Online Track and Vertex Reconstruction on GPUs for the Mu3e Experiment

Dorothea vom Bruch

March 28th 2017

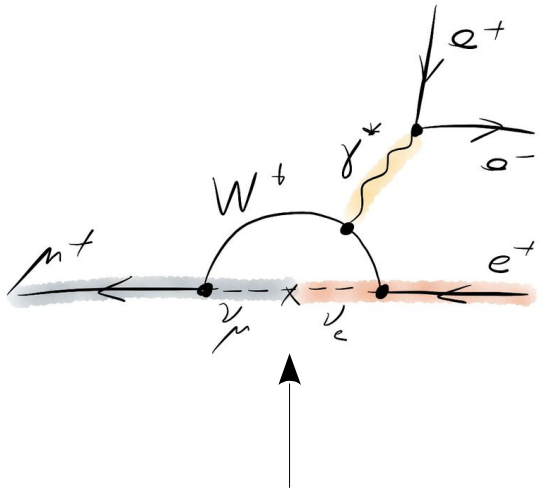
DPG Frühjahrstagung 2017, T46: Elektronik

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

The Mu3e Experiment

Search for charged lepton flavour-violating decay $\mu^+ \rightarrow e^+e^-e^+$ with a sensitivity in branching ratio better than 10^{-16}



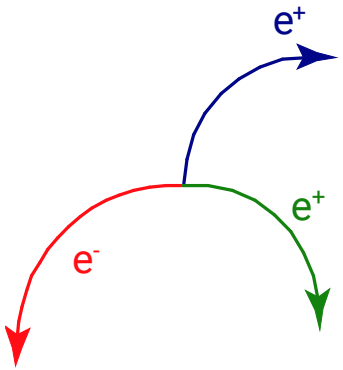
Branching ratio suppressed in Standard Model to below 10^{-54}

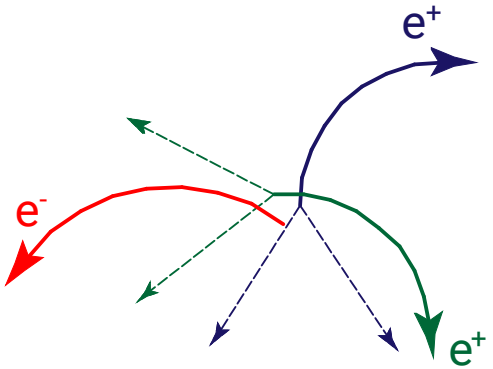
Any hint of signal → new physics

- Supersymmetry
- Grand unified models
- Extended Higgs sector
- ...

Current limit on branching ratio: 10^{-12} (SINDRUM, 1988)

