A Pixel Front-End ASIC in 0.13 μm CMOS for the NA62 Experiment with on Pixel 100 ps Time-to-Digital Converter

S. Martoiu , A. Rivetti, *Member, IEEE* , A. Ceccucci, A. Cotta Ramusino, S. Chiozzi, G. Dellacasa, M. Fiorini, S. Garbolino, P. Jarron, J. Kaplon, *Member, IEEE*, A. Kluge, F. Marchetto, E. Martin Albarran, G. Mazza, *Member, IEEE,* M. Noy, P. Riedler and S. Tiurianemi

***Abstract*—The paper describes the design of a front-end chip for hybrid pixel detectors optimized for good timing resolution (200 ps rms) and high event rate (150 kHz per pixel). Each channel consists of a fast transimpedance amplifier with 5 ns peaking time, a constant fraction discriminator (CFD), and a Time-to-Digital Converter (TDC). In order to cope with the rate requirement, a multi-event buffering scheme employing both analog and digital pipelines is implemented in each cell. This development is part of the R&D activity for the silicon tracker of the NA62 experiment at CERN. The architecture of the chip and the design of the critical building blocks are discussed in the paper.**

***Index Terms*—Constant Fraction Discriminator, Pixel Detector, Time-to-Amplitude Converter, Time walk.**

I. Introduction

THE NA62 experiment [1] will exploit the CERN Super Proton Synchrotron (SPS) accelerator to probe the very rare decay of a K^+ meson into a π^+ and a neutrino-antineutrino pair. NA62 will use a DC beam with an average particle rate of 800 millions particles per second having a momentum of 75 GeV/c. A potential interesting event is flagged by a final state containing only a π^+, whose track must be correlated in time with the one of the parent Kaon with a resolution of 150 ps rms. The momentum of all the incoming particles must also be measured before they enter the decay volume. The kinematics information will be used in the offline analysis to suppress part of the strong background, since the decay channel of interest has a branching ratio of 10^{-12}. The detector that will perform the momentum measurement and provide the time-stamp information with an average resolution of 150 ps rms is called "Gigatracker" [2] and is formed by three stations of hybrid silicon pixels detectors. Each station consists of a single sensor covering an area of 60 mm x 27 mm. The sensor will be read-out by ten front-end chips connected to it via bump-bonding. The pixel size is 300 μm x 300 μm and the final ASIC will contain 1800 independent channels organized in a matrix of 45 rows per 40 columns. The read-out will be triggerless and each channel, when firing, will generate a stream of 32 bits. This leads to an average data rate of 26 Gbits/s for each station. Since the particle density will peak in the central region, the chip located here will have to sustain a throughput of 6 Gbits/s. The maximum power consumption allowed is $2W/cm^2$, resulting in 2mW per pixel. However, in allocating the power budget to the different blocks one must take into account that a significant power will be required by the data transmission system. Therefore, the power available for the front-end electronics is of the order of 1mW per channel. The key challenge is providing within this power budget a time stamp with accuracy better than 200 ps rms. The task is further complicated by the event rate, which is on average 45 kHz per pixel, but peaks in the central region at 140 kHz. The design of the ASIC for the Gigatracker is now in the R&D phase and two different architectures are being explored. The first one is based on a "minimal" pixel cell formed by a transimpedance amplifier followed by a leading edge discriminator [3]. The discriminator output is sent downstream to the end of the column, where it is used to latch into local registers a timing word provided by a DLL based TDC. Both the leading and the trailing edges of the signal are stored, so the time-over-threshold information is available for off-line correction of the time-walk. The second architecture follows a complementary approach, trying to exploit the relatively big pixel area to perform most of the signal processing, including the time-to-digital conversion, already at the pixel level. A detailed analysis of the advantages and drawbacks of both solutions can be found in [4], [5]. In this paper we focus on the design of the TDC per pixel option. Section II describes the overall architecture. Section III to V discuss respectively the design of the front-end amplifier, the constant fraction discriminator and

Manuscript received November 13, 2009.

