A front-end chip development for the sLHC CMS Silicon Strip Tracker

To cite this article: H Chanal et al 2012 JINST 7 C02065

View the article online for updates and enhancements.

Related content
- 2-D PT module concept for the SLHC CMS tracker
  G Hall, M Raymond and A Rose
- Development of the ABCStar front-end chip for the ATLAS silicon strip upgrade
  W. Lu, F. Anghinolfi, L. Cheng et al.
- Macro Pixel ASIC (MPA): the readout ASIC for the pixel-strip (PS) module of the CMS outer tracker at HL-LHC
  D. Ceresa, A. Marchioro, K. Kloukinas et al.
A front-end chip development for the sLHC CMS Silicon Strip Tracker

H. Chanal,\textsuperscript{a,b} Y. Zoccarato\textsuperscript{c,b,1} and D. Contardo\textsuperscript{c}

\textsuperscript{a}LPC Clermont Ferrand, Université Blaise Pascal, CNRS/IN2P3, France
\textsuperscript{b}MICRHAU pole de MICroélectronique RHône Auvergne,\textsuperscript{2} Université de Lyon, Université Blaise Pascal, CNRS/IN2P3, France
\textsuperscript{c}Institut de Physique Nucléaire de Lyon (IPNL), Université de Lyon, Université Lyon 1, CNRS/IN2P3, France

\textit{E-mail: herve.chanal@clermont.in2p3.fr}

\textbf{ABSTRACT:} The FEAFS chip has been designed for the upgrades of the CMS Silicon Strip Tracker, which is planned in view of the LHC high luminosity upgrade. Its primary function is to provide a 40 MHz selective readout of particle hits that will be used for the generation of the 100 kHz hardware trigger of the experiment within a latency of $6.4 \mu$s. To achieve this goal, the chip identifies clusters of limited number of activated strips and correlated in position, in a given window, in two closely superimposed sensors connected to the same chip. Finally, trigger and DAQ data are transmitted off detector via a common link. The FEAFS chip has been developed in IBM 0.13 $\mu$m technology. This paper presents the design of the chip and test results.

\textbf{KEYWORDS:} VLSI circuits; Trigger concepts and systems (hardware and software); Digital electronic circuits

\textsuperscript{1}Corresponding author.
\textsuperscript{2}http://micrhau.in2p3.fr/.
1 Introduction

The CERN Large Hadron Collider (LHC) has successfully started to deliver proton-proton collisions for physics at 7 TeV, at the beginning of 2010. A plan to increase the luminosity of the accelerator in the next decade has emerged, aiming to an ultimate luminosity of about $5 \times 10^{34}$ cm$^{-2}$s$^{-1}$.

Under these conditions, the CMS tracker will have to be completely rebuilt to cope with the higher particle multiplicity and to sustain increasing radiation doses. Moreover, in order to preserve the current CMS performance, it would be very important that the tracker contributes to the generation of the first trigger level (trigger level 1). To achieve this goal, the Silicon Strip Tracker data flow must be reduced to a level compatible with the expected bandwidth of future optical links ($\sim 5$ Gb/s per module) with a chip of minimum power consumption so that the overall value per module never exceed 5 W.

The proposed solutions for data reduction are based on a local measurement of the curvature of the particle trajectory in the 3.8 T magnetic field of the experiment. It will allow the on detector read-out electronics to identify the hits related to particles with a transverse momentum larger than few GeV/c, and transfer the corresponding data to the control room trigger electronics, which will combine it with the data from other sub-detectors. In the CMS solenoid magnetic field, the
deviation of the measured trajectory with respect to a radial straight line pointing to the primary vertex is inversely proportional to its transverse momentum and proportional to the radius of the measured point. Such a deviation can be estimated by measuring the width of the signal produced in a single sensor, the cluster width method, and from the transverse position difference of the hits produced in two adjacent sensors with a radial distance of about 1 mm, which are connected to the same front end chip. Details on the layout and simulated performance of the new silicon Tracker can be found in [3] and [4].

A full front-end chip for CMS silicon strip sensors at sLHC will include an analogical part, pre-amplifier and comparators, followed by a digital one. The FEAFS chip is proposed for this second part, dedicated to the data reduction and transfer to an optical link system driving the signal to the control room electronic boards. In this present version, the FEAFS chip is limited to 2 groups of 64 channels, each one simulating the input of a sensor as described above. For each of them the first stage of the chip identifies clusters of strips above threshold. In the second stage, the cluster addresses in the two groups are compared in order to find correlation. In a third part of the chip the digital outputs of all channels are buffered in a pipeline allowing the storage during the 6.4 µs latency of the first level (L1) trigger. These data are referred in the following as the raw data. The fourth part of the chip is the data link handling the L1 trigger and the shared output between the trigger and raw data paths.

2 Architecture

The architecture of the FEAFS prototype circuit, which is pure digital circuit, is presented in figure 1.