Mu3e Signal





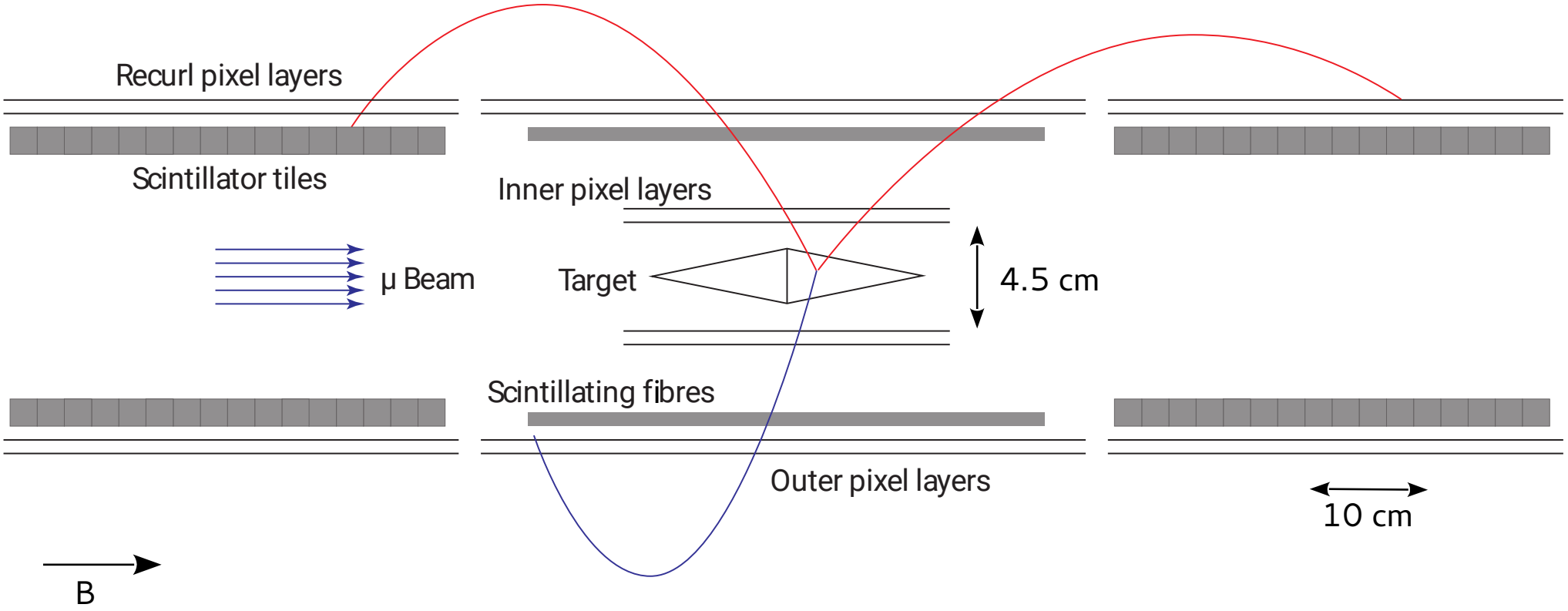
Signal

- Coincident in time
- Single vertex
- $\sum \vec{p}_i = 0$
- E = m_μ

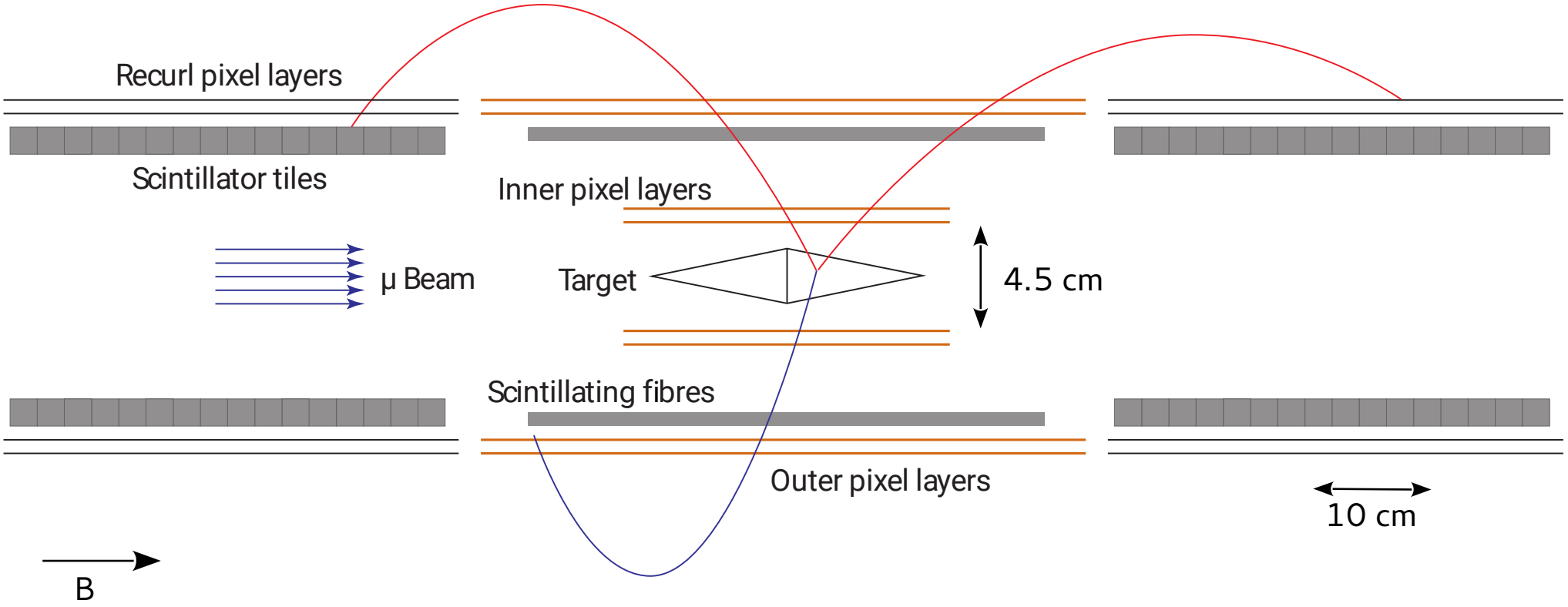
Random Combinations

- Not coincident in time
- No single vertex
- $\sum \vec{p}_i \neq 0$
- E ≠ m_μ

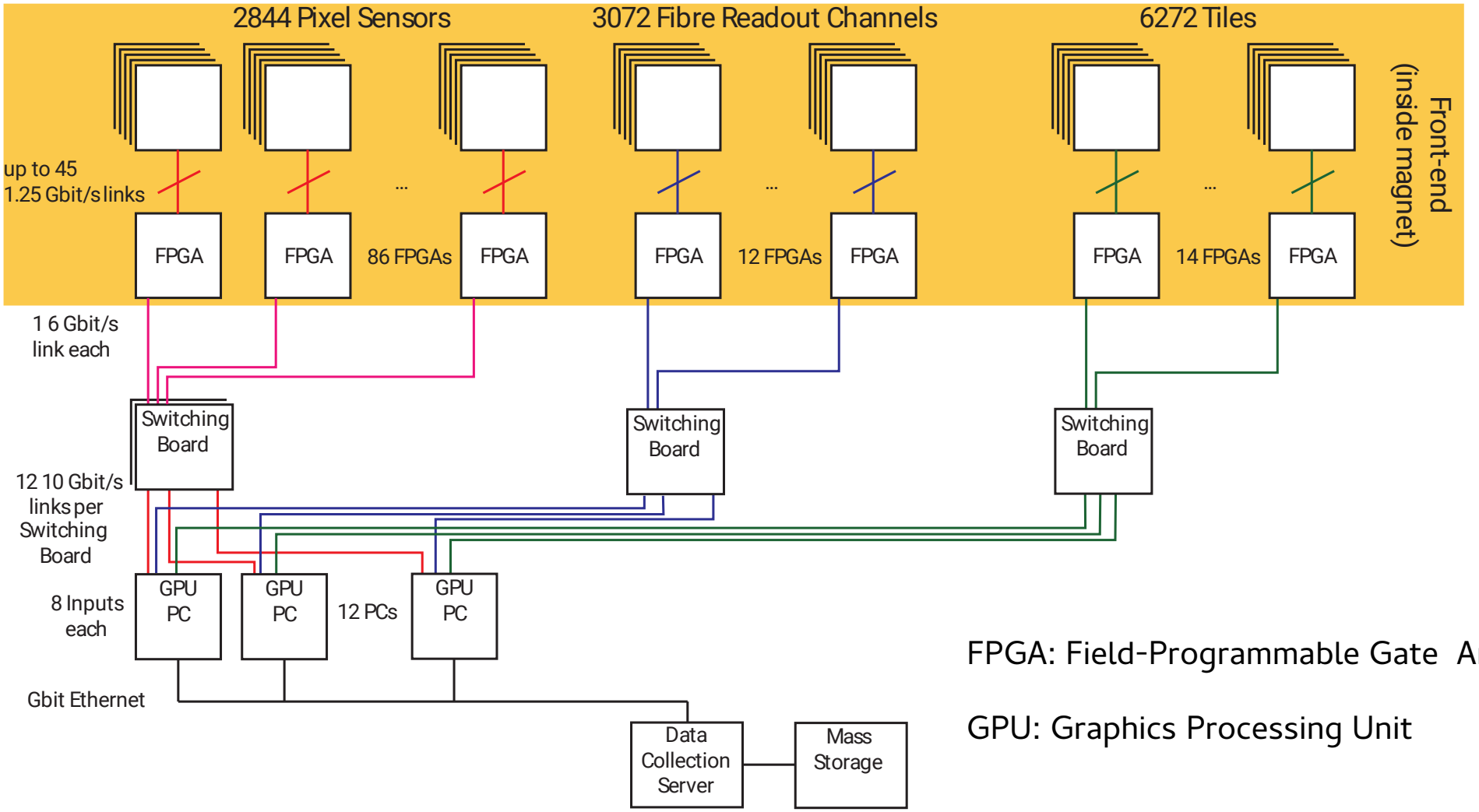
The Mu3e Detector



The Mu3e Detector



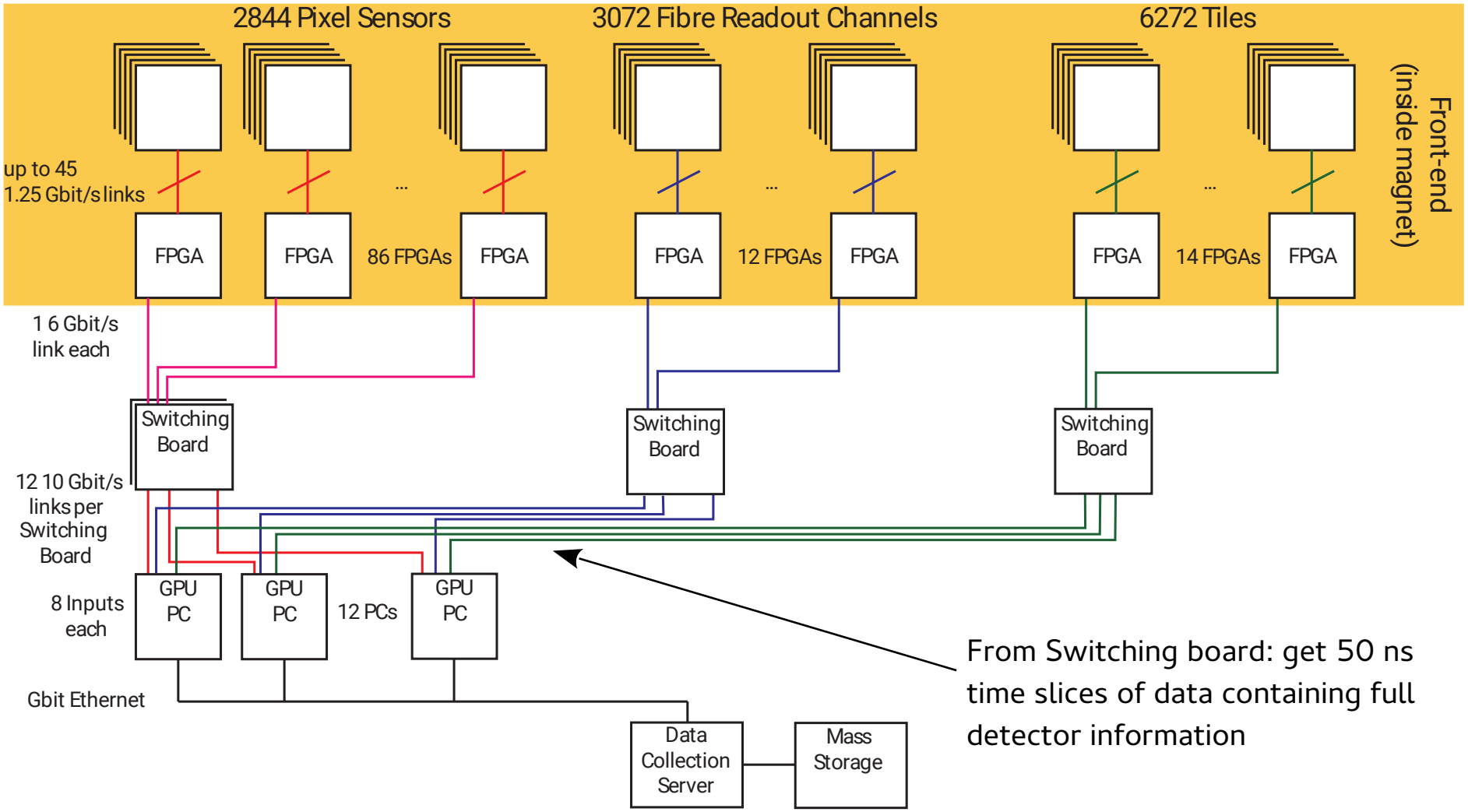
Readout Scheme



FPGA: Field-Programmable Gate Array

GPU: Graphics Processing Unit

Readout Scheme



Readout Rate

At a rate of 10^8 muons / s

Triggerless, zero-suppressed readout

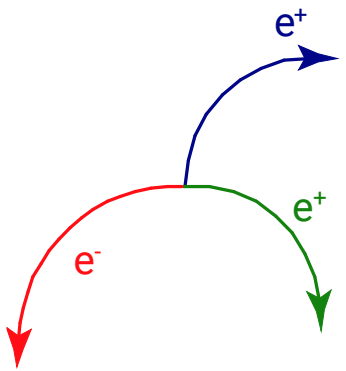
	Data rate [Gbit / s]
Pixel detector	40
Fiber detector	20
Tile detector	negligible
Total	**~ 60**

Need factor ~80 reduction to reach 100 MB/s

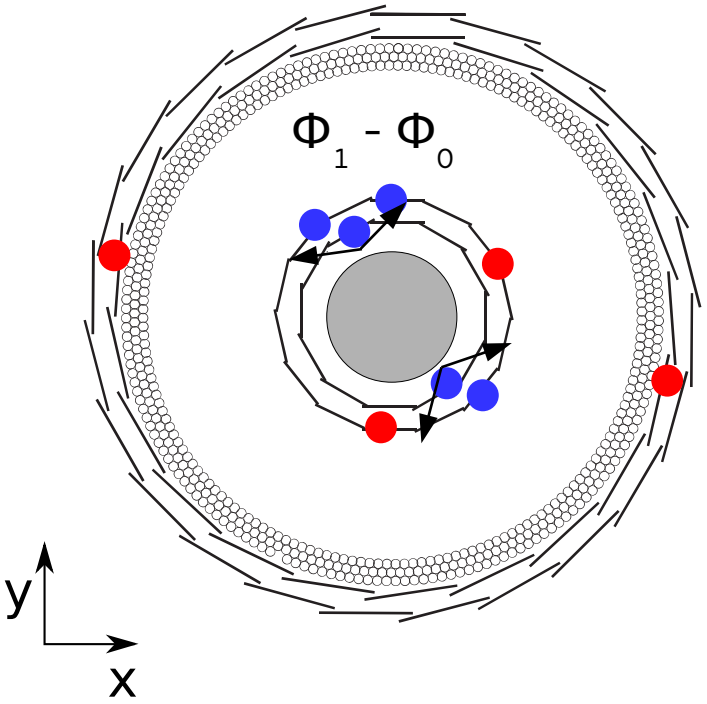
Selection Process

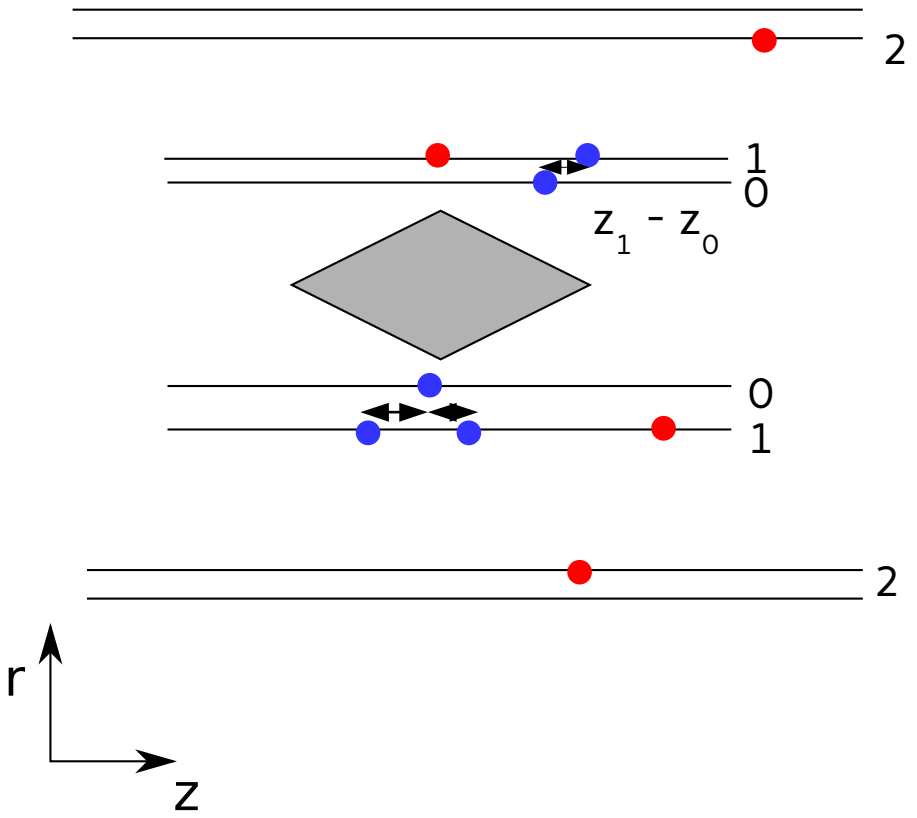
How do we find the three signal tracks?

1) Selection Cuts
2) Track fitting
3) Vertex search



Geometrical Selection





In subsequent layers, cut on:

- z-difference of hits
- Φ-difference of hits

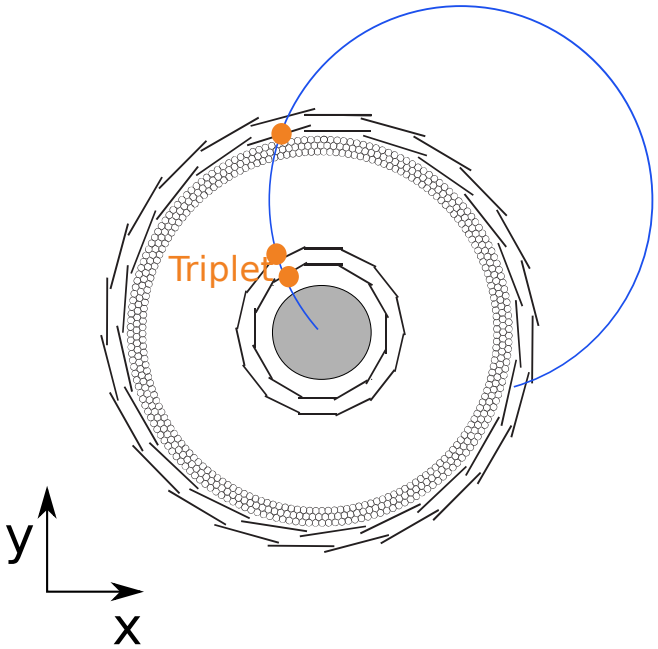
After all cuts:

Reduce 3-hit combinations by factor 50

Multiple Scattering Fit

- Electrons: 12 – 53 MeV/c
- Resolution dominated by multiple Coulomb scattering
- Ignore hit uncertainty

→ Talk by A. Kozlinskiy
(T 116.1, Thursday, 16:45)



- Three consecutive hits: "triplet"
- Multiple scattering at middle hit of triplet
- Minimize multiple scattering

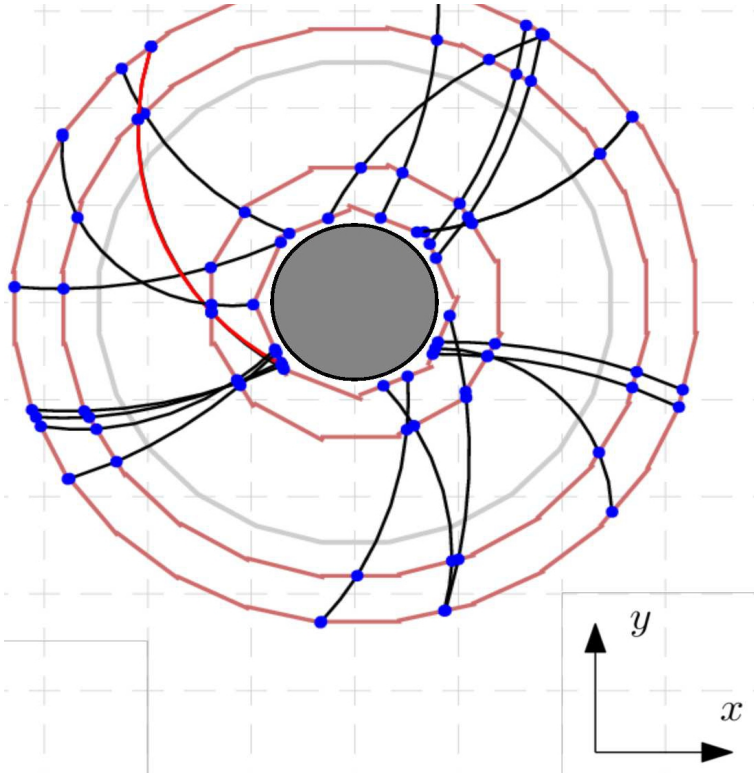
$$\chi^2 = \frac{\Phi_{MS}^2}{\sigma_{MS,\Phi}^2} + \frac{\theta_{MS}^2}{\sigma_{MS,\theta}^2}$$

Fitting

- Fit hits in first three layers
- Propagate to 4th layer
- Select hit in 4th layer closest to propagated position
- Redo fit with a second triplet, cut on χ^2

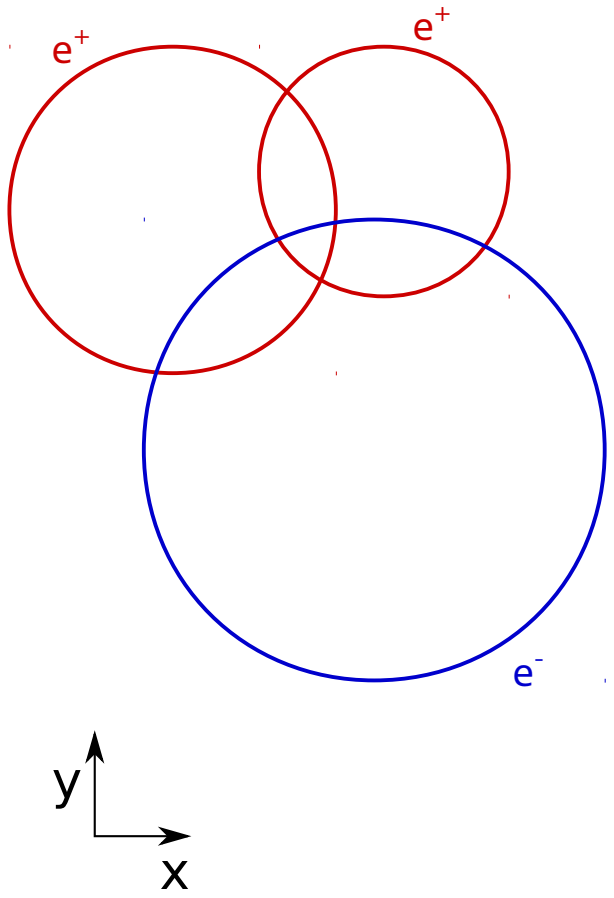
After all selections:

- 98.5 % of true 4-hit MC tracks selected
- 74 % of 4-hit tracks are true MC tracks



