FPGA-Accelerated Pattern Recognition for the ATLAS Event Filter at the HL-LHC

IEEE-RT 2026, Elba, Italy, May 2026

Priya Sundararajan (University of California, Irvine)
on behalf of the ATLAS TDAQ Collaboration

ATLAS HETEROGENEOUS EVENT FILTER TRACKING FOR HL-LHC

The HL-LHC increases luminosity by **10**×, producing ≥15 million Higgs bosons/year after 2029. Higher detector occupancy and readout rates place extreme demands on the ATLAS EF trigger farm. The Level-0 trigger passes events to the EF for **Regional Tracking** at 1 MHz, followed by **Full-Scan ITk Tracking** at 150 kHz, reducing the output to **10 kHz**.
F150i is an FPGA-accelerated tracking pipeline designed to evaluate heterogeneous EF farm requirements from an FPGA perspective. It performs pixel/strip clustering, partitions full-detector pixel clusters into 1280 η-ϕ regions, and applies Inside-Out pattern recognition to seed tracks per region — **offloading 50% of tracking to FPGA and reducing power consumption**. Physics performance is benchmarked against the CPU baseline using the FPGATrackSim software framework.

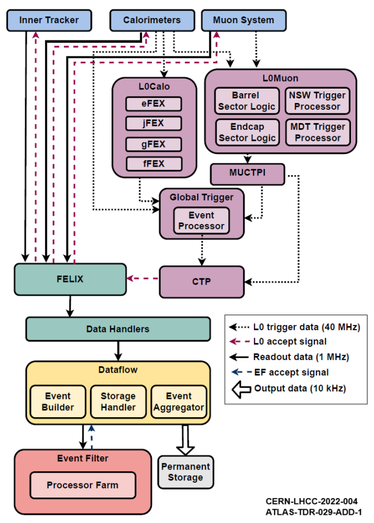

Figure 1: TDAQ Phase II Architecture

The Inside-Out algorithm firmware architecture (Figure 2) operates in two stages:

- **Phi-Binning:** Pixel clusters are stored in matching ϕ-bins; processing begins once the full event is binned.
- **Z-Binning, Pairing and Grouping:** Hits are binned in z across slices within a region, sorted by layer, and paired from outer to inner layers; groups with ≥4 hits are retained as seed candidates.

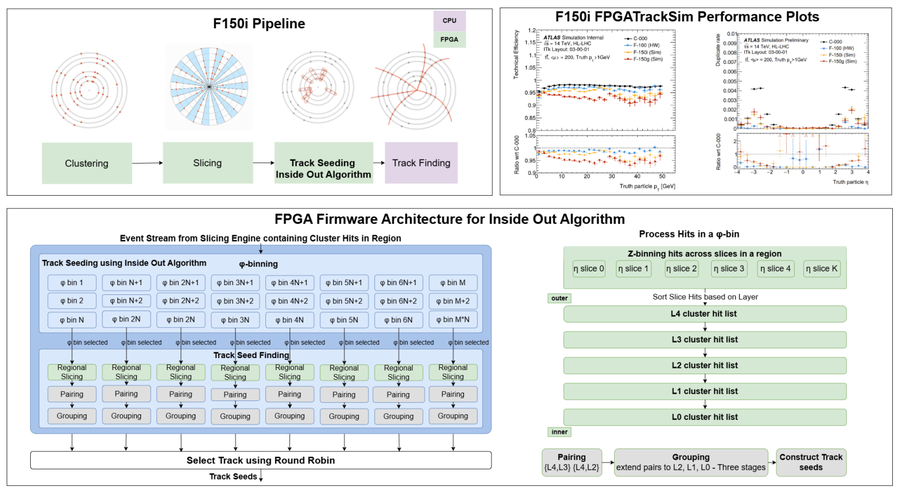

Figure 2: F150i Pixel Seeding Pipeline. The firmware supports up to 140 ϕ-bins per region, parameterized by N (bins sharing memory, processed sequentially) and M (bins processed in parallel).

