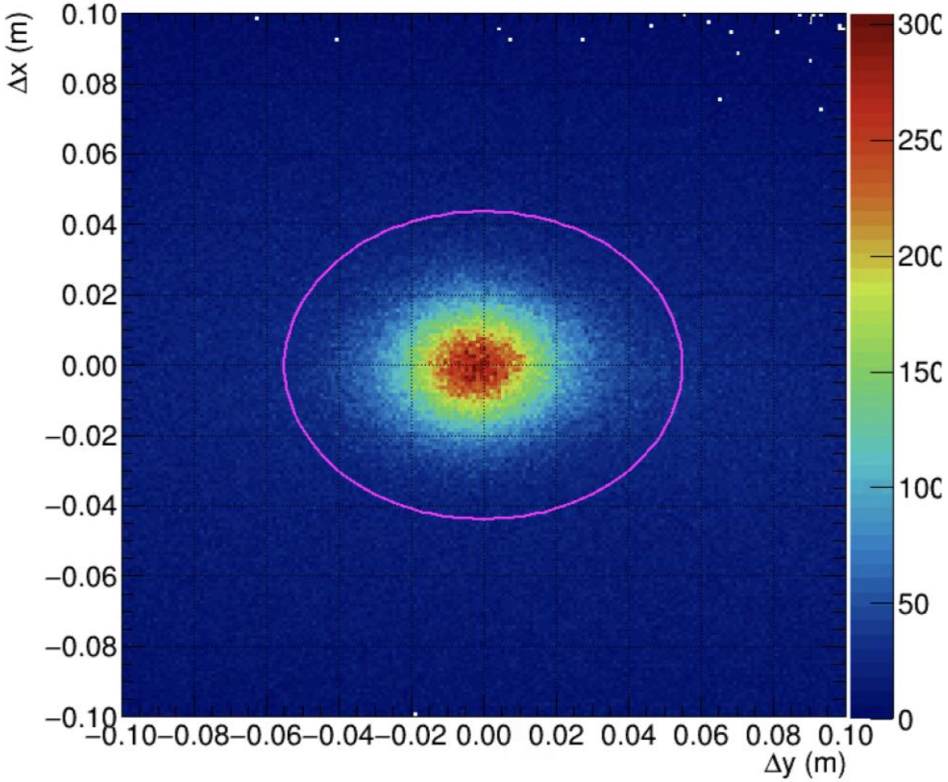
Track Reconstruction

- 3D track formation
  - Divide each tracking layer into a 2D uniform rectangular grid ($5 \times 5\ mm^2$) and make a list of 2D hit candidates in each grid bin
  - Start to make tracks by requiring all 'N' layers (1 hit in each layer). Then N-1, N-2, … down to 3 hits on track.
  - Loop over all grid bins with the hits in the "front" layer. Form a straight line from grid bin center to "back constraint point". Calculate error matrix from grid bin width and back constraint width and project to back GEM layer to define a range of grid bins that must be considered in the back layer. Now for a given "front" layer bin, we have a set of "back" layer bins.
  - For a given combination of front and back layer bins, consider ALL combination of one GEM hit in front and back layers
  - Draw straight line between the front layer GEM hit and back layer GEM hit, and project to all intermediate layers, with special resolution in the range of intrinsic GEM resolution. Now we consider hits within bins of the intermediate layers that intercept the straight-line projection and consider the hits of the bins within a tolerance to the projection (edge tolerance – 1.5 mm)
  - Now loop over all possible combinations of one hit per layer, find the combination with best χ^2/dof of straight-line fit in 3D, consistent with some other cuts on "hit quality" (timing, ADC correlation, optics, etc.)
- Finding constraints for FT and FPP trackers
  - Get front and back constraints for FT from ECal/CDet. For FPP, front constraint from ECal/CDet with proton track projected to analyzer mid-plane and back constraint from HCal cluster
- Two modes of tracking currently developed: 1) FT first and then FPP 2) FPP first and then FT (faster; only one v_z bin with smallest 'sclose' w.r.t FPP track)