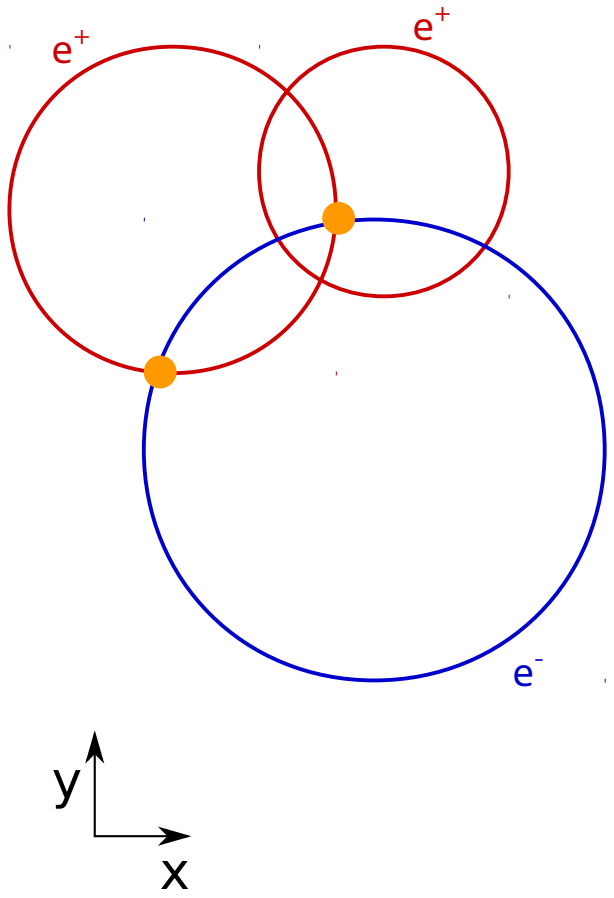
Vertex Estimate: XY-Plane

- Study each combination of two e^+, one e^-
- In xy-plane: find intersections of track circles
- Calculate weights of intersections based on uncertainties due to
  - multiple scattering
  - pixel size



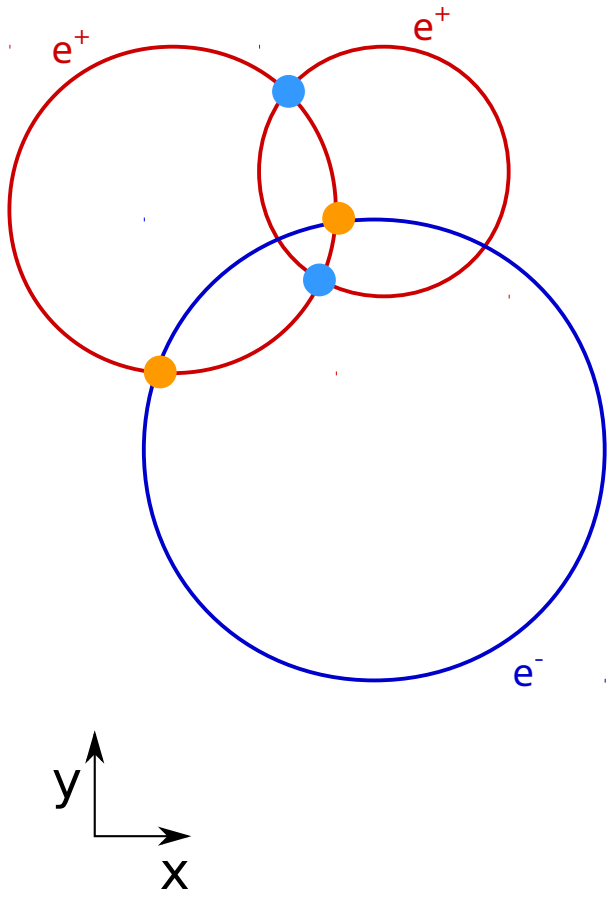
Vertex Estimate: XY-Plane

- Study each combination of two e^+, one e^-
- In xy-plane: find intersections of track circles
- Calculate weights of intersections based on uncertainties due to
  - multiple scattering
  - pixel size



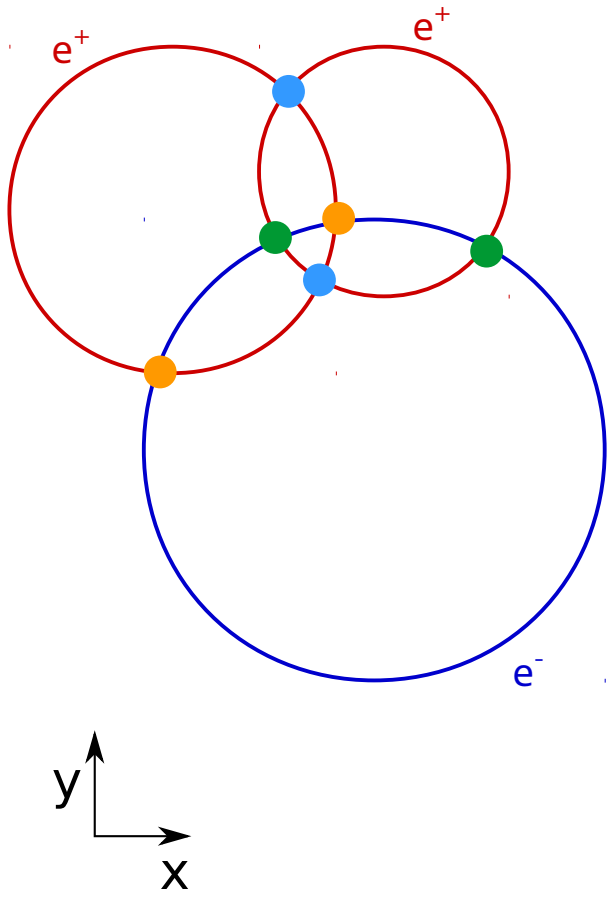
Vertex Estimate: XY-Plane

- Study each combination of two e^+, one e^-
- In xy-plane: find intersections of track circles
- Calculate weights of intersections based on uncertainties due to
  - multiple scattering
  - pixel size



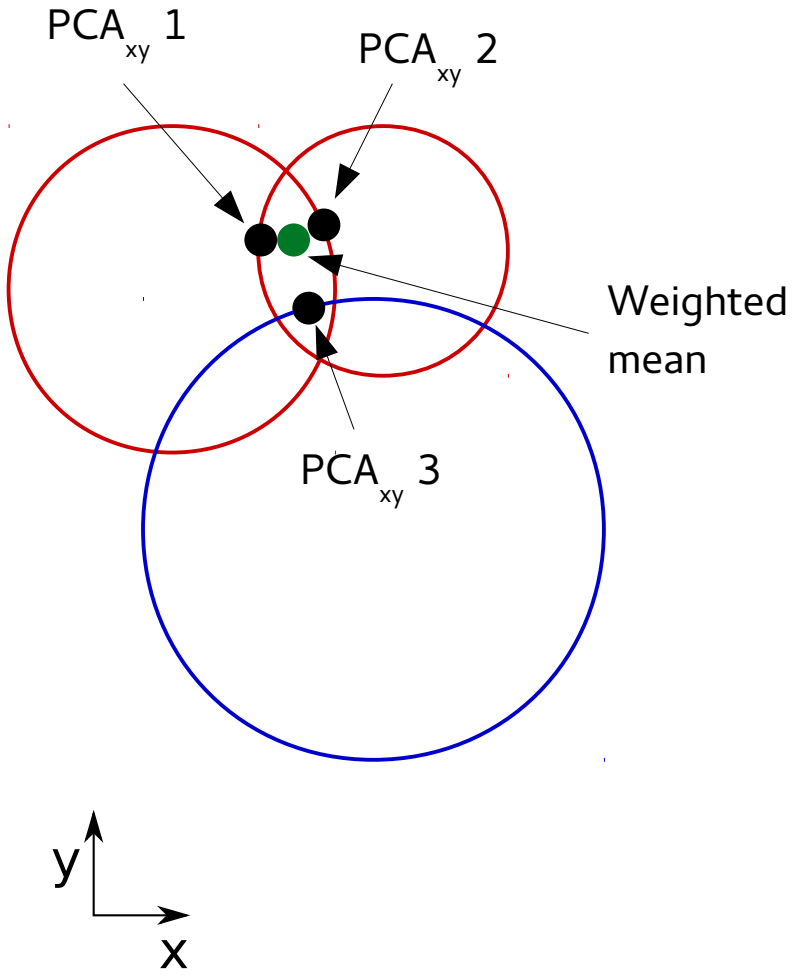
Vertex Estimate: XY-Plane

- Study each combination of two e^+, one e^-
- In xy-plane: find intersections of track circles
- Calculate weights of intersections based on uncertainties due to
  - multiple scattering
  - pixel size



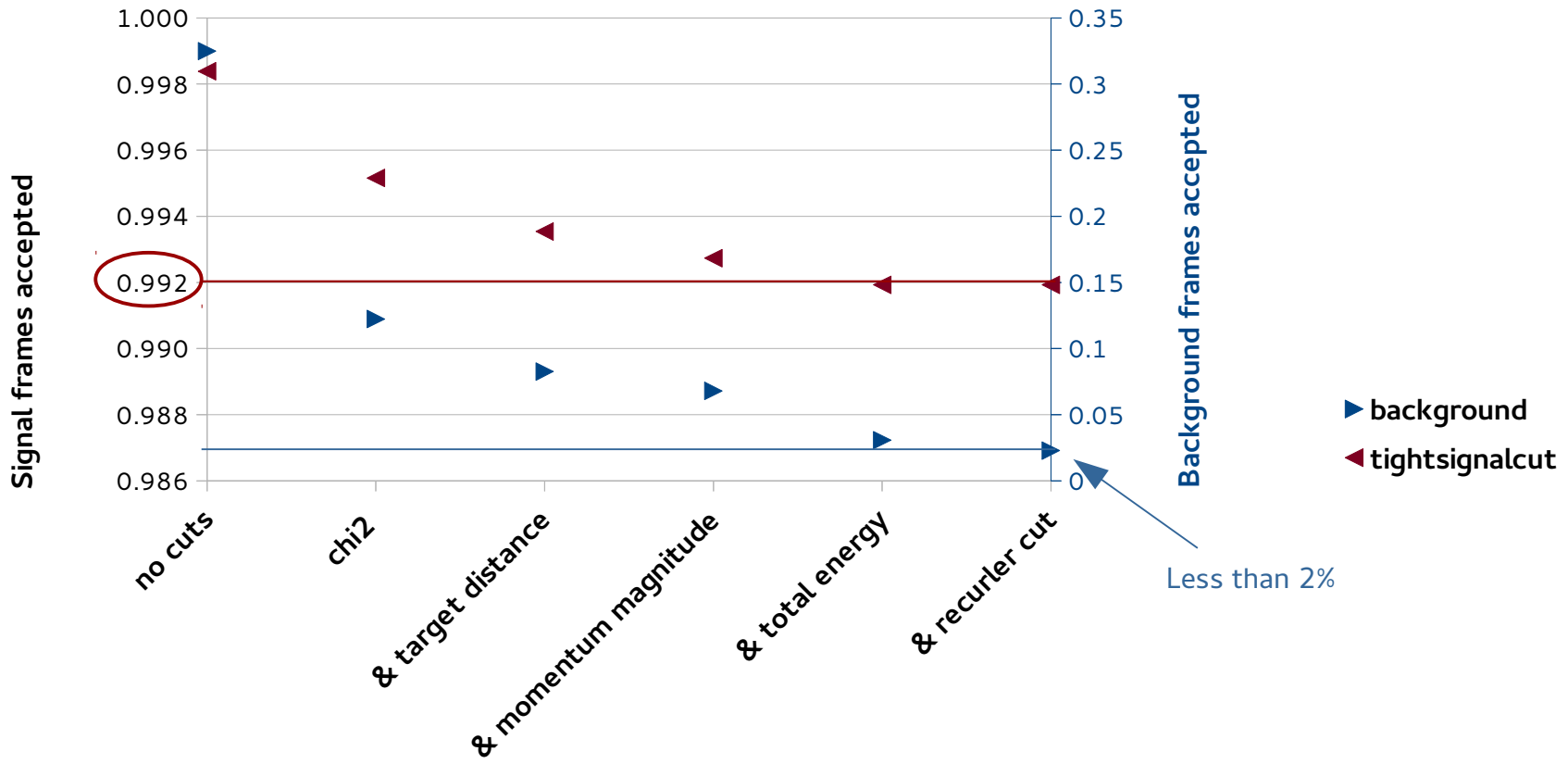
Vertex Estimate

- Calculate weighted mean of intersections from three different tracks
- Find point of closest approach (PCA_{xy}) to weighted mean in xy-plane on each track
- Calculate z-position PCA_z and weight at PCA_{xy}
- Find weighted mean in z-coordinate
- Achieve vertex resolution of ~400 µm sigma



Cut Effects

Signal reference: full offline track reconstruction and offline vertex fit



Sufficient data reduction

Fast Reconstruction on GPU

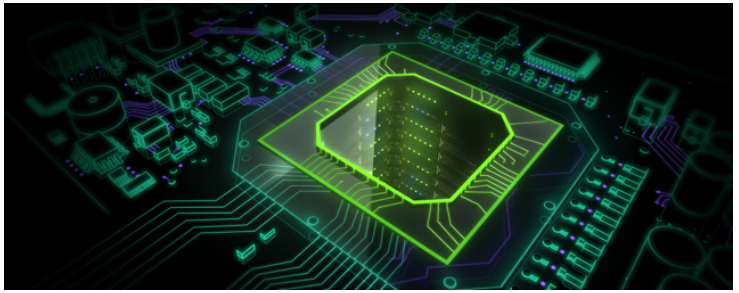
- Use time slices of 50 ns for track & vertex search

