A prototype of pixel readout ASIC in 65 nm CMOS technology for extreme hit rate detectors at HL-LHC

A prototype of pixel readout ASIC in 65 nm CMOS technology for extreme hit rate detectors at HL-LHC

A. Paternò,[b,c,1] L. Pacher,[a,b] E. Monteil,[a,b] F. Loddo,[d] N. Demaria,[b] L. Gaioni,[e,f] F. De Canio,[e,g] G. Traversi,[e,f] V. Re,[e,f] L. Ratti,[e,g] A. Rivetti,[b] M. Da Rocha Rolo,[b] G. Dellacasa,[b] G. Mazza,[b] C. Marzocca,[d,h] F. Licciulli,[d] F. Ciciriello,[d,h] S. Marconi,[i,j,o] P. Placidi,[i,j] G. Magazzù,[k] A. Stabile,[l] S. Mattiazzo,[m] and C. Veri[n]

[a]*Dipartimento di Fisica, Università di Torino, Via Pietro Giuria 1, Torino, Italy*

[b]*INFN Sezione di Torino, Via Pietro Giuria 1, Torino, Italy*

[c]*Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, Torino, Italy*

[d]*INFN Sezione di Bari, Via Orabona 4, Bari, Italy*

[e]*INFN Sezione di Pavia, Via Agostino Bassi 6, Pavia, Italy*

[f]*Dipartimento di Ingegneria e scienze applicate, Università di Bergamo,*
Viale Marconi 5, Dalmine (BG), Italy

[g]*Dipartimento di Ingegneria Industriale e dell'Informazione, Università di Pavia, Pavia, Italy*

[h]*Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Via Orabona 4, Bari, Italy*

[i]*INFN Sezione di Perugia, Via Alessandro Pascoli, Perugia, Italy*

[j]*Dipartimento di Ingegneria, Università di Perugia, Via G. Duranti 93, Perugia, Italy*

[k]*INFN Sezione di Pisa, Largo Bruno Pontecorvo 3, Pisa, Italy*

[l]*INFN Sezione di Milano, Via Celoria 16, Milano, Italy*

[m]*Dipartimento di Ingegneria dell'Informazione, Università di Padova, Via Gradenigo 6B, Padova, Italy*

[n]*Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Via per Monteroni, Lecce, Italy*

[o]*CERN, Switzerland*

E-mail: andrea.paterno@to.infn.it

---

[1]Corresponding author.

ABSTRACT: This paper describes a readout ASIC prototype designed by the CHIPIX65 project, part of RD53, for a pixel detector at HL-LHC. A 64×64 matrix of $50 \times 50 \mu m^2$ pixels is realised. A digital architecture has been developed, with particle efficiency above 99.5% at 3 GHz/cm^2 pixel rate, trigger frequency of 1 MHz and 12.5μsec latency. Two analog front end designs, one synchronous and one asynchronous, are implemented. Charge is measured with 5-bit precision, analog deadtime below 1%. The chip integrates for the first time many of the components developed by the collaboration in the past, including the Digital-to-Analog converters, Bandgap reference, Serializer, sLVS drivers, and analog Front Ends. Irradiation tests on these components proved their reliability up to 600 Mrad.

Contents

1 Motivation

The high luminosity of the Phase 2 Upgrade of the LHC represents a challenge for physicists and chip designers alike. Although the requirements of the detectors can only be estimated at this moment, most of the electronics, especially in the inner tracker, will certainly need to be upgraded in order to withstand the higher rates and radiation levels. The goal of the CHIPIX65 collaboration is to realize an ASIC prototype for pixel detectors in 65 nm CMOS technology. The CHIPIX65 collaboration [1] is also part of the RD53 collaboration at CERN; as such, the CHIPIX65 demonstrator was designed as an intermediate step before the full-scale RD53 chip. Both designs, however, share the requirements, which are:

- 40 MHz bunch crossing frequency

- small pixels ($50\,\mu\mathrm{m}^2 \times 50\,\mu\mathrm{m}^2$), large chips ($2\,\mathrm{cm} \times 2\,\mathrm{cm}$)

- 200 event Pile-Up $\rightarrow$ 3 GHz/cm^2 hit rate $\rightarrow$ 75 kHz/pixel particle rate

- 1 MHz trigger rate, $12.5\,\mu\mathrm{s}$ trigger latency

- Low power consumption

Characteristics. The chip contains an active matrix of 64×64 pixels, featuring two different front-end designs. The pixels are grouped into a novel region-based digital architecture which manages 16 pixels at a time. The chip have been assembled with a Digital-on-Top, top-down hierarchical flow. Many of the silicon-proved IP blocks developed in the RD53/CHIPIX65 framework have been integrated in the design.

2 The Design

The design can be divided into an active area, and a chip periphery. The active area contains the pixel matrix, which is structured as a digital sea, where the hit management logic lies, with so-called analog islands, containing the Analog Front Ends. These perform analog processing of the signal coming from the sensor, transforming it into a hit discrimination binary information to be passed to the digital circuitry. The pixel matrix contains 4096 pixels, divided into 64 rows and 64 columns. Half of the columns feature a synchronous Analog Front End developed by INFN Torino, while the other half an asynchronous Analog Front End developed by INFN Bergamo/Pavia. [1]

The front ends are grouped in 4×4 blocks called Pixel Regions, which also contain a shared digital architecture which stores the data from the pixels and performs trigger matching. A column of 16 Pixel Regions is called a Macro Column. As there are 64 pixels in a row, there are 16 Macro Columns in the pixel matrix. In a Macro Column, every Pixel Region is connected to the following one, passing by configuration bits and receiving the other's output data.

