POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Low-Power, Short Dead-Time ASIC for SiPMs Readout with 200 MS/s Sampling Rate

(Article begins on next page)

25 April 2024

A Low-Power, Short Dead-Time ASIC for SiPMs Readout with 200 MS/s Sampling Rate

Silvia Tedesco

Department of Electrical, Electronics and Communications Engineering
Politecnico di Torino
INFN Torino
Turin, Italy
Email: silvia.tedesco@to.infn.it

Abstract—The design of a low-power, 64-channels front-end ASIC for Silicon Photomultipliers is presented. The chip is being developed in a 65 nm CMOS technology and it is optimised for space applications. In each channel, the current pulse delivered by the sensor is amplified, converted into a voltage and sampled at 200 MS/s by an array of 256 cells, each containing a storage capacitor and a single-slope ADC. If a trigger signal is received, the analog samples are digitised in parallel and sent off-chip, otherwise the memory cells are overwritten. The ADC resolution can be programmed in the 7-12 bit range, trading-off dead time with amplitude resolution. The target power consumption is 5 mW/channel. The chip can thus take snapshots of relatively rare events at high sampling rate with low power. The analog memory can be partitioned in shorter slots that work in a time-interleaved configuration. In this way, the input data stream, which usually follows a Poisson distribution, can be derandomized. The chip is scheduled to be submitted for fabrication in the second quarter of 2022. In the paper, the design concept is presented and the ongoing verifications are discussed.

Index Terms—SiPM, low-power, multichannel, analog memory

I. INTRODUCTION

Silicon Photomultipliers (SiPMs) are solid-state photodectors based on Single-Photon Avalanche Diodes (SPADs) organized in matrices [1]. These sensors offer low-voltage operation, higher robustness to magnetic fields and mechanical stress and higher photodetection efficiency compared to traditional vacuum tube Photomultipliers (PMTs).

Thanks to their excellent timing performance, SiPMs are employed in medical imaging applications such as Positron Emission Tomography (PET) [2] or Single Photon Emission Computed Tomography (SPECT) [3]. Moreover, SiPMs can provide high energy resolution measurements. Therefore, they are becoming increasingly popular in High Energy Physics (HEP) and space applications [4], [5]. SiPMs are considered to instrument the focal plane of future space-born telescopes that will detect the Cherenkov and fluorescence light produced in the Earth atmosphere by Ultra-High Energy Cosmic Rays (UHECRs) and Cosmic Neutrinos (CNs) [6]–[8]. Since Cherenkov radiation produces fast and large transients, a sampling frequency of minimum 100 MHz and a dynamic range of at least 10 bits are required to conveniently extract the relevant signal features. In addition to the detection of

pulses produced by highly energetic cosmic particles, which might be extremely rare[1], a periodical monitoring of the sky background is also of interest. SiPM-based cameras are thus expected to take high resolution snapshots at a rate of a few tens of Hertz. The front-end chip should incorporate all the relevant processing chain in order to allow for a compact and low mass system. Power consumption must also be kept as low as possible. With these constraints, it would be inefficient to have a fast and high resolution ADC digitising continuously the sensor signal. It is instead preferable to sample the input signal at high rate and to perform a full A/D conversion only if a given time window is validated by a trigger. This paper presents a 64-channel ASIC which is being developed to satisfy the above requirements. In order to have some margin, the chip is designed for a 200 MHz sampling frequency and a 12 bit dynamic range. Each channel is equipped with an array of 256 sampling and digitisation cells. A/D conversion is carried out by single slope ADCs embedded in each cell, while digital data are transmitted off-chip through Double Data Rate differential links. The sampling array can be configured to work as a single buffer that can capture a time window of 1.28 μs or it can be split into shorter segments that work in a time interleaved configuration. A maximum of 8 independent segments with 32 sampling cells each can be provided. This multi-buffer mode allows for event derandomization. Each channel can work independently (sparse mode) or all channels can be synchronised to provide a coherent snapshot of the full focal plane (imaging mode).

The paper is organized as follows: Section II describes the ASIC architecture in more detail, with a focus on the analog memory design. The trigger strategy is also discussed. The simulation results are shown in Section III, while conclusions are drawn in Section IV.