G. Dellacasa, S. Garbolino, F. Marchetto and A. Rivetti are with INFN-Sezione di Torino, via Pietro Giuria 1 – 10125 – Torino –Italy

S. Martoiu was with INFN, Sezione di Torino. He is now with CERN, Geneva, Switzerland.

S. Chiozzi and A. Cotta Ramusino are with INFN- Sezione di Ferrara, Via Saragat 1 – 44100-Ferrara, Italy.

A. Ceccucci, M. Fiorini, P. Jarron, J. Kaplon, A. Kluge, M. Noy, P. Riedler are with CERN, Geneva, Switzerland.

E. Martin Albarran is with Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Email: rivetti@to.infn..it. Phone:+39 011 6707380.

the time-to-digital converter, while the digital logic is briefly treated in Section VI.

II. ASIC ARCHITECTURE

The ASIC operates with a master clock of 160 MHz. A Gray counter counts the clock pulses and the resulting word is sent to all the pixels, providing the coarse time measurement. In each pixel the front-end stage is formed by a fast transimpedance amplifier followed by a constant fraction discriminator. When the CFD fires the output of the coarse counter is stored into a local register. The CFD pulse starts in addition a voltage ramp, which is stopped on a suitable clock edge and stored into a capacitor. The resulting voltage, which is proportional to the time elapsed between the occurrence of the CFD pulse and a known clock transition, is then digitized and provides the fine time resolution.

The clock period is 6.25 ns, so to obtain a time bin of 100 ps 6 bits in the local TDC would be sufficient. However, one must take into account the possibility of timing misalignment between the coarse and the fine time logic which might occur when the CFD fires in close proximity of a clock transition. Furthermore, one must guarantee to the circuits generating the voltage ramp an adequate time for settling. For these reasons the TDC logic works at half the master clock frequency (80 MHz). Some relevant timing waveforms are shown in Fig. 1. The ramp is stopped on the first leading edge of the 80 MHz clock following a trailing edge. In this way a time offset is added and the risk of halting the ramp immediately after the occurrence of the CFD pulse is avoided. The maximum duration of the ramp therefore spans 3 cycles of the master clock or 18.75 ns. With a time bin of 97.5 ps this requires 192 codes, hence the ramp is digitized with 7.5 bits equivalent resolution. With such an arrangement the LSB of the coarse counter and the MSB of the fine counter overlap, thus giving the possibility of a digital error correction.

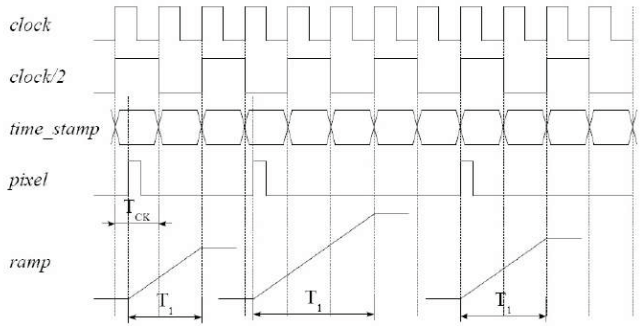

Fig. 1. Relevant timing waveforms used in the ASIC.

To reduce the power consumption the digitization of the voltage is performed with a Wilkinson ADC working on the master clock frequency, therefore the maximum conversion time is 1.2 µs. Since the maximum event rate is 140 kHz per pixel, a multi-event buffering scheme in needed to keep the event loss to an acceptable level. Simulations have shown that if four buffers are used the event loss is less than 0.2 %. The multi-buffer has a digital component (the registers needed to store the coarse counter output) and an analog one (the circuits that generate the voltage ramp). The use of derandomization at the pixel level simplifies the design of the read-out logic that can be tailored on the average event rate, which in the worst case is 5.5 MHz per column. On the other hand the design of such a complex pixel entails some issues. First, the logic circuitry inside the pixel will be exposed to a high particle flux, so the registers have to be protected against Single Event Upsets. This increases the amount of digital gates present in the pixel. Second, these gates are located in close proximity of sensitive analog blocks. To minimize the risk of coupling digital noise into the analog part the front-end amplifier and the CFD employs a fully differential topology.