  → Process $20 \cdot 10^6$ time slices per second

- Plan for 12 filter farm PCs with one GPU each

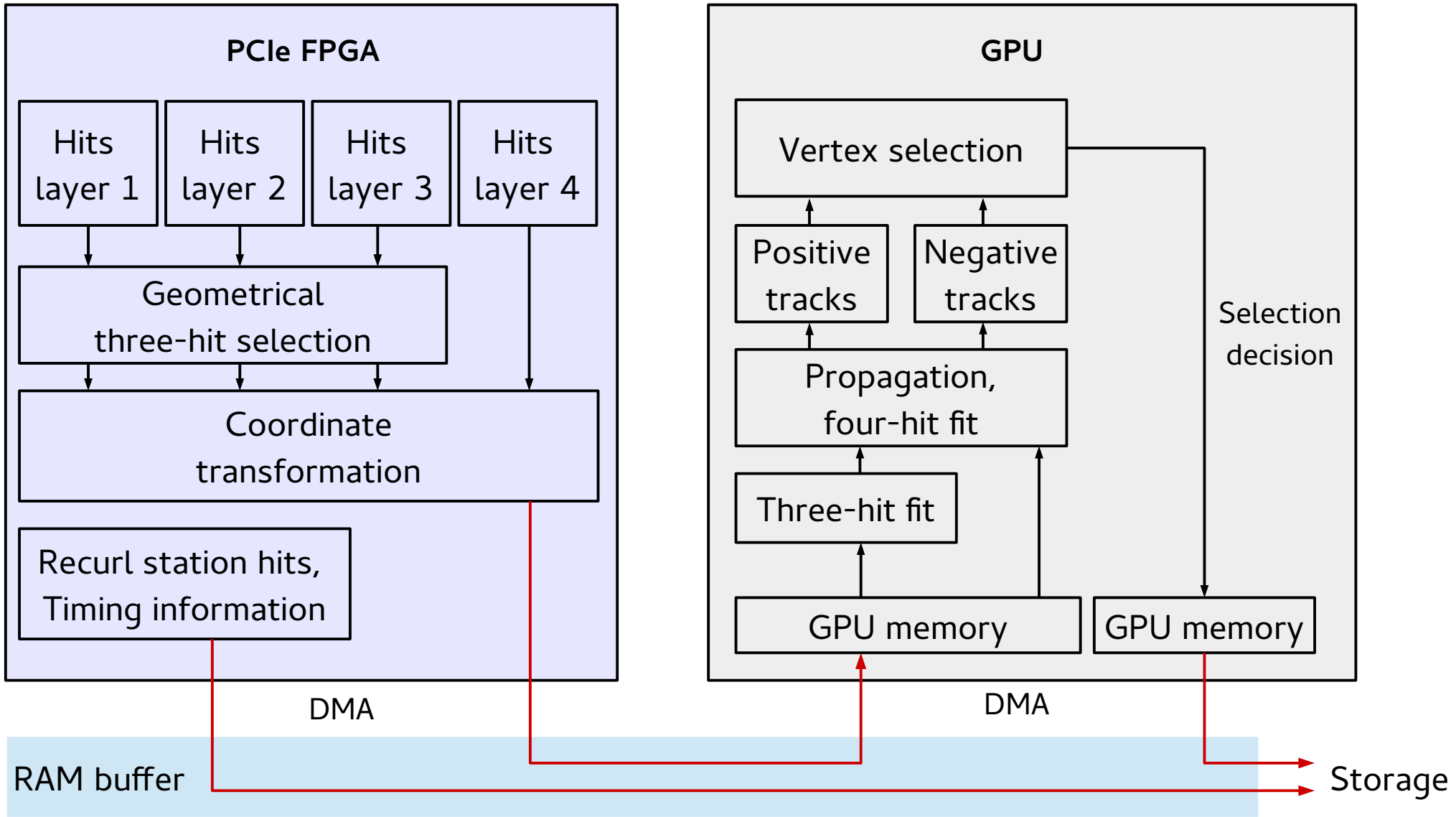
  → Process at least $1.7 \cdot 10^6$ time slices per second

→ **use GPUs**

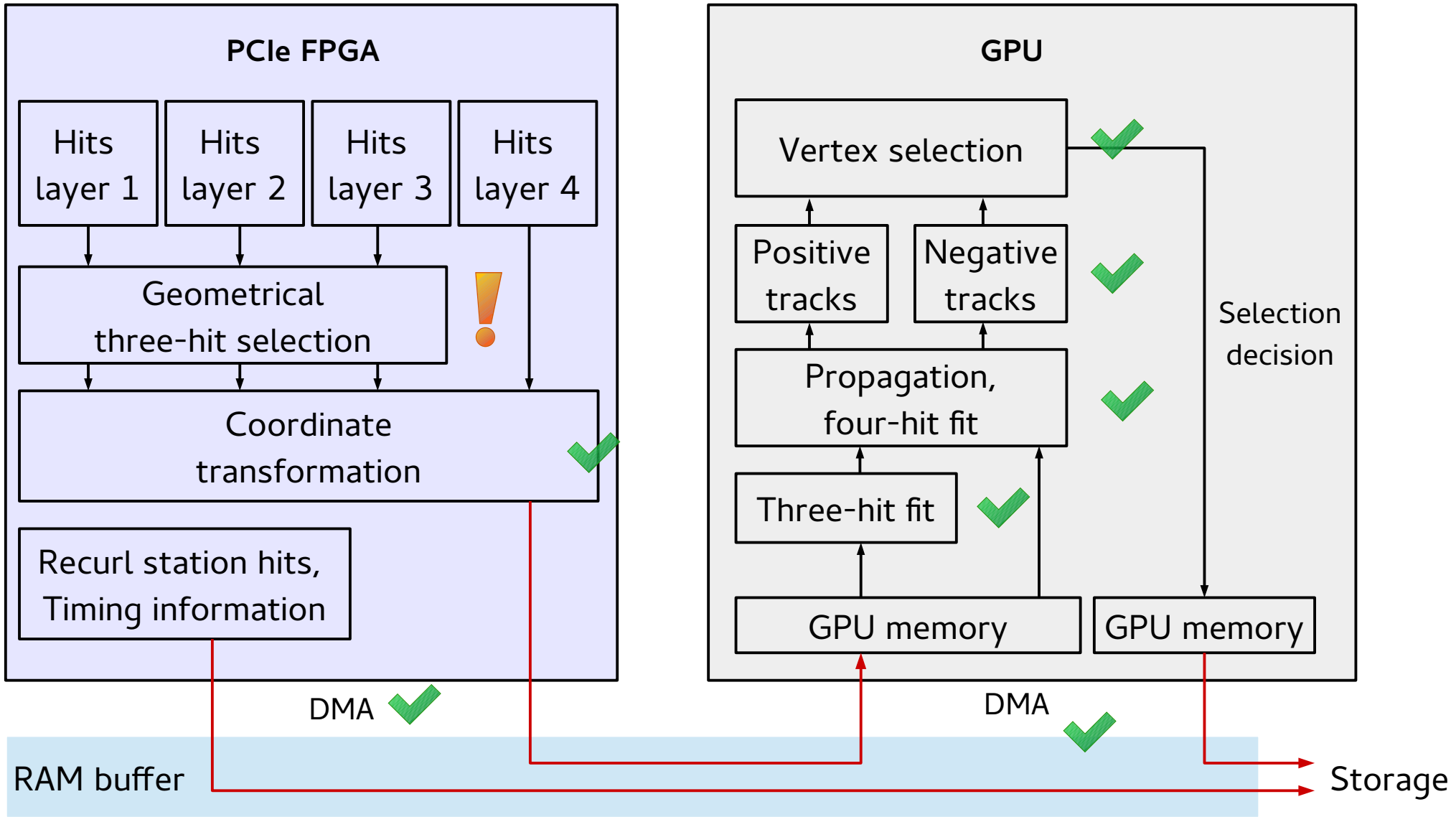


- Thousands of cores
- Optimal parallel performance
- Best suited for many floating-point operations / second

Selection on GPU



Selection on GPU



Performance

Optimizations performed:

- Memory layout and access pattern
- Register usage
- Grid dimensions

Currently process $2 \cdot 10^6$ time slices / s on one nvidia GTX 1080 at a muon stopping rate of $7 \cdot 10^7$ Hz

Backup

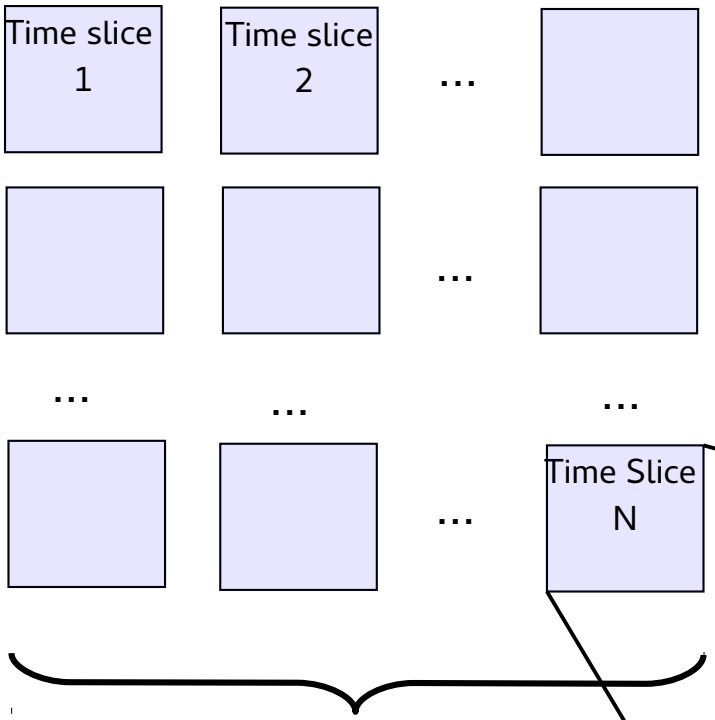
Other Mu3e Talks:

- L. Huth: **Test beam results for neutron and proton irradiated MuPix7 prototypes**, T26, Monday, 17:45
- H. Augustin: **The MuPix8**, HK 18, Tuesday, 11:00
- T. Kar: **Large Area Monolithic Pixel Detectors for HL–LHC & Future High Rate Experiments**, HK 18, Tuesday, 11:15
- J. Kroeger: **Flexprint Design Studies for the Mu3e Experiment**, T46, Tuesday, 12:15
- U. Hartenstein: **Track Based Alignment for the Mu3e Detector**, T89, Wednesday, 17:00
- A.-K. Perrevoort: **Searches for New Physics with the Mu3e Experiment**, T78, Wednesday, 17:35
- A. Herkert: **A Thin Silicon Pixel Tracker for the Mu3e Experiment**, T94, Wednesday, 18:30
- S. Dittmeier: **Readout of the Mu3e pixel detector**, T94, Wednesday, 18:50
- A. Kozlinskiy: **Track reconstruction for the Mu3e experiment**, T116, Thursday, 16:45

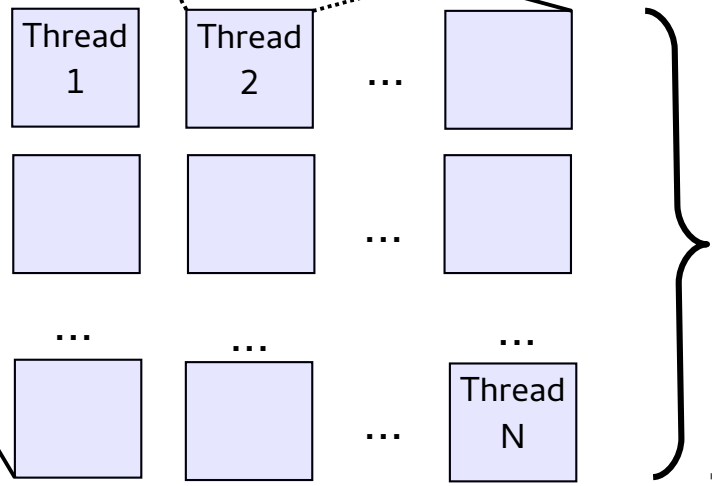
Institutions

- University of Geneva
- Heidelberg University
- Karlsruhe Institute of Technology
- Mainz University
- Paul Scherrer Institut
- ETH Zurich
- University of Zurich

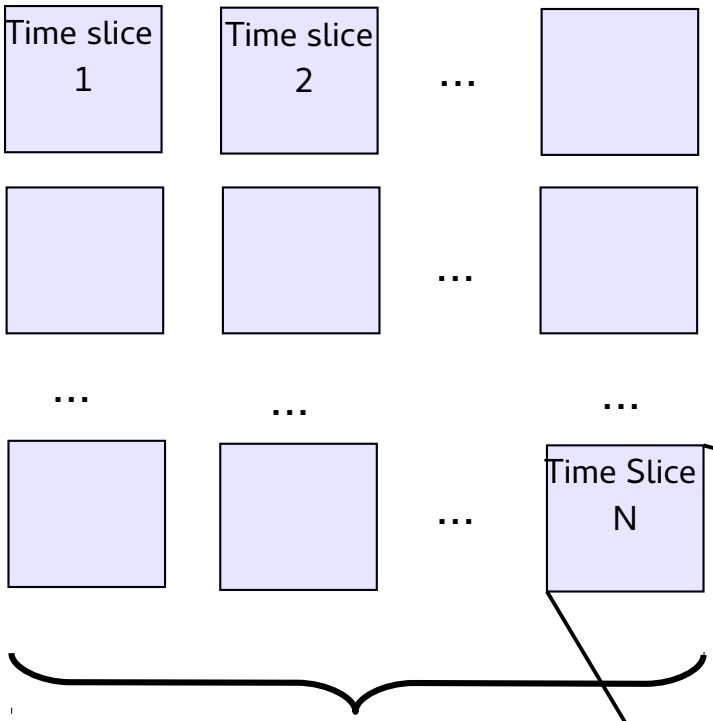
Parallelization Track Fit



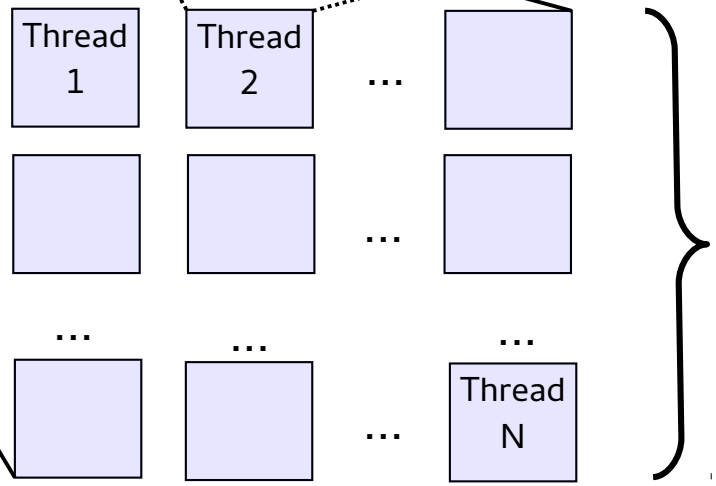
- Fit for one combination of three hits
- Propagation to 4th layer
- Loop over hits in 4th layer: check if hit exists in proximity of propagated track, re-fit
- Wait for all cores in one time slice to be done with previous steps



