An FPGA-based Track Finder for the L1 Trigger of the CMS Experiment at the HL-LHC

D. Cieri[*][†]**, L. Calligaris, K. Harder, K. Manolopoulos, C. Shepherd-Themistocleous, I. Tomalin**

STFC - Rutherford Appleton Laboratory, UK
E-mail: davide.cieri@cern.ch

R. Aggleton, F. Ball, J. Brooke, E.Clement, D. Newbold, S. Paramesvaran

University of Bristol, UK

P. Hobson, A. Morton, I. Reid

Brunel University, London, UK

P. Vichoudis

CERN, Geneva, Switzerland

G. Hall, G. Iles, T. James, T. Matsushita, M. Pesaresi, A. Rose, A. Shtipliyski, S. Summers, A. Tapper, K. Uchida

Imperial College London, UK

L. Ardila-Perez, M. Balzer, M. Caselle, O. Sander, T. Schuh, M. Weber

Karlsruhe Institute of Technology, Germany

A new tracking detector is under development for use by the CMS experiment at the High-Luminosity LHC (HL-LHC). A crucial component of this upgrade will be the ability to reconstruct within a few microseconds all charged particle tracks with transverse momentum above 3 GeV, so they can be used in the Level-1 trigger decision. A concept for an FPGA-based track finder using a fully time-multiplexed architecture is presented, where track candidates are reconstructed using a projective binning algorithm based on the Hough Transform followed by a track fitting based on the linear regression technique. A hardware demonstrator using MP7 processing boards has been assembled to prove the entire system, from the output of the tracker readout boards to the reconstruction of tracks with fitted helix parameters. It successfully operates on one eighth of the tracker solid angle at a time, processing events taken at 40 MHz, each with up to 200 superimposed proton-proton interactions, whilst satisfying latency constraints. The demonstrated track-reconstruction system, the chosen architecture, the achievements to date and future options for such a system will be discussed.

---

[*]Speaker.

[†]Also at University of Bristol, UK. 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"

1. Introduction

The High-Luminosity upgrade of the Large Hadron Collider (HL-LHC) [1] is scheduled to take place around 2024. The upgraded collider will be capable to deliver an increased instantaneous luminosity of about 5-7.5 $\times 10^{34}$ cm^{-2} s^{-1}, producing an average of 140-200 overlapping pp collisions (pileup or PU) per 40 MHz bunch crossing (BX).

The Compact Muon Solenoid (CMS) general-purpose detector [2] will undergo a major upgrade (CMS Phase-II Upgrade) [3], in order to operate at the HL-LHC. The experiment will completely replace its Outer Tracker (OT) system, which will be composed by newly designed "p_T-modules", able to reject on detector signals not compatible with an high-p_T particle hypothesis. *Stubs* produced by these modules will be then injected into the novel hardware-based *Level-1 (L1) Track Trigger*, which will reconstruct tracks out of them. Tracks compatible with particles with a large p_T (> 2-3 GeV) will then be exploited by the upgraded L1 Trigger to take a final decision. The use of tracks at L1 will become unavoidable, in order to maintain the trigger acceptance threshold within the defined maximum of 750 KHz. The L1 Track Trigger system will have a time budget of 4 μs to run its track finding algorithm.

2. The Time Multiplexed Track Trigger

The architecture that will be used to build the Track Trigger system has been matter of discussion within the CMS collaboration for many years. An intriguing proposal consists on adopting a fully time multiplexed architecture based on FPGA processing cards, similar to the one currently employed in the Stage-2 L1 Calorimeter Trigger [4], in which each event or BX is processed by a single processing unit.

The Outer Tracker is expected to produce an average of 10.000 stubs per bunch crossing at a PU200, with peaks of 20.000 stubs. A total of about 256 Data, Trigger and Control (DTC) boards will be installed to service the full outer tracker, each of them transmitting to the Track Finding system a data bandwidth of 600 Gbps, spread out across 36 links running at 16.3 Gbps. A preliminary cabling system between the detector and the DTCs divided the OT into eight regions (octants) along the azimuthal angle ϕ.

The DTCs will perform the time multiplexing task, transmitting formatted stubs belonging to the same BX to the downstream Track Finding Processor (TFP) boards in a round-robin fashion. To deal with boundaries between detector octants, each TFP will receive data from two neighbour octants but it will reconstruct tracks only in an artificial processing octant, corresponding to an approximate 22.5 degrees rotation with respect to the detector octant. Eventually each TFP will process data from one eight of the tracker and one slice in time. A baseline system with a time multiplexed factor of 18 will therefore be made of 144 boards (Figure 1).

