Trigger level track reconstruction in CMS with a fully time-multiplexed architecture using a Hough transform implemented in an FPGA

INFIERI Summer School
USP, São Paulo, Brazil
23 Jan - 5 Feb 2017

C. Amstutz[1], L. E. Ardila Perez[1,4,5], F. A. Ball[2], M. N. Balzer[1], J. Brooke[2], M. Caselle[2], L. Calligaris[4,5], D. Cieri[2,4,5], E. J. Clement[2], G. Hall[3], T. R. Harbaum[1], K. Harder[4], P. R. Hobson[6], G. M. Iles[3], T. James[3], K. Manolopoulos[4], T. Matsushita[7], A. D. Morton[6], D. Newbold[2,4], S. Paramesvaran[2], M. Pesaresi[3], I. D. Reid[6], A. W. Rose[3], O. Sander[1], T. Schuh[1], C. Shepherd-Themistocleous[4], A. Shtipliyski[3], S. P. Summers[3], A. Tapper[3], I. Tomalin[4], K. Uchida[3], P. Vichoudis[8], M. Weber[1]

[1]Karlsruhe Institute of Technology,Karlsruhe, Germany
[2]University of Bristol, Bristol, UK
[3]Imperial College London, London,UK
[4]Rutherford Appleton Laboratory, Didcot, UK
[5]Supported by the EU FP7-PEOPLE-2012-ITN project nr 317446, INFIERI, "Intelligent Fast Interconnected and Efficient Devices for Frontier Exploitation in Research and Industry
[6]Brunel University London, Uxbridge, UK
[7]Österreichische Akademie der Wissenschaften, Vienna, Austria
[8]CERN, Geneva, Switzerland

The High-Luminosity LHC

The **Large Hadron Collider** will be upgraded to increase its peak instantaneous luminosity up to values between 5×10^{34} and 7×10^{34} $cm^{-2}s^{-1}$. This upgrade is referred to as **High Luminosity LHC** (HL-LHC). The number of concurrent p-p collisions (**pile-up**) in a bunch crossing will reach values between **140** and **200**, significantly larger than the design conditions for the present detectors.

The **CMS collaboration** plans to upgrade its detector (**Phase-2 upgrade**), to attain good physics performance in the conditions expected at the HL-LHC.

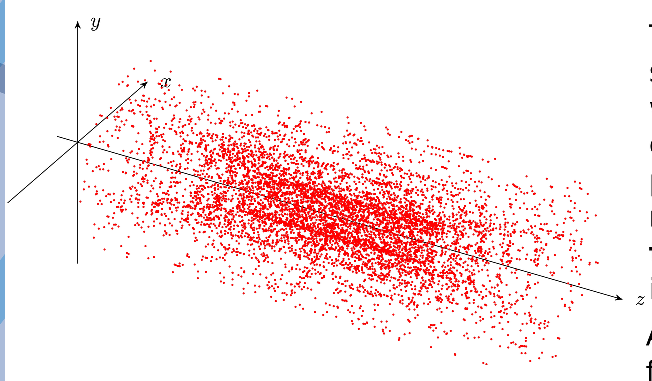


Fig. 1: Doublets of hits detected by the upgraded CMS silicon tracker in a busy event with 140 pile-up collisions.

The current implementation of the first stage (**L1**) of the **trigger** system would experience a large performance degradation in HL-LHC conditions. CMS plans to improve this **performance** by **reconstructing charged particle tracks in hardware,** using this information in the L1 decision.

An overview of the various proposals for such a track trigger implementations is presented in this conference in a poster by Sudha Ahuja titled "Level-1 track trigger for the upgrade of CMS detector at HL-LHC".

The Upgraded CMS Outer Tracker

The Phase-2 outer tracker is composed of **double layers** of silicon detectors (Fig. 2).

Correlation logic (Fig. 3) in the detector electronics selects pairs of **hits compatible** with the bend of a track with a p_T > 2-3 GeV/c. These hit pairs (named **stubs**) are sent off-detector for L1 track reconstruction. This selection reduces the bandwidth to L1 electronics.

