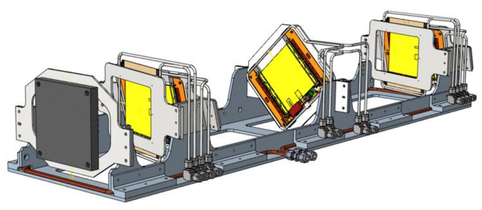
Results for Online Track-fitting in Hardware at 40 MHz

Michael McGinnis [michael.mcginnis@cern.ch], on behalf of the MUonE Collaboration

Supported by NSF award PHY-2111556

MUonE in Brief

- MUonE[1]: using elastic scatter of muons on atomic electrons to measure leading-order hadronic contribution to the muon magnetic moment (a_μ^{HLO})
- Select signal events via outgoing scattering angles, requiring **precise tracking.**

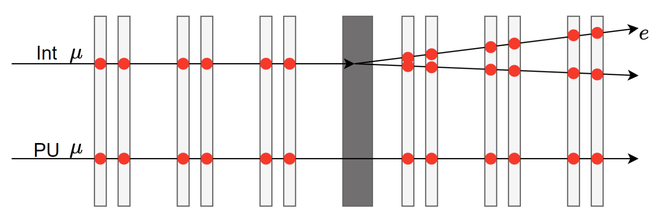


- Tracker composed of 40 stations
  - Each with a thin scattering target and six double-layered strip (2S) modules
  - 2S modules: 90 μm pitch strip sensors, designed for CMS in the HL-LHC[2]
  - Order: (XY)(UV)(XY), (UV) are rotated 45°
- 2S modules readout at 40 MHz to Data Acquisition System (DAQ), with 4x10 Gb/s links to long term storage
  - 40 stations will produce ~600 Gb/s

→**Event selection crucial to reduce data rates**

Event Selection via Hit Patterns

- Select scattering events via increased hits in station following target
- Single interaction:
  - One track and at most one noise hit in both stations
  - At least three additional hits, with one added hit per module, in second station
- Single interaction + N pileup events selected by requiring N additional tracks in both stations
- Selections tested with test beam data, and rate estimated with 40 MHz readout
- **Can save 1μ interaction + pileup at ~0.7 MHz**



Pattern	Rate (MHz)
1μ int.	0.32
Int + 1 PU	0.18
Int + 2 PU	0.06
Int + 3 PU	0.01
Int + ≥3 PU	0.12

Online Track Fitting for Event Selection

Is it possible to perform online track-fitting and vertexing on an FPGA to reconstruct and select for μe scatters?

Currently focusing on implementing online track-fitting in hardware.

- Tracks are straight lines, x and y fits are independent
  - Associate x and y fits with shared UV hits
  - Assuming no multiple scattering

→Use Linear Least Squares (LLS)

LLS 'problem': $m = F\pi + \epsilon$ ← Unknown measurement errors: $Var(\epsilon) = G^{-1}$

- m: hit measurements
- F: Jacobian, from z
- π: fit parameters: slope, intercept

1. Unknown errors: $\begin{pmatrix} m \\ F \end{pmatrix} \rightarrow \sqrt{G} \begin{pmatrix} m \\ F \end{pmatrix}$
2. F generally noninvertible → use QR Decomposition: $F = QR$

   Q orthogonal ↔ $Q^{-1} = Q^T$; R right triangular

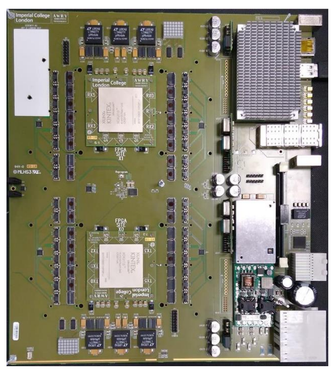
1 + 2 → Instead of calculating π directly, find estimator $\tilde{\pi}$

- $\tilde{\boldsymbol{\pi}} = \boldsymbol{R^{-1}Q^T G_{sqrt} m}$
- $\boldsymbol{Cov(\tilde{\pi})} = \left(\boldsymbol{F^T G F}\right)^{-1}$