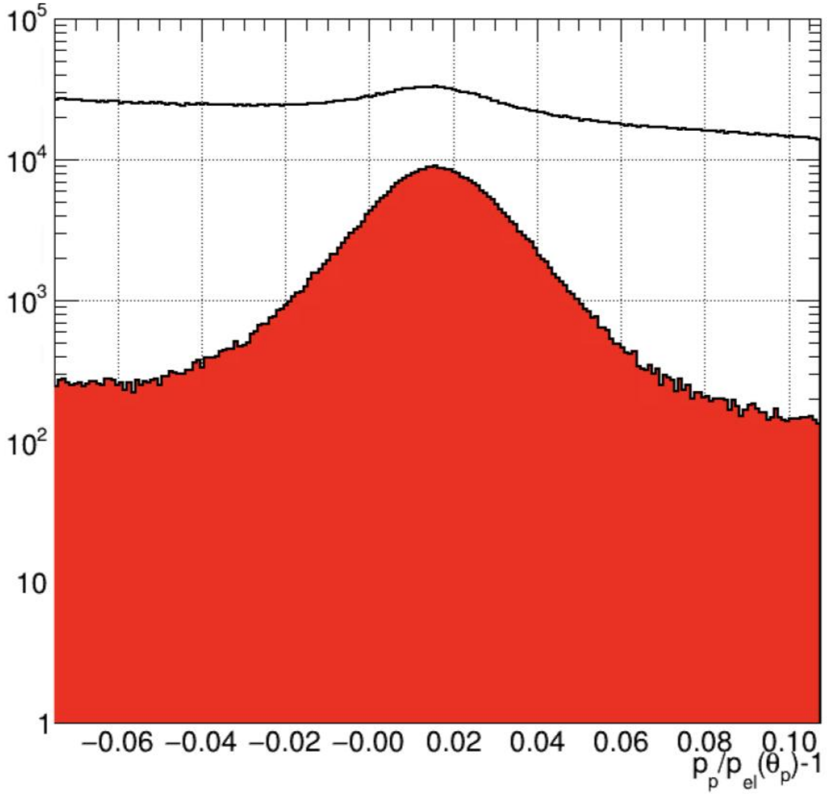


Elastic Event Extractions (GEP kin-3; 11.1 GeV^2)

Difference between ECal clus. pos. and predicted elastic e' hit pos on ECal by proton kinematics



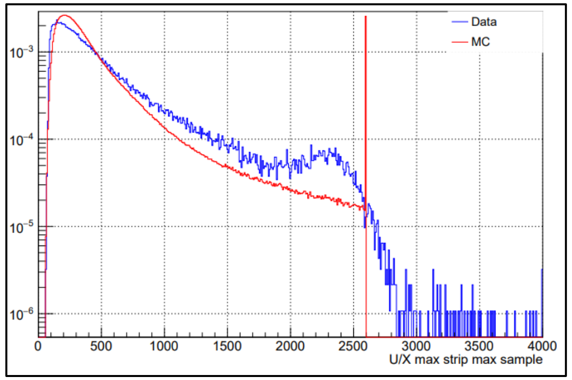
Relative difference between reconstructed proton momentum via SBS magnet momentum analysis and and polar scattering angle of the proton + elastic kin. assumption

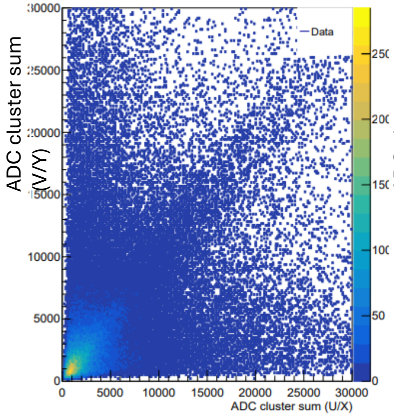


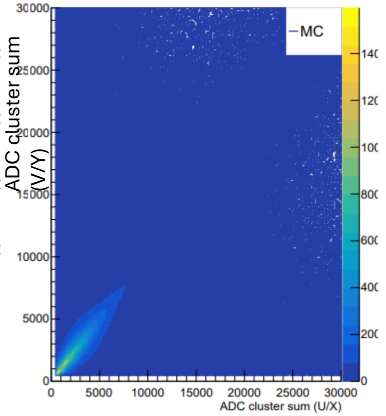
Plots: A. Puckett

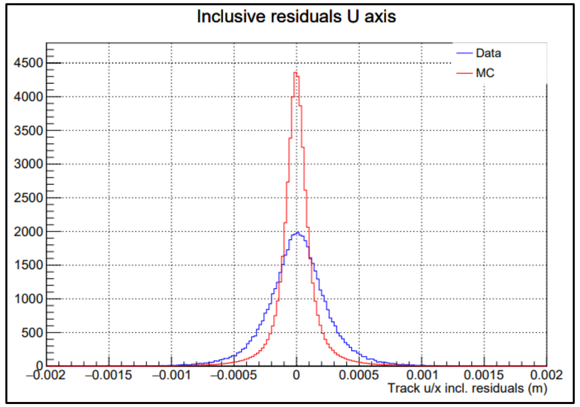
Improvements Envisioned

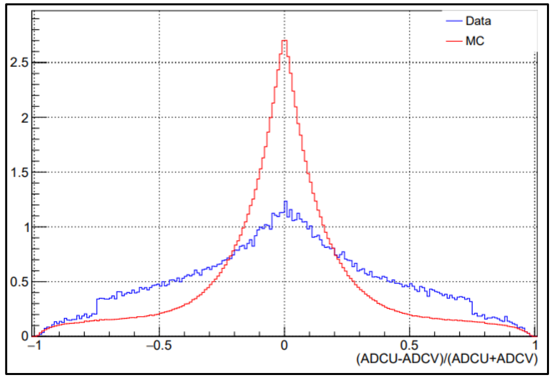
- Train a ML model with simulated data for a more *accurate/efficient* 2D hit formation
  - Currently we rely on strip ADC pulses' shape, timing, ADC strength, and U-V correlation criteria for 1D cluster and 2D hit formation
  - Hundreds of "knobs" to adjust and prohibitively complicated to find the right combination for a given experiment condition
  - Make use of possible "unseen" correlations by the existing algorithm
  - Need MC truth to train. Currently working on improving our MC machinery to make its results agree better with the real GEM data
  - Plan to seek expert guidance in the very near future