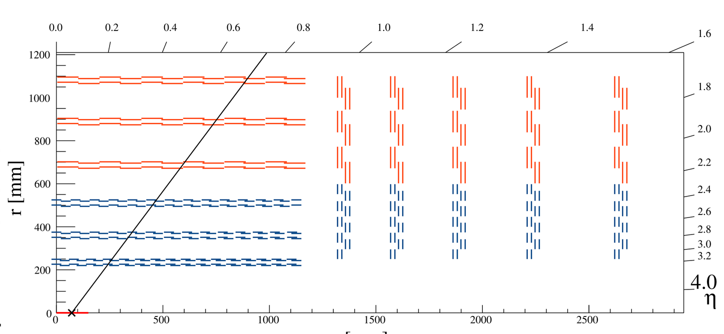


Fig. 2: r-z quadrant of the silicon outer tracker

The stub bend information is sent off-detector, so it can be used as a rough estimate of p_T in downstream filtering stages.

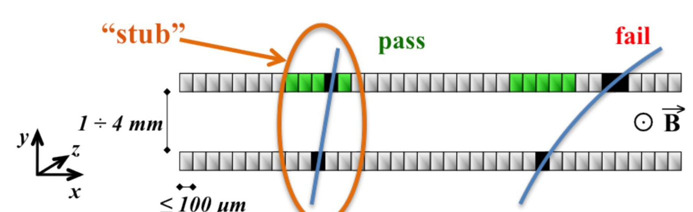


Fig. 3: Stubs are pairs of hits compatible (green band) with a p_T > 3 GeV/c track.

Detector buffers across CMS are designed for a **maximum** L1 trigger **latency** of **12.5 μs**, so the reconstruction of tracks has to take place in a **few μs**.

Hardware Demonstrator

We are currently running **two setups** employing MP7 cards: a smaller setup is based at **Rutherford Appleton Laboratory, UK** and is used for firmware development, a larger one is based at **CERN** and is used to implement in hardware the demonstrator for the proposed track finding approach.

Currently, 11 MP7 cards are available in the CERN setup, as shown in Fig. 4 below. The cards can be managed and run **remotely** via Ethernet connection. Eight cards are daisy-chained together, following the scheme shown in Fig. 6, implementing the track reconstruction demonstrator. The remainder of the cards are currently run in isolation or in pairs to test new ideas and firmware improvements for the components of the system. The flexibility offered by this MP7-based setup was extremely useful to speed up development.

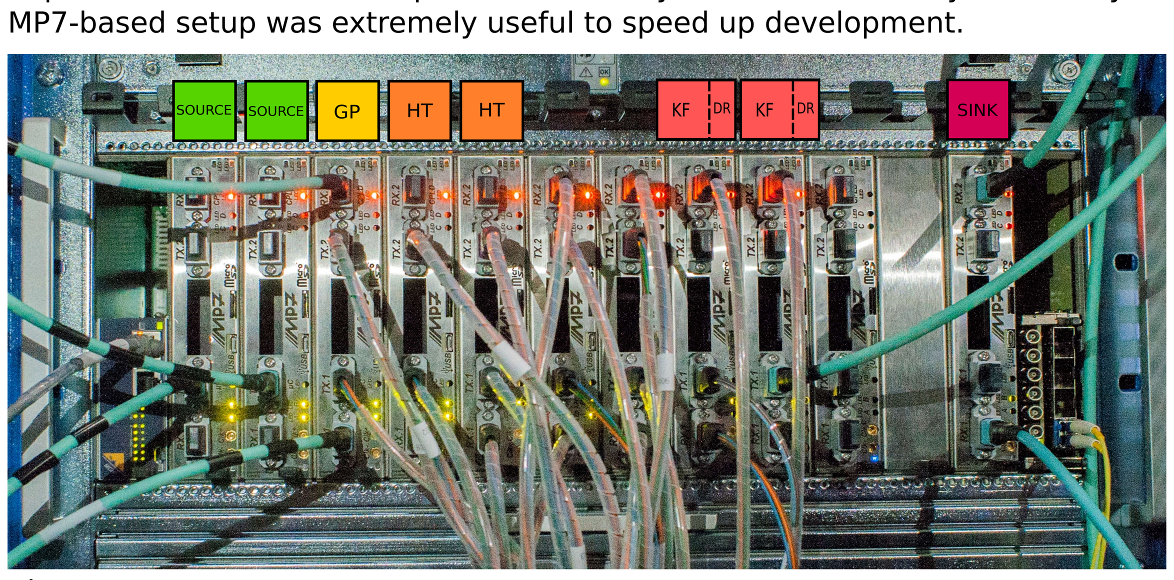


Fig. 4

Data read out from either the sink cards or the internal MP7 data buffers is compared with the expected values from a **simulation** based on the CMS Software framework (**CMSSW**). A **pattern writer** and an **unpacker** software translate - respectively - simulated stubs to and from the MP7 data format.

