Fast Linearized Track Fitting on Parallel Hardware for Track Triggers

Joachim Zinßer

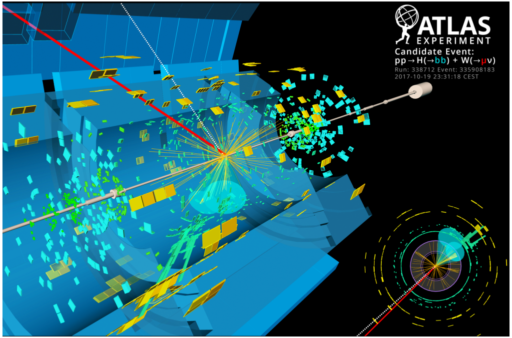
Physikalisches Institut Heidelberg

6th February 2024

The High Luminosity LHC – Scientific Aims

- Explore the *Electroweak Symmetry Breaking*.
- Measure properties of the *Higgs Boson*.
- Measure parameters of the *Standard Model* with high precision.
- Try to look *beyond* the Standard Model.
- Flavor Physics.
- Study of *Heavy Ion Collisions*.

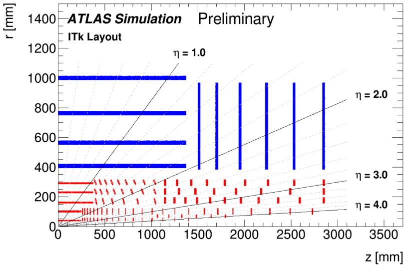
[ATLAS 2018, p. 9,10]



The ATLAS detector observing a Higgs Boson decay into bottom quarks [?].

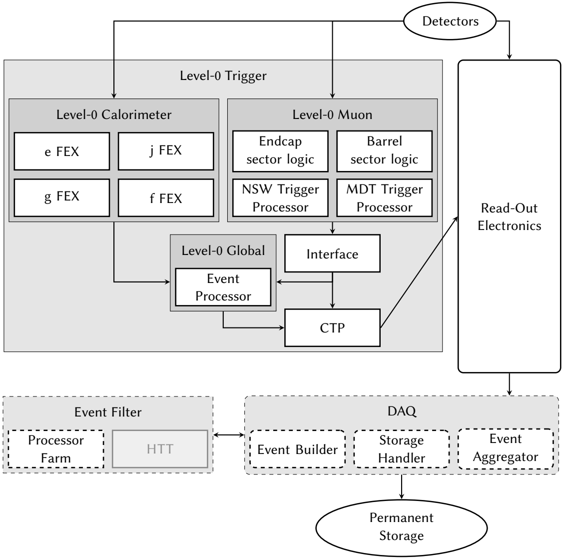
ATLAS – Phase II Upgrade

- Detectors degrade due to radiation.
- Opportunity for Upgrade:
  Installation of the new *Inner Tracker* silicon detector.
- Upgrades of the calorimeters and spectrometers.
- Increase in collisions per crossing from ≈ 40 to ≈ 200
  $\rightarrow$ Necessity to filter interesting events more efficiently
  $\rightarrow$ Trigger system Update



Sketch of the Inner Tracker (ITk) [Moening 2019, p. 5]

Summary of ATLAS-TDAQ Dataflow



- sharp corners: Trigger
- round corners: Data Acquisition
- clean lines: Hardware
- broken lines: Software
- transparent: Discontinued

