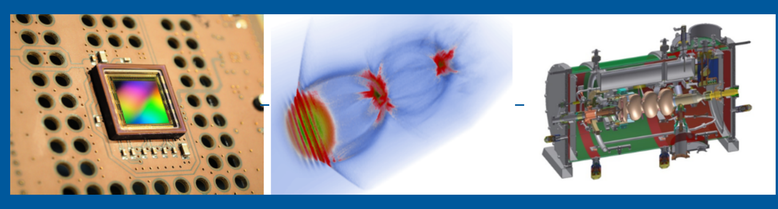
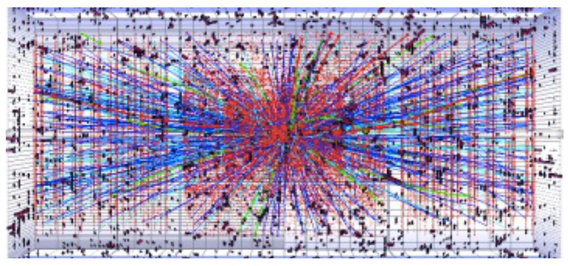
Programme Matter and Technologies

CMS L1 Track Trigger Hardware Demonstrators



Luis Ardila-Perez, Timo Dritschler, and Thomas Schuh

Motivation: High Luminosity LHC

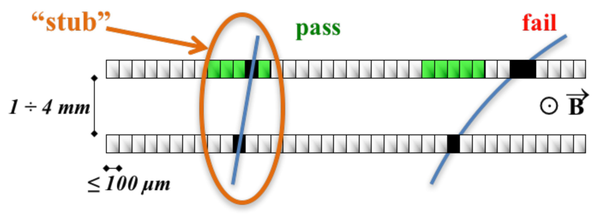


Pileup 200 simulation

- In 2026 - **LHC** will be **upgraded** in **luminosity**
- **Tracker will be replaced** due to radiation damage and for high occupancy conditions
- **New design** will allow tracker **read out at 40 MHz**

CMS Tracker Upgrade

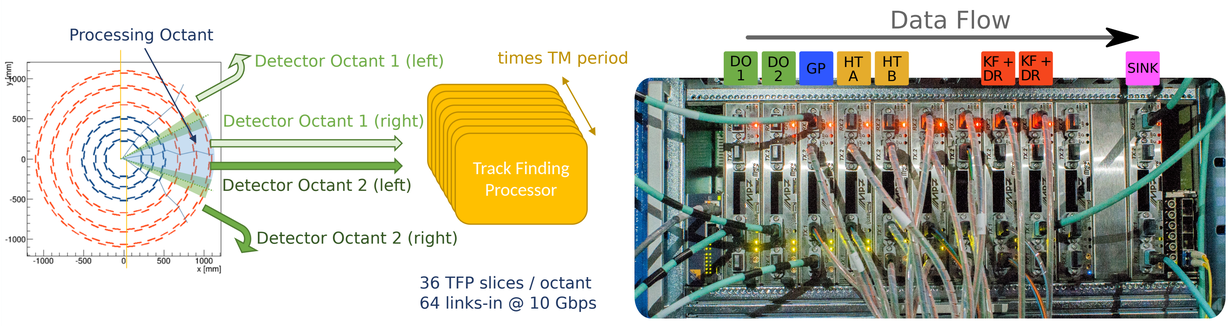
- High $\mathbf{p_T}$ tracks are indicative of interesting physics (decays of high mass particles)
- **Novel tracking modules** utilise two **1.6-4.0 mm spaced silicon sensors**, to **discriminate p_T > 2-3 GeV**
- Rate **reduction O(100)** inside detector module



The Track Finding Processor (TFP)