In this ASIC, the clock jitter must be maintained below the required time resolution and the clock distribution is performed using integrated transmission lines available in the selected fabrication process. The chip has been designed and produced in a 0.13 µm CMOS process. One of the aims of our work was to design a prototype as realistic as possible while maintaining the costs to an acceptable level. The floorplan of the implemented ASIC is shown is Fig. 2. The chip contains two full columns having the final number of pixels (45). To keep a reasonable form factor the columns are folded into three slices of 15 cells each but the critical signals are routed in such a way that the wires have the same length expected for the final ASIC (1.4 cm). One shorter column containing only 15 pixels has been also implemented together with spare samples of the critical analog building blocks.

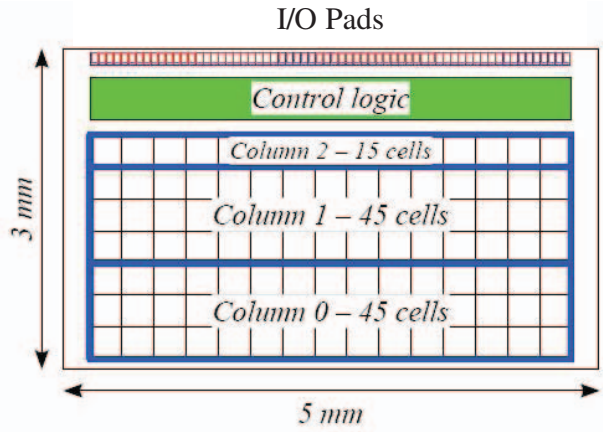

Fig. 2. Floorplan and size of the designed prototype.

III. FRONT END AMPLIFIER

A. Circuit topology

The front-end amplifier, reported in Figure 3, is a transimpedance amplifier with fully differential outputs. The baseline choice for the sensor material is p-in-n, however the front-end amplifier has been designed to be compliant with signal of either polarity to maintain the possibility of a different option in future upgrades of the Gigatracker. The

preamplifier has a nominal peaking time of 5 ns, a gain of 80 mV/fC and is designed to be linear up to 10 fC. The heavy radiation load to which the sensors will be exposed will increase significantly their leakage current, so a compensation system has been provided. This system consists of a low pass filter, formed by resistor R4 –R5 and capacitor C2-C3.

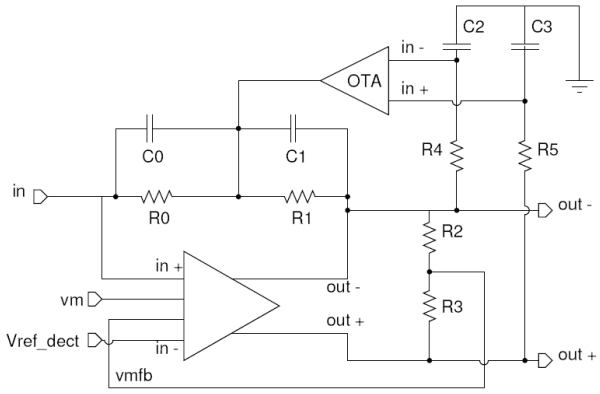

Fig. 3: The front-end amplifier.

The filter is followed by an Operational Transconductance Amplifier (OTA) that sinks (or sources) the detector dark current through R0, avoiding that it passes also through R1. In this way the two output nodes are kept at the same potential. The circuit can tolerate a leakage current of ± 200 nA while maintaining the shift in the DC output voltage below ± 10 mV. Resistors R4 and R5 are implemented as long diode-connected PMOS transistors, while C3 and C4 are MOS capacitor. The output of the OTA is tied to the mid point of the feed-back network as a compromise between compensation efficiency and noise. Resistors R0 to R3 and capacitors C0-C1 are implemented as linear components (polisilicon resistors and metal-to-metal capacitors, respectively).

