Highly Parallelized Pattern Matching Execution for the ATLAS Experiment

S. Citraro for the AM system team

***Abstract* — The Associative Memory (AM) system of the Fast Tracker (FTK) processor has been designed to perform pattern matching using the hit information of the ATLAS experiment silicon tracker. The AM is the heart of FTK and is mainly based on the use of ASICs (AM chips) designed on purpose to execute pattern matching with a high degree of parallelism. It finds track candidates at low resolution that are seeds for a full resolution track fitting. The AM system implementation is named "Serial Link Processor" and is based on an extremely powerful network of 2 Gb/s serial links to sustain a huge traffic of data.**

This paper reports on the design of the Serial Link Processor consisting of two types of boards, the Local Associative Memory Board (LAMB), a mezzanine where the AM chips are mounted, and the Associative Memory Board (AMB), a 9U VME board which holds and exercises four LAMBs.

We report on the performance of the prototypes (both hardware and firmware) produced and tested in the global FTK integration, an important milestone to be satisfied before the FTK production.

***Index Terms* — Pattern matching, Image processing, Parallel processing, Trigger circuits, Field programmable gate arrays, Application specific integrated circuits**

I. Introduction

THE implementation presented in this paper was developed for the Fast TracKer Processor (FTK) [1], which is an approved ATLAS [1] trigger upgrade. The FTK processor executes a very fast tracking algorithm organized in a pipelined architecture. A key role in the FTK architecture is played by high performance field programmable gate arrays (FPGAs), while most of the computing power is provided by full-custom ASICs named Associative Memories (AM) [3]. A very large number of hardware dedicated devices is organized in pipelines connected by thousands of specialized links: ~8000 dedicated custom chips (AM chips) that perform pattern matching and ~2000 FPGAs for all the other needed functions. Fig. 1 shows the organization of this large computing system. The black arrows show the connections on fibers and the yellow rectangles show high frequency connections between mezzanine cards and motherboards or between two boards in the same crate slot.

The most power consuming part is the Core Algorithm shown at the center of the figure in the blue box. It is organized into 128 Processing Units (PUs) (see Fig. 1) that process the tracker data in parallel, working on different sections (towers) of the detector.

Each PU executes the two main FTK algorithms, the Pattern Matching (PM) and the Track Fitting (TF). They are executed in pipeline and interfaced by the Data Organizer that provides reduced resolution hits to the PM and well organized full resolution hits to the TF.

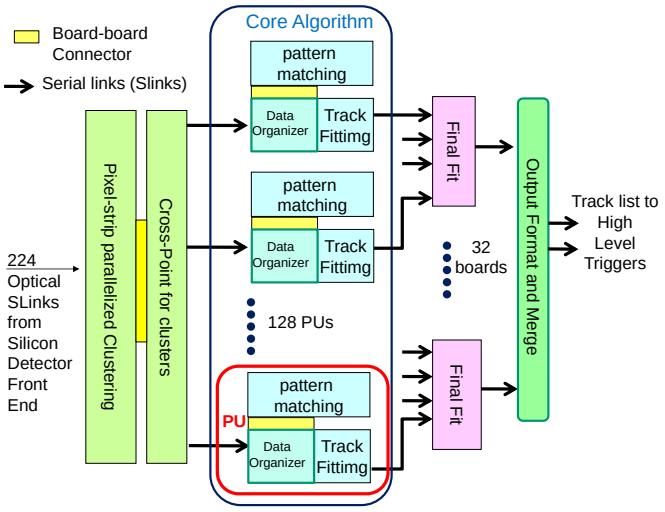

Fig. 1: the FTK architecture

The AM system implements the first stage of the pipeline, the pattern matching. It is based on the use of a large bank of pre-stored patterns of trajectory points, the AM bank. The whole AM system stores 1 billion (10^9) AM patterns. It recognizes track candidates at low resolution to match the demanding task of tracking at the detector readout rate. The second stage receives track candidates and high resolution hits to perform full resolution track fitting at the AM output rate.

The AMBSLP prototype boards received support from Istituto Nazionale di Fisica Nucleare and the European Commission FP7 People grant FTK 324318 FP7-PEOPLE-2012-IAPP.

S. Citraro, N. Biesuz, H. Nasimi, M. Piendibene, C.-L. Sotiropoulou and G. Volpi are with the Department of Physics "Enrico Fermi" of the University of Pisa and INFN Pisa Section, Polo Fibonacci Largo B. Pontecorvo, 3, 56127, Pisa, Italy (email: saverio.citraro@for.unipi.it, nicolo.vladi.biesuz@cern.ch, hknasimi@gmail.com, marco.piendibene @pi.infn.it, c.sotiropoulou@cern.ch, guido.volpi@cern.ch).

P. Giannetti is with INFN Pisa Section, Polo Fibonacci Largo B. Pontecorvo, 3, 56127, Pisa, Italy (email: paola.giannetti@pi.infn.it).

P. Luciano is with the University of Cassino and Lazio Meridionale, Gaetano di Biasio, 43, Cassino, 03043, Italy and INFN Pisa Section (email: pierluigiluciano@pi.infn.it).

Powerful highly parallel dedicated hardware provides excellent performance, reaching resolutions, efficiencies and fake track rejection typical of offline algorithms. The latencies are short (few tens of microseconds), power usage is low (the AM chip, a device able to execute 1 Million comparisons each 10 ns, has a power consumption below 3 W), and the system is small (4 racks of electronics are able to perform a task that would need a farm of thousands of commercial CPUs [4]).

II. AM System Implementation

The PU is made of a 9U VME card, the AM board assembled with 64 AM chips, and a Rear Transition Module (RTM), named AUX card, which is placed in the same slot of the VME core crate (see fig. 2) and contains both the Data Organizer and the Track Fitter. The AUX card communicates with the AM board through a high density high speed connector providing the input data and collecting the fired patterns.