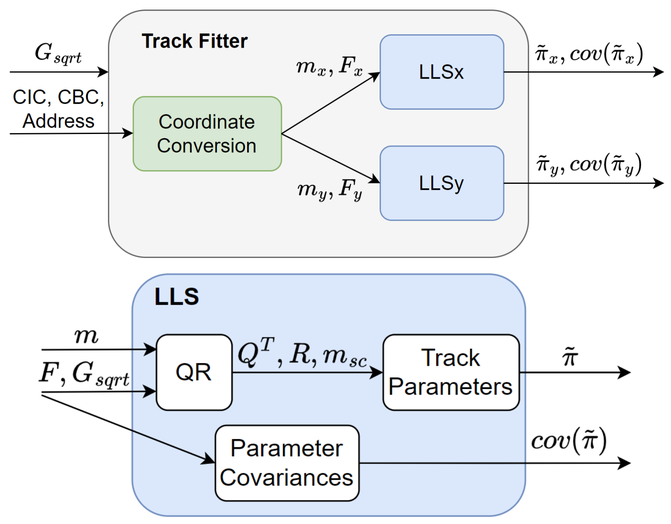
More information in Fruhwirth[3]

Hardware Implementation

- Implemented alongside DAQ in FPGAs mounted on Serenity card
  - An ATCA card developed for CMS tracker for the HL-LHC[4]
  - Up to 2 FPGAs, here Xilinx KU15p

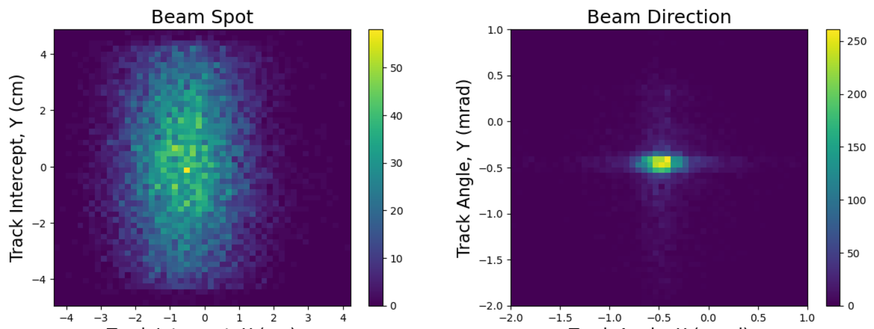


- Use High-Level Synthesis[5] (HLS) to synthesize C/C++ into Register-Transfer Level (RTL) description to program FPGA
  - Verifies firmware via a cosimulation
  - Estimates latency and resource use
  - Produces an IP block to be placed in firmware

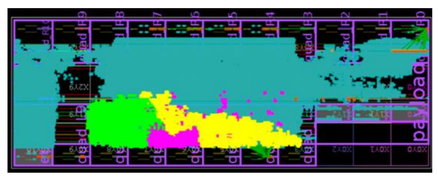


Online Results from Test Beam

- The online fitter ran in a test beam in fall 2023
  - CERN M2 beam, 160 GeV/c muons at peak rate of 50 MHz
  - Fitting tested in a single station
- Preselections applied due to latency and for simplicity
  - Fit only events with a single hit in each module
  - Fit events in first BX out of eight