3. The Track Finder Processor

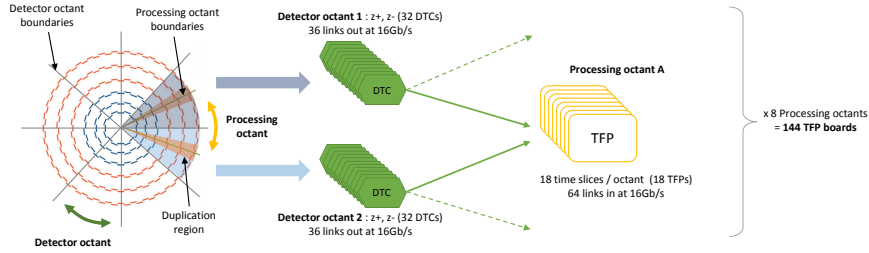The Track Finder Processor will implement the track finding algorithm, that can be divided into four logic components:

Figure 1: The TMTT baseline system architecture. DTCs from two neighbour detector octants time-multiplex and duplicate stubs across processing octant boundaries, before sending them to the TFPs. [4]

- **Geometric Processor (GP)**. It takes in input stubs form the DTCs and it converts them into a useful data format, to ease the load on the downstream stages. Stubs are also assigned to finer 36 sub-sectors in (η, ϕ) to increase parellisation;

- **Hough Transform (HT)**. It performs the first stage of the track finding process by looking for tracks in the r-ϕ plane.

- **Track Filter and Fitting (TFF)**. It rejects fake tracks and removes not consistent stubs from HT candidates, assuming a straight line trajectory on the r-z plane. Track parameters are then estimated using the surviving sub-set of stubs.

- **Duplicate Removal (DR)**. It exploits the precise fit information to reject duplicate tracks produced by the HT.

Two independent algorithms have been developed for the TFF tasks. A Kalman Filter (KF) [4] similar to official CMS offline fitter, and a Seed Filter plus Linear Regression fitter.

3.1 The Seed Filter plus Linear Regression fitter

The Seed Filter plus Linear Regression (SF+LR) fitter is a two-step algorithm, which cleans up the HT track candidate before computing the final track parameters. The filter step is represented by the Seed Filter, which identifies and eliminates stubs assigned to tracks by the HT that lie tens of centimetres from the track in the r-z plane. The operation of the Seed Filter is shown in Figure 2 with an example track candidate in the barrel region.

The stubs consistent with a track candidate surviving the SF stages are then used to compute the final track parameters. Those are calculated by means of the Simple Linear Regression technique, which is applied independently to fit straight lines on both the r-ϕ and r-z planes.

4. The TMTT Hardware Demonstrator

A hardware demonstrator has been built at CERN Tracking Integration Facility (TIF) to measure and validate tracking performance on real hardware, within the $4\,\mu s$ latency constraints. The demonstrator is made up of several MP7s boards [5], five of them replicating the functionality of a single TFP, installed in a microTCA crate. Because of hardware limitations, the demonstrator
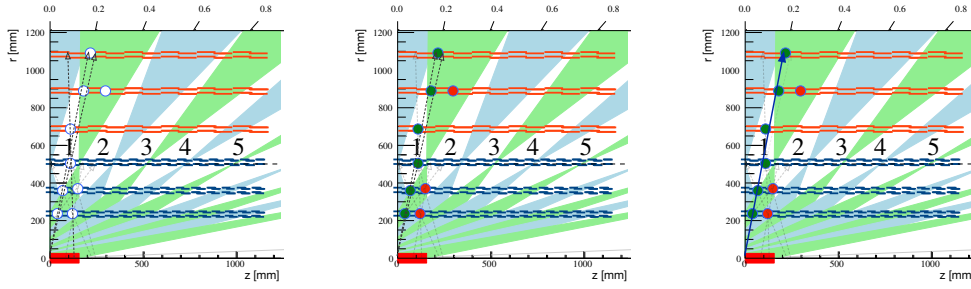
Figure 2: Operation of the SF algorithm. Left: seeding line are computed trough pairs of stubs in different Pixel-Strip (PS) modules. Middle: Lines are extrapolated to remaining layers, and not compatible stubs are rejected (red dots). Right: Only the seeding line with the most stubs is kept.

system has been running with a higher TM period of 36, using 10 Gbps optical links for communications between boards.