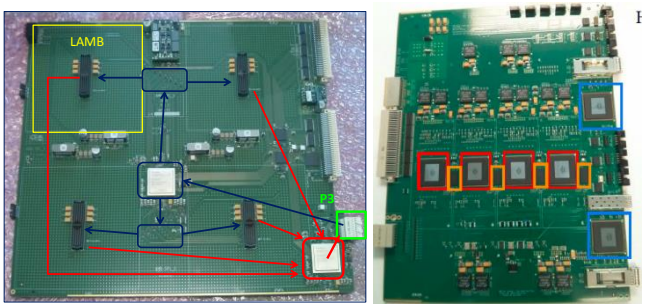

Fig. 2: AMBSLP version 2 and AUX card

The design of the AM system is a challenging task, due to the following factors: (1) the high pattern density (8 million patterns per board), which requires a large silicon area: (2) the I/O signal congestion at the board level, which requires the use of serial links; and (3) the power limitation due to the cooling system: as we are fitting 8 000 AM chips in 8 VME crates and 4 racks, the power should not exceed 250 W per AM board.

A. AM Integration on a mezzanine: power dissipation and serial link communication optimization

The current LHC event complexity requires each 9U-VME board to hold 8 million patterns, equivalent to 64 AM chips. In order to simplify the I/O operations and increase the system modularity, the AM chips are grouped into daughter cards, each one with 16 chips, called Little Associative Memory Boards (LAMB, Fig. 3).

Previous versions of the AM chip used parallel buses for I/O ([5]). This design, however, had some serious drawbacks: distributing 16×8 lines to 16 chips led to extreme complexity in the design of the mezzanine. The parallel data distribution gave upper limits on data speed, not compatible with further system upgrades, and no space was available on the board to optimize the routing and decrease cross-talk as well as noise. In addition, several cooling issues appeared, due to the insufficient dissipation capabilities of the used package (TQFP208 [6]).

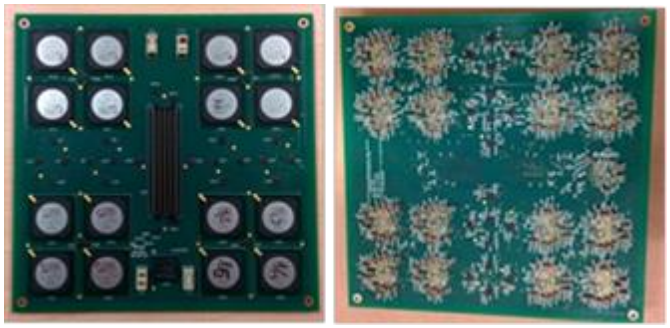

Fig. 3: Top and bottom sides of the LAMB mezzanine. The 16 AM chips are visible on the left.

In order to solve these problems, a switch from parallel buses to high speed serial buses was decided. By using this strategy, we produced free space on the PCB and new opportunities for the layout, routing and power dissipation.

The new package of the AM chip was optimized for high dissipation capability. A Ball Grid Array package with Flip Chip interconnection and a high number of balls (FCBGA 23x23, 529 pins) were chosen. We included a heat slug for high dissipation capability.

Most of the 529 pins have been assigned to the 3 power domains and ground. A small number of pins, optimally routed, are used for the serial I/O: 8 input links to receive input data from the detector, one per layer used in the pattern matching, 2 links to receive pattern addresses from other AM chips, and an output to send out the addresses of patterns fired in the chip itself. In total, the AM chip needs 11 serial links. The AM chip serial communication interface is a Silicon Creations' IP. Its use and test is described in reference [7]. The main features of the Silicon Creation serializers/deserializes (SERDES) are: (a) data rate at least 2 Gbit/s to match 16 bit at 100 MHz; (b) 8b/10b encode/decode capabilities; (c) separate serializer and deserializer macro (the AM chip has 10 input buses, but one output bus for patterns); (d) 32bit I/O buses; (e) driver and receiver circuits compatible with LVDS standard; (f) comma detection and word alignment; (g) BIST capabilities for fast debugging; (h) Low power [7].

We have produced 200 AM chips in a Multi-Project Wafer run, with the final functionality and package but a much smaller die, where only 2048 patterns/chip could fit. We have used those chips to build the first AM system. Sixteen AM chip locations are arranged into 4 quartets. Each quartet has a low jitter oscillator in the center, necessary for the 11 serial links handled by each AM chip. The 100 MHz LVDS clock is distributed to the 4 AM chips by a small fan-out chip, the black square in the middle of the quartet. Below the connector there is a small FPGA and its FLASH Memory, used for AM configuration.

Particular care has been devoted to the PCB routing, especially for the many serial links (~ 200 links), in order to keep the differential impedance fixed at 100 Ω, minimizing the cross-talk. This is a 12 layer PCB where signal and power-GND planes are alternated. The serial links are all routed into internal layers, so that they are isolated between two metal planes. Details of the serial link routing optimization and their quality tests are reported in [8].

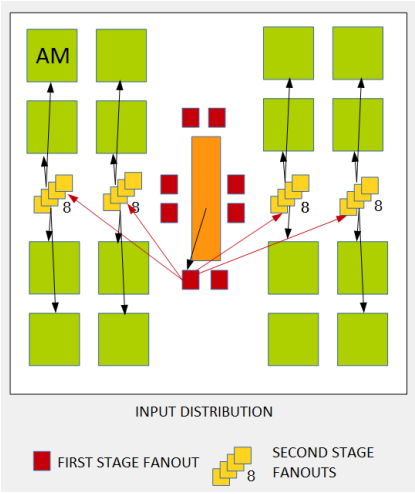

Fig. 5. Input data distribution to AM chips.