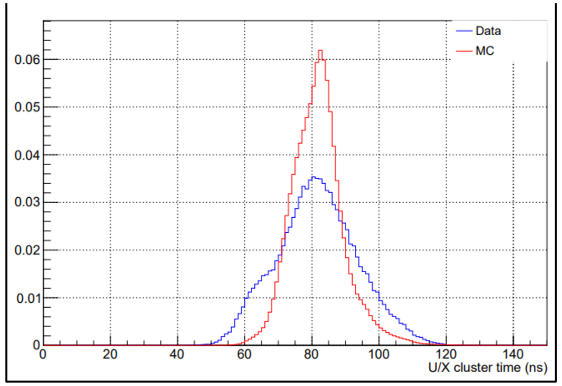


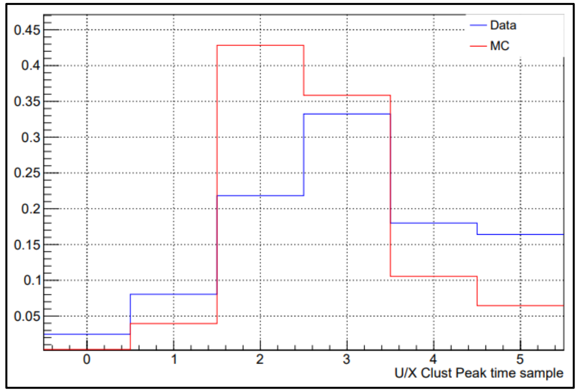












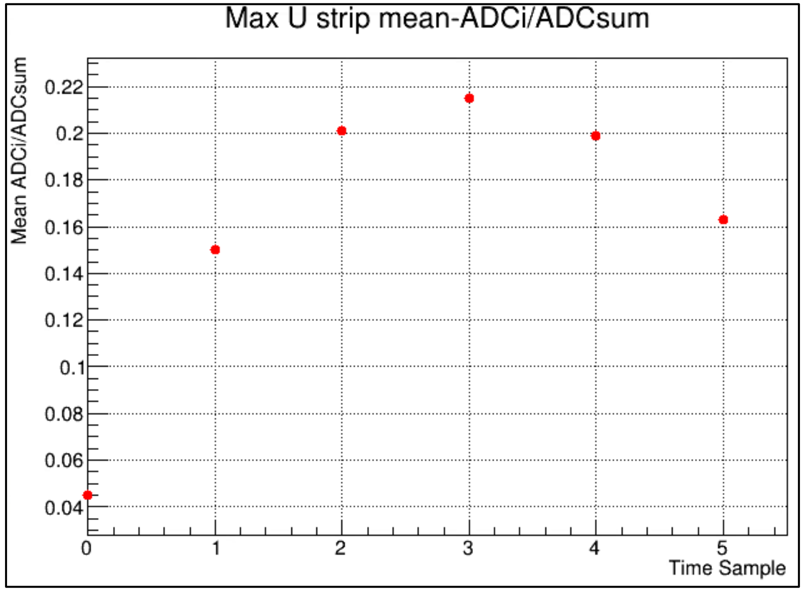
Summary

- Improving event (track) reconstruction is the most vital and challenging analysis task for GEP
- Current machinery is able to find tracks with GEM occupancies reaching 50%; but improvements/additions needed to enhance efficiency
- We think the GEM 1D clustering and 2D hit formation is where the biggest efficiency lost happen
- Plans to incorporate AI/ML → work ongoing to improving GEM simulation to generate more realistic training data
- Other traditional algorithmic improvements are also proposed
- *Stay tuned for future analysis developments GEP results*

Backup

Improvements Envisioned Cont.

- For GEM strip timing calculations, rely mostly on the last 3-4 time-samples (out of 6) as the good MIP signals correlated with the trigger typically peak at around the 4th time sample → i.e., use high signal/background time samples



Improvements Envisioned cont.

- Transform GEM "strip-space" information to a "pixel-space" and do $(i,j)^{th}$ 'U-pixel' to 'V-pixel' comparisons to identify 2D hits → the goal is to preserve the 2D special correlation of GEM hits and avoid brute-force U and V 1D cluster matching which *could be* (unverified) a bottle-neck / inefficient in high occupancy scenarios