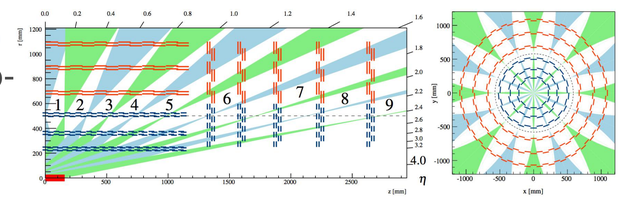
- Total available L1 time will be 12.5 μs, but **only ~4 μs is available for track finding**
- Processes **very high data rates (~20,000 stubs per event)** down to the O(10) genuine/interesting tracks expected on average
- **Track Finding Processor (TFP)** receive data links from adjacent detector octants in φ
- Fully **time-multiplexed system**: Processing of subsequent events done on parallel independent nodes
- Each TFP processes **1 of 8 φ sectors and 1 event out of 36**
- One TFP become the **demonstrator slice unit**
- **Highly scalable system**

FPGA-based Hardware Demonstrator

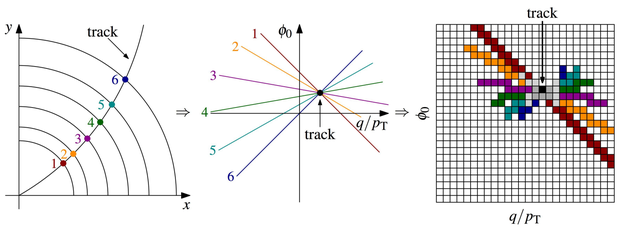


Data Flow

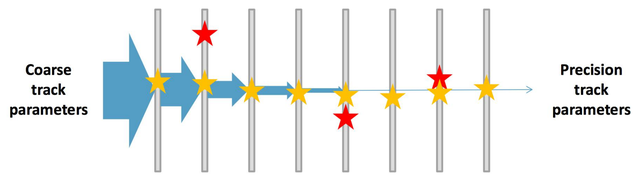
Geometric Processor (GP)
Processes stub data, and sub-divides the octant into 36 finer sub-sectors



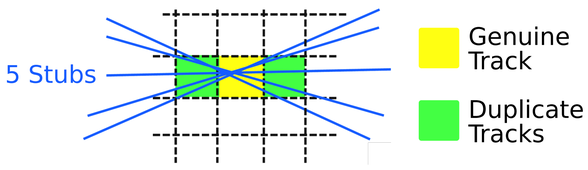
Hough Transform (HT)
Track finder that identifies groups of stubs consistent with a track in the r-φ plane



Kalman Filter (KF)
A candidate cleaning and precision fitting

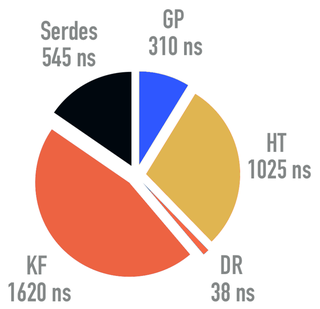


Duplicate Removal (DR)
Uses precise fit information to remove duplicate tracks generated by the HT



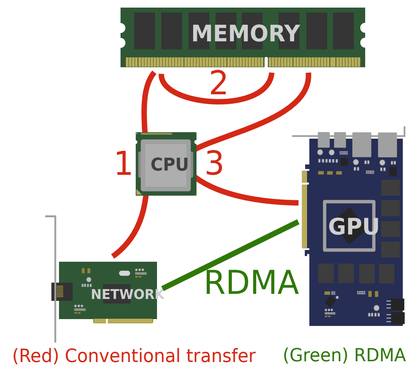
- **Total latency < 4 μs**
- TFP demonstrator slice could be **realized with today's hardware** (Virtex7 FPGA in μTCA board)
- **KIT contributions** to the project include the **HT**, **DR**, and the **improved GP** for the future demonstrator