Data must be distributed at 2 Gbit/s on each serial link with a very large fan-out: 8 words from 8 detector layers have to reach in parallel the 8 million of patterns at each clock cycle. The large input fan-out on the LAMB is obtained through 2 levels of serial fan-out chips and a very powerful data distribution tree inside each AM chip. Fig. shows the distribution of the input data through 40 1:4 fan-outs. The 8 red ones around the central connector (orange box) replicate each of the 8 incoming buses 4 times to make them available to each subgroup of 4 AM chips organized into vertical sections, as shown by the blue dotted lines.

The second level of fan-outs (16 yellow little squares on the top side and 16 on the bottom side of the PCB, not visible in the figure) replicates again each bus 4 times, one for each single AM device in the subgroup. The placement of chips on the LAMB has been studied and optimized to minimize the crossing of the serial links.

Fig. shows how the output words are collected from the 16 AM chips organized in 4 independent quartets. Each AM device has the capability to receive outputs from other two AM chips and merge them internally with patterns that fired in the chip itself. Each quartet has a single output that goes directly to the connector.

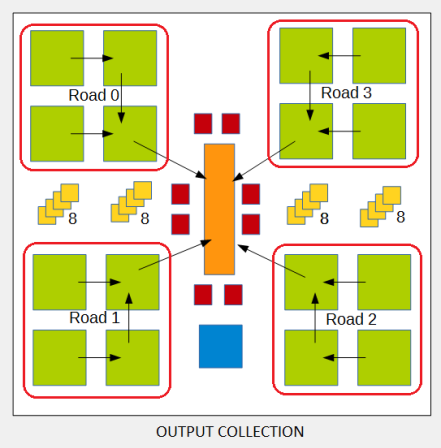

Fig. 6. Output data collection from AM chips.

B. *Assembling four LAMBs on the motherboard*

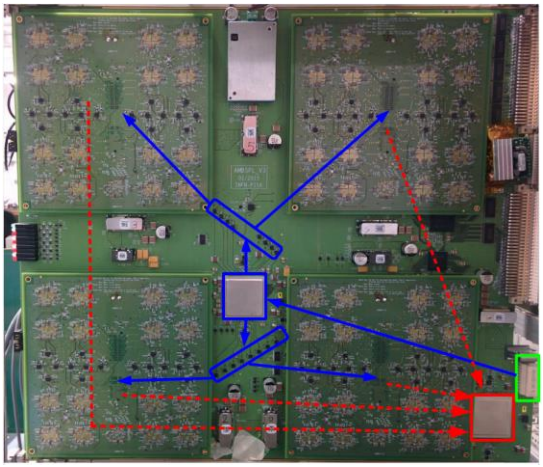

Figure 7. The data traffic in the motherboard. AMBSLP version 3.

A 9U-VME board has been designed to hold 4 LAMBs. Figure shows the motherboard. The LAMB and the motherboard communicate through a high frequency and high pin-count connector placed in the center of the LAMB. A network of high-speed serial links handles the data distribution from the input (the high-density connector in the green box on the bottom-right side of Fig. 7, called P3) to the 4 LAMB connectors and back to the P3 connector. Twelve input serial links (in blue) carry the silicon data from the P3 to the LAMBs, and 16 output serial links (4 links from each LAMB represented by a red arrow in the figure) carry the fired patterns from the LAMBs back to the P3. Events are loaded to the board at a maximum rate of 100 kHz corresponding to a maximum input bandwidth of 1.6 GB/s. An even larger number of output words per link can be collected and sent back to the P3. Each board can read out up to 8000 matched patterns per average event, for a maximum output bandwidth of ~ 3.2 GB/s, thanks to 16 parallel output links (4 links per LAMB).

The data traffic is handled by 2 Xilinx Artix-7 XC7A200T FPGAs, which have 16 Gigabit Transceivers (GTP) [9], each one providing ultra-fast data transmission. The FPGA in the blue box in Fig. 7 handles the input data, while the FPGA in the red box near the P3 handles the output data. Two separate Xilinx Spartan-6 FPGAs (XC6SLX16 and XC6SLX45T) implement the data control logic and configuration. The 12 input serial links are merged into the 8 buses received by each AM chip, one bus for each detector layer used for pattern matching.

The AMBSLP motherboard is a PCB VME 9U (366mm x 400mm) and is made of 12 layers. Half of them are used to distribute power mainly to the LAMBs, 50 Watt each layer. The needed power is generated from 48 V using a large number of DC-DC converters from GE Critical Power: a device visible at the AMBSLP top in Figure , which generates 33 A at 12 V in order to power a set of smaller devices that generate the needed currents at 1 V, 2.5 V, 1.2 V and 1.8 V. On the other layers, we routed the differential lines following

the same design rules adopted for the LAMB, and all the buses used for control and configuration.

C. *Data Flow, Event Synchronization and Processing*

The AM system is designed to be part of a data driven pipeline and perform the pattern matching task (see Fig. 1).

A simple communication protocol is used for data transfers inside the AM system (between FPGAs and AM chips) and with neighboring boards. The data flow through serial links connecting one source to one destination. The protocol is a simple pipeline transfer driven by control words. Control words can be idle words and alignment words. An 8b/10b encoding [7] is used in the serial data stream in order to provide clock recovery, i.e. a 32-bit word is transmitted as 40 bits. The idle word is transmitted when no valid data is available. Input words in each processing step of the pipeline are pushed into a de-randomizing FIFO buffer. All the words not identified as control words are pushed into the FIFO (write-enable signal asserted to the FIFO). The FIFO is popped by whatever processor sits in the destination device. To maximize speed, no handshake is implemented on a word-by-word basis. Data can flow at the maximum rate compatible with the link bandwidth. A hold signal (HOLD) is used instead as a loose handshake to prevent loss of data when the destination is busy. If the destination processor does not keep up with the incoming data, the FIFO produces an "almost full" signal that is sent back to the source as HOLD signal. The source responds to the HOLD signal by suspending data flow. Using "almost full" instead of Full gives the source sufficient time to stop. The standard clock frequency is 100 MHz for 16-bit words or 50 MHz for 32-bit words, which corresponds to 2 Gbit/s for serial transmission.