B. Transistor-level design

The transistor level schematic of the core amplifier used in the front-end is shown in Fig. 4. The circuit employs a telescopic cascode architecture for maximum bandwidth. To increase the DC gain, part of the current necessary to bias the input devices is provided by two additional current sources and does not flow into the loads. All NMOS transistors are insulated from the chip substrate using triple well devices.

The load transistors M4-M5 are driven by the common-mode feed-back amplifier (M6 to M9). The gain of this stage must be kept low enough in order to prevent oscillations in the common mode feed-back loop. Since the preamplifier has to drive the resistive load of the CFD filter, a class AB output stage, shown in the inset of Fig. 4 is also provided.

The front-end amplifier dissipates 200 µW from a 1.2 V power supply and the simulated input referred noise is 250 electrons rms with a total input capacitance of 200 fF.

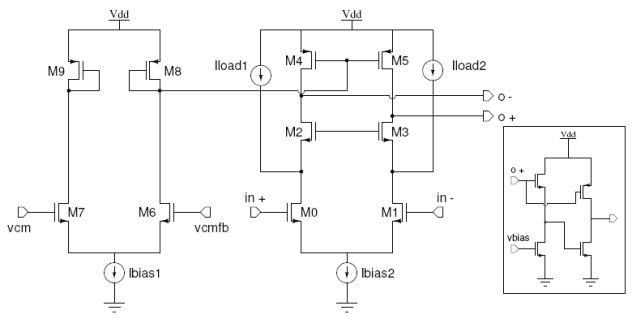

Fig. 4: Transistor level implementation of the core amplifier used in the front-end. The class AB output stage is shown in the inset.

IV. THE CONTANT FRACTION DISCRIMINATOR

The Constant Fraction Discriminator (CFD) [6], [7] is the block providing the precise timing pulse. Shown in Fig. 5, the CFD is formed by a passive RC filter followed by a three stage discriminator. The choice of a CFD stems from the fact that in this circuit the time-walk is compensated with analog preprocessing, so a single time measurement is needed. This is an advantage since the Time-to-Digital Converter is relatively slow and performing a double measurement on both the leading and trailing edges will result in a significant burden.

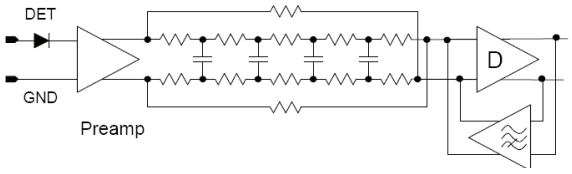

Fig. 5: The constant fraction discriminator. The two cross-connected resistors generate the signal fraction while the cascaded low-pass filters generate the delay.

A. The CFD filter

A constant fraction discriminator works by comparing a delayed and inverted copy of the input signal with an attenuated fraction of the same signal. In an integrated implementation the delay line can be emulated with a RC filter. Since the circuit has a fully differential topology, the signal fraction can be obtained by cross-connecting two resistors between the outputs of the filter and the outputs of the front-end amplifier [8]. From the point of view of jitter minimization the zero crossing should occur at the point of maximum slope. In our design this requires to set the fraction to one half of the signal. Statistical fluctuations in the signal generation process in the sensor may however result in variation of the signal shape, which would lead to an additional jitter component. To minimize this effect it is beneficial to reduce the fraction at which the zero crossing occurs. In the circuit, the attenuation fraction can be programmed by a configuration bit to be either 50% (which minimizes the noise contribution to the jitter) or 30 % (which

reduces the effect of signal shape variations).

.

B. *The zero crossing detector*

After the filter the signal is firstly amplified by a circuit consisting of two cascaded differential pairs with resistive load before being presented to the zero-crossing detector. A typical problem is that this circuit has a nominal threshold equal to the baseline, so it fires continuously also on noise. To prevent these spurious hits from propagating to the rest of the system the output of the zero crossing detector is usually put in logical AND with the one of a leading edge discriminator. The input of this stage is connected to the non-delayed output of the front-end amplifier and a threshold well above the baseline is set in order to suppress the noise.

