SoLID Track Reconstruction Feasibility Study

Ole Hansen

Jefferson Lab

SoLID Collaboration Meeting
November 8, 2014

SoLID Tracking Considerations

- High rates $\mathcal{O}(1\ \mathrm{MHz/cm}^2)$, high raw occupancies (50%) $\rightarrow$ fairly difficult environment
- Tracks not straight
- GEM readout axes not parallel between planes (in PVDIS design)
- Real-time tracking for level-3 trigger $\rightarrow$ want fast algorithm

Choice of Reconstruction Algorithm

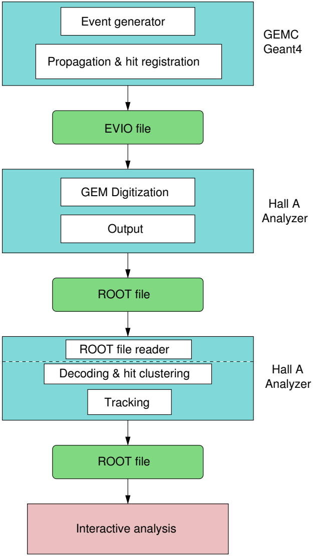
- Curved tracks, non-parallel coordinate axes $\rightarrow$ progressive algorithm (Kalman filter). Slow.
- Very preliminary version exists for SoLID (Xin Qian), being further developed by Duke group (Zhihong Ye, Weizhi Xiong).
- Little expertise in Hall A, but available in other halls $\rightarrow$ consult
- This is a multi-year development effort

This Talk: Track Reconstruction Feasibility Study

- Use existing TreeSearch reconstruction (BigBite)
  - Available now
  - Well tested & integrated in Hall A analyzer
  - Shown to work with SBS GEM trackers at $\geq$ SoLID occupancies
  - BUT: requires straight tracks, parallel axes
- To use TreeSearch, simplify the problem in simulation:
  - Rotate GEM strips in software $\rightarrow$ parallel coordinate axes
  - Simulate DIS signal without magnetic field $\rightarrow$ straight tracks
  - Background (with field) added separately $\rightarrow$ can vary background level
  - Caveat: Background may be underestimated. TreeSearch rejects curved tracks.
  - Expect this still to demonstrate *feasibility* of track finding

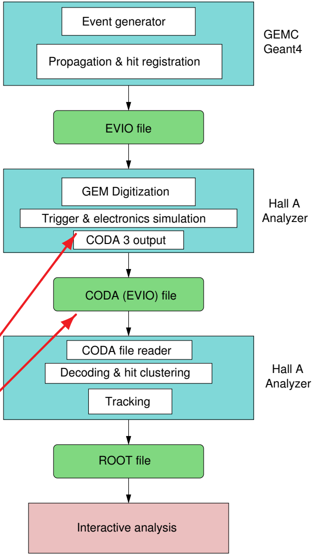
Track Reconstruction Simulation

- `solgemc` EVIO files as digitization input (S. Riordan)
- GEM digitization based on SBS work (E. Cisbani, R. Holmes)
  - APV25 pulse shape
  - Background added with randomized time offset
  - No other detectors digitized yet
  - Generated data (tracks, vertices) passed through
- ROOT file interface
- Tracking



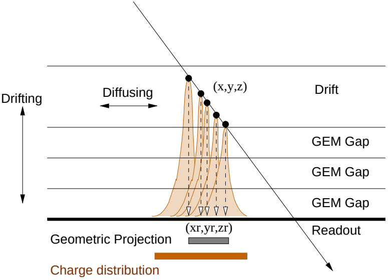
Track Reconstruction Simulation (“data challenge” ready)

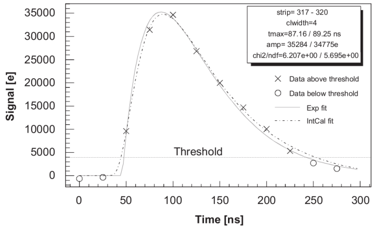
- `solgemc` EVIO files as digitization input (S. Riordan)
- GEM digitization based on SBS work (E. Cisbani, R. Holmes)
  - APV25 pulse shape
  - Background added with randomized time offset
  - No other detectors digitized yet
  - Generated data (tracks, vertices) passed through
- ROOT file interface
- Tracking
- Should eventually use actual DAQ format (CODA 3) for analyzer input