The End Event (EE) word, marked by a specific control word, separates data belonging to different events on each transmission link. Each device will assert an EE word in its output stream after it has received an EE word in each input stream and it has no more data to output. The EE word has a special format used to tag the event and to report the parity and any error flags.

The AM system has many independent input streams, and events are subdivided into these streams. Data arriving from different layers of the detector have to be synchronized, since the same event can arrive on different links at different times. Input FIFOs perform this task. Their depth covers fluctuations in the device processing time and arrival time of input data.

Each event is processed in two phases inside the AMBSLP: (1) first of all the detector data (hits) have to be loaded inside the AM to be compared with the whole pattern bank and (2), when the event is completely loaded, the matched patterns (roads) have to be read out of the system. The AM architecture is organized in such a way that the two phases can be executed in parallel on contiguous events, since the results of the comparison of one event can be kept inside the chip when a new event is downloaded. The event processing is controlled by two Finite State Machines in the two FPGAs shown in Fig. 7 (input and output FPGAs), one for the hit and the other for road transit: when hits of the N+1 event are sent to the AM chips, the roads of the event N are read out. At the end of this interval of time that we call the "event half-processing" an INIT signal is sent to all the AM chips, so that the event N+1 is saved in the readout section of the chips, the matches of the pattern bank are reset and a new event can be downloaded.

Fig. 8 shows a synthetic drawing of the logic of the two FPGAs that distribute the input data and collect the outputs.

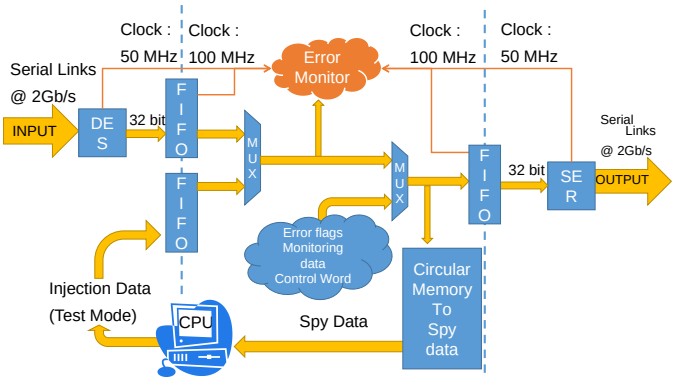

Fig. 8: the logic of both input and output FPGAs.

When the FSM starts to process a new event, words are popped from the input FIFOs for the various input streams. The incoming data are sent to the AM chips and in parallel results of the previous event are sent to the output streams. When the EE word is received on an input stream, no additional data is read from that FIFO until the EE word is received on all the other input streams. The device can issue a HOLD signal if a FIFO becomes "almost full", causing back pressure, but the goal is to have the FIFO deep enough to limit back pressure as much as possible. The EE words from the input streams are checked to make sure they contain the same event tag. Upon detection of different event sequences, a severe error is issued and the system must be resynchronized. Once the event is completely read out from the input FIFOs and the all the roads of the previous event have been sent to the output, the event is closed by sending an EE word to all the output streams with the same event tag detected in the input streams.

D. *System Monitoring*

The AM processes a large quantity of data, a small part of which ends up in the event record. If an error occurs, the proper diagnosis of its source requires access to data at every step in the pipeline. To accomplish this, we have implemented the Spy Buffer system (see Fig. 8), a circular memory used as a logic state analyzer for input and output of the system. This memory is continuously written with the data processed by the board. The write operation is stopped when a Freeze signal is asserted, in order to preserve the data already written. After Freeze is set, no data can be written into the memory and the content of the memory can be read through VME access for diagnosis of error conditions or for standard monitoring functions. For each Spy Buffer there is a status register that contains a pointer to the first free memory location, an overflow bit that indicates if the memory has been written more than once, and the Freeze bit. Spy Buffers are 4-8k locations deep, allowing the storage of 4-8 average events.

The comparison of the content between a sender's output buffer and a receiver's input buffer makes it possible to check the quality of the data transmission. We check data processing comparing the board's input and output with emulation software. The memories also serve as sources and sinks of test patterns for testing single boards or a small chain of boards, as a standalone system.

Other important quantities are monitored:

- The fraction of time each FIFO is Empty and Half-full.
- For each "event half-processing" we measure the time needed to load hits in the AM chips (input time) and the time needed to send roads to the output (output time) for each input/output stream. These measurements are sent to the output and included in the event to be written in the permanent storage.
- The event half-processing rate that in average has to be 100 kHz is also monitored counting the INITs issued per second.

E. System Control and configuration

The AM system is fully controlled and monitored by a CPU using the VME standard. The VME slave interface, implemented in a Spartan-6 FPGA, allows writing/reading functions to/from registers, memories and FIFOs, using random access and block transfer modes.

The most important implemented function is the configuration of the AM chips, in particular uploading the patterns that have to be stored in their memory. The AM chips are configured through a JTAG port. The 64 chips are organized into 32 chains of 2 AM chips each. The chains are handled in parallel to minimize configuration time. The VME 32-bit wide data transfer is segmented into 4 bytes, one byte assigned to each LAMB. On each LAMB, 8 JTAG chains are handled by a small Spartan 6 FPGA, an interface between the VME and the AM chips. Pattern uploading time for a single board is ~30 seconds using block transfer.