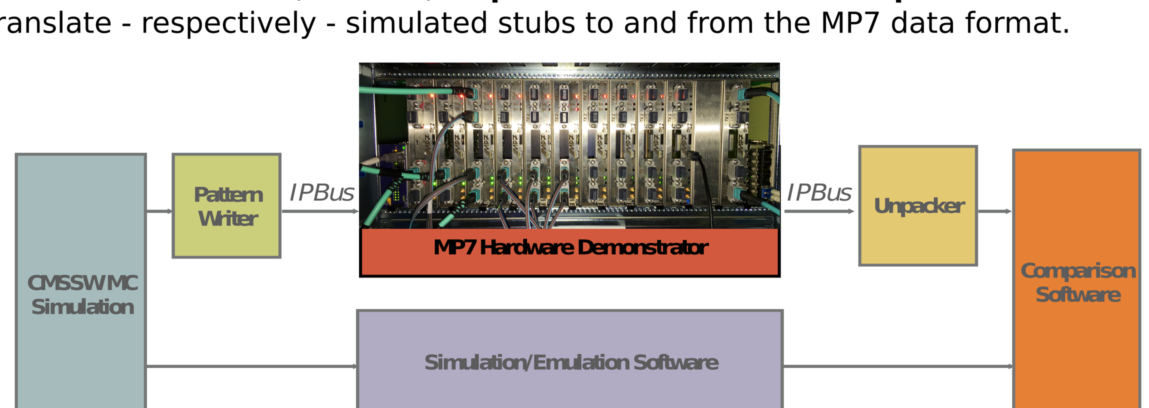


Fig. 5

Track reconstruction at L1 implemented in FPGA

We propose a **track reconstruction system** implemented in firmware running on **FPGAs**.

The system uses a **Hough Transform** (HT) in the **r-ϕ plane** to find tracks, a **Kálmán filter** (KF) to remove fake tracks and misassigned stubs and concurrently perform a precise 3D fit of track parameters, and a filter (DR) to remove **duplicate** tracks identified by the Hough Transform. Duplicates are sets of similar tracks sharing stubs and arising from the discrete and coarse binning of the HT histogram. The system employs **time multiplexing** to spread load and relax the latency requirements over multiple instances of the reconstruction system running in parallel.

A hardware **demonstrator** for the system (Fig. 7) has been completed and will be presented in a detailed publication in the near future. It is implemented on a chain of **MP7** processor cards and aims to demonstrate track reconstruction in one octant in ϕ of the CMS Outer Tracker (Fig. 6) for one of the 36 time slots of the time multiplex cycle. We expect that in the future more capable electronics will allow to fit the entire device on one card equipped with 2 or 3 FPGAs.

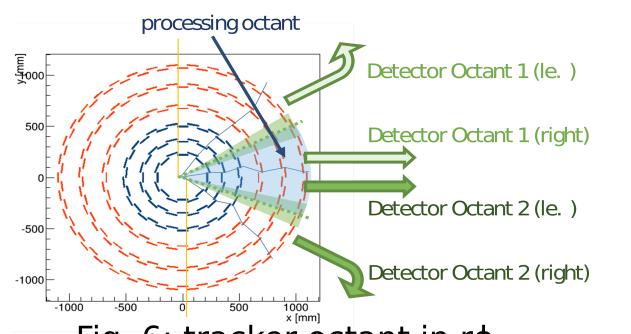


Fig. 6: tracker octant in rϕ

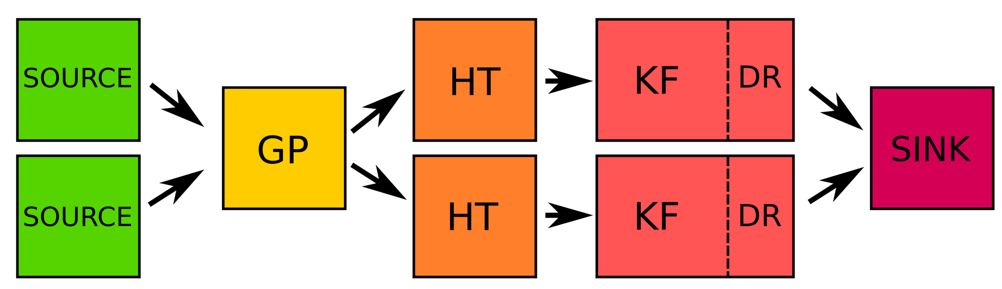
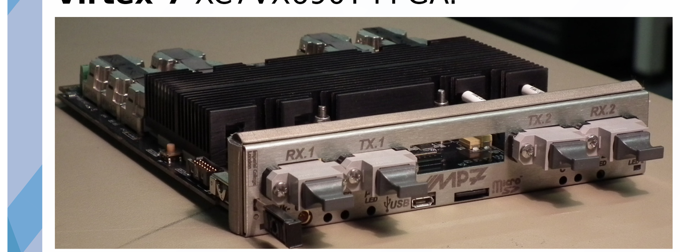