GEM & APV25 Digitization (by E. Cisbani, adapted by Rich Holmes)

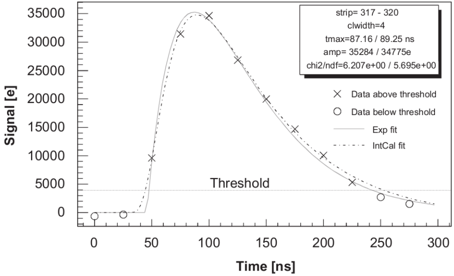
- GEMC outputs raw hits (energy deposition ΔE) in GEM layers
- Avalanche simulation:
  - Poisson-distributed number of ion pairs calculated from ΔE
  - Use geometric distribution for ionization probability along path
  - Assume constant-velocity diffusion and drift
  - Gaussian distribution of charge deposited on strips
- GEM response tuned to match COMPASS observations
- Shape output amplitude: $v = A\tau \exp(-\tau)$, record 3 samples in 25ns intervals
- Currently seems to produce clusters that are too small and slightly asymmetric between u and v $\longrightarrow$ improve





APV25 Pulse Shape Deconvolution & Noise Filtering

S. Gadomski *et al.*, NIM A320, 217 (1992)



- For a first-order RC circuit, the original signal amplitudes s_k can be recovered from only three measured values v_k:

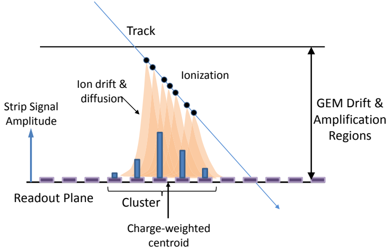
$$s_k = w_1 v_k + w_2 v_{k-1} + w_3 v_{k-2}$$

$$w_1 = e^{x-1}/x, w_2 = -2e^{-1}/x, w_3 = e^{-x-1}/x, \text{ where } x = \Delta t/\tau$$

$$\text{Integrated amplitude:} \quad A \approx \sum_{k=1}^{3} s_k$$

- Reject noise by cutting on ratios, $r_1 = v_3/v_1$ and $r_2 = v_2/v_1$, requiring rising slope

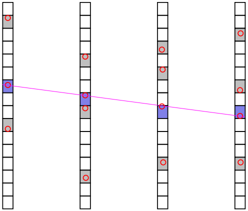
GEM Hit Clustering



- Signals on adjacent readout strips typically belong to a single track crossing
- Sum signals to get
  - total hit amplitude
  - charge-weighted position centroid
- Currently use simple algorithm:
  - Look for local peak
  - When sequence "peak-valley-peak" is seen, split cluster at "valley"
  - Regardless of shape, limit clusters to a maximum size
- Improvements possible
  - Match hits by their pulse shape, *i.e.* timing centroid
  - Redo clustering after preliminary tracking (*e.g.* better cluster splitting)
  - . . . possibly more
- Clustering could also be integrated directly into a progressive tracking algorithm

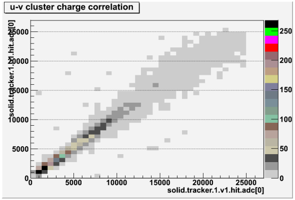
TreeSearch Algorithm

M. Dell'orso and L. Ristori, NIM A287, 436 (1990)

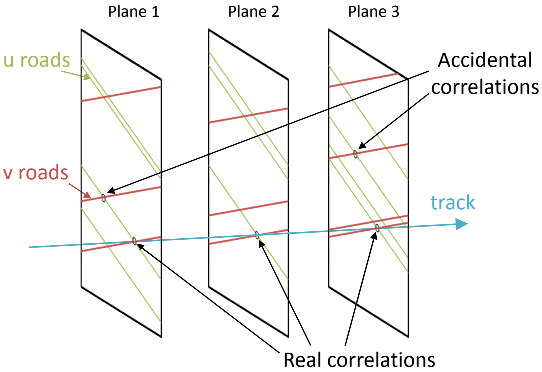


- Recursive template matching algorithm (global, non-progressive)
- Advantages
  - Very fast ($\mathcal{O}(\log N)$) and memory-efficient ($\mathcal{O}(10\text{ MB})$)
  - Independent of other detectors $\rightarrow$ no seed needed
- Limitations
  - Works in 2D only (one readout coordinate, "projection")
  - Only suitable for straight tracks
- Used by HERMES, Qweak, etc.
- Configure "virtual planes" to pre-select regions of interest
  - Calorimeter hits (emulated using MC hit in last GEM plane)
  - Target area