Another important part of the configuration is the very large number of serial I/O interfaces. The AM chips alone use 640 receivers and 64 transmitters that require proper initialization. These links are also used to download patterns, so they are the first step of the configuration procedure.

III. RESULTS

A. Event Processing Validation

To test the global functionality of the system, we use a comprehensive test called "Random Test". It generates a random bank and a bunch of random events (random inputs) in order to run it for a long time and test rare conditions that could escape standard specific systematic tests. For each bunch of events, it simulates the AM system to predict which patterns will fire. We will also use it for diagnostic purposes during real data taking to debug errors on the boards in the shortest time as possible, so that a minimum number of events from the detector are lost. For the "Random Test" we use a test board able to provide the input data and receive the output matched patterns at the real experimental rate. We perform these steps:

- We generate random patterns and we download the bank in the chips.
- We generate random data, enriched with words that fire patterns.
- We simulate the data flow of the AM system and calculate the patterns that are expected to fire, taking into account the contents of the bank and the data sent to the input.
- We download the input words to the test board through VME and we let them flow to the AM system at full speed.
- Fired patterns are sent back to the test board and saved into its spy buffers. We read them by VME.
- Finally, we compare these patterns with the expected ones.

The Board has been successfully tested using these events in a long test of 3 days without any error. It will be installed in the ATLAS experiment to take data for the first time at the beginning of 2016.

B. The system Temperature evaluation and the cooling tests

The final AM system will be composed of 512 LAMBSLPs installed on 128 AMBSLPs. They will be contained in eight 9U VME core crates, each one with 16 AMBLSPs. Since the back side of the core crates is occupied by the AUX boards, the power supply (PS) of each core crate must be positioned in the rack over or below the crate, increasing the rack space necessary to host a core crate and the resistance to air flow.

Fig. 9 represents the proposed rack layout for the final AM system. A custom PS, particularly transparent to the air flow, has been built by CAEN. A single 9U box is able to power 2 crates and 4 fans trays, two for each crate.
We have reserved enough space for two fan trays per crate and three heat exchangers.

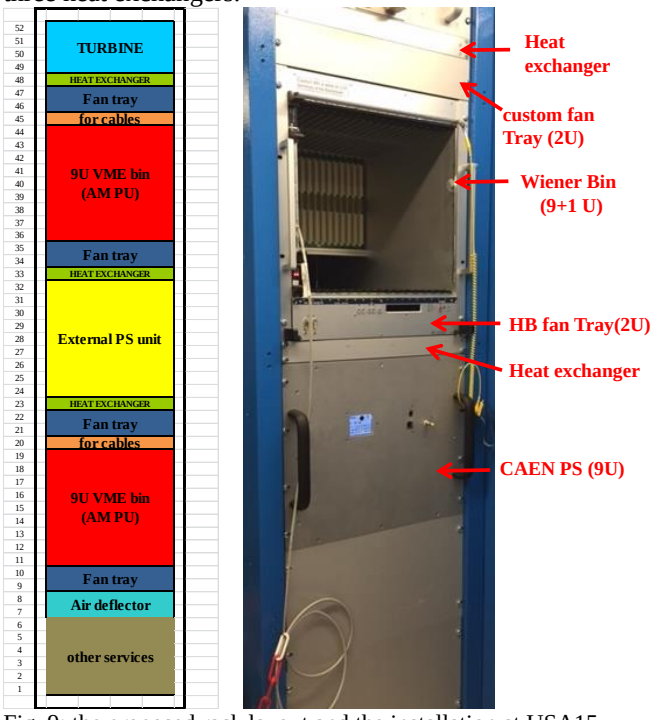

Fig. 9: the proposed rack layout and the installation at USA15.

The power dissipated on each LAMB is really high: 16 AM chips, each one dissipating roughly 2.4 W at 1 V (core) and 0.5 W at 2.5 W (I/O), and 44 fan-out chips for additional 7 W on 2.5 V. The rack structure has been designed with the goal of a silicon temperature not higher than 80^0 C. The system has been optimized using:

- ANSYS computational fluid dynamics (CFD) simulation software that allowed us to predict, with reliability, the impact of fluid flows on our product
- cooling tests that provided us temperature measurements in a similar system.

The simulation has a key role to predict the temperature difference between the die inside the package and the PCB where the chip is mounted.

The very high number of pins and balls included between the die and the substrate, oversized with respect to the electrical need, have been activated in order to have a good thermal connection between the PCB, the substrate and the die. We also optimized the PCB for maximum dissipation, increasing the thickness of the metal layers and covering with a metal surface connected to the GND the areas where the 16 AM chips' GND pins have to be soldered. The board design was exported from the Cadence Allegro PCB Designer to ANSYS computational fluid dynamics (CFD) simulation software that allowed us to predict, with reliability, the impact of fluid flows on our product. We used the ODB++ Intelligent Data Format to export the project from Cadence to ANSYS, providing all the necessary inputs for an accurate temperature study: PCB layers, stack-up and component's information. Together with the use of the bill of material, we could generate the physical model of the boards. Particular attention was dedicated to the AM chip model. In Fig. 10 the AM chip model (A), the simulation conditions (B), and the best simulation results (C) are shown.