The Event Filter

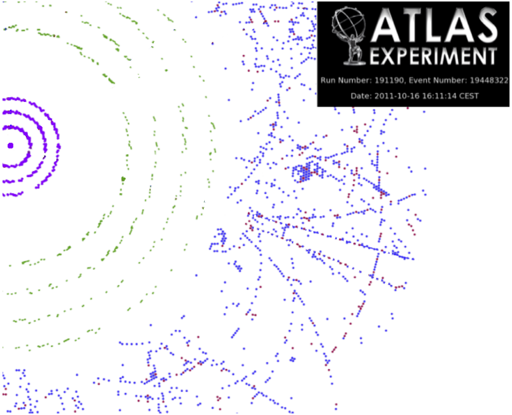
- Refines trigger objects from the Level-0 Trigger.
- Uses offline-like algorithms that scale linearly with luminosity.
- Required tracking rate: 1 MHz regional, 150 kHz global.
- Implementation – 3 options: CPU only, heterogeneous CPU-GPU, CPU + HW accelerators.
- Decision based on:
  - technical complexity
  - estimated performance
  - opportunities for improvement
  - associated risks

Motivation for Hardware Track-Fitting

- High pileup conditions (200 interactions / bunch crossing) [ATLAS 2018, p. 39].
- Analyze up to 13 detector layers simultaneously [?, p. 15,23].
  $\rightarrow$ Very easy to parallelize hardware.
- High frequency 1 MHz Level-0 trigger rate [ATLAS 2018, p. 13]. $\rightarrow$ Custom hardware can be the fastest.
- Custom hardware is expensive to develop but cheap to produce.
- Custom hardware saves energy-costs.

The Goal of the Track Fitter

- Goal: Estimate the helix-trajectory for a group of hits in the *Inner Tracker* detector.
- Calculates a linear approximation of the five helix parameters.
- Identifies viable track-candidates via the goodness of their fits.

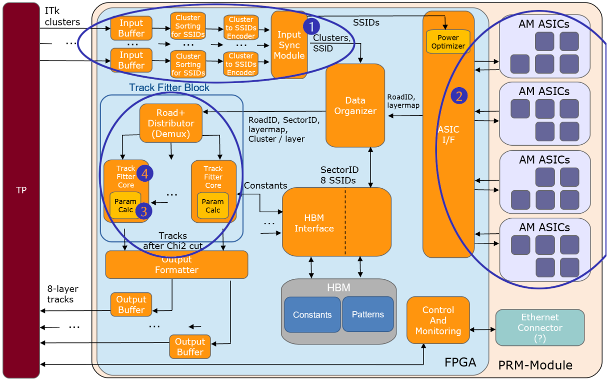


A close-up view perpendicular to the beam direction of an ATLAS event [ATLAS 2017].

Principle Steps of Track Reconstruction

1. Data Preparation
2. Track Finding
3. Track Fitting
4. Quality Assessment
5. Vertex Reconstruction

Possibly repeat steps 2 – 5.



Firmware Diagram of HTT's Pattern Recognition Board.

Track Fitting

Use n measured values m_n to find a set of parameters $p_i^{\text{fit}}(\vec{m})$ so that:

$$\sum_i \left(p_i^{\text{fit}} - p_i^{\text{true}}\right)^2 \rightarrow \text{Minimum}$$

Usually p_i^{true} are unknown. Compare the track model $\vec{f}$ with the measurement $\vec{m}$:

$$\left(\vec{f}\left(\vec{p}^{\text{fit}}\right) - \vec{m}\right)^{\mathsf{T}} V^{-1} \left(\vec{f}\left(\vec{p}^{\text{fit}}\right) - \vec{m}\right) \rightarrow \text{Minimum}$$

Model: Charged Particle in Homogeneous B-Field

- A homogeneous magnetic field $\vec{B}||\vec{z}$.
- Lorentz Force of a particle with mass m and charge q is $\vec{F} = q\left(\vec{v} \times \vec{B}\right)$
- Neglect energy-loss of the particle.
- Solve differential equation $\rightarrow$
  - No change of v_z.
  - Rotation in x-y-plane $\vec{v}_\mathrm{T}$.
- Integrate the solution $\vec{v}$ to get the helix-trajectory $\vec{x}$.