INSIDE-OUT ALGORITHM IMPLEMENTATION AND FUNCTIONAL VALIDATION ON U250 PLATFORM

The F150i inside-out pixel seeding algorithm is implemented on the AMD Alveo U250 (Xilinx UltraScale+) using a High-Level Synthesis (HLS) flow. Auto-generated RTL is validated through hardware emulation using OpenCL-based host code (Figure 3) and Xilinx XSIM simulator (Figure 4) prior to deployment. Inside-Out kernel fits in a single SLR running at 175 MHz (Figure 5).

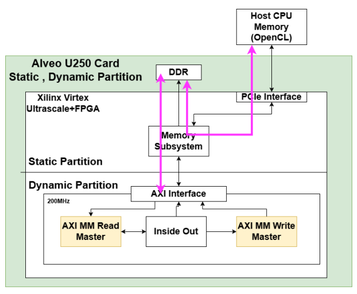

Figure 3: Static - Dynamic Partition

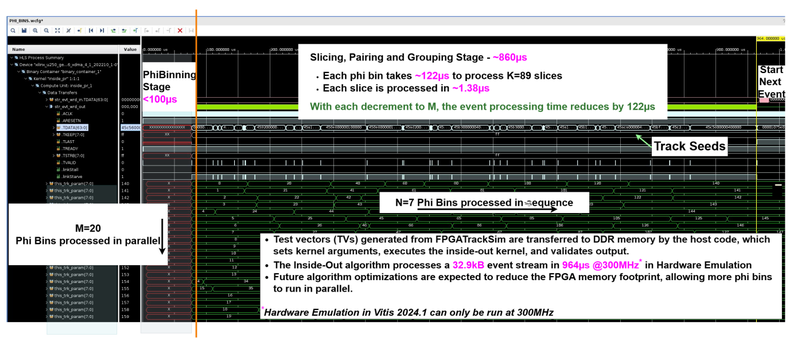

Figure 4: Hardware Emulation Waveforms

Site Type	SLR0 (%)	SLR1 (%)	SLR2 (%)	SLR3 (%)
LUTs	69.63	31.26	1.42	62.40
Registers	46.77	20.57	1.70	31.59
Block RAM Tile	71.21	60.04	0.00	80.80
URAM	22.50	5.00	0.00	90.00
DSPs	79.52	0.75	0.00	23.31

Table 1: F150i FPGA Resource Utilisation across SLRs

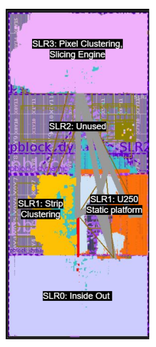

Figure 5: Floorplan

SOFTWARE FRAMEWORK AND HARDWARE TESTS

The F150i FPGA pipeline is fully integrated into the ATLAS Athena framework via three core packages: **FPGATrackSim** (CPU implementation), the **Dataformat Package** (bit-level kernel interface stream definitions), and the **Bytestream Maker** (converts ROOT output into pipeline-ready streams). Together they enable online detector-readout-to-FPGA translation and standalone kernel validation via test vector (TV) generation.

F150i Athena Validation (Figure 6) uses the CERN EF Track Testbed, where the OpenCL interface integrates the U250 into the reconstruction pipeline, validating hardware outputs against FPGATrackSim.

Testbed: 4U Supermicro 4125GS-TNRT1 (AMD EPYC 9174F), 2×U250, 1×U55C.