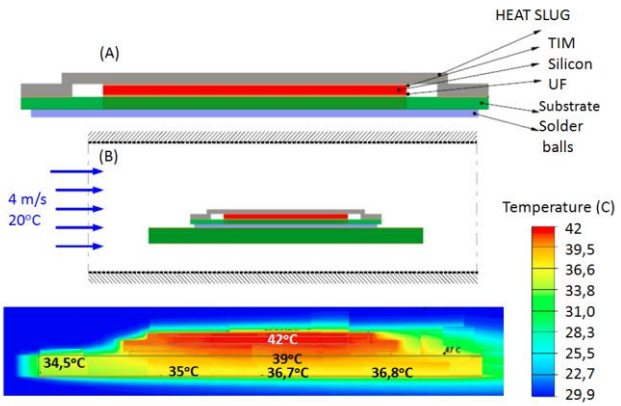

Fig. 10: (A) the AM chip package model; (B) the simulation conditions; (C) the simulation results.

We performed the analysis of the influence of the copper distribution on the PCB. The AM chip detailed component level thermal simulation shows that the temperature of the die is almost the same of the temperature of the top of the package (42 ^{o}C in a uniform air flow of 4 m/s) and just few degrees higher than the temperature on the PCB (39 ^{o}C on the top of the PCB and 36,7 ^{o}C on the bottom side). This is an important result.

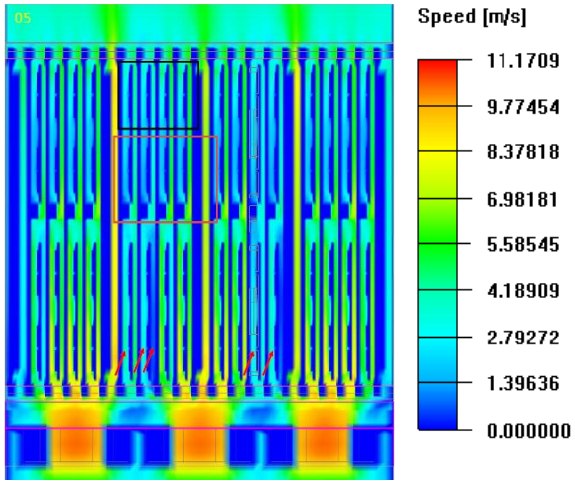

Fig. 11: Contour plot of flow intensity in a central section of the crate. Three fans are visible in the bottom of the image. Air flow speed is lower in the regions between fans. The red arrows indicate slot positions where the temperatures are expected to be higher.

The simulation was also used to predict the position of points of insufficient air flow in the crate. The flux of air in fact appears quite not uniform (see Fig. 11), especially if a single fan tray placed below the bin is used to produce the air movement. We also observed not convenient placements of tall devices (DC-DC converters, see Fig. 2) stopping the air flow in the top half AMBSLP exactly where AM chips are located, as shown in Fig. 12. The temperature in the half top board is clearly higher in the regions covered by the four tall DC-DC converters. For this reason in the subsequent AMBSLP version the DC-DC converters have been moved in positions that do not cover the AM chips (see Fig. 7).

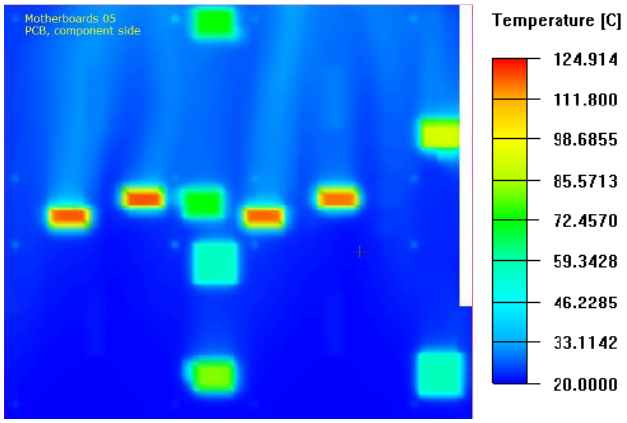

Fig. 12: Motherboard PCB temperature contour plot.

The simulation with a single fan predicted too high temperatures (~99^o C) also for the new AMBSLP in the upper part of the board in particular slots. For this reason we decided to perform real measurements in the final racks at USA15, where FTK will be commissioned, to compare two different setups: (a) one single fan tray placed below the bin, (b) two fans tray, one below and one above the bin.

The test setup uses four old boards built in the past for the Slim5 project [10], [11], the AMBslim shown in the Fig. 13. They are assembled with AMchip03 [5] built for the CDF

experiment. The consumption of the old chip is adequate to simulate the consumption of the final FTK chip [7], AM06, not yet available. It is ~2.5 W per chip when all the inputs swap at each clock cycle. Most of the power is provided from 48 V (the core) part of it from 3.3 V (the I/O).

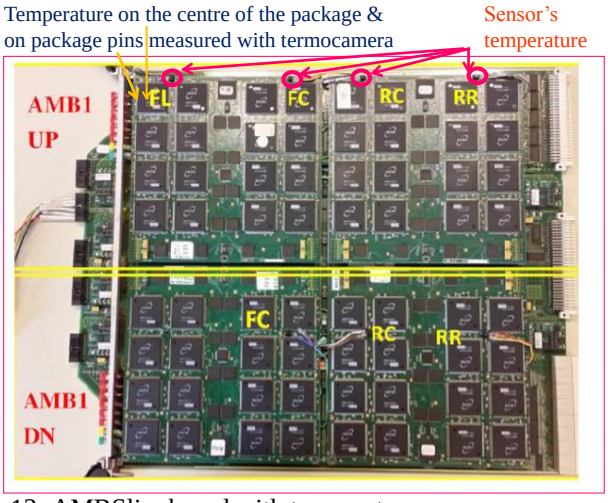

Fig. 13: AMBSlim board with temperature sensors

We have 2 types of boards: (a) type-1 with 96 AMchips for a total consumption of 280-240 W, (b) type-2 with 64 AMchips for a total consumption of 185-190 W. For type-1 the two LAMB mezzanines near the front panel have chips on both top and bottom sides (64 chips near the front panel, 32/LAMB) while the two LAMBs on the rear side, near the VME connectors, have chips only on the top side (32 chips in total, 16 chips/LAMB). Type-2 has chips only on the top side everywhere, like the uniform future board, but the total measured number of watts is somehow lower than the expected future 240 W.