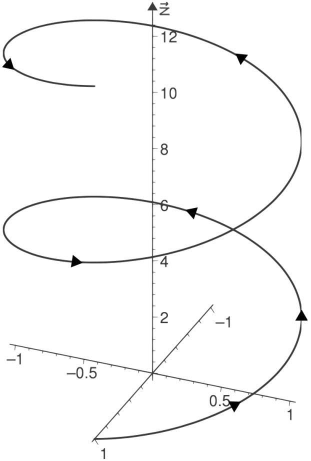
The Helix Trajectory

The Helix-Track

$$\vec{x}(t) = \begin{pmatrix} r\sin\left(\omega_B\, t + \psi_0\right) + x_0 \\ r\cos\left(\omega_B\, t + \psi_0\right) + y_0 \\ v_z\, t + z_0 \end{pmatrix}$$

With the

- cyclotron-frequency $\omega_B = \frac{qB}{\gamma m}$
- radius $r = \frac{v_z}{\omega_B}$
- rotational velocity $v_T = \sqrt{v_x^2 + v_y^2}$ and the
- translation $\vec{x}_0$.

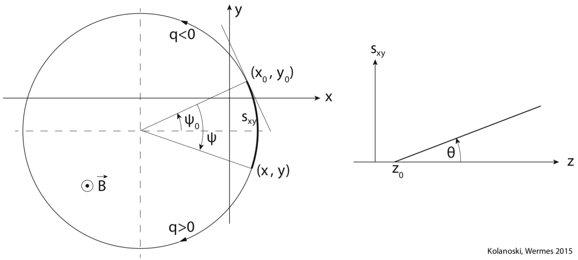


A helix [?].

The Helix Trajectory

The detector does not resolve the time-development of the track.

$$\vec{x} = \begin{pmatrix} x_0 + r\left(\cos\left(\psi_0 - \psi\right) - \cos\psi_0\right) \\ y_0 + r\left(\sin\left(\psi_0 - \psi\right) - \sin\psi_0\right) \\ z_0 + \frac{r\psi}{\tan\theta} \end{pmatrix}; \text{with } \psi = \omega_B t \text{ and } \psi = 0 \text{ at } \vec{x} = \vec{x}_0.$$

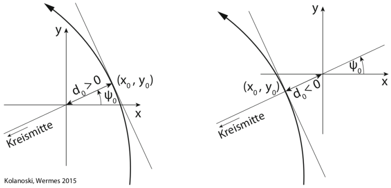


Projections of the helix in the *x*-*y*-plane and a *z*-plane [Kolanoski 2017, p. 393]

Track Parameters of a Helix

- With concentric barrel-detectors ψ_i, r_i and z_i are measured.

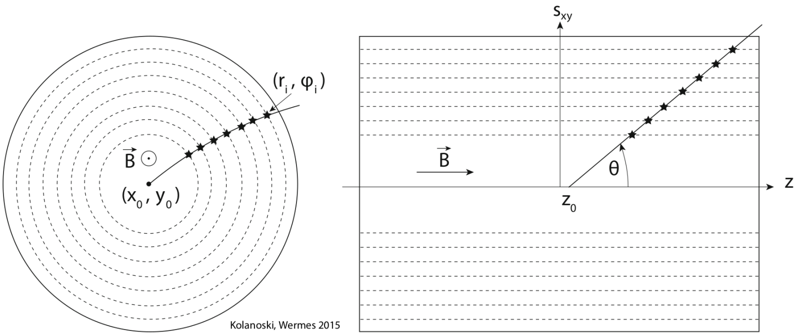
- Five parameters can be estimated (example):
  - The curvature $\kappa = -\frac{\text{sig}(q)}{r}$.
  - The angle ψ_0.
  - The distance d_0.
  - The inclination θ of the track against $\vec{z}$.
  - The z coordinate.



Sketch of the origin in the r-ψ-plane [Kolanoski 2017, p. 399].

- The origin of the helix cannot be reconstructed.
  But it is assumed to be the nearest point to the center in the r-ψ-plane.
- Therefore, $x_0 = d_0 \cos \psi_0$ and $y_0 = d_0 \sin \psi_0$.