Parallelization Track Fit

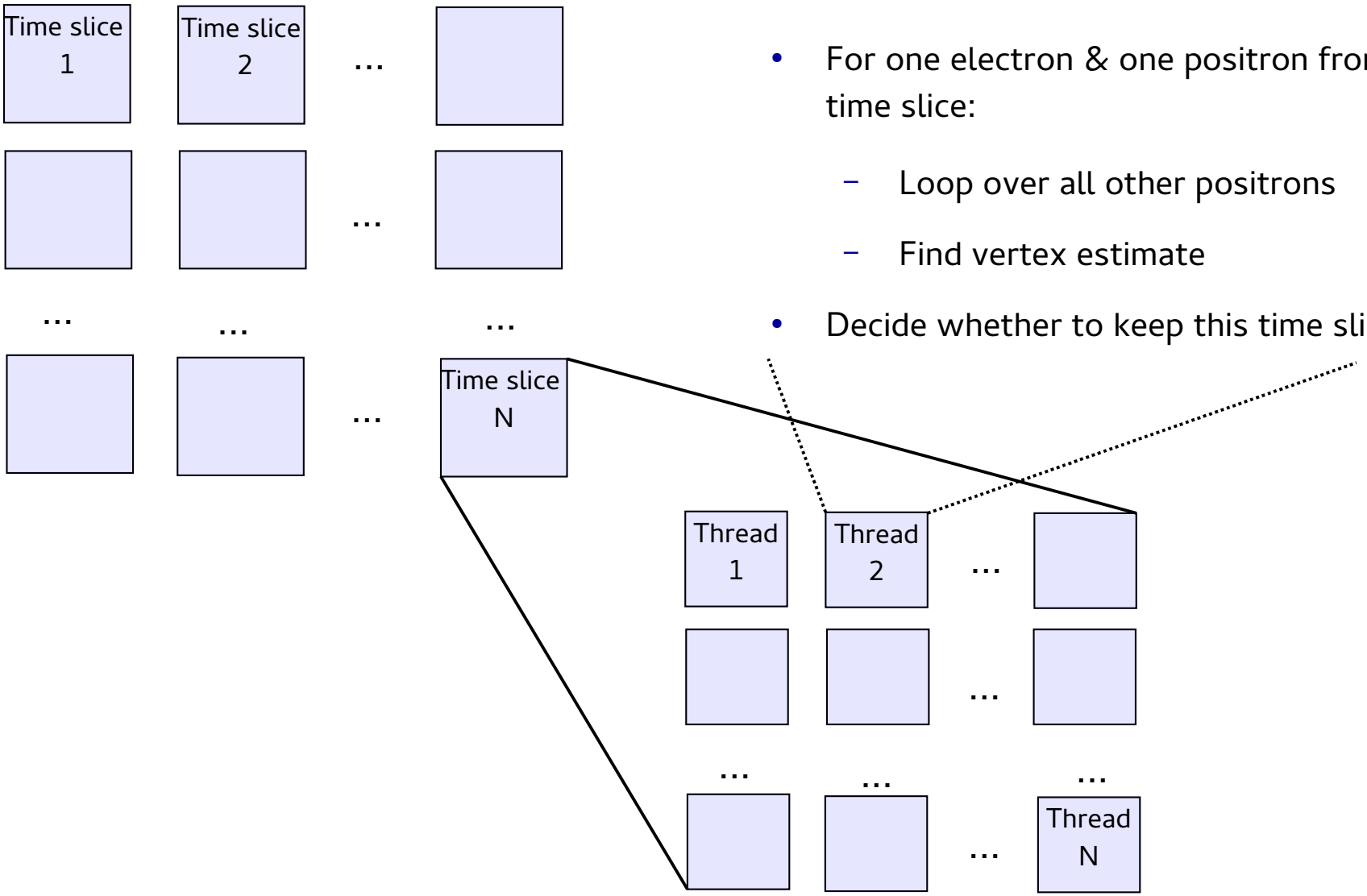


- Fit for one combination of three hits
- Propagation to 4th layer
- Loop over hits in 4th layer: check if hit exists in proximity of propagated track, re-fit
- Wait for all cores in one time slice to be done with previous steps



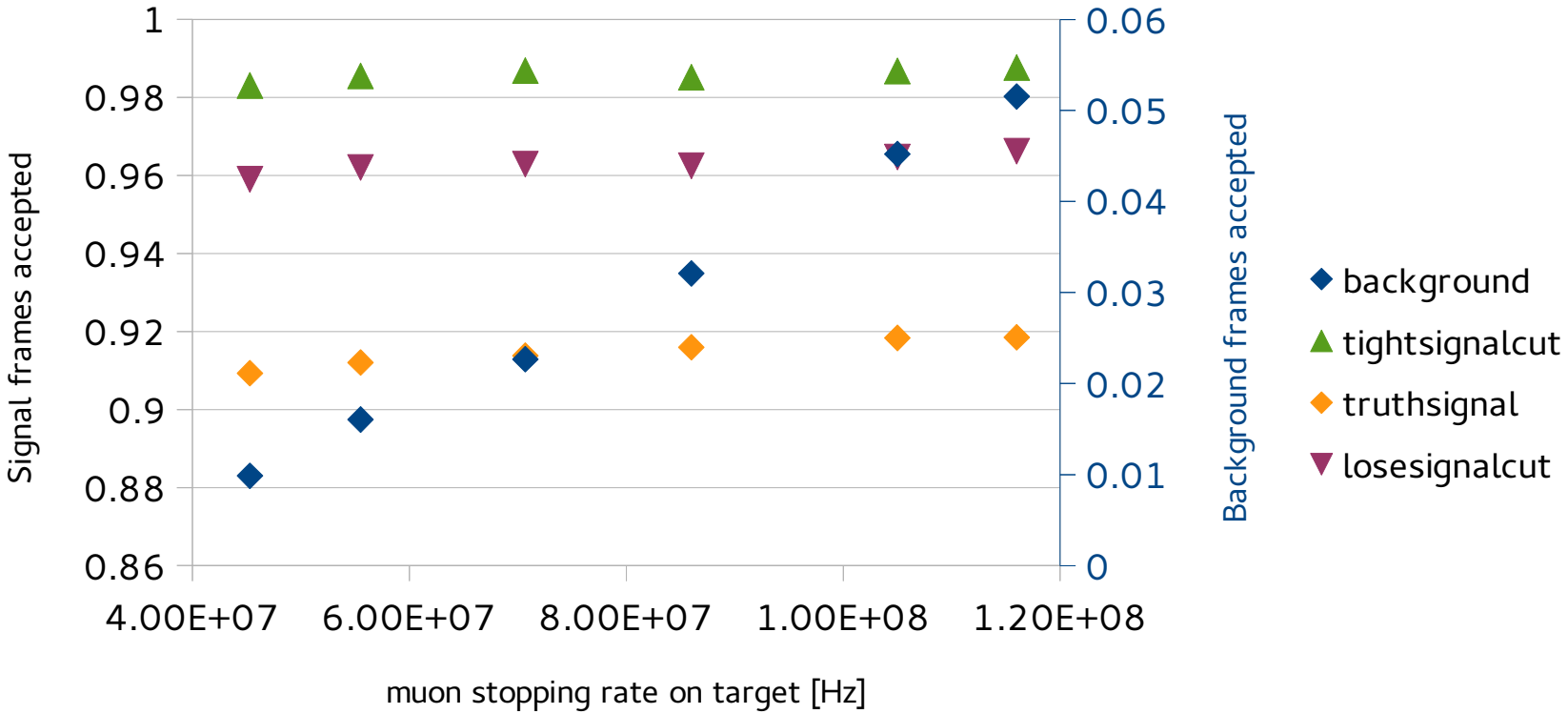
Total of 12.6 million threads to be distributed among 2560 cores

Parallelization Vertex Selection

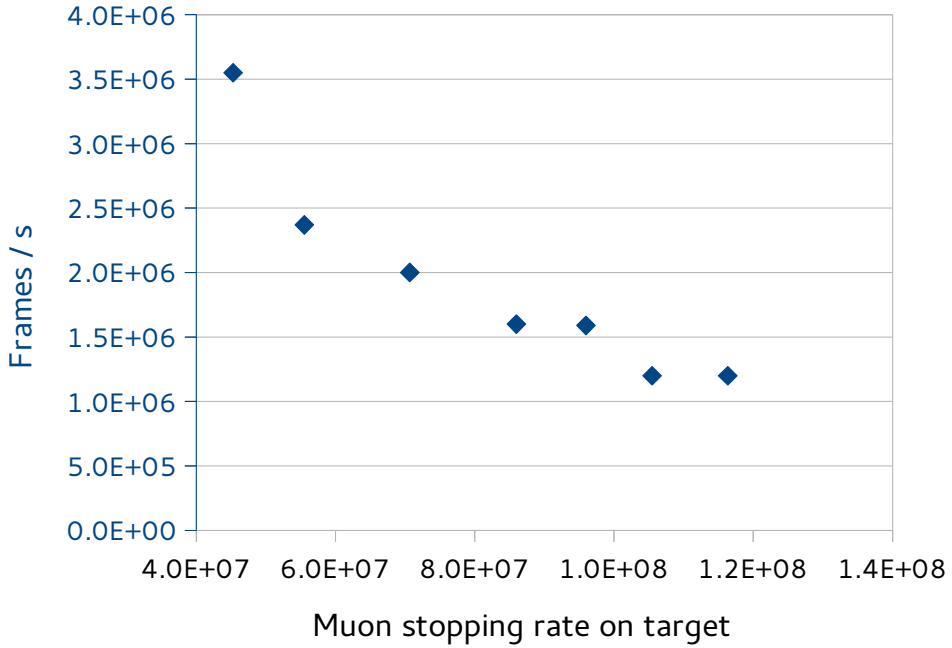


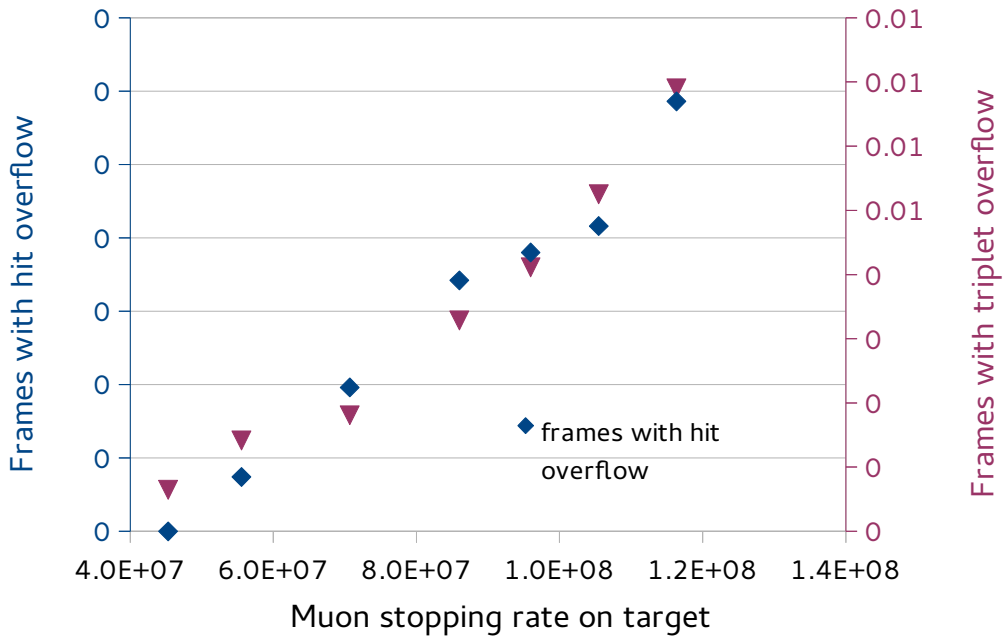
- For one electron & one positron from this 50 ns time slice:
  - Loop over all other positrons
  - Find vertex estimate
- Decide whether to keep this time slice