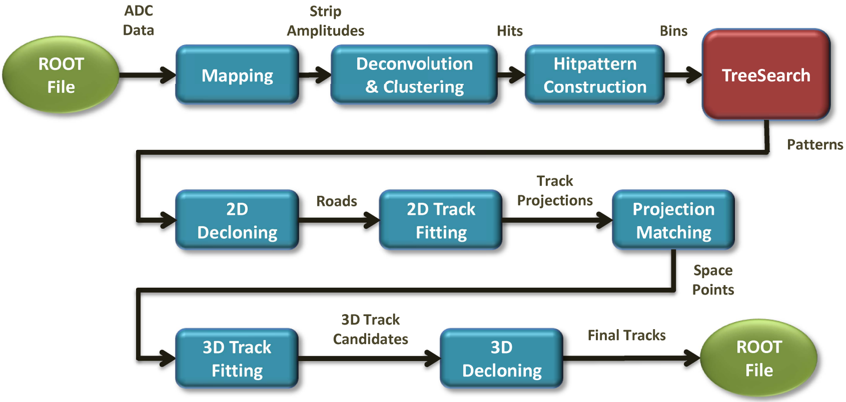
3D Matching



- Correlate roads from different projections via hit amplitude in shared readout planes
- Pair roads with the best overall correlation to get space points for 3D track fits
- Noise may destroy amplitude correlations → must allow for mismatches
- Here: Require 2/5 matches



TreeSearch Track Reconstruction Chain (GEM version)



MC Data Sets

(simulations by Seamus Riordan)

- Configuration
  - 40 cm LD_2 target in 11 GeV beam
  - PVDIS detector setup with 5 GEM planes
  - CLEO baffles (old)
- "Signal Runs"
  - Generator: DIS
  - Only interactions of primary particle recorded
  - Using μ^- primary particles
  - Field off to get straight tracks
- "Background Runs"
  - Same configuration as for signal runs, except field always on
  - Simulated 198 M background events (= electrons passing through target)
  - Production rate $\approx$ 40 M/hr

Background Simulation

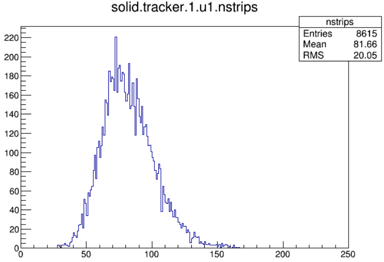
- Adding background to signal runs in digitization step
  - $\approx$ 86 M electrons pass through target in a 275 ns time window at 50 μA $\equiv$ 100% background
  - To reduce analysis time, fold background from 30 sectors into signal sector with random time offset per sector
  - Obviously not enough background events for any significant number of signal events $\longrightarrow$ re-use events, but with different time randomization
- Status
  - 100k signal events digitized with 0%, 1%, 5%, 10%, 25%, 50% and 100% background added
  - At 100% bg level, digitization rate is $\approx$ 500 signal events/hr/CPU core, or $\approx$ 7 CPU core seconds/event (on Xeon E5-2650v2 “Ivy Bridge”)

Trigger Emulation/Event Selection

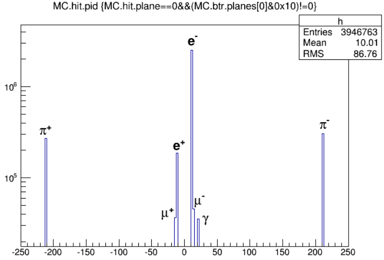
- Trigger not part of simulation yet
- Trigger emulation for this study:
    - Select only events where signal track passes through all GEM planes
    - Signal tracks not passing through all planes either
        - would do so if the field was on; or
        - would be outside of the acceptance and so would be blocked by baffles or not make a calorimeter hit
    - background events never make a trigger
    - some good signal events may also not make a trigger
- This is neither an acceptance study nor a study of background trigger rates!
- We are only characterizing track reconstruction for this subset of clean triggers. Less clean signal triggers and background triggers may have other (presumably less favorable) reconstruction characteristics, but we would typically want to reject such events, not reconstruct them. If they do reconstruct, they may become a contamination of the signal, but that's a different study that cannot be done at present, lacking a realistic trigger simulation.