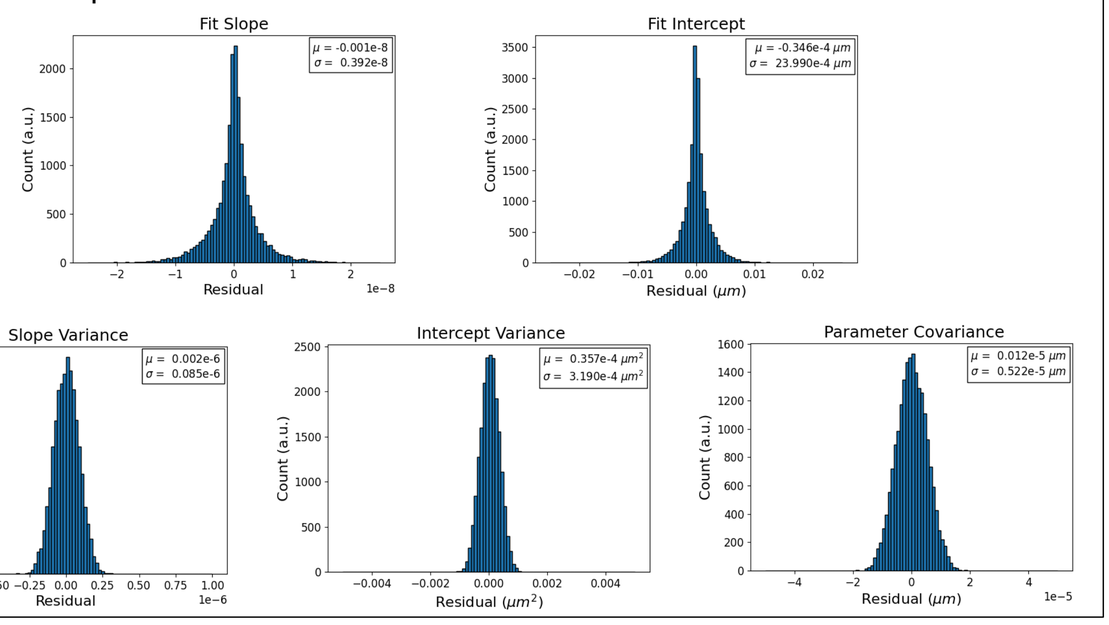


- **Latency of ~1200 cycles at 320 MHz (3.75 µs)**
- The total track fitter required **7.75 % of the resources of a Xilinx KU15P FPGA**
  - Green, yellow are LLSx and LLSy, pink is the Coordinate Conversion and infrastructure. Blue is the rest of the DAQ.



Validating Online Results

- Verify online fitting by comparing to fitting done offline via software
- Using 10k fits from 2023 test run
- Residuals of fitter outputs show agreement
- **Intercept standard deviation is orders of magnitude smaller than 2S module uncertainty of 26 μm**
  - Other distributions have standard deviations similarly small compared to appropriate quantities



Further Development

- Implement pipelining to decrease the time required between two fits (II)
- Replace HLS functions such as QR with custom optimized versions

→**Both tested and show good agreement.**

- **With an II = 6 and 320 MHz clock, fits calculated at ~50 MHz, in line with the peak rate for the CERN M2 beam**

	HLS	HLS +Pipeline	Custom	Custom +Pipeline
Latency (Cycles)	1369	852	701	431
II (Cycles)	1372	470	704	6
LUT Use (%)	7.74	10.82	4.78	4.56
DSP Use (%)	9.30	22.41	6.55	6.66

In progress:

- Further improvements to algorithm such as combining QR and Track Parameters blocks and optimizing data transfer between blocks.
- Implement track finding and **vertexing**
- Integrate pipelined, custom version of the fitter with DAQ for **live testing in summer 2024 test beam**

References

1. G. Abbiendi et al., Measuring the leading hadronic contribution to the muon g-2 via µe scattering, Eur. Phys. J. C 77 (2017) 139, arXiv: 1609.08987 [hep-ex].
2. The CMS Collaboration, The Phase-2 Upgrade of the CMS Tracker, tech. rep., Geneva: CERN, June 2017, url: https://cds.cern.ch/record/2272264.
3. R. Fruhwirth and A. Strandlie, Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors, Particle Acceleration and Detection, Springer, Nov. 2020, isbn: 978-3-030-65771-0.
4. A. Rose et al., Serenity: An ATCA prototyping platform for CMS Phase-2, in proceedings of the Proceedings of Science, vol. 343, 2018.
5. Xilinx, Vitis High-Level Synthesis, version v2022.2, 2022, url: https://www.xilinx.com/products/design-tools/vitis/vitis-hls.html.