Muon Stopping Rate Study I

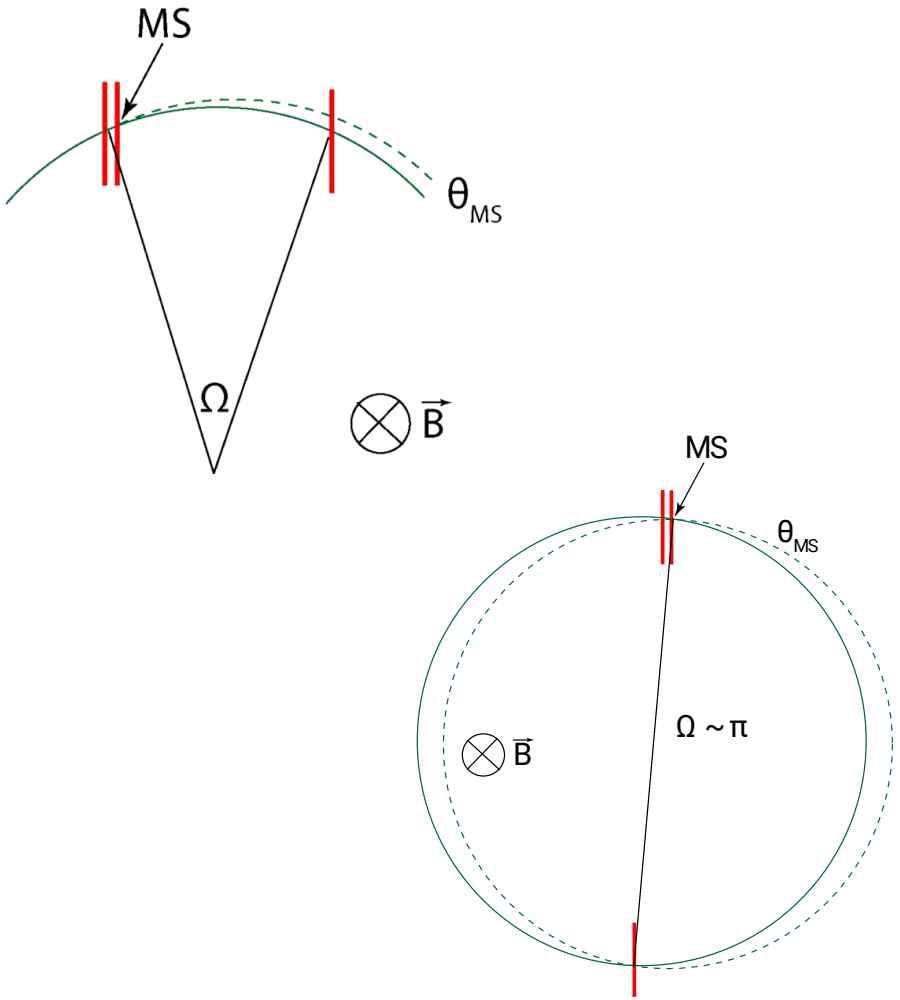


Muon Stopping Rate Study II





Multiple Scattering



- Muons decay at rest

  → momentum < 53 MeV/c

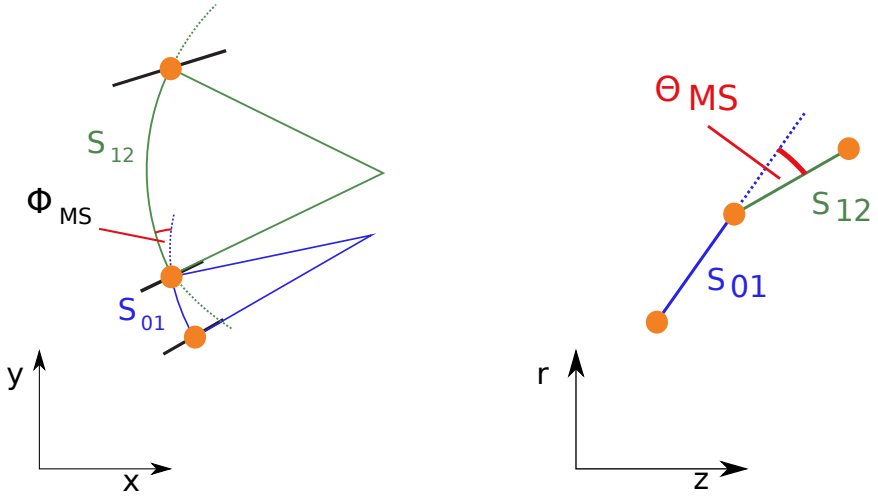
- Momentum resolution to first order:

  $\sigma_p/p \sim \theta_{MS}/\Omega$

- Use recurling tracks for momentum measurement

→ **Minimize material budget**

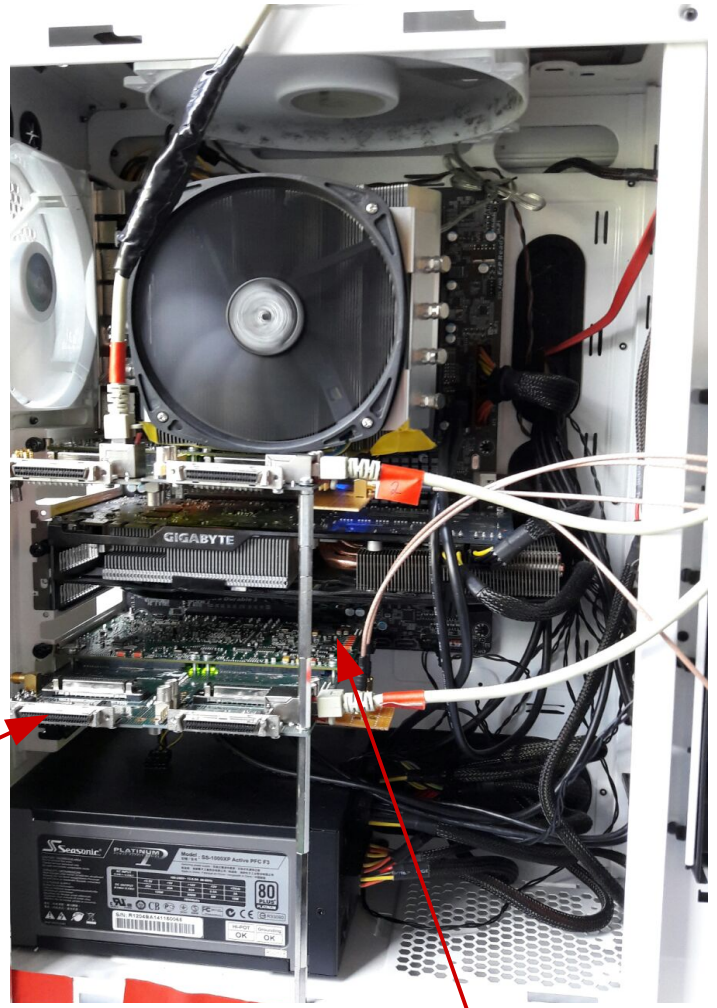
Multiple Scattering Fit



$$\chi^2 = \frac{\Phi_{MS}^2}{\sigma_{MS,\Phi}^2} + \frac{\theta_{MS}^2}{\sigma_{MS,\theta}^2}$$

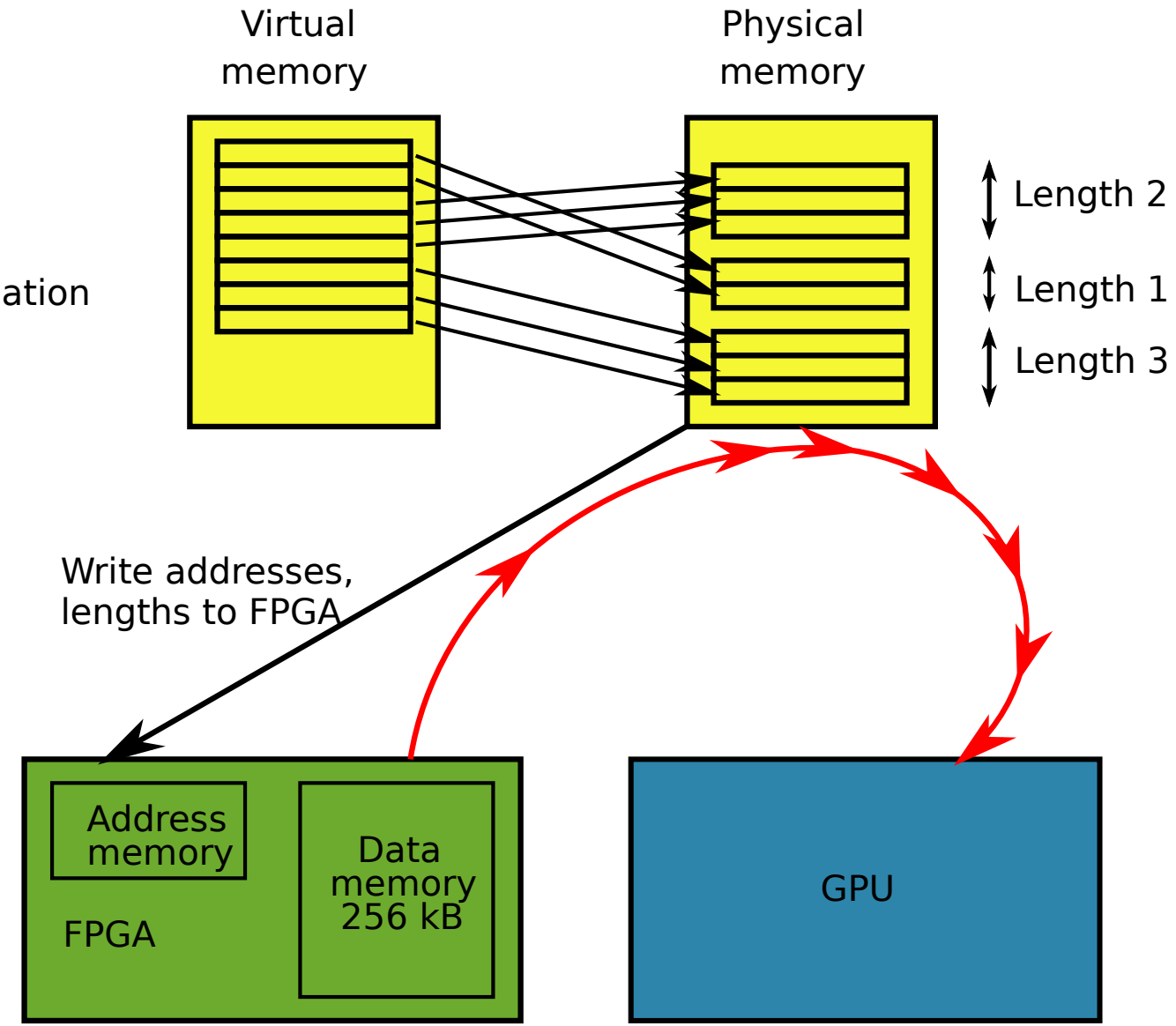
Data Transfer

- Transfer data from FPGA to RAM via direct memory access (DMA)
- Tested at 1.5 GB/s: BER ≤ $4 \cdot 10^{-16}$ (at 95% confidence level)
- Tested on beam test campaigns
- Will be used for readout of next MuPix prototype