Strip Occupancy at 100% Background

Number of strips above ADC threshold *after* noise cut



PID of hits at 100% background



Occupancies from strips passing noise cut			
Plane	Mean # active	Total # strips	Occupancy (%)
u1	81.7	753	10.8
v1	88.3	627	14.1
u2	73.3	945	7.8
v2	75.6	659	11.5
u3	68.3	921	7.4
v3	72.5	657	11.0
u4	56.4	1271	4.4
v4	58.2	1271	4.6
u5	54.5	1309	4.2
v5	56.9	1309	4.3

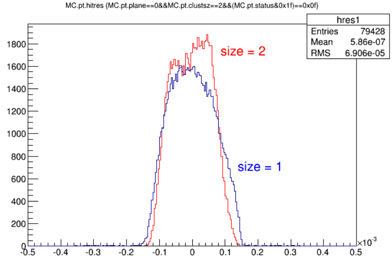
- First plane sees many slow electrons ($p < 1$ MeV)
- u-v asymmetry is mostly artifact of the rotation of strip directions $\rightarrow$ asymmetry in # and lengths of strips
- SoLID-PVDIS occupancies below SBS's (SBS has up to 20%)

FYI: Raw Strip Occupancies at 100% Background

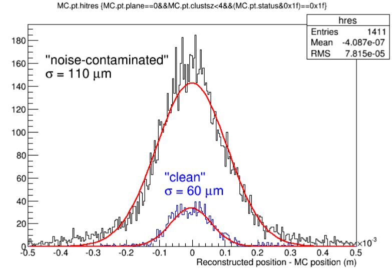
Occupancies without noise cut			
Plane	Mean # active	Total # strips	Occupancy (%)
u1	306.4	753	40.7
v1	335.3	627	53.5
u2	300.8	945	31.8
v2	306.2	659	46.5
u3	279.6	921	30.4
v3	294.4	657	44.8
u4	236.6	1271	18.6
v4	245.3	1271	19.3
u5	231.9	1309	17.7
v5	242.4	1309	18.5

Hit position resolution from cluster reconstruction

Hit resolution vs. cluster size, no background



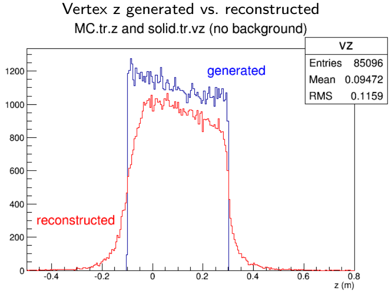
Hit resolution clean vs. noise-contaminated cluster, 100% bg.

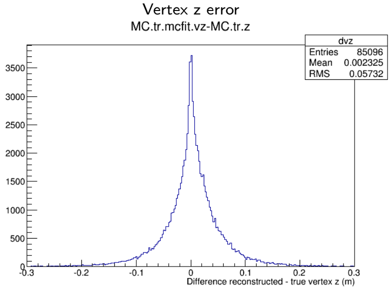


- Clusters too small (typ. 1 or 2 hits) $\rightarrow$ need to improve digitization
- size = 1 clusters have only slightly worse resolution than size = 2 $\rightarrow$ somewhat surprising
- "Noise-contaminated" means that at least one strip of a signal cluster contains a contribution from noise.
- Signal cluster reconstruction largely intact even with high noise, but resolution degrades about $\times 2$.
- Estimate average $\sigma \approx 90\ \mu$m. Use this value for estimating χ^2 of fits.
- Hit digitization and cluster building not carefully optimized yet.