II. ASIC ARCHITECTURE

The mixed-signal ASIC is developed in a 65 nm CMOS technology. The block diagram of a single channel is shown in Fig. 1. The key building blocks are described hereafter.

---

[1]Planned space experiments expect to detect 0.1-10 UHECR per day and up to 1 or a few neutrinos per life mission.
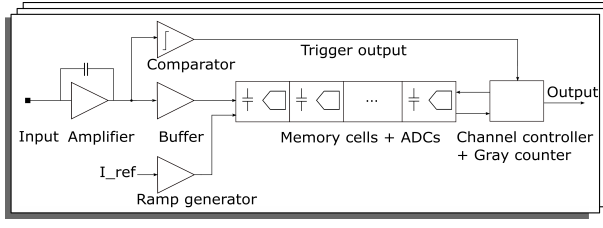
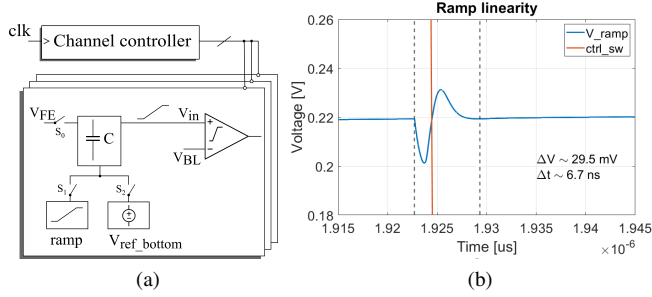Fig. 1. Channel block diagram.

(a)   (b)

Fig. 2. (a) Memory cell block diagram. (b) Simulation.

A. Input Amplifier

The input stage employs a common gate topology with an auxiliary amplifier for gain boosting to provide low input impedance and low power [9]. The preamplifier is followed by a leading edge discriminator, with a threshold that can be tuned on a per-channel basis with a Digital to Analog Converter (DAC).

B. Analog Memory

The analog memory is used to store the samples coming from the front-end. It includes 256 cells per channel. The block diagram of a single cell is shown in Fig. 2a. In contrast with the most common topology [10], each cell embeds a single-slope ADC formed by a comparator and a simple digital logic (not shown in the figure) which is used to store the Gray counter value. The Gray counter is common to all cells in a channel.

During the sampling phase, the capacitor is connected between the reference voltage value V_{ref_bottom} and the front-end amplifier output. After 5 ns, the sampling ends and the top plate is left floating. If a trigger is received, the cells are set up for digitisation. The single-slope digitisation can be implemented in several ways. The most common approach is to apply a linear ramp to the comparator input not connected to the sampling capacitor. The comparator flips when the ramp reaches the sampled value, thus triggering the storage of the Gray counter outputs into the local registers. The drawback is that the common mode at which the comparator fires depends on the stored signal. This can cause signal dependent delays from cell to cell, thereby affecting the overall system linearity. Another alternative is to keep the input of the comparator fixed at the reference voltage V_{BL} and to recharge the storage capacitor towards V_{BL} with a constant current source. The comparator therefore always fire at the
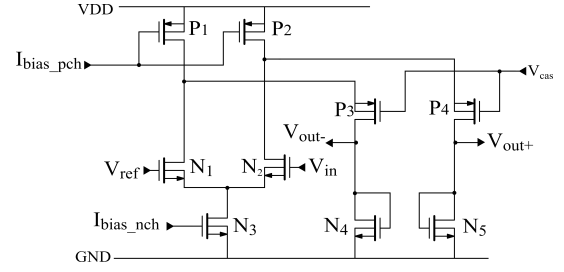

Fig. 3. Folded cascode preamplifier.

same common mode voltage regardless of the value of the stored signal. However, gain uniformity is an issue, because due to unavoidable mismatches between the current sources each cell will have its own gain. As a result, non-linearities can be introduced in the reconstructed signal. Therefore, a single ramp generator is used in each channel thus ensuring the same gain for all the 256 cells. The ramp generator is connected to the bottom plate of the sampling capacitors through the switch S_1. During digitisation, the top plate of the sampling capacitor is disconnected from the amplifier, therefore it follows the ramp as well because the charge in the capacitor must be conserved. In this way, all comparators flip when the top plate reach again the reference voltage V_{BL} thus avoiding possible errors due to common mode and gain variations between cells. The worst case consists in the simultaneous flips of all the cells. This situation is presented in Fig. 2. Here, the red line represents the transition time of the cells and the blue one the voltage ramp. It can be noticed that the transient settling time is acceptable since it is about 1-2 clock cycles. This can be easily managed introducing a delay between the switching of the bottom plate and the start of the ramp.