The leading edge comparator fires before the zero-crossing, "arming" in this way the logical gate and allowing the

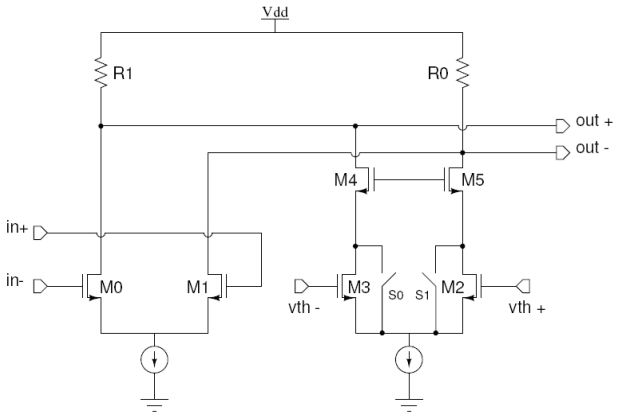

Fig. 6: Schematic of the zero crossing detector used in the CFD.

propagation also of the walk-free signal. One problem in using this approach in our design is that in a final implementation the ASIC will have 1800 channels. Having a stage that fires continuously is then a primary concern. For these reasons we have preferred to employ an alternative approach, based on the use of a "dynamic" hysteresis. Reported in Fig. 6, the circuit consists of two differential pairs sharing a common resistive load. The working of the circuit can be better understood with the help of Fig. 7, which displays the most relevant waveforms. M0 and M1 receive from the previous stage the bipolar waveform shown in the bottom part of Fig. 7. M2 and M3 are used to set a differential threshold that unbalances the circuit and avoids triggering on noise. This threshold is visualized by the red horizontal line. When the threshold is crossed the leading edge transition occurs. This information is used to short-circuit M2-M3, thus disabling the threshold. The high-to low transition occurs then at the zero-crossing point. Since the circuit is powered by a single rail power supply, this level is physically at about 0.6 V. After the zero-crossing a differential-to-single ended converter generates a CMOS pulse that is propagated to the rest of the circuitry. The CFD was

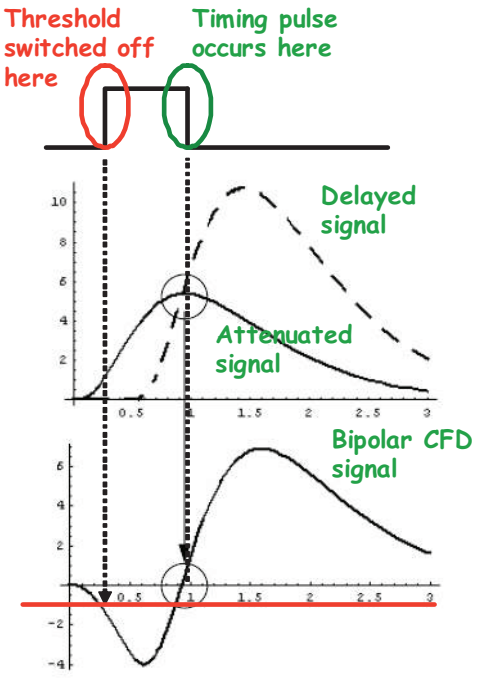

Fig. 7: Relevant waveforms for the circuit of Fig. 6

previously prototyped as a separate chip. The experimental results have shown that the circuit achieves an optimal walk correction for an input signal of 24000 electrons. As it can be seen in Fig. 8, the time walk is always smaller than 300 ps in the dynamic range of interest.

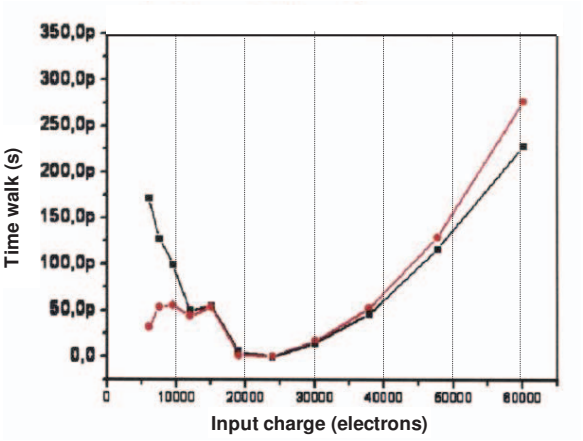

Fig. 8: Measured performance of the CFD.