Fig. 7: demonstrator layout

Two MP7 cards ("**SOURCE**") act as data source, emulating upstream electronics from two detector half-octants, making up a **processing octant**. The stubs fed into the system are generated by a Monte Carlo simulation of the conditions and detector configuration expected at the HL-LHC. An MP7 implements a geometric processor ("**GP**"), sorting stubs in 36 subdivisions of the octant (2 in ϕ x 18 in η) and assigns them to independent processing **segments** running on 2 MP7 Hough Transform ("**HT**") track finders (each hosting 18 segments). The output of each HT is processed by an MP7 ("**KF/DR**") implementing the Kálmán filter and duplicate removal algorithm then remove duplicate tracks. Finally a card acts as data sink, saving the output for analysis. All connections between cards are established through the MP7 optical infrastructure.

MP7: Imperial Master Processor, Virtex-7

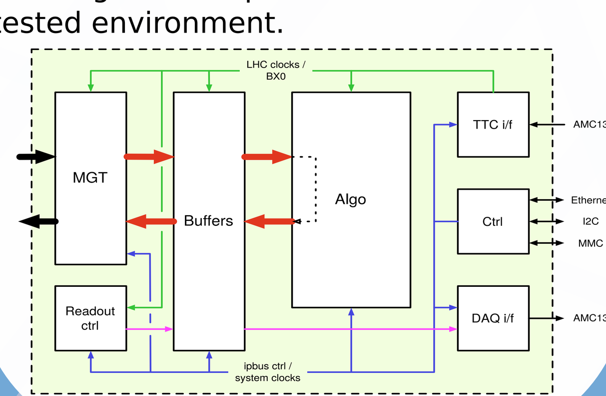
The **MP7** is a generic stream processor card adhering to the **μTCA** standard. The latest version (MP7-XE), used in our project, employs an onboard **Xilinx Virtex-7** XC7VX690T FPGA.



The card is equipped with 72 input and 72 output **optical serial connections**, each running at 10.3 Gb/s. Connection to the μTCA backplane allows communication via **Ethernet** and the programming of the card using the **IPBus** protocol.

The MP7 is a mature and tested product. It is **currently** used to run parts of the **calorimeter** and **muon** triggers in CMS.

An important feature of the MP7 common firmware infrastructure is that it **isolates** the **payload** algorithm away from details of the hardware. This feature eases development of new algorithms by allowing to swap them inside a well-tested environment.



Hough Transform track finding

The **trajectory** of a charged particle in the magnetic field of CMS (which is aligned to the z-axis) is bent in the **r-ϕ plane** (Fig. 8). If the radius of curvature is large compared to the size of the tracking volume, as expected for high p_T tracks, the following relation holds for stubs:

$$\phi_T \simeq \phi_{\mathrm{stub}} - C \cdot (q/p_T) \cdot (r_{\mathrm{stub}} - T)$$

where T is a reference radius (in our case 58cm), ϕ_T is the azimuthal angle of the track at radius T, (q/p_T) is the charge to p_T ratio, C a constant proportional to the magnetic field, r_{stub} and ϕ_{stub} are the stub radius and azimuthal angle.

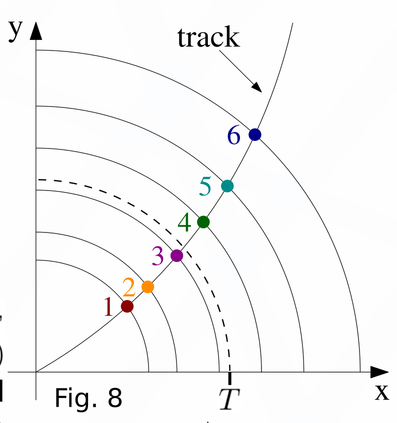


Fig. 8

The linear relation means that **stubs** can be represented by straight **lines**, as shown in Fig.9 . If a set of stubs is consistent with a real track, they will **meet** at the coordinates of this track.