The ADC comparator is composed of two stages: a folded-cascode preamplifier and a dynamic latch. A simplified schematic of the preamplifier is shown in Fig. 3. This stage acts like a buffer between the sampling capacitor and the latch thus minimising kickback noise. The preamplifier can work in two modes, called power up and power down. Since the digitisation is enabled only when a trigger is acquired, the circuit does not need continuously the full bias current. Hence, transistors P_1, P_2 and N_3 have been split up so that they form two branches: the first one provides 1/4 of the bias current while the second one provides the remaining 3/4. Furthermore, a switch is added in series in each branch. Therefore, when the digitisation is off, the current is provided only through the first branch. This scheme allows for a swift transition between the power down and the power up modes, while saving a significant power during the sampling phase.

The schematic of the positive feedback latch is shown in Fig. 4. This architecture allows to minimise kickback noise towards the input stage [11]. When the comparator takes the decision and the two outputs assume different values, the XNOR output switches off N_5 and N_6 thus disconnecting the input from the latch. Moreover this stage consumes power only in the transition phases.
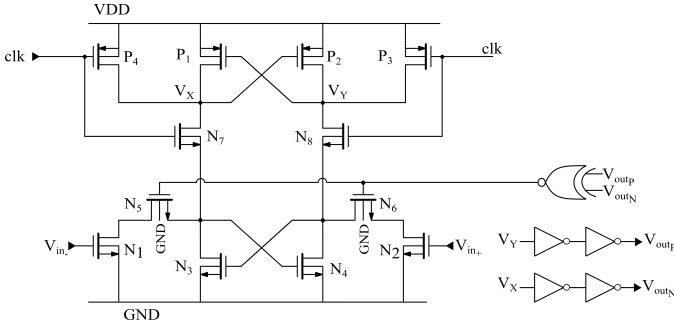
Fig. 4. Positive feedback latch.

The maximum conversion time is $2^N T_{clk}$, where N is the number of bits and T_{clk} is the clock period. Since $N = 12$ in the full resolution case and the clock frequency is equal to 200 MHz, the maximum conversion time is $T_{conv\,max} = 20.48\,\mu s$. By counting on both clock edges, the conversion time can be reduced to 10.24 μs.

The digital output word is composed of a 16-bit header, which includes 4 bits for alignment, 6 bits for the timestamp and 6 bits for the channel address, followed by the digitised data. Assuming that the maximum resolution has been chosen, a total of 3088 bits per channel must be sent off-chip in case all the 256 cells are used. To send data out, one can use more low-speed serializers in parallel or a faster serializer. However, even a 10 Gb/s serializer would require 20 μs to send off a full data set. The use of intelligence on chip could allow data selection and reduction. In applications where rare event detection is targeted, dead-time is not however the primary issue, whereas keeping as much information as possible for offline analysis is an advantage. Therefore it has been decided to implement a eight channel DDR serializer working with a master clock of 400 MHz, which can be easily provided to the chip. The transmission dead-time thus becomes 32 μs, which is fully adequate for our purposes.

C. Derandomization

Each channel embeds a digital control logic in which Finite State Machines (FSMs) have been implemented in order to manage the different operation modes. If a reset is sent, all the cells are set in idle state. Then, the cells are sampled in rolling shutter mode with a sampling period of 5 ns. If no trigger signal is received, the process is repeated thus overwriting the cells. Otherwise, a proper number of cycles are counted in order to place the event in the middle of the acquisition window. After that, the analog memory enters the warm-up stage which sets up the circuit for digitisation. Afterwards, the conversion phase is carried out by enabling the global Gray counter and powering up the converter of each cell. When the conversion is completed, the data are serialised and sent off-chip.