Linearization of the Track Model



- Helix Projects to a circle $y = y_0 + \sqrt{r^2 - (x - x_0)^2}$.
- Taylor for large radii $y = \left(y_0 + r - \frac{x_0}{2r^2}\right) + \frac{x_0}{r}x - \frac{1}{2r}x^2 + \ldots$.
- Approximation of the Track $y = a + bx + cx^2$ that has three functions a, b and c of the parameters x_0, y_0 and r.
- With data y_i at the positions x_i linear fit of a, b and c is possible.

Track Fitting Workflow

Offline preparation:

1. Perform helix fit on all possible tracks.
2. Estimate goodness of the fit (χ^2-method).
3. Linearize the fit and the χ^2 for small variations in the data.
4. Combine similar patterns into sectors to save storage space.
5. Save those constants $\vec{S}_i$, h_i, $\vec{C}_j$, and q_j in a database.

In the Track Fitter:

1. Retrieve the constants.
2. Calculate the goodness of a track candidate $\vec{x}$ with $\chi^2 = \sum_{i=0}^{I} \left(\vec{S}_i \cdot \vec{x} + h_i\right)^2$.
3. For good tracks, calculate the helix parameters $p_j = \vec{C}_j \cdot \vec{x} + q_j$.

Full Scale Requirements

Requirements of the baseline Hardware Track Trigger as part of the Event Filter System [Camplani 2023]:

- 48 Hardware Track Trigger (HTT) units with twelve Pattern Recognition Mezzanines (PRM) each.
- Each PRM holds $375\,\text{MB}$ [p. 8] of fit constants, equivalent to 600 thousand sectors.
- Each PRM contains four Track-Fitters (TF) on its FPGA.
- Each TF calculates the χ^2 of track candidates at a rate of $714\,\text{MHz}$ [p. 9].
- Each TF calculates the helix parameters of a track at a rate of $57\,\text{MHz}$ [p. 9].
- To achieve these rates, the TF must operate at a clock rate of at least $200\,\text{MHz}$ [p. 15].
- Hardware trigger latency must be $10\,\mu\text{s}$ or less [p. 2].

$\rightarrow \chi^2$ rate of about $1.65\,\text{THz}$, fit rate of about $131\,\text{GHz}$, and about 350 million sectors in total.

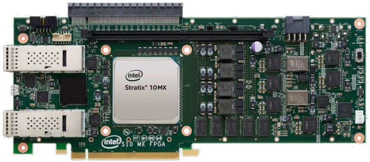
Why is this algorithm well suited for parallel hardware?

- Regions, Sectors, and individual tracks can be processed independently.
- The computations use exclusively multiply-accumulate operations.

Comparison: FPGA – GPU

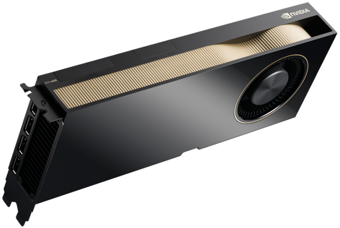
FPGA

- Architecture based on logic gates.
- More power efficient than GPU.
- Extremely fast at logic and basic arithmetic.
- Developers with HDL-skills necessary for best results.
- Maximum customizability.



GPU

- Architecture based on parallel processors.
- More power efficient than CPU.
- Very fast if job and device fit perfectly.
- Can be programmed with common languages.
- Options are limited by the architecture.