Stubs produced with Phase-II CMS geometry simulation are injected into the demonstrator via IPBus. Figure 3 shows a schematic view of the demonstrator system, where each block represents a separate MP7 card. One board is used for the GP, two for the HT and two other boards implement both functionalities of the TFF and the DR.
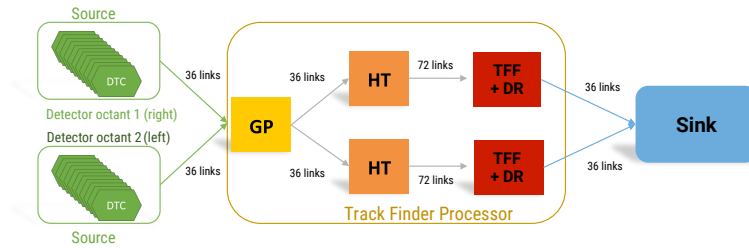

Figure 3: Block diagram of the demonstrator system. Each block represents a separate MP7, running a specific step in the track finding algorithm. 5 boards replicate the functionality of a TFP. 2 *source* boards are placed in front of the TFP, emulating the DTCs task; and one *sink* board is installed at end, storing the produced tracks.

Figure 4 shows the performance in top pair production events at a PU of 200, measured in both hardware and emulation using a track finding chain with a SF+LR fitter. An excellent matching has been observed between the two. An excellent efficiency ($\sim 95\%$) to reconstruct tracks with $p_T > 3$ GeV has been observed. The computed track parameters have been found to have also good resolutions, comparable to that obtained in the offline reconstruction.

The overall latency of the system has been measured in hardware, as the time difference between the first stub coming in and the last track coming out, and it has been found equals to 3.57 µs. Table 1 reports the resource utilisation of the entire demonstrator TFP, as implemented in the Xilinx Virtex-7 XC7VX690T FPGA, for the two approaches with KF and SF+LR fitters.

5. Conclusions

The built demonstrator system has proven in hardware the feasibility of a fully time-multiplexed
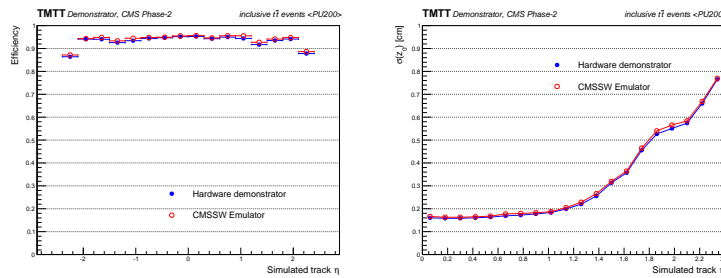
Figure 4: Track finding efficiency (left) and longitudinal impact parameter z_0 resolution (right) as a function of the simulated track pseudo-rapidity η, in top pair production events with PU200, as measured in hardware and emulation. [4]

Table 1: Total resource usage for the demonstrator TFP (with time-multiplexed factor n=36), as implemented in the Xilinx Virtex-7 XC7VX690T FPGA. Also the available resources for a Virtex-7 and a Kintex Ultrascale 115 FPGA are reported.

	LUTs $[10^3]$	**DSPs**	**FFs** $[10^3]$	**BRAM** (36 Kb)
SF+LR TFP Total (excl. infra)	877	8554	1030	2844
KF TFP Total (excl. infra)	763	8472	820	3186
Virtex-7 690	433	3600	866	1470
Kintex Ultrascale 115	633	5520	1266	2160

track finding system, which operates under HL-LHC conditions. The system has been found to correctly reconstruct tracks with high efficiency, with a total latency well below the assigned time budget of 4μs. The system is easily configurable and scalable, giving the opportunity to develop different algorithms or to use different segmentation to deal with larger or smaller input data rates.

Further studies are currently ongoing, with plans to build another demonstrator system utilising newer FPGAs and faster optical links.

References

[1] B. Schmidt, *The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments*, J. Phys. : Conf. Ser. **706** (2016) 022002. 42 p.

[2] CMS Collaboration, *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, J. Instrum. **3** (2008) S08004. 361 p.

[3] D. Contardo, M. Klute, J. Mans, L. Silvestris and J. Butler, *Technical Proposal for the Phase-II Upgrade of the CMS Detector*, Tech. Rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun, 2015.

[4] R. C. Aggleton, L. E. Ardila Perez, F. A. Ball, M. N. Balzer, G. Boudoul, J. J. Brooke et al., *An FPGA-Based Track Finder for the L1 Trigger of the CMS Experiment at the High Luminosity LHC*, Tech. Rep. CMS-NOTE-2017-009. CERN-CMS-NOTE-2017-009, CERN, Geneva, Jan, 2017.

[5] K. Compton, S. Dasu, A. Farmahini-Farahani, S. Fayer, R. Fobes, R. Frazier et al., *The mp7 and ctp-6: multi-hundred gbps processing boards for calorimeter trigger upgrades at cms*, Journal of Instrumentation **7** (2012) C12024.