We perform 2 types of measurements: (a) we use temperature sensors glued on the LAMB PCBs, as you can see in Fig. 13. The most important are the ones in the red circles on the top of the board, where less cold air arrives (in particular the one in the left top corner is predicted by the T-simulation with a single fan to be the hottest corner of the board); (b) we measure the temperature on the top left corner AMchip package using a termo-camera: on the pins and in the center of that package. The temperature on package was ~20^{o} higher than on pins, and the temperature on pins is ~6^{o} higher than on PCB. The large temperature differences are due to the fact that the AMchip03 has a plastic PQFP208 package with much worse dissipation capability compared to the new flip chip BGA package chosen for AM06.

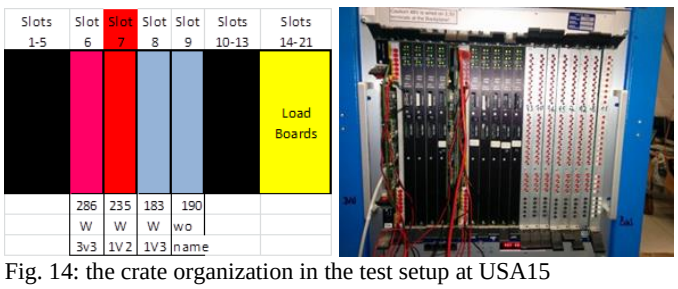

Fig. 14: the crate organization in the test setup at USA15

The test setup has 4 AMBslims, one near the other (see Fig. 14). On the left are the two type-1 boards, on the right the two type-2 ones. We fill the rest of the crate with old CDF boards (black in the figures) with 64 AMchip03 each one, and load boards full of resistors (yellow in the figure), however no one of them is connected at power, they are there just to produce the correct resistance to the air. The only boards connected to power are the 4 AMBslims. The crate measures a total current of 15 A on 48 V and 48 A on 3.3 V.

We test all the slots between 7 and 14 placing the red board in the sketch of Fig. 14 on each of these slots and repeating each time the measurements. The 4 boards are moved of one slot at time on the right, maintaining the relative position: the violet board, on the left of the 4-board packet, has the maximum consumption of 286 W. In the middle of the packet are the two boards with sensors, the type-1 (red) on the left with a consumption of 235 W, and a type-2 on the right with a total consumption of 183 W.

Fig. 15 shows clearly that different slots are not equivalent. In the figure are reported all the FL sensor measurements from slot 7 to slot 14 for the red type-1 board and from slot 8 to slot 15 for the blue type-2 board. The measurements are reported below the figure for each slot. Maximum temperature is near the center of the wheel (slot 10 in the figure). The red board has systematically higher temperatures since in the left top corner it has 2 chips for a total local delivery of power of 5 W/cm^2, while the blue board has a single chip on the top and half consumption in that specific area. However the two types of boards have the same trend showing the maximum temperature at the center of the wheel (slots 10-11).

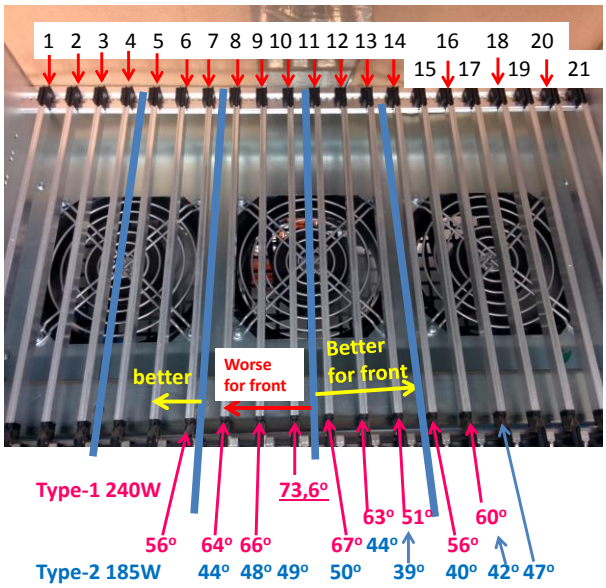

Fig.15: Temperatures measured in each one of the central slots of the bin for the FL sensor of the type-1 and type-2 boards always positioned in the middle of the 4-AMBSlim packet.

Since these temperatures are the ones on the PCB, the temperatures on the package, roughly 26 ^{o}C higher, are not acceptable (~99 ^{o}C) for the type-1 board and near the limit for the type-2 in the slots 10-11 (~76 ^{o}C). So we have inserted a second fan tray on the top of the bin to perform equivalent

measurements in these new conditions. The fan tray below the bin is the Hyper Blower (HB) fan from Wiener (see Fig. 9) while the fan on the top is a custom fan produced at INFN Pavia. The HB fan controlled by a Wiener PS could not reach fan speeds higher than 5000 rpm, while the custom fans could run at speeds above 6000 rpm.

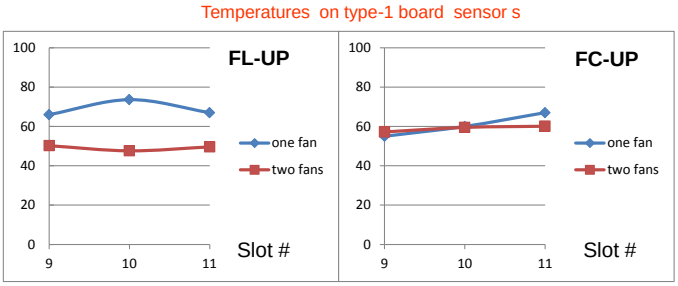

Fig 16: Comparison of temperatures measured with a single and a double fan tray in the most critical points of the type-1board at the center of the 4-AMBSlim packet.