Vertex Reconstruction

- Straight tracks → simple reconstruction: Find point of closest approach between two lines, track and beam
- Expect ≈ 3 mm from tracking resolution (see later)
- Observe poor resolution ≈ 3 cm → straight tracks punch through baffles (my bad, wrong signal data file selected)
- Still good enough for applying coarse vertex cut



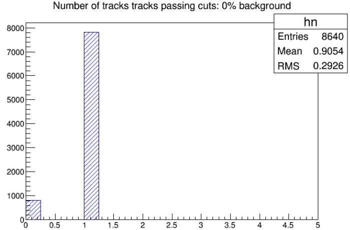


Definitions

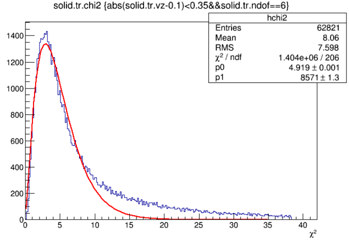
- "Tracking Efficiency"
  - Fraction of events with at least one reconstructed track that would appear acceptable in a real experiment
  - Idealized fully efficient calorimeter and PID cuts already implied by the initial event selection
  - This means: Track must pass vertex-z cut
  - As close to realistic event selection as we can get with this simulation
- "Accurately Reconstructed Track"
  - A track from the acceptable sample, defined above, with at most two misidentified clusters
  - *NB:* A cluster represents a measured coordinate (u or v)
  - This means: Track must have at least 6/8, 7/9, or 8/10 clusters that were generated by the signal track
- "Ghost Track Rate"
  - Ratio number not accurately / number accurately reconstructed tracks
- "Track Parameter Residual"
  - Difference between a track coordinate (r, φ, θ' or ϕ') at first tracker plane and corresponding coordinate of true signal hit in that plane

Tracking Results — No Background

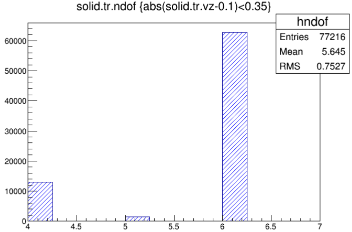
Multiplicity of accepted tracks



χ^2 of accepted tracks with $N_{dof} = 6$



Degrees of freedom of track fits (max = 6)



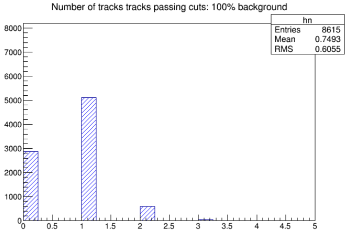
- Tracking efficiency

$$\frac{7823}{8640} = 90.5\%$$

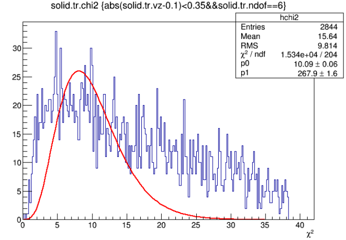
- Inefficiency almost entirely due to χ^2 cuts

Tracking Results — 100% Background

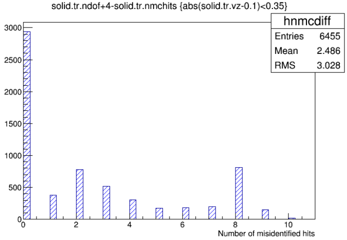
Multiplicity of accepted tracks



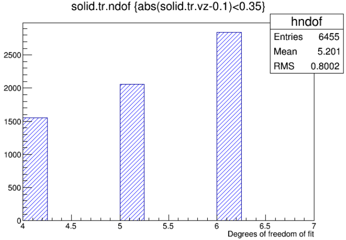
χ^2 of accepted tracks with $N_{dof} = 6$



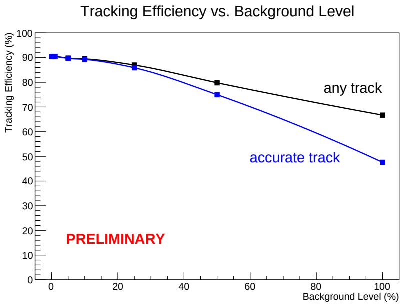
Number of misidentified hits

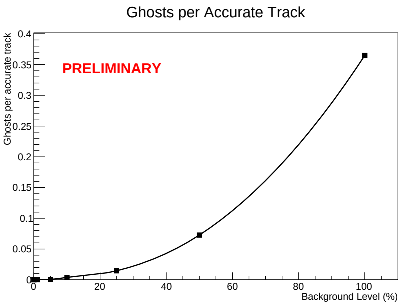


Degrees of freedom of track fits (max $= 6$)