The cells in the memory are further partitioned into 32-cell units referred to as sections. Each one can acquire a signal independently of the others and an additional grouping can be selected to use segments of 32, 64 or 256 cells. This segmentation ensures a convenient flexibility to the ASIC depending on the application, namely, on the time-length of the event. This solution allows to reduce the probability of loosing an event because the channel is processing the previous one. The derandomization allows to accept a maximum event rate which is the inverse of the dead time with a negligible event loss. With 8 buffers per channel, the maximum event rate per channel can be from 40 kHZ (with the ADC set to 12 bit resolution) to 1.5 MHz (ADC set to 7 bit resolution).

In addition, the chip was designed to operate in two modes named imaging and sparse mode, respectively. The first configuration sets the channels to work in parallel considering the selected partitioning. If any channel detects a trigger, the global acquisition state is frozen and the cells are consecutively prepared for readout. In other words, the system allows to take an entire frame of the current event with the chosen segmentation. By contrast, the sparse mode enables a set of dedicated FSMs whose number is based on the selected configuration of segments. For instance, if 32-cell units are configured, the number of FSMs is equal to 8. In this way, the channels of the ASIC work independently of each others. When a block is ready for the readout, a module manages this task taking into account the whole requests of 8 channels. A pointer determines if the content of a channel/segment is enabled for the reading, then the output path for the stored data is reserved.

D. Trigger Configuration

The signal due to the interaction of a neutrino in atmosphere appears on the focal plane of the detector as a point-like flash lasting order of 100 ns. However, a similar signal is expected to be generated by low energy cosmic rays interacting directly with the sensors at much higher rates. The way to distinguish the rare neutrino candidate in this strong 'noise' is to adopt a bi-focal optical system which focuses the light on two corresponding pixels. The correlation between corresponding pixels can thus be exploited to reject direct cosmic rays which would still activate only one pixel. Moreover, this technique allows to reduce the impact of SiPM dark count rate as well, in a very effective way. The trigger configuration is thus organised in rows. Each pixel is connected to an AND gate together with its second neighbouring pixels. The outputs of the AND ports are joined into a OR gate whose output provides the trigger. The first and last pixels of each row can be connected to the appropriate pixels pair of the previous and the next ASICs respectively to avoid edge effects, as shown in Fig. 5. The trigger generated inside the ASIC can be used to trigger internally the readout (self-trigger mode) or it can be sent to an FPGA to combine trigger primitives provided by several ASICs to make a final decision. The external trigger mode can also be exploited to periodically acquire baseline data to monitor the background.
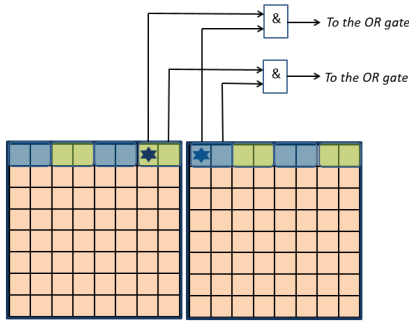
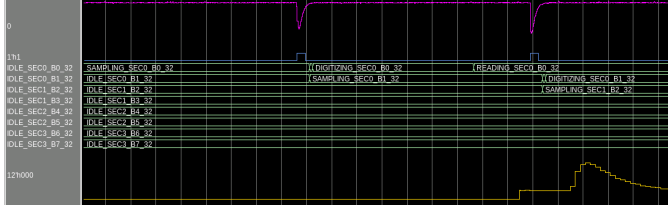Fig. 5. Trigger configuration between two matrices of SiPMs.

Fig. 6. Timing diagram of the channel with a 32-cells segmentation. The pink signal represents the input and the blue one the trigger. Then, the FSM is presented. Finally, the yellow waveform shows the output data.

Fig. 7. Input and decoded data. The straight line between two decoded pulses is not a physical signal but it is due to the graphical representation.

III. SIMULATIONS

Both the testbench and the Device Under Test (DUT) have been developed in SystemVerilog (SV). The DUT consists in the digital block in which the analog components are integrated in the Place and Route (P&R) stage in a Digital-on-top integration flow. For a first verification, the analogue components have also been modelled in SV to speed up the simulation time. The timing diagram of a single channel is shown in Fig. 6. The channel has been configured in 32-cell segmentation. The functioning of the FSM can be observed: the eight segments work independently from the others. The digital controller ensures that there is no overlapping between the states of the sections. Actually, the operational phases of each slice are carried out in a work queue. The decoded data are complementary to the inputs since the Gray counter starts counting from zero.