V. THE TIME TO DIGITAL CONVERTER

The time-to-digital converter has a time bin of 97.5 ps. To understand the details of the design it is first necessary to describe the circuit used to generate the voltage ramp, which is shown in Fig. 9. The circuit contains a cascode common source amplifier, a linear metal-to-metal capacitor and a set of

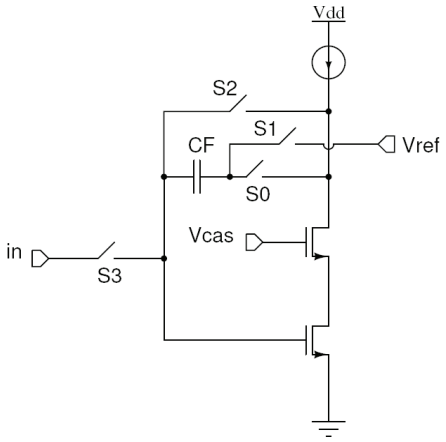

Fig. 9: The Time to Amplitude Converter.

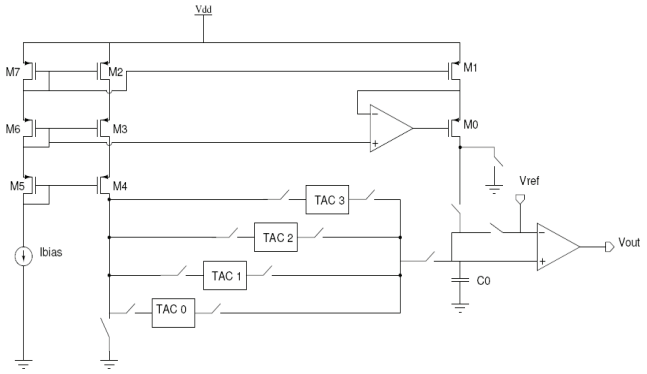

Fig. 10: Simplified schematic of the full TDC.

four switches. When S0 is open and S1-S2 are closed, the circuit is in the reset mode. When the vice-versa is true, the circuit is in the normal operating mode. The CFD triggers the closing of S3, which connects the input node to a constant current source. The current enters into the input node, therefore the output voltage is pulled down. Four of these circuits are located in every pixel, so four voltage ramps can be generated one after the other, thereby providing the analog multi-buffering capability. The schematic of the whole TDC is shown in Fig. 10. Transistors M2-M4 form the current source creating the ramp. In the idle state this current is diverted to ground. When the CFD fires the current source is connected to one of the four TAC. When the ramp is completed the output of the TAC is sampled on capacitor C0. One has to note that since the output voltage of the TAC will be less then Vref, the sampling will actually discharge C0. Therefore, C0 must be recharged to restore the baseline. This implies that one needs here a current source of the same polarity of the one that generates the voltage ramp. The ratio between the sampling capacitor and the feedback capacitor CF used in the TAC can be exploited to reduce the ratio between the current sources. For instance, in a straightforward 6 bits implementation, the current restoring the baseline on C0 should be 64 times smaller than the one used in the TAC. However, if C0 is four times CF the ratio required is reduced to sixteen, allowing for a significant saving in the total circuit area.

In our case we want to have a time bin of 100 ps and cover a dynamic range of 18.75 ns. This is equivalent to requiring a resolution of 8 bits over a range of 25 ns. If one would achieve this resolution just by scaling the currents a ratio of 256 would be required. However the time needed to restore the baseline is measured with the master clock cycle (6.25 ns) gaining a factor of four. Another factor of four is obtained by making C0 four times bigger than CF, therefore the ratio necessary between the two current sources is only sixteen.