Latency Breakdown



Heterogeneous Hardware Demonstrator

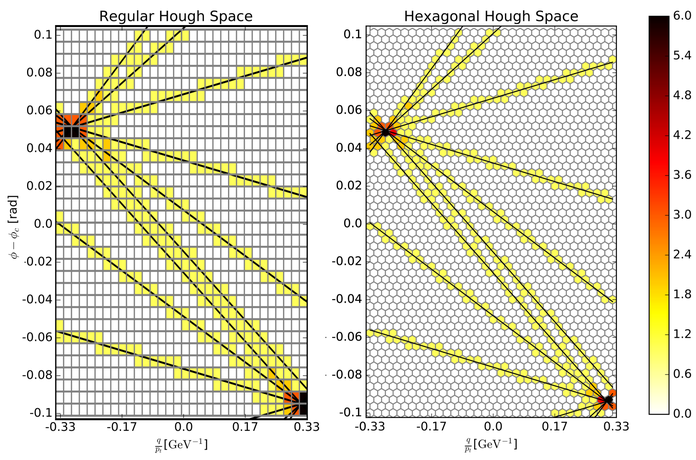
- Remote Direct Memory Access (RDMA) can be used to **directly connect GPUs with FPGAs**
- This allows for **highly flexible FPGA-based DAQ** hardware combined with high performance computing **GPUs**



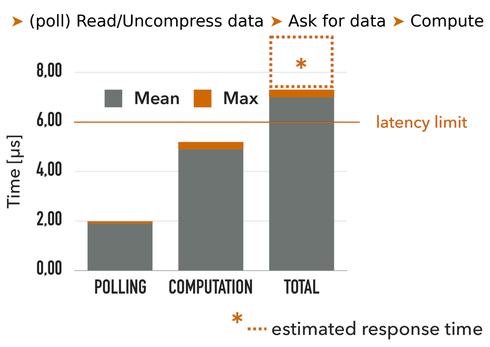
(Red) Conventional transfer (Green) RDMA transfer

r-φ Floating Point Hexagonal HT

- Hexagonal bins in Hough space
- This **Suppresses fake** candidates by **80%**
- Runtime comparable to FPGA approach
- **Only 1 bin activated per row** in Hough space

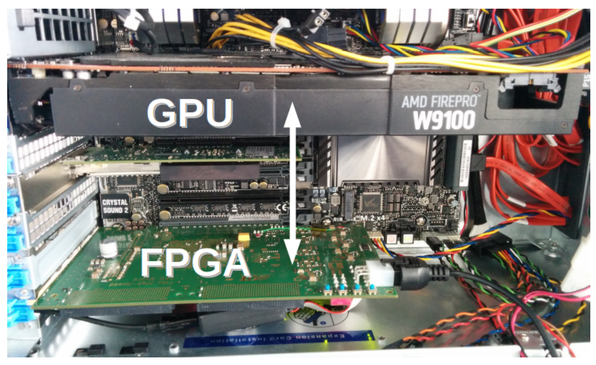


- less algorithmic branching
- Computational time of around **4 μs**
- More **complex algorithms** are possible: hexagonal approach



Outlook of Heterogeneous Demonstrator

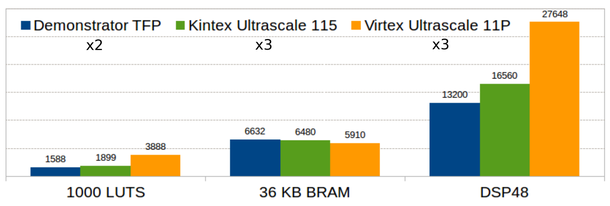
- RDMA enables **low-latency** data transfers
- Paired with efficient algorithms, GPUs can be used to realize **tight feedback loops** with less than **10 μs latency!**
- We illustrated the performance of such systems by a **prototype implementation** of the L1 CMS Track Trigger HT



- Look at performance of newer GPUs
  - High Bandwidth Memory **(HBM)** **2-4x** better throughput
- Investigate new transfer technologies
  - PCIe 4.0 (**2x** faster)
  - NV-Link (**5-10x** faster)

Future FPGA Demonstrator Development

- Next step is a system using the Xilinx **KU115 FPGA** and **16.3 Gbps** optical links with a target **latency of 2 μs** using the **ATCA** form factor
- Targeting **18 time multiplex period needs 2x processing capacity**
- Total sum of logic needed would be **O(3) KU115 FPGAs**



DESY JÜLICH FORSCHUNGSZENTRUM HI Jena Helmholtz Institute Jena HIM Helmholtz Institute Mainz GSI HZB Helmholtz Zentrum Berlin HZDR KIT Karlsruhe Institute of Technology HELMHOLTZ ASSOCIATION