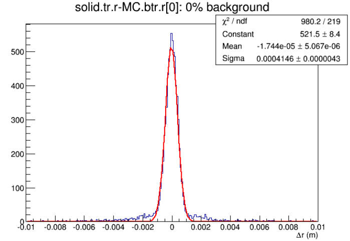
Tracking Performance vs. Background Level



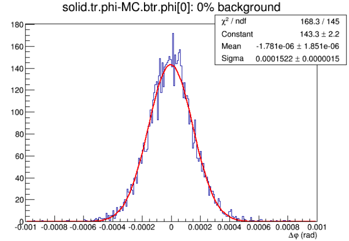


Residuals of Accepted Tracks — No Background

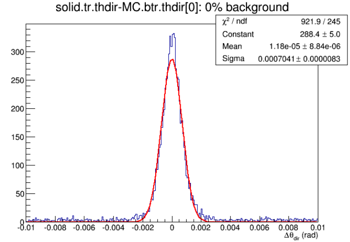
r-coordinate of crossing point in first GEM plane



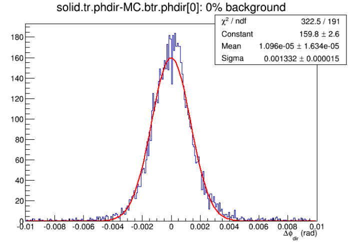
φ-coordinate of crossing point in first GEM plane



θ_{dir}: Polar angle of momentum



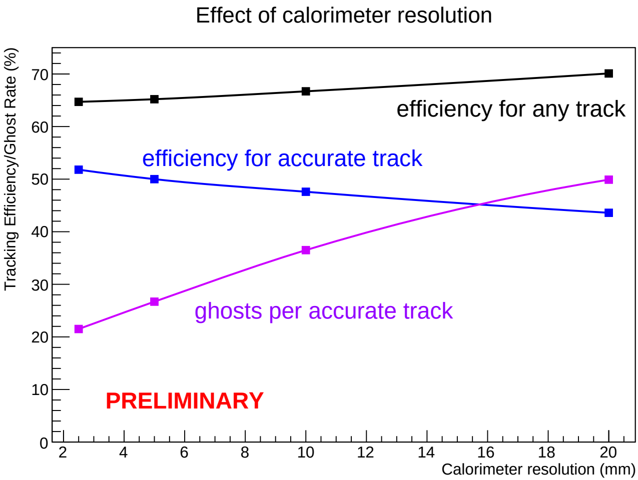
ϕ_{dir}: Azimuth of momentum



Resolutions vs. Background

Parameter	Fit to true hits	Reconstructed Tracks at Background Level						
		0%	1%	5%	10%	25%	50%	100%
σ_r (mm)	0.14	0.41	0.42	0.44	0.48	0.60	0.76	0.96
σ_φ (mrad)	0.15	0.15	0.15	0.15	0.16	0.16	0.17	0.19
$\sigma_{\theta'}$ (mrad)	0.65	0.70	0.71	0.72	0.74	0.79	0.84	0.95
$\sigma_{\phi'}$ (mrad)	1.40	1.33	1.33	1.33	1.35	1.38	1.42	1.63
σ_{vz} (mm)	41	39	39	39	39	42	45	88

Tracking at 100% Background vs. Calorimeter Resolution



Conclusions

- Roughly realistic data analysis framework has been set up, including calorimeter emulation and vertex reconstruction.
- Tracking efficiencies and ghost rates at $\leq 50\%$ background level acceptable, marginal at 100% background.
- Improvements quite possible, so acceptable reconstruction at 100% background may be within reach.
- Quality of digitization, hit signal analysis and cluster finding seems to have critical impact on the track reconstruction at high background level.
- More recent baffle design suggests overall lower background, which should be evaluated next.

Outlook

- Re-run simulation & analysis with current baffle design (in progress)
- Make GEM digitization more realistic & check against data
- Include other detectors in digitization & analysis
- Develop progressive tracking algorithm (started). Learn from Halls B & D who have such algorithms.
- Demonstrate curved track reconstruction feasibility, performance etc.
- Study tracking with SIDIS simulation data. Requires curved track reconstruction.