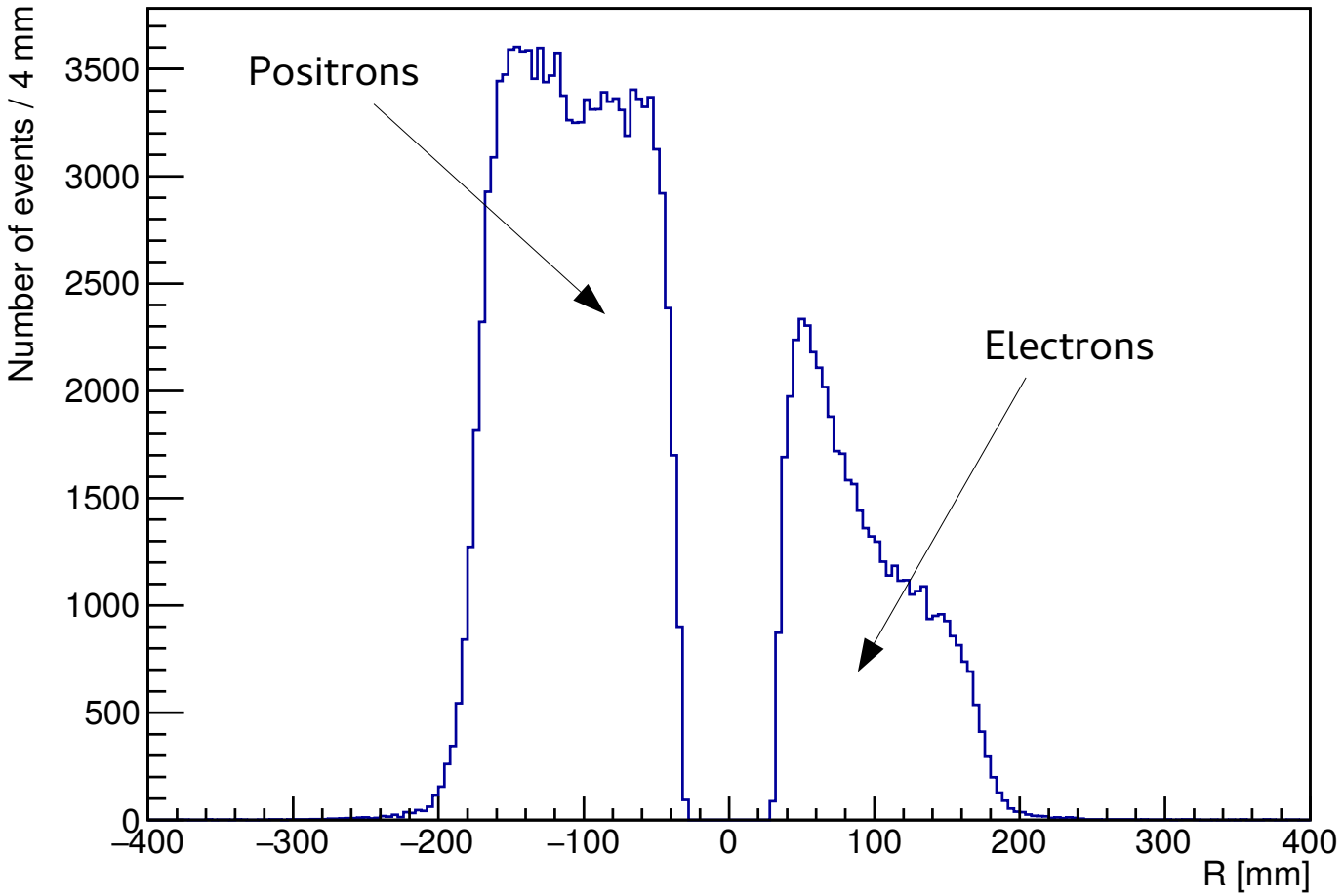


DMA: Implementation

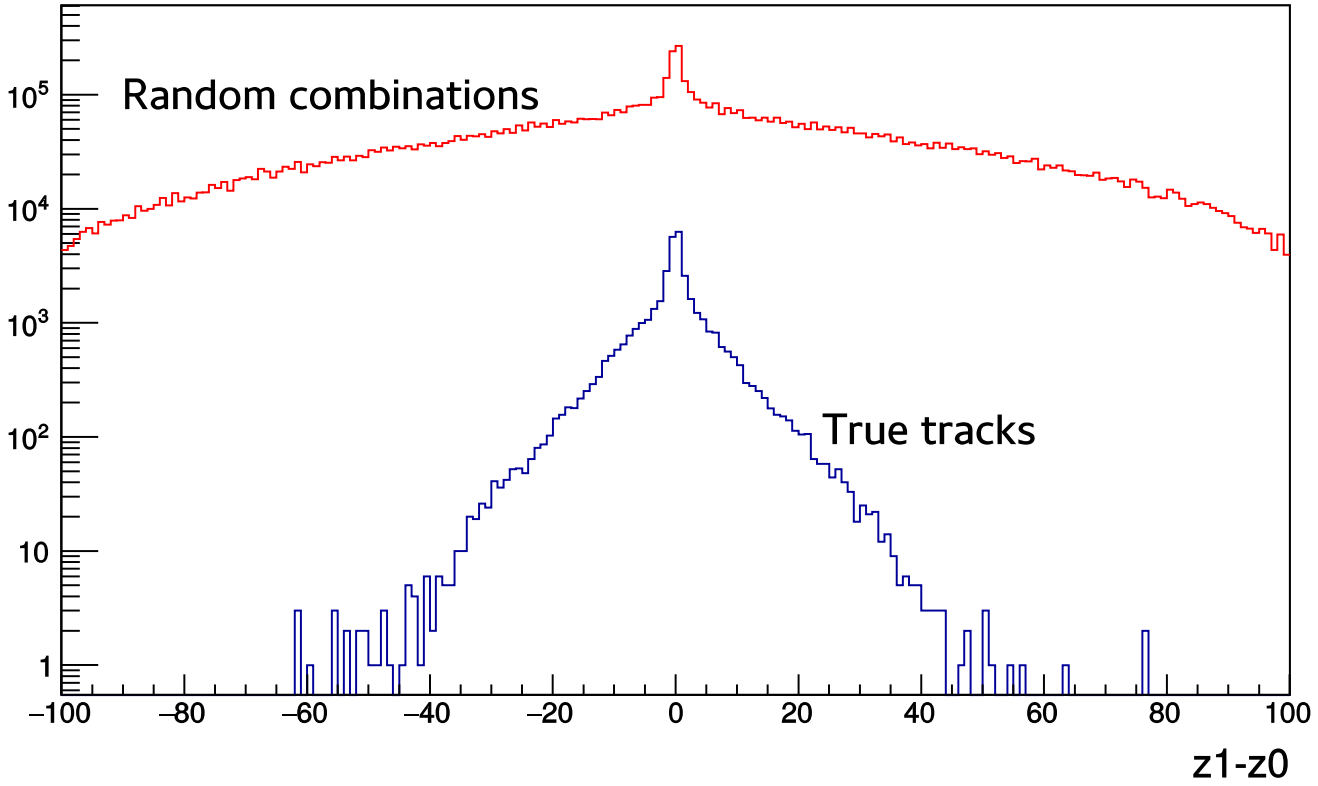
CUDA API:
memory allocation



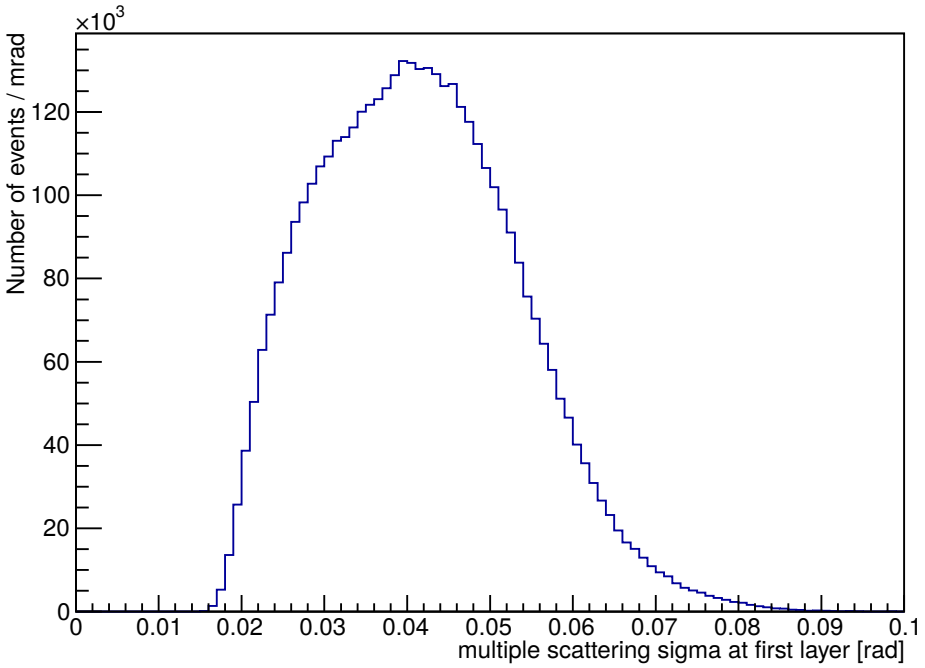
Radius Distribution

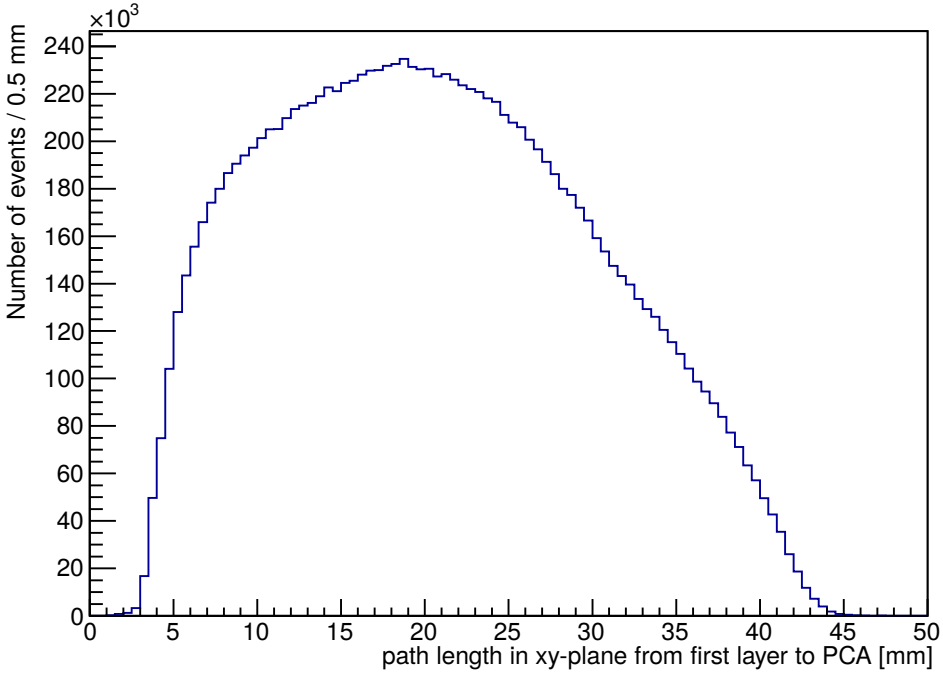


Z distance


Random combinations
True tracks
10^5
10^4
10^3
10^2
10
1
−100
−80
−60
−40
−20
0
20
40
60
80
100
z1-z0

Uncertainty at Intersection



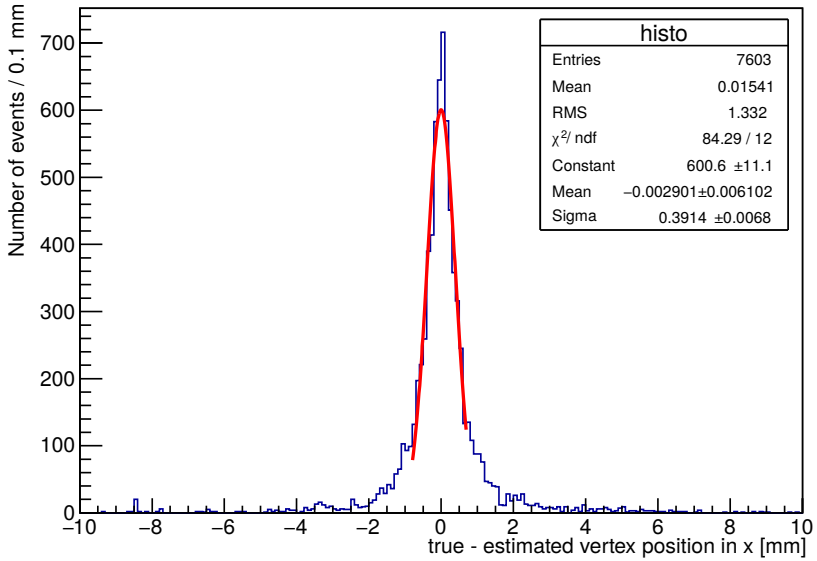


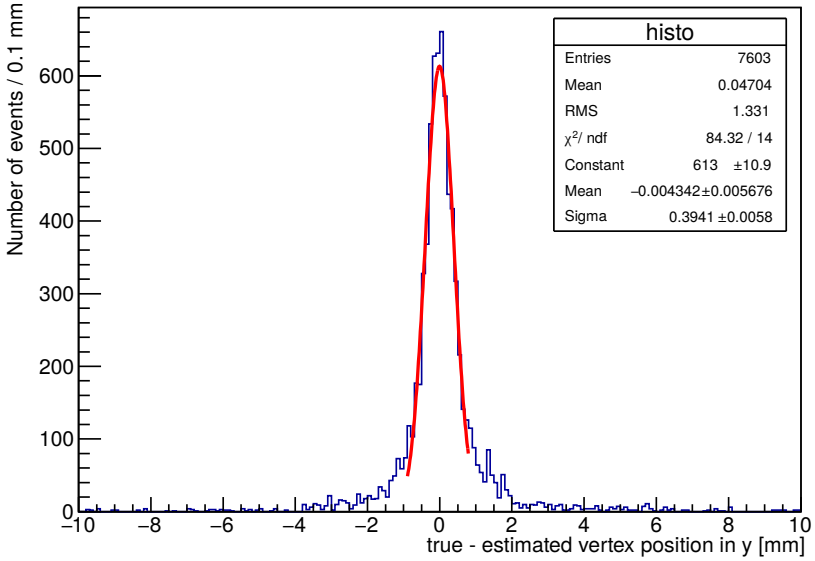
$$\sigma_{MS,PCA} = \sigma_{MS,\,first\,layer} \cdot s \approx 0.8\,mm$$

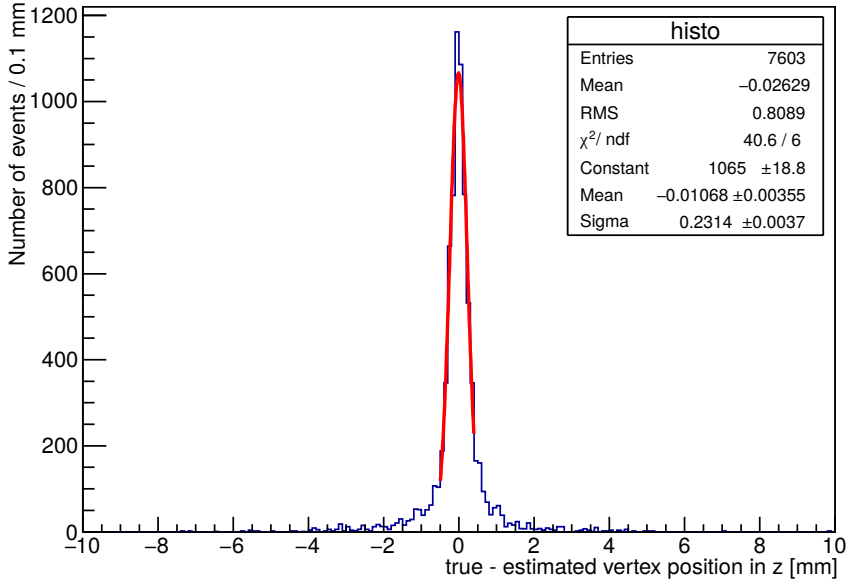
$$\sigma_{pixel} = 0.08\,mm/\sqrt{12} = 0.02\,mm$$