Comparison FPGA – GPU

Intel Stratix 10 FPGA [Intel Website]

- 3960 Digital Signal Processors
- 8600 Gflop s^{-1}
- 16 GB HBM 2
- 512 Gbit s^{-1}
- PCIe 3.0
- $ 9000 today

Nvidia Tesla T4 [GPU Database]

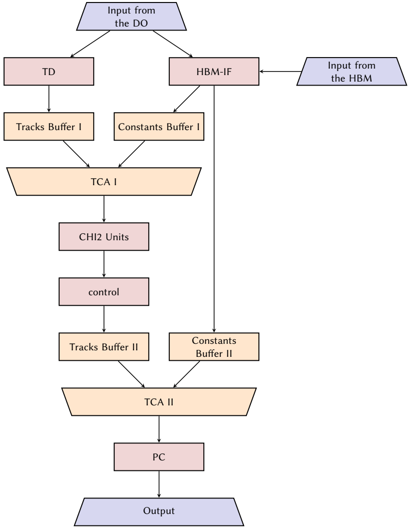
- 2560 Cores (Shading Units)
- 8141 Gflop s^{-1}
- 16 GB GDDR6
- 320 Gbit s^{-1}
- PCIe 3.0x16
- € 1790 today

Nvidia RTX A6000 [GPU Database]

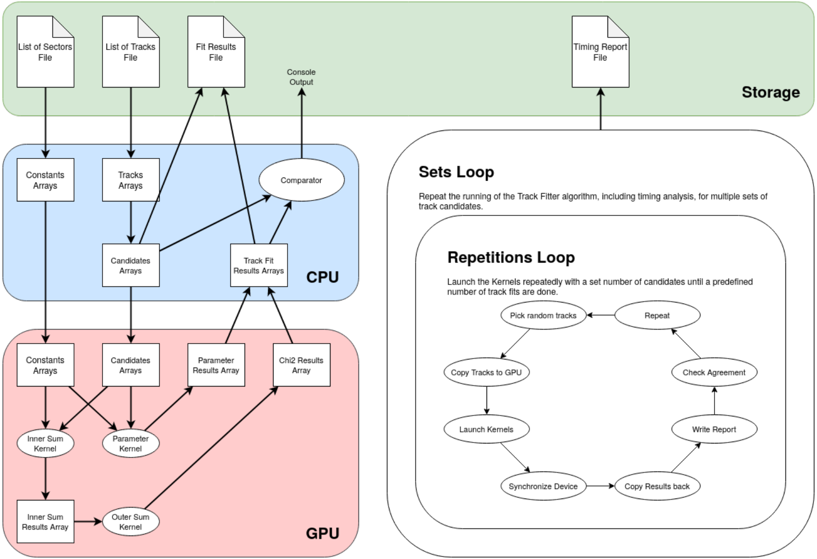
- 10752 Cores (Shading Units)
- 38 710 Gflop s^{-1}
- 48 GB GDDR6
- 768 Gbit s^{-1}
- PCIe 4.0x16
- $ 4649 at launch

The FPGA Track Fitter Implementation

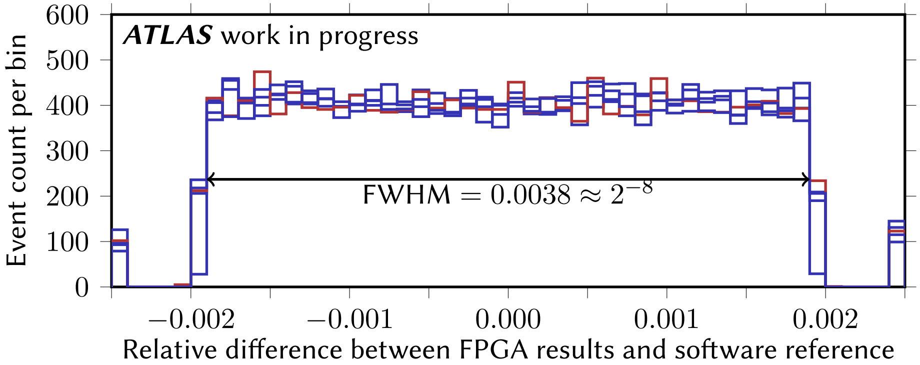
- **TD** Decodes possible tracks from the pattern. Distributes the matched patterns into four channels.
- **TCA** Aligns the tracks with their corresponding χ^2 constants.
- **CHI2** Performs the goodness-test.
- **ctrl.** Funnels the output of four CHI2 into a single channel.
- **TCA** Aligns the tracks with their corresponding fit constants.
- **PC** Calculates the helix parameters.