VI. THE DIGITAL LOGIC

The digital logic can be divided in two major components: the in-pixel logic, located inside the active area, and the End of Column logic (EoC), which takes care of the read-out of the pixels. The logic located into the pixel cell contains mainly registers to store the coarse time values and the results of the TDC conversions. Small state machines control the switching between the buffers and manage the interface with the EoC logic. In the EoC, a finite state machine sends a read-enable to the pixels which is passed by one cell to the next with a token ring scheme. When its read-enable is high, the pixel puts on the output bus its data content. In addition, a busy signal alerts the End of Column controller if there are data in at least one of the 45 pixels in a column. In order to avoid ambiguities in the time reconstruction the event belonging to the same cycle of the coarse counter are grouped together. This rearrangement is performed by adding an extra bit to the coarse time stamp. This information is used to tag events of type “0” and events of type “1” putting them in two different FIFOs. A merger circuit reconstructs the right order of the data frame, adding a header and a trailer. The header contains the information on the frame counter value. To separate the frames the controller uses additional information provided by the pixel logic. Each pixel sets the “old-data” flag if it still contains data when the frame changes. This signal is propagated to the whole column via a fast or. In the present prototype the data are sent out independently for each column through a serial line with a speed of 160 Mbit/s.

Due to the high particle flux, the digital circuits must be protected against Single Event Upset (SEU) and the following solutions have been adopted:

- Hamming encoded state machines and registers with auto correction in the pixel cell.
- Hamming encoded state machines in the End of Column logic.
- Triple Modular Redundancy with auto correction for single flip flops in the EoC.

A single event correction, double error detection policy has been generally adopted. In the EoC FIFOs the data are

Hamming encoded at input and are checked and corrected at the output of the FIFO. In fact, the data will stay in the FIFO for an average time of 12 μs and it has been calculated that the probability of more than one upset during this time is negligible. Therefore, protecting all the registers in the FIFO would add a heavy burden without a significant advantage.

The End of Column Logic occupies an area of 1500 μm x 300 μm.

VII. Summary and Outlook

We have developed a prototype of a front-end ASIC for hybrid pixel detectors designed to measure the time of occurrence of an event with a resolution of 200 ps rms or better. The analog part of the circuit is based on a transimpedance amplifier, a constant fraction discriminator and a Time-to-Amplitude Converter and dissipates less than 1 mW per channel from a 1.2 V power supply. The circuit has been designed and produced in a 0.13 μm CMOS technology. At the time of writing the test set-up is in being finalized and the measurements are expected to begin before the end of the year. After a first phase of electrical characterization the ASIC will be bump bonded to a silicon sensor and the timing performance will be studied in detail using both lasers and particle beams.

References

[1] NA62 collaboration, NA62 proposal, CERN-SPSC-2005-013 SPSC-P-236

[2] M. Fiorini et al., "The P326 (NA48/3) Gigatracker: Requirements and desing concept", *Nucl. Instr. Meth.* (A572), pp. 290-291, March 2007.

[3] E. Martin Albarran et al., "The 5ns Peaking Time Transimpedance Front-End Amplifier for the Silicon Pixel Detector in the NA62 GigaTracker", 2009 IEEE Nuclear Science Symposium Conference Records

[4] A. Kluge et al., "An ultra fast 100ps, 100 μm, 3D-pixel imager", Proceedings of the 2008 SPIE conference. Volume 7249, 724909 (2009).

[5] G. Dellacasa et al., "Readout electronics for the NA62 Gigatracker system", Proceedings of the "Topical workshop on electronics for Paricle Physics", 2008.

[6] H. Spieler, "Semiconductor Detector Systems", Oxford Science Publications, ISBN-978-0-19-852784-8, 2007, pp. 185-187.

[7] D. M. Binkley et al, "A 10-Mc/s, 0.5-μm CMOS Constant-Fraction Discriminator Having Built-In Pulse Tail Cancellation", *IEEE Trans. Nucl. Sci.*, vol 49, no.3, June 2002, pp. 1130-1140.

[8] S. Martoiu et al., " A Low-Power Front-End Prototype for Silicon Pixel Detectors with 100ps Time Resolution", 2008 IEEE Nuclear Science Symposium Conference Records, N44-5