Take both into account when calculating weights

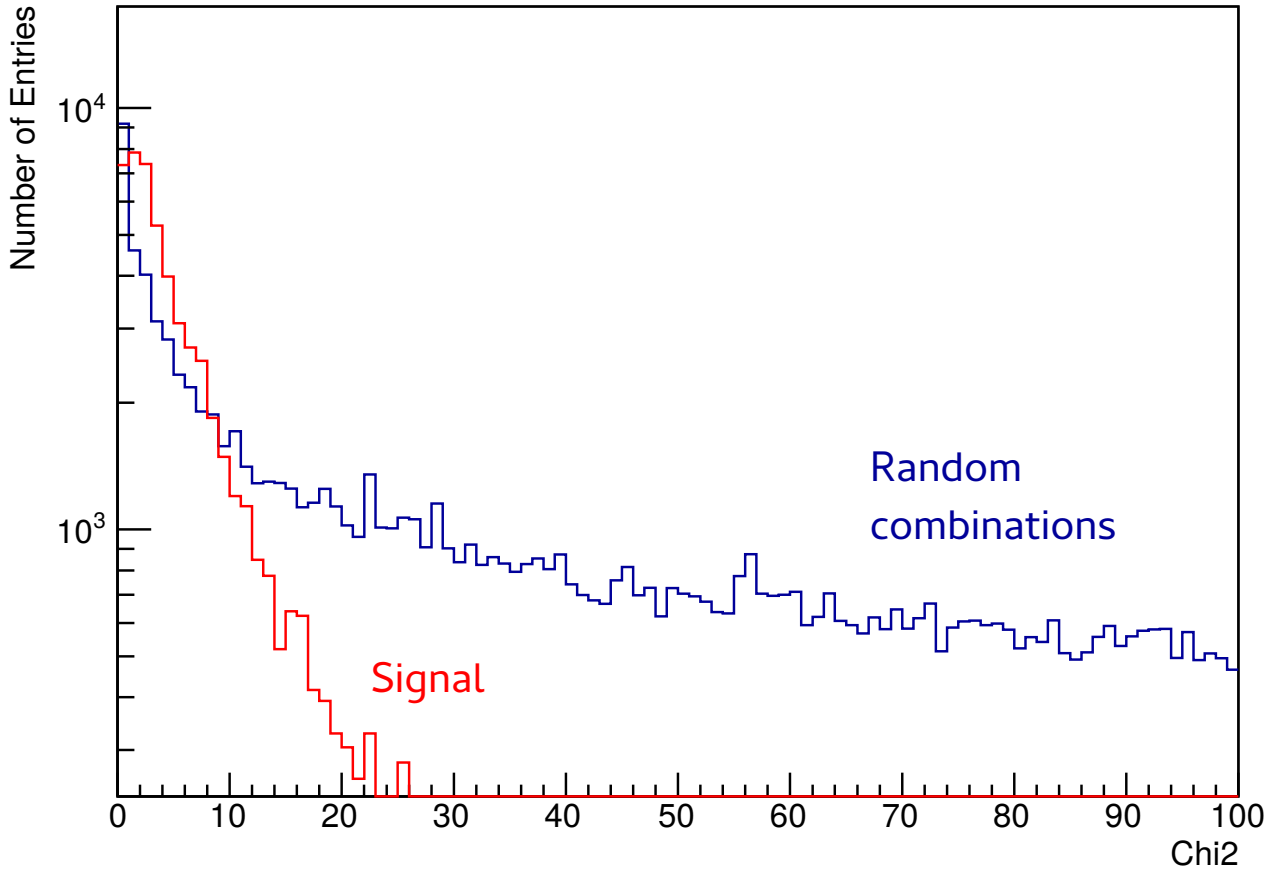
Vertex Position Distribution



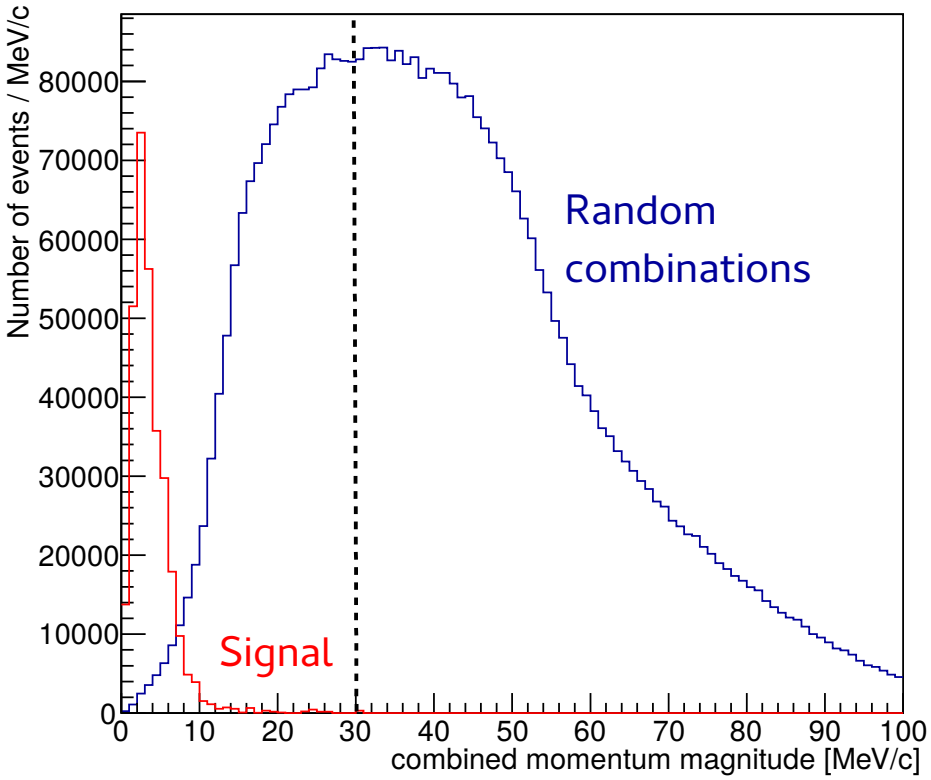


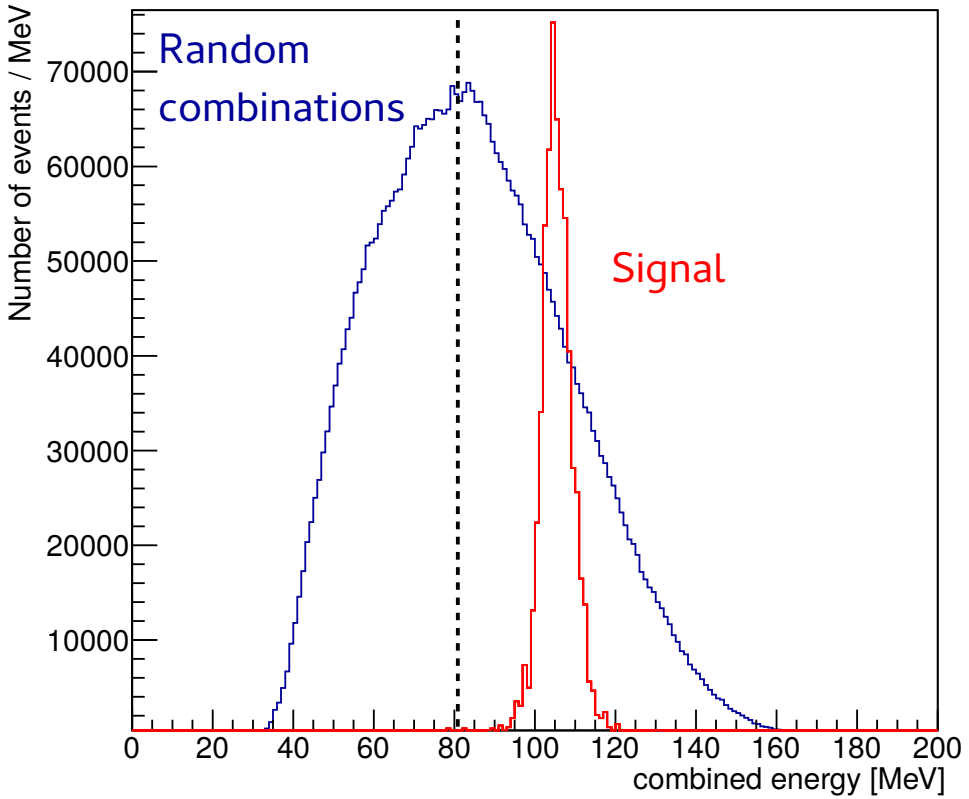


χ^2 Distribution

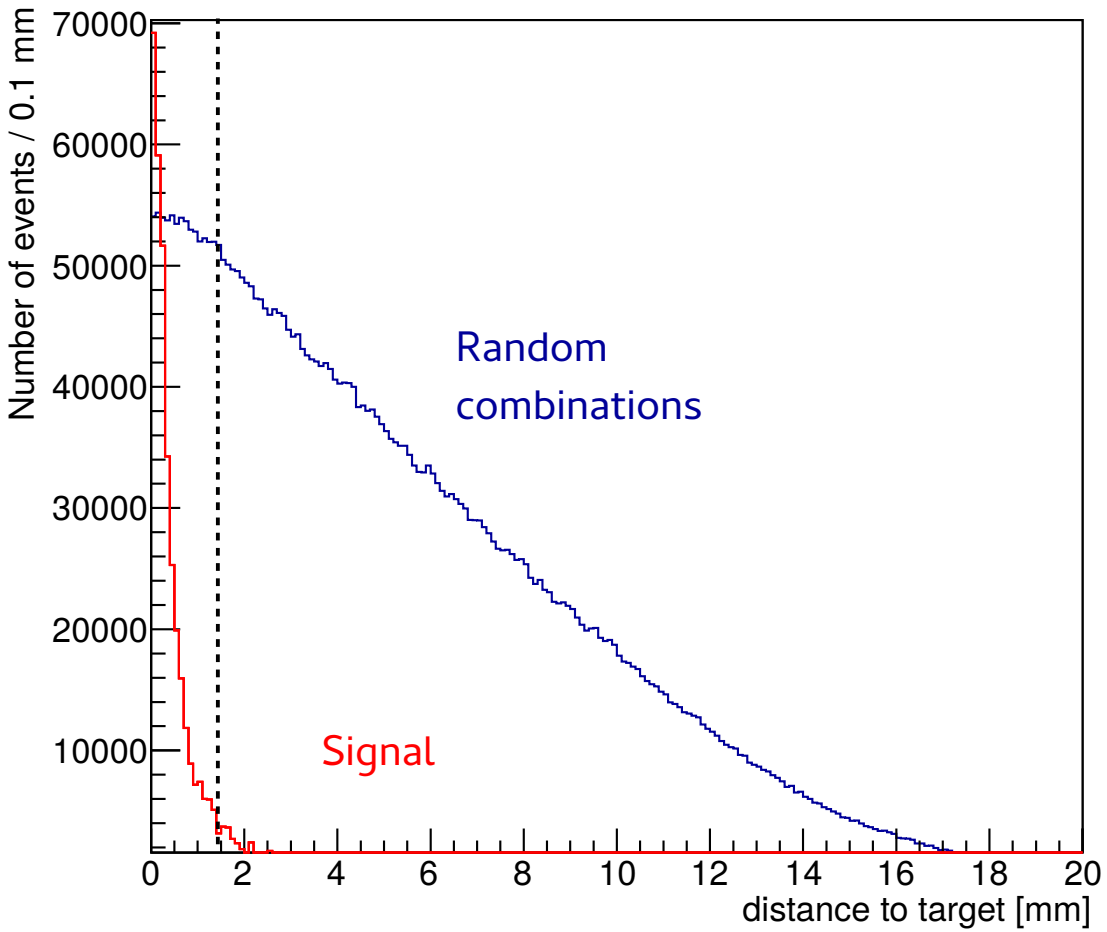


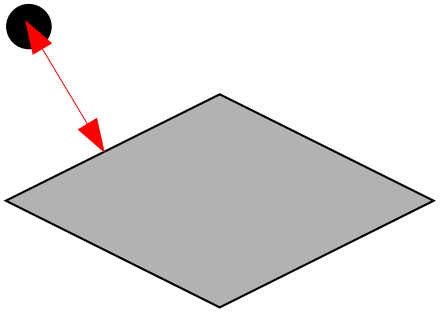
Combined Momentum and Energy





Distance to Target

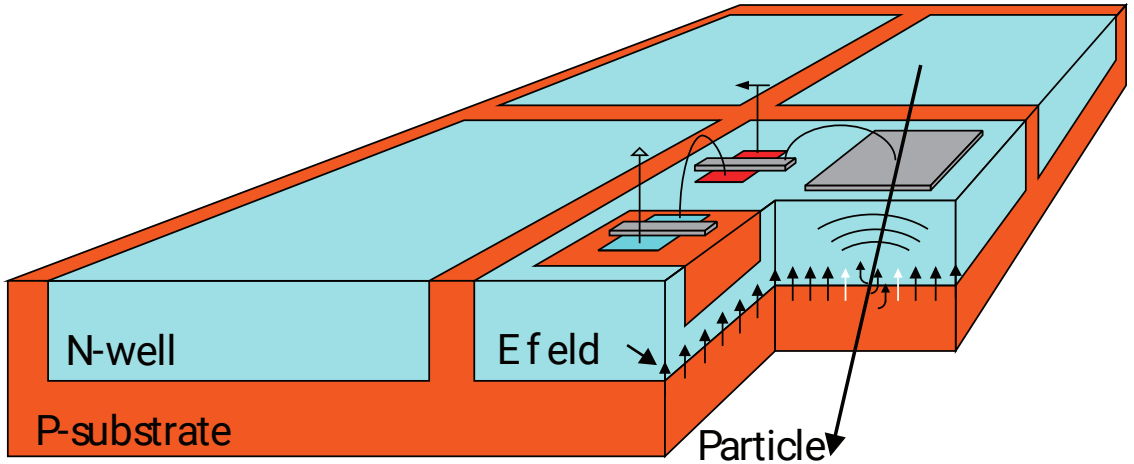




Pixel Detector

- High Voltage Monolithic Active Pixel Sensors (HV-MAPS)
- Fast charge collection via drift
- Thinned down to 50 µm
- Pixel size: 80 µm x 80 µm
- Chip size: 2 cm x 2 cm
- Thickness chip & readout:

  $\mathcal{O}(0.1\ \%)$ radiation length

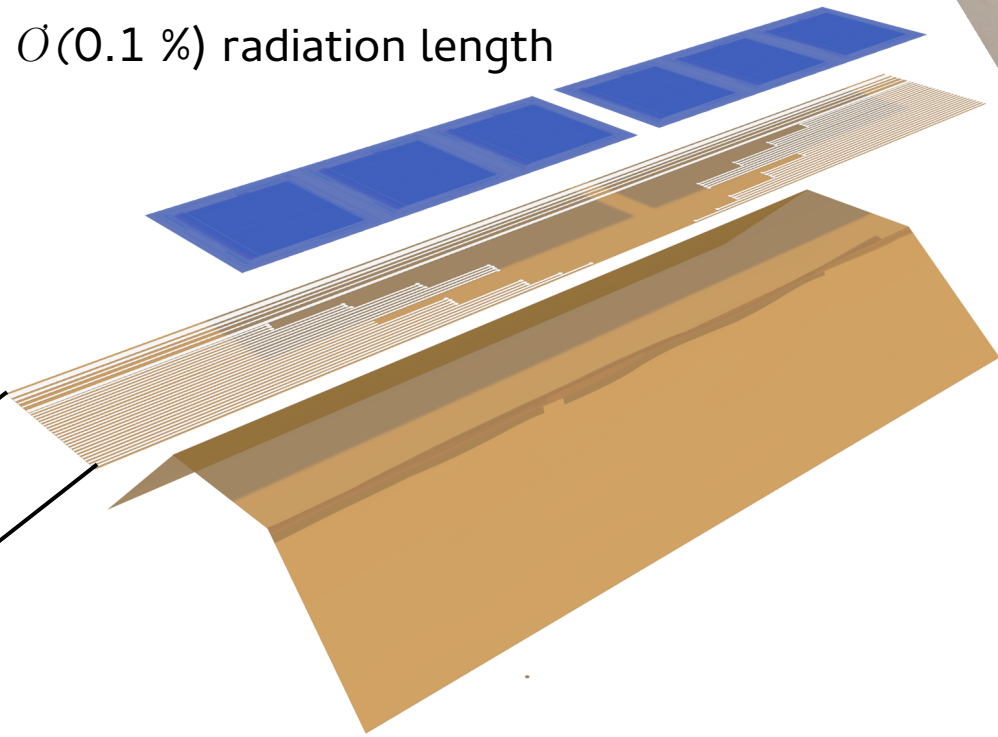


I. Peric, P. Fischer et al, NIM A 582 (2007) 876

Mupix: Mechanics

- 50 µm silicon
- ~ 50 µm flexprint: Kapton, aluminum, copper
- 25 µm Kapton foil

→ $\mathcal{O}(0.1\ \%)$ radiation length



Sensitivity Study