The data coming from the comparator is stored in a pipeline in the raw data path and processed in the trigger path. The latter has to find the clusters, apply a selection based on their size and finally require that the separation between two clusters belonging to two sensors, which are radially adjacent and coupled to the same readout chip, does not exceed a given value. A selection of hits based on these criteria greatly reduces the dataflow. The corresponding operations will be further described in the next section. The raw data path consists of a variable length pipeline, which has
the capacity to store all data while waiting for the L1 trigger decision. Finally, the trigger and raw data path are merged on a 4 bits shared link. This part will be described in section 4. The chip slow control is handled by an I2C like bus. It allows to configure the various thresholds and to configure the running modes. The current chip can be operated at 20 or 40 MHz.

3 Trigger path

3.1 Architecture

As shown in figure 1, the 128 digital inputs coming from the comparators are divided in two sets of 64 entries corresponding to each group of entries coming from the two “radially coupled” sensors. The processing begins by computing the size and the address of each input cluster. The clusters are defined by a sequence of logical ‘1’ separated by logical ‘0’. The algorithm is based on the approach previously developed for the D0 experiment [1]. The 128 input bits are split and analyzed in groups of 4 channels by a block called “Preliminary Cluster Search Block” (PCSB). This PCSB (based on a combinatorial logic) determines the size of the two possible clusters on the 4 input channels. If a cluster is found by a PCSB, the cluster address is directly given by the PCSB numbering. If a cluster is overlapping several PCSB, a special Merger Block (MB) computes and merges the final cluster address. The final address of an overlapping cluster corresponds to the PCSB finding the wider part of the cluster.

A wake-up signal act as a clock enable for the next blocks if a cluster is found in one of the two groups of 64 bits.

In the second processing block, the clusters sizes are compared to a programmable threshold. Only those with a size below the threshold are kept.

The third block is a zero suppression block based on priority encoders with a maximum output size of 6 clusters. This choice is discussed in the next subsection.

The fourth block is used in the coincidence finding processing. When this mode is enabled, the 6 addresses from the two set of 64 strips are compared. A cluster of address $c_1$ in the first set of strip and a cluster of address $c_2$ are kept only if, with a programmable window threshold $t_w$ and offset $t_0$, the following condition is fulfilled:

$$|c_1 + t_0 c_2| t_w$$

(3.1)

$t_0$ is a programmable parameter allowing alignment of the two sensors and compensating difference in radius from center to edge, due to their planarity.

Lastly, given the expected mean cluster occupancy ($< 1\%$) and in order to reduce the data throughput, a priority encoder reduces the number of transmitted cluster from 12 to 4.

3.2 Cluster loss

The number of cluster lost in the priority encoder processing as been studied as a function of the mean cluster occupancy (before the size selection) in the cases of 4 and 6 clusters finally accepted in the trigger path. The simulation was performed with a simplified implementation of the cluster size, leading to an underestimated rate of clusters with a large size and therefore an over pessimistic result on the loss. In figure 2, it can be seen that there is no loss when accepting 4 trigger clusters.
Figure 2. Cluster loss resulting from the limited number of clusters accepted for the final trigger.

below the expected mean cluster occupancy of 1%. On the other hand, if the occupancy increases, a cut at 6 might be better suited.

This preliminary study justifies the choice of preserving 6 clusters after the size selection and 4 after the space correlation. A more realistic simulation of the cluster size and review of the real mean occupancy will be needed to definitely decide the value of these parameters.

4 Raw data path

This bloc is devoted to the readout of all non zero suppressed data of the Tracker for events selected at the level 1 trigger. It is build around a shift register allowing the data storage during the level 1 trigger latency. The maximum L1 latency in CMS is 6.4 µs, hence for an LHC clock frequency of 40 MHz the shift register length must be at least 256. The final latency will be chosen by an extra adjustable delay on the L1 trigger to synchronize with the data (this extra delay will be added externally).

5 Shared data link

5.1 Asynchronous FIFO

As the throughput of the output bus depends on the strip activity, it is necessary to implement a FIFO in the chip design, both for the read-out and trigger paths. Its depth has been fixed to 16 words. The ability to have two different asynchronous clock domains will allow adjusting the output bus frequency in order to prevent overflows.

The architecture and implementation of such a FIFO are described in [2]. The use of a FIFO on the trigger data path leads to a loss of the timing reference since the data latency is strongly dependant on the FIFO state (empty, full, almost full...). A simple way to preserve the time
reference is to add a timestamp to each event before writing it in the FIFO. However, such a solution leads to a higher data throughput on the output bus and to a larger bus size. Instead, it was decided to send the FIFO write signal to the receiver chip that will have to log the corresponding timestamp. This solution requires adding a dedicated line to the output bus.

5.2 Output bus management

In order to minimize the power consumption of the ASIC, only one output bus has been implemented to send the data from both FIFOs (trigger and readout). Two different blocks allow performing this function (cf. figure 1).

5.2.1 Communication controller

This block performs the output bus arbitration with the goal to give higher priority to the trigger data. Arbitration is made depending on the source of the data (trigger or readout) and also on the state of the FIFO (following the data fill rate). The four following functional modes are implemented:

- Normal mode: both FIFOs are not full. The trigger data are sent with a higher priority as soon as they come. The readout data are sent between two trigger packets.