The GPU Track Fitter Implementation

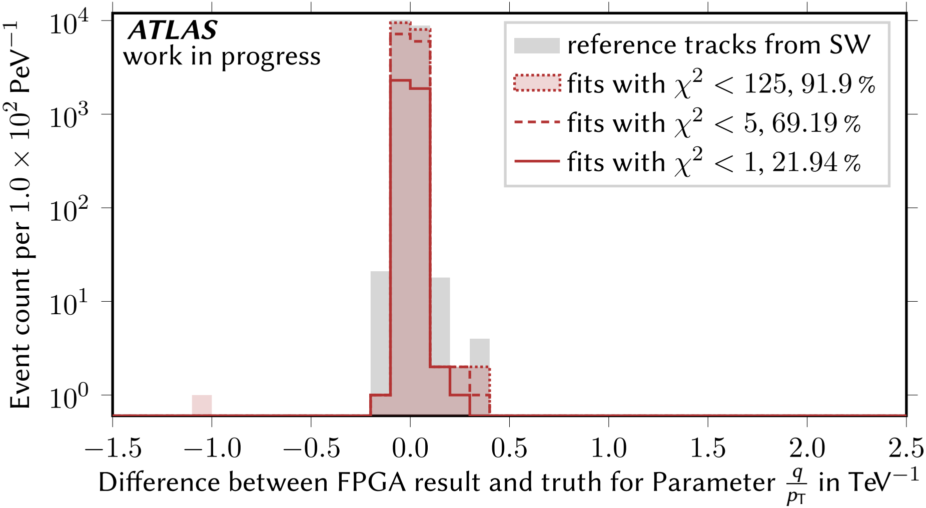


Numerical precision of the Track-Fitter

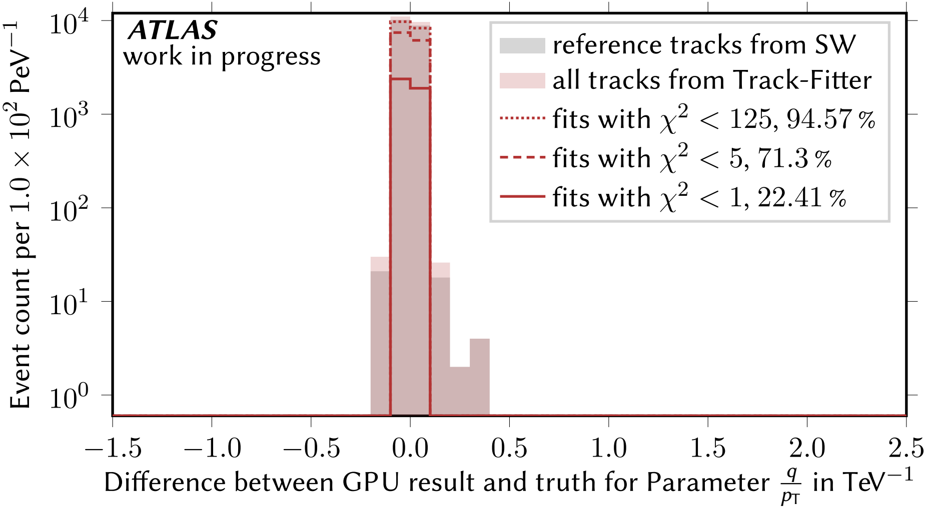


The GPU and the reference software operate with the same numerical precision.