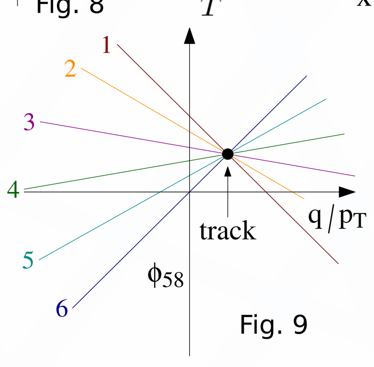


Fig. 9

Track finding uses a **histogram** that **counts** the number of stub lines that have crossed each bin. Parts of a stub line which have a (q/p_T) ratio not compatible with the strip distance measured by the double layer (see Fig. 3) are excluded. Bins crossed by stubs in at least **five** distinct tracker **layers** (Fig. 10) are identified as candidate tracks.

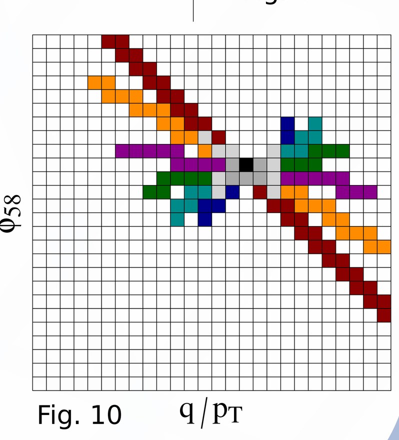


Fig. 10

Three different firmware architectures have been tested to implement the histogram, with increasing improvements in FPGA resources utilization. The latest design, which we refer to as "**Daisychain**", currently used in development, is able to fit **up to 18 segments** in one MP7 card, meeting timing requirements.

Kálman Filter

The HT in r-ϕ is set at a working point that allows finding tracks with high efficiency. A downstream filter against combinatorial stub background can improves performance on fake rate for tracks reconstructed in busy events.

A Kálman filter (KF) is an iterative algorithm to estimate a set of parameters, describing the state of a system for which a model has been provided, from a set of observations containing statistical noise and other inaccuracies (Fig. 4).

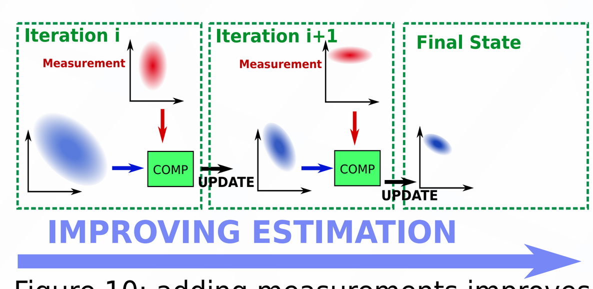


Figure 10: adding measurements improves the estimate on the model parameters

The matrix calculations needed by a Kálman filter can be complicated and care should be taken to avoid using the division operation, as it is computationally very expensive on FPGA logic. Our implementation was enabled by a Java-to-HDL compiler by Maxeler Technologies, producing an optimized VHDL source from a high level algorithm description.

For a real-time application such as a trigger the components of the track reconstruction system should provide a guarantee on the maximum latency for processing an event. The KF we are using allows to stop the accumulation of new measurements into the fit after a tunable amount of time, trading precision for fixed latency.

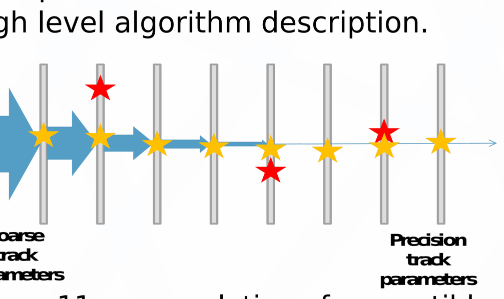


Figure 11: accumulation of compatible stubs (yellow) improves the KF fit. Incompatible stubs (red) are discarded.