- De-rated mode: the “readout FIFO” is not full, but the “trigger FIFO” is full. As in the normal mode, higher priority is given to the trigger data, but the readout data FIFO output is stopped. During this mode all the new incoming triggers are lost (a signal called “trigger-off” is activated).

- Busy mode: the “trigger FIFO” is not full, but the “readout FIFO” is full. The data coming from the “readout FIFO” are read with the highest priority. During this mode all the new incoming readout data are lost (a signal called “busy” is activated).

- Survival mode: both FIFO are full. A word is alternatively read from each FIFO starting with the “trigger FIFO”. All new incoming trigger and readout data are lost, and the signals “busy” and “trigger-off” are asserted.

5.2.2 Multiplexer

Following the priority given by the “communication controller” this block performs the effective selection of the two possible sources of data (cf. figure 4 for the raw data frame and figure 3 for the trigger frame). All data are sent on the same 4 bits bus output, ID first. The first ID bit is set to 1/0 respectively for readout and trigger packets. The trigger data are sent as a single packet with a minimum of 3 words and a maximum of 8 words (the first word of each packet containing its size). The readout data are sent as 8 different packets of 5 words: the first word of each packet contains a sequence number. The readout data are split in several packets in order to allow interleaving with the higher priority trigger packets.
6 Performance evaluation

The FEAFS chip has been submitted for fabrication in the 130 nm IBM process using the VCAD standard cells library provided by CERN. It has been designed using the Cadence tools. The layout is shown in figure 5. In order to limit the size of this prototype, 32 multiplexed inputs have been used instead of 128 leading to a circuit size of $2 \text{ mm}^2$. It uses almost 60 k standard cell, most of them in the readout data pipeline that has currently been limited to a size of 135 registers.

The chip has been delivered beginning of 2011 and tested last summer. Two types of test were performed using the development board DE2-115 from Terasic (cf. figure 6).

- A random bit pattern, emulating the event data with variable occupancy, is applied at the input of the FEAFS chip by the DE2-115 FPGA. The FEAFS output bus is compared with an emulated FEAFS design in the DE2-115 FPGA. This setup is used to test the FEAFS chip at full speed on a long run period.
Figure 6. Test bench using DE2-115 board from Terasic.

Table 1. Power consumptions of the main functional blocks.

<table>
<thead>
<tr>
<th>Functional block</th>
<th>Power consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strips PAD</td>
<td>29.8 mW</td>
</tr>
<tr>
<td>Pipeline of the raw data</td>
<td>15 mW</td>
</tr>
<tr>
<td>Clocks PAD</td>
<td>8.4 mW</td>
</tr>
<tr>
<td>FIFO raw data</td>
<td>2.5 mW</td>
</tr>
<tr>
<td>Cluster search</td>
<td>5.5 mW</td>
</tr>
<tr>
<td>Communication controller</td>
<td>12.8 mW</td>
</tr>
</tbody>
</table>

- A specific pattern written in DE2-115 FPGA RAM is applied to the chip. The FEAFS output is recorded for a software analysis. This setup is used to test the chip at reduce speed on a specific relatively short sequence.

The chip response is in perfect agreement with the simulations. The following sections summarize the main results obtained during these tests.

6.1 Power consumption

The power consumption was measured for a strip occupancy of about 3%, at nominal frequency (40 MHz for the cluster search function and 100 MHz for the output link) and at nominal voltage power supply (1.8 V). The total power consumption of the FEAFS ASIC is 74 mW (75 mW was measured in simulation). The power consumptions of the main functional blocks are summarized in table 1:

30 mW are due to the 32 multiplexed inputs strip sequenced at the frequency of 160 MHz, this contribution will completely disappear when the analogue part will be integrated in the same chip.
6.2 Cluster finding

To completely test the functionality of the cluster finding function, test vector simulated clusters with occupancy of 3% had been applied to the FEAFS chip. Figure 7 represents the clusters found by the cluster finding function when the correlation function is disabled (each cluster is represented as a function of this position on the 2 sensors).

Figures 8, 9, and 10 represent the good functioning of the correlation function for several values of the parameters offset and window in equation (3.1) of the section 3.1.

For all these tests, the results obtained are in agreement with simulations, demonstrating the good functioning of FEAFS chip.

7 Conclusion

A chip including the main digital features (Cluster ID and selection, stub identification, pipeline, FIFO, multiplexer, serial line) of the Front-End read-out for silicon modules of a future CMS tracker at sLHC, with trigger capabilities, has been developed in the 130nm IBM technology. It is fully functional and a new version is now proposed with an improved data format and with adap-
tation to the GBT (GigaBit transceiver system from CERN). The power consumption needs also to be minimized; new pipeline (using RAM instead of D flip-flop), output interface and clustering algorithm are under development. Further simulations with realistic sLHC events are ongoing to adjust the design parameters in order to minimize cluster losses and trigger latency.

References