Diversion of the Transversal Momentum $\frac{q}{p_T}$



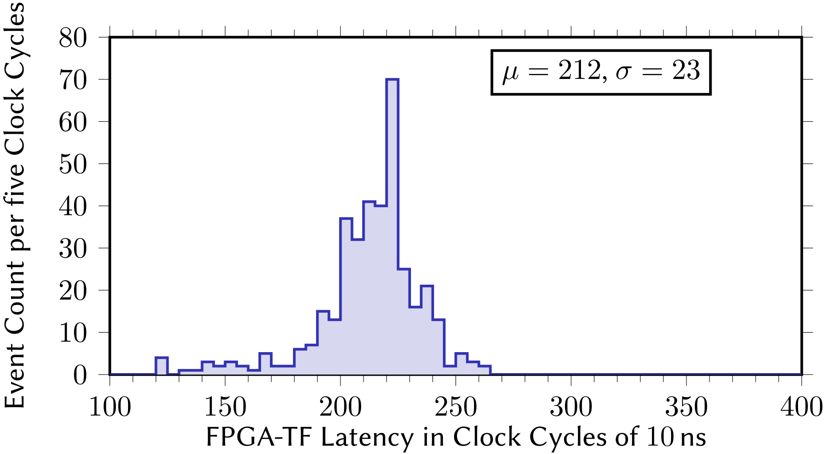
Diversion of the Transversal Momentum $\frac{q}{p_{\mathrm{T}}}$



Resolution of the other Parameters

Parameter	Reconstructed	Limit
Transversal Momentum $\frac{q}{p_\mathrm{T}}$	$6.036(67)$ PeV^{-1}	4 PeV^{-1}
Distance d_0	$875.1(30)$ µm	4 µm
Angle ϕ	$3.473(23)$ mrad	4 mrad
Distance z_0	$2.663(20)$ mm	31 µm
Pseudorapidity η	$5.004(38)$ mη	4 mη

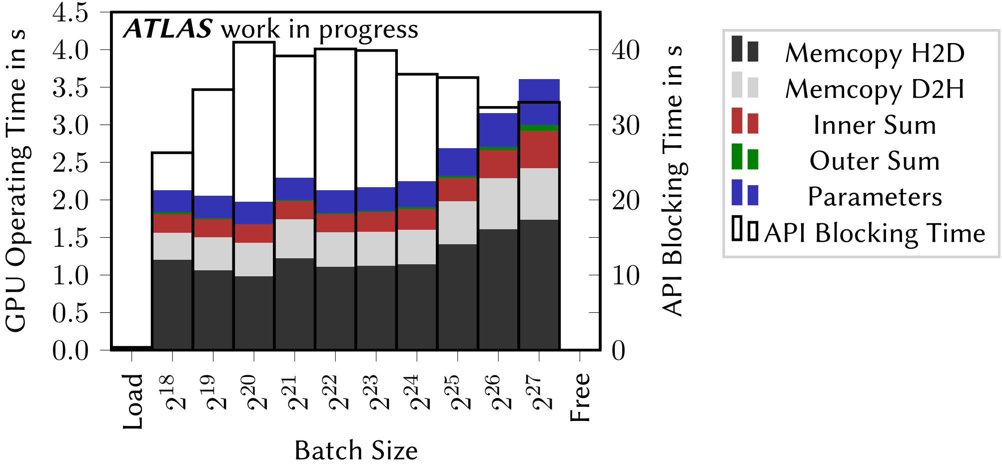
Latency of the Track-Fitter



The latency of the FPGA-TF is 2.1 µs the latency of the GPU-TF ranges from 450 µs to tens of seconds.

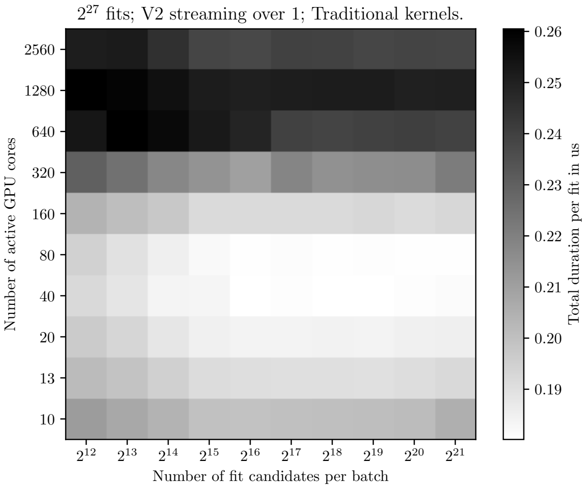
Duration of Various GPU Activities

178772 sectors; 2^{27} fits; no streaming.