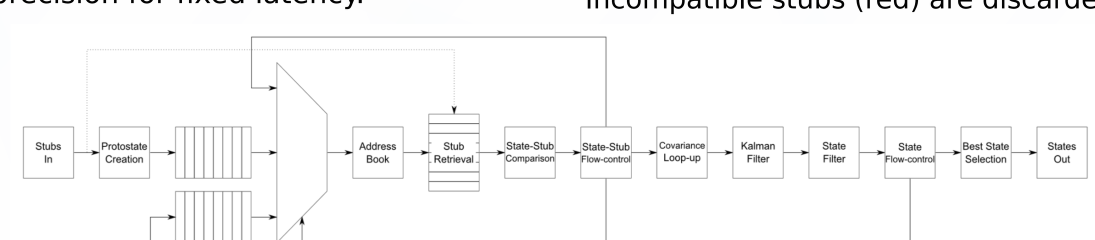


Figure 12: a logical scheme of one Kálman filter processor

Results

The hardware demonstrator implementation has been completed and the device shows an excellent performance in terms of efficiency (Fig. 13 and Table 1) for various test samples. The total latency required for finding tracks in any event has been measured to be 3918 ns, meeting the target of 4 μs. A future prototype of the track reconstructor, based on a smaller number of larger FPGAs will benefit from the closer placement of the processing components and the reduction of optical links, reconstructing events faster. The possibilities offered by this flexible FPGA solution for finding tracks at L1 have only been partly explored, and it is expected that in the near future the algorithms can be tuned to further improve the performance, in particular for electrons, which are subject to large deviations while they traverse the tracker material. In the latter case a more sophisticated KF fit is expected to achieve higher efficiencies.

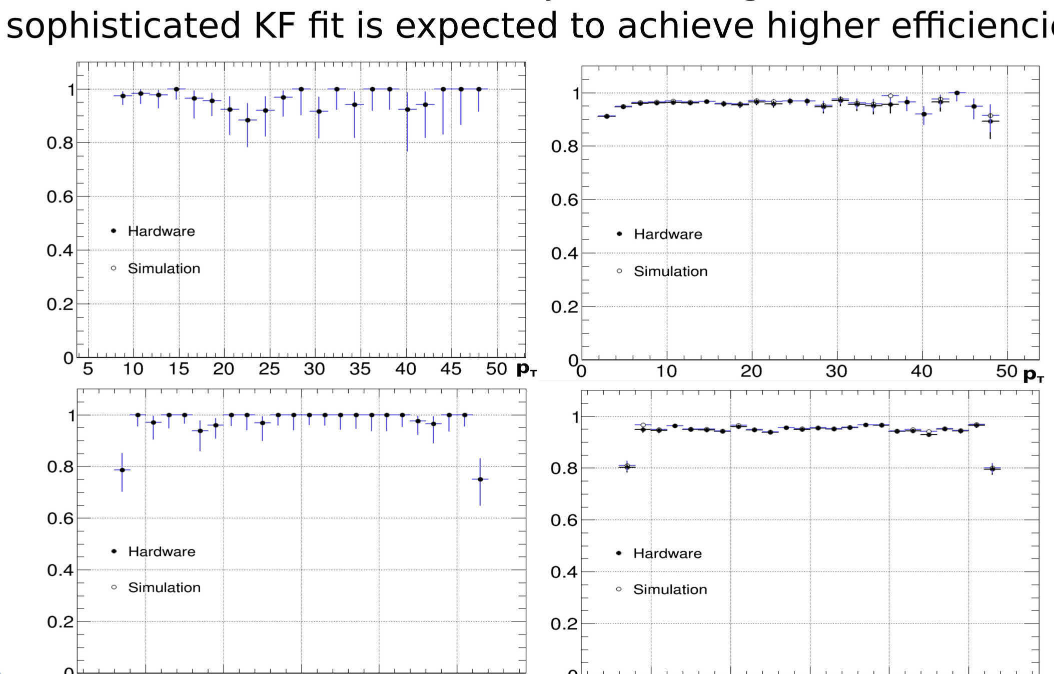


Fig. 13, from left to right: efficiency at PU200 for individual muons, charged tracks in di-top events and electrons in di-top events as a function of p_T in GeV/c (top) and η (bottom).

Sample (all @ PU 200)	Expected efficiency (simulation)	Measured efficiency (hardware)
Charged tracks in top quark pair events	94.8	94.5
Muon gun (8 to 100 GeV/c)	97.1	97.1
Electrons in top quark pair events	81.8	81.4
Charged tracks in top quark pair events (cooling loop fault)	94.5	94.2

Table 1: Track finding efficiencies for different samples, as expected from simulation and as measured in the hardware demonstrator.

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° [317446] INFIERI ``INtelligent Fast Interconnected and Efficient Devices for Frontier Exploitation in Research and Industry".