Track Header Comparison

FIELD	FPGATRACKSIM (EXP)	FPGA (ACTUAL)	STATUS
FLAG	238.0	238.0	MATCH
TYPE	0.0	0.0	MATCH
ETA_REGION	0.0	0.0	MATCH
PHI_REGION	0.0	0.0	MATCH
PHI_BIN	12.0	12.0	MATCH
Z_BIN	30.0	30.0	MATCH
SECOND_STAGE	0.0	0.0	MATCH

Table 2: Header Comparison

Comparing GTracks (FPGATrackSim Vs FPGA) in phi-bin=12, z-bin=30
Reference:
PHI1 / Z1 / R1 → FPGATRACKSIM (Expected)
PHI2 / Z2 / R2 → FPGA (Actual)

GHIT#	PHI1	PHI2	DPHI	Z1	Z2	DZ	R1	R2	DR
0	0.489	0.489	0.000	10.094	10.094	0.000	33.727	33.727	0.000
1	0.482	0.483	0.001	19.234	18.172	-1.062	104.859	96.531	-8.328
2	0.478	0.477	-0.001	25.875	26.820	0.045	156.637	164.051	7.414
3	0.472	0.472	0.001	35.320	35.320	0.000	230.723	230.723	0.000

Table 3: Hit Comparison

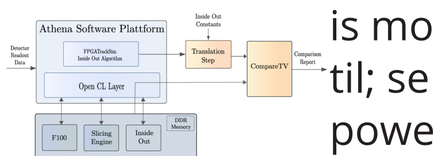

Figure 6: Validation in Athena

F150i Power Measurements: U250 card power is monitored via AMD xbutil; server and FPGA card power dashboard is maintained using Grafana service (Figures 7, 8).

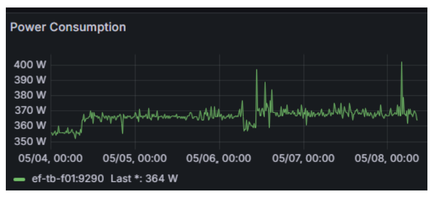

Figure 7: 4U Server Power

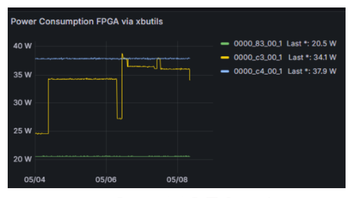

Figure 8: U250 Power, label:0000.C3.00_1

Average execution time of F150i kernels (See Table 4). Clustering on full detector; 1280 η-ϕ regions segmented in $\eta \times \phi = 0.2 \times \pi/16$. Seeding* measured per region; scale to full detector by multiplying ×1280.

Kernel	Time (ms)				Note
	$t\bar{t}$200	JZ0	JZ9	$t\bar{t}$140	
Pixel CPU→FPGA	1.05	1.12	1.05	0.78	/event
Strip CPU→FPGA	0.39	0.34	0.34	0.28	/event
Pixel Clust+L2G+EDM	13.6	13.4	13.5	8.3	/event
Strip Clust+L2G+EDM	1.68	1.69	1.71	1.32	/event
Slicing*	3.4	3.4	3.4	2.9	1 region out
IO seeding*	1.66	1.65	1.7	1.43	/region
Pixel FPGA→CPU	4.6	4.6	4.6	3.3	/event
Strip FPGA→CPU	2.7	2.8	2.8	2.2	/event
Seeds FPGA→CPU*	0.11	0.09	0.12	0.11	/region

Table 4: F150i kernels: Average Execution time

IO seeding time is stable across QCD multi-jet and top pair events, averaging ∼1.65 ms/region (2.1 s for the full detector). Deploying an extra inside-out kernel on available SLR2 resources reduces this to ∼1.05 s for the full detector (see figure 5).

Conclusions: 50% of tracking is offloaded to FPGA; the U250 draws <40 W versus 364 W for the 4U server. Deploying multiple U250s per server via FPGA-as-a-Service helps meet the 1.8 MW ATLAS EF power budget [ATLAS-TDR-029-ADD-1]. Planned optimizations include: coarser ϕ-binning and optimized LUTs to reduce Block RAM; parallel binning with slice-and-group; higher firmware clock frequency; and migration of time-critical HLS functions to RTL.