Speed Optimization of the GPU-TF

- 134 217 728 fits
- Memcopies and kernels called asynchronously
- Calculations done in individual kernel-calls
- Utilize 80 GPU Cores
- Supply track candidates in batches of 32 786



Throughput of the TF

- The track fitter can operate under a clock frequency of $175\,\mathrm{MHz}$ [Camplani 2023, p. 16] (target $200\,\mathrm{MHz}$).
  $\rightarrow \chi^2$-rate of $1.44\,\mathrm{THz}$ and parameter rate of $115\,\mathrm{GHz}$
- The highest achieved fit rate on the T4 GPU is:
  - $5.55\,\mathrm{MHz}$ with respect to the total duration.
  - $90.43\,\mathrm{MHz}$ with respect to the sum of all GPU activities.
- With respect to the total duration, 3×10^6 T4 GPUs necessary to achieve target rate.
- Expect about a factor of four from using the RTX but this has to be tested.

Summary and Outlook

Summary:

- The linearized track fitting algorithm has been implemented on FPGA and GPU.
- The TF has been verified to work on both types of hardware.
- The FPGA-TF performance has been evaluated in simulation.
- The GPU-TF performance has been measured.

Outlook:

- Find the last bug in the TF firmware.
- Run the algorithm on the faster A6000 GPU.
- Optimize the resource utilization and the kernel execution on GPUs.

The ATLAS Experiment:
"Event Display from Upgrade Physics Simulated Data",
CERN, 2021.

The ATLAS Collaboration:
"Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties",
Journal of Instrumentation, Volume 12, May 2017.

The ATLAS Collaboration:
"Technical Design Report for the Phase-II Upgrade of the ATLAS Trigger and Data Acquisition System",
CERN-LHCC-2017-020, June 2018.

The ATLAS Collaboration:
"HTT ATLAS Electronics Specification for Pattern Recognition Mezzanine (PRM)",
ATL-COM, DAQ-2018-039, December 2019.

The AM Design Team:
"Phase-II Associative Memory ASIC Specifications and Technical/Scientific Report",
CERN-OPEN-2018-003, November 2019.

Paolo Francavilla:
"PRM HW Status",
PRM Meeting, Presentation, March 2021.

Hermann Kolanoski, Norbert Wermes:
"9 Spurrekonstruktion und Impulsmessung" in "Teilchendetektoren",
Springer-Verlag, Berlin und Heidelberg, 2016.

Mathias Braagaard:
"Implementation of a linearized χ^2-Square [sic] method for the ATLAS Pattern Recognition Mezzanine",
DTU Kongens Lyngby, early 2020.

Magnus Høegh Jensen:

"Implementation of Particle Track Parameter Calculations for the ATLAS Pattern Recognition Mezzanine",
DTU Kongens Lyngby, early 2020.

Veronique Boisvert, Stephanie Majewsk:
"Task Force 21.9. Samples",
ATLAS Twiki, February 2022.

The ATLAS Collaboration, Klaus Moening:
"Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC",
ATL-PHYSI, PUB-2019-014, March 2019.

Camplani Alessandra, Dittmeier Sebastian, et al.:
"Intel Stratix 10 FPGA Design for Track Reconstruction for the ATLAS experiment at the LHC",
JINST 18 P06029, June 2023.

The Intel Corporation:
"Product Specifications: Intel Stratix 10 MX 2100 FPGA",
Intel, 2017.

W1zzard, T4CFantasy, et al.:
"GPU Database – Specs",
TechPowerUp, 2024.