Fig. 16 compares the hottest temperatures in the crate for the two configurations of the test stand, cooled by one or two fans. It is clear that two fans make the temperature uniform across the slots and reduce the maximum temperature at 60°C on the PCB, ~86 °C on the package, even in the areas where a power of ~5 W/cm^2 is released (front of type-1 board).

C. *Pattern Matching Performances*

Our hardware satisfies the system requirements. Candidate tracks are found exploiting the detector readout time, few clock cycles after the arrival of the corresponding detector channels belonging to the track.

This powerful highly parallel dedicated hardware has been demonstrated using the experiment simulation [7] to provide excellent performance, reaching resolutions, efficiencies and fake track rejection typical of the best tracking algorithms. For this reason, the use of the system in offline reconstruction applications has also been proposed [10], with the advantage of a low power usage (250 W/board). The system in fact is very compact and requires simplified infrastructures compared to the ones necessary for the huge CPU farms executing an equivalent task. Four racks of electronics, for a total power of ~40 kW, are able to reconstruct events with an average latency of ~100 μs [7], while offline tracking requires several seconds when performed on events containing 60 p-p collisions [4].

One interesting technology which recently has attracted the attention of the high energy physics community for real-time applications is graphic processing. Both ATLAS ([10], [14]) and CMS [15] are studying the performance of real time tracking at LHC executed on modern Graphic Processing Units (GPUs). Even if the comparison with the CPU performances is promising, the latency to execute tracking is at least tens of milliseconds for simplified algorithms and reduced detector occupancies, with a fast grow above hundreds of milliseconds when the occupancy increases. In conclusion, our hardware dedicated approach is today thousands of times faster than any available commercial computing device.

The short latencies, reachable by the parallelized AM system, push both CMS and ATLAS to study its possible application at L1 [16] for the future accelerator upgrades, when the LHC luminosity will cause hundreds of pile-up collisions and will require much faster and more efficient trigger selections.

IV. Conclusions

The presented powerful, highly parallelized pattern matching system exploits dedicated hardware to provide excellent timing performance, reaching resolutions, efficiencies and precision typical of algorithms executed on CPU farms. The system has been thoroughly tested and has been proved to be robust from both algorithmic and hardware point of view. It has been developed for the ATLAS real-time event processing for particle tracking from proton-proton collisions, but it can be adapted to generic image processing applications.

The planned future evolution includes miniaturization for its use as a coprocessor in any kind of image reconstruction, included high precision reconstruction of LHC events, based on pattern recognition on massive data throughput (namely “big data” problems).

References

[1] A. Andreani, et al., “The FastTracker Real Time Processor and Its Impact on Muon Isolation, Tau and b-Jet Online Selections at ATLAS,” IEEE Trans. on *Nuclear Science*, vol.59, no.2, pp. 348-357, April 2012.

[2] The ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation* 3 S08003, 2008.

[3] M. Dell'Orso and L. Ristori, "VLSI Structures Track Finding", *Nucl. Instr. and Meth. A*, vol. 278, pp. 436-440, 1989.

[4] The CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* 9 P10009, 2014.

[5] A. Annovi, et al., “VLSI Processor for Fast Track Finding Based on Content Addressable Memories”, in IEEE Trans. on *Nuclear Science*, vol. 53, no. 4, pp. 2428-2433, August 2006.

[6] IDT, “IDT Thermal Considerations in Package Design and Selection,” Application Note, available online: http://www.idt.com/document/apn/842-thermal-considerations-package-design-and-selection

[7] A. Andreani, et al., “Characterisation of an Associative Memory Chip for high-energy physics experiments,” in *Proc. I2MTC*, 2014, Montevideo. pp. 1487 – 1491.

[8] S. Citraro et al., “Highly Parallelized Pattern Matching Execution for Event Real-Time Reconstruction”, submitted to IEEE TNS.

[9] Xilinx Inc, “7 Series FPGAs GTP Tranceivers,” User Guide, available online:http://www.xilinx.com/support/documentation/user_guides/ug482_7Series_GTP_Transceivers.pdf

[10] S. Bettarini et al., “The SLIM5 low mass silicon tracker demonstrator”, Nuclear Instruments and Methods in Physics Research A 623 (2010) 942–953

[11] G. Batignani, et al., The associative memory for the self-triggered SLIM5 silicon telescope, in: IEEE Nuclear Science Symposium, conference record, Dresden, Germany, October 19–25 2008, pp. 2765–2769.

[12] A. Annovi, et al., “Associative memory design for the Fast TracKer processor (FTK) at Atlas,” in IEEE *NSS/MIC*, 2009, Orlando, pp. 1866 – 1867.

[13] D. Emeliyanov, et al., “GPU-based tracking algorithms for the ATLAS high-level trigger,” in *Journal of Phys. Conf.*, Ser. 396, 012018, 2012.

[14] J. Mattmann, et al., “Track finding in ATLAS using GPUs,” in *Journal of Phys. Conf.*, Ser. 396, 022035, 2012.

[15] V. Halyo, et. al., “GPU Enhancement of the Trigger to Extend Physics Reach at the LHC,” *Journal of Instrumentation* 8 P10005, 2013.

[16] A. Annovi, et al., “Associative Memory for L1 Track Triggering in LHC Environment,” in IEEE Trans. on *Nuclear Science*, Vol. 60, No. 5, pp. 3627 – 3632, 2013.