The graph shown in Fig. 7 has been obtained with the same segmentation as the previous one (32-cells) and an 8-bit resolution is chosen. In this simulation, an extreme situation is presented: in fact, the input pulse do not arrive with a Poisson distribution, but they are uniformly distributed in time, so when all the slices are busy, the incoming pulses are lost.

IV. CONCLUSION

A low-power 64-channel front-end ASIC prototype in a 65 nm CMOS technology for SiPMs readout has been presented. The incoming signal is amplified and stored into an analog memory. If a trigger signal is generated, the data are converted into a digital word by using a local 12-bit single-slope ADC and sen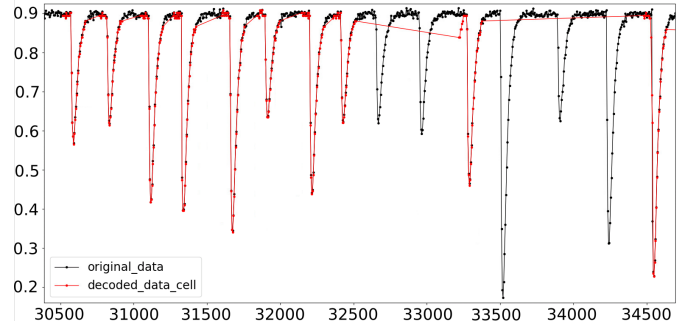t off-chip. The ASIC incorporates a total of 16384 single slope ADCs. The use of an analog memory makes the device suitable for burst-mode applications thus reducing the power consumption. Moreover, the flexibility of the chip has been increased with the analog memory segmentation and the possibility to configure the operation in imaging or sparse mode. The expected power consumption is 5 mW/channel and the final dimensions will be $6mm \times 4mm$. The design is now in the final layout and verification phase and it will be submitted for production in the second quarter of 2022.

REFERENCES

[1] Sanaei Behnoush, et al., "Characterization of a new silicon photomultiplier in comparison with a conventional photomultiplier tube", J. Mod. Phys, 2015, 6.4: 425-433.
[2] Di Francesco, A., et al., "TOFPET2: a high-performance ASIC for time and amplitude measurements of SiPM signals in time-of-flight applications." Journal of Instrumentation 11.03 (2016): C03042.
[3] Paolo Trigilio, Paolo Busca, Riccardo Quaglia, Michele Occhipinti, Carlo Fiorini, "A SiPM-readout ASIC for SPECT applications", IEEE Transactions on Radiation and Plasma Medical Sciences, 2(5), 404-410.
[4] Zhenxiong Yuan,et al., "KLauS: A Low-power SiPM Readout ASIC for Highly Granular Calorimeters", IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (pp. 1-4), 2019.
[5] Bagliesi, M. G., et al., "A custom front-end ASIC for the readout and timing of 64 SiPM photosensors." Nuclear Physics B-Proceedings Supplements 215.1 (2011): 344-348.
[6] James H Adams Jr, et al. "White paper on EUSO-SPB2." arXiv preprint arXiv:1703.04513 (2017).
[7] Valentina Scotti, Giuseppe Osteria, and JEM-EUSO Collaboration. "The EUSO-SPB2 mission." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 958 (2020): 162164.
[8] A. V. Olinto, et al. "The POEMMA (Probe Of Extreme Multi-Messenger Astrophysics) observatory." Journal of Cosmology and Astroparticle Physics 2021.06 (2021): 007.
[9] Paolo Carniti, Marcello De Matteis, Andrea Giachero, Claudio Gotti, Matteo Maino, Gianluigi Pessina, "CLARO-CMOS, a very low power ASIC for fast photon counting with pixellated photodetectors", Journal of Instrumentation, 2012.
[10] G. Anelli, F. Anghinolfi, and A. Rivetti, "A large dynamic range radiation-tolerant analog memory in a quarter-micron CMOS technology", IEEE Trans. Nucl. Sci., 48:435–439, 2001.
[11] Huang, Yan, Horst Schleifer, and Dirk Killat. "Design and analysis of novel dynamic latched comparator with reduced kickback noise for high-speed ADCs." 2013 European Conference on Circuit Theory and Design (ECCTD). IEEE, 2013.