In the Chip Periphery, there is a dedicated readout stage for each Macro Column, called Macro Column Drainer. These modules are connected to a shared logic, which handles the Pixel Regions configuration and buffers all the output information, feeding it to the serializer. The Chip Periphery also contains the bias cells, the monitoring circuitry, and the output pad drivers.

Figure 1 shows the assembled design, with a close up of the Pixel Matrix.
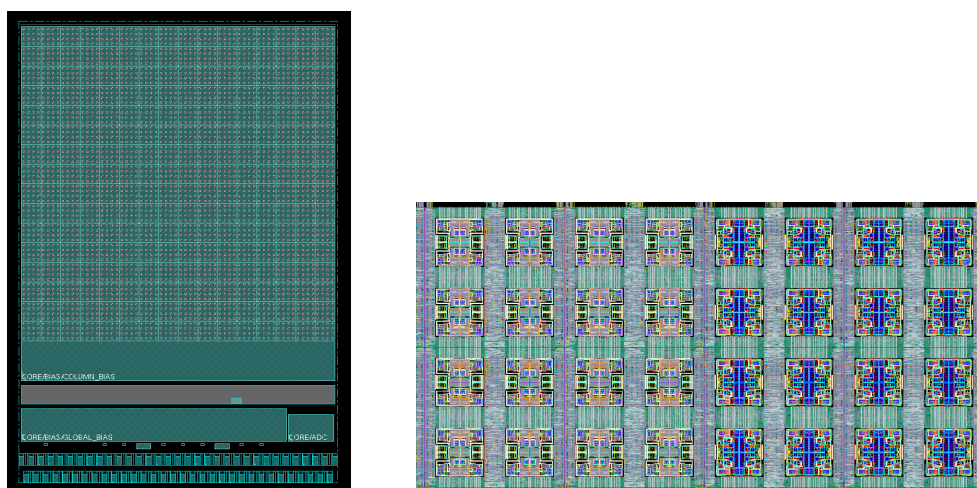
(a) The whole chip. From the top, the Pixel Matrix, Column Biases, Readout circuitry, Global Bias and the padframe.

(b) Close-up of the Pixel Matrix, where the two Front Ends can be distinguished.

Figure 1. Views of the assembled design.

2.1 The Bias Network

The Bias Network consists of 16 current Digital-to-Analog Converters (DACs) and a set of Column Bias cells which distributes these currents to every column of pixels. It generates and distributes the bias voltages and currents needed by the frontends for their operation. The 16 DACs are used to

generate 9 biases needed for the Torino FE, 6 for the Bergamo/Pavia FE, and one for the calibration circuit; for these reasons, they are also called Global Bias cells. These are then mirrored for every column of pixel by the Column Bias cells.

The Global Bias cells are 10-bit segmented current-steering DACs: each is composed of a thermometric DAC (for the 8MSBs) and a binary DAC (for the 2 LSBs). This solution represents a tradeoff between area and power consumption, and performance in terms of differential nonlinearity. In order to properly deal with the high radiation environment, no minimum-size transistors were used. Instead, a set of custom digital cells has been designed. This IP block has been tested using both X-rays and protons up to 1 Grad. It worked under irradiation with acceptable increase in DNL and INL [6].

In order to work, the Global Bias cells need a reference current. This is provided by a BandGap Reference circuit (BGR). Unlike conventional BGRs, based on the voltage drop across a p-n junction and PTAT currents, and shown to be sensitive to TID and TDD, the adopted solution is based entirely on MOSFETs: it is capable of working after having received a high dose of radiaton, and has been tested with X-rays up to 580 Mrad. It provides a PVT-independent DC voltage, trimmable using a dedicated 5-bit DAC, to compensate for process variations and mismatches [7].

These bias voltages can be monitored with a 12-bit ADC that has been integrated on-chip. Adopting a dual-slope integrating architecture, the ADC is capable of achieving high accuracy, compromising on the conversion rate, which is about 5 kSample/s. A linear transconductor converts the input voltage, in the dynamic range from 0 to 900 mV, into a current, which is integrated onto a 70 pF Metal-on-Metal capacitor for 212 clock cycles. The output code is obtained by counting the clock cycles needed to discharge the integration capacitor at a constant current. An automated calibration procedure has been also implemented to finely adjust the gain and the discharge current.

2.2 Pixel Analog Front Ends

The demonstrator features two Front End designs: a synchronous one and an asynchronous one. [2, 3, 8] One half of the pixel matrix embeds the synchronous design, while the other one the asynchronous design. For both architectures two small prototypes have been submitted and tested, and their functionality tested also after irradiation. Both architectures feature a single stage preamplifier to minimize the power dissipation and a Krummenacher feedback to provide both leakage current compensation and constant current feedback capacitor discharge, along with a calibration circuit to inject an input charge. A schematic for both Front End flavors is shown in figure 3.

The synchronous architecture uses a telescopic cascode preamplifier, AC coupled to a synchronous discriminator composed of a Differential Amplifier and a positive feedback latch. The input stage is a Charge Sensitive Amplifier. It is implemented as a single-ended, high open-loop gain inverting amplifier with capacitive feedback. Its output is directly fed to the discriminator: no additional signal shaper has been included. This discriminator represents the most innovative component in the design: implemented as a discrete-time voltage comparator, it uses a low-gain differential amplifier with a fast regenerative latch stage. An external clock signal is required to periodically enable/reset positive feedback in the latch, thus introducing synchronous operations. This latch can be turned into a local oscillator with a frequency up to 800 MHz by an asynchronous feedback loop. Together with a stronger discharge current, this feature can be exploited to enable fast TOT computation. Measurements show very promising results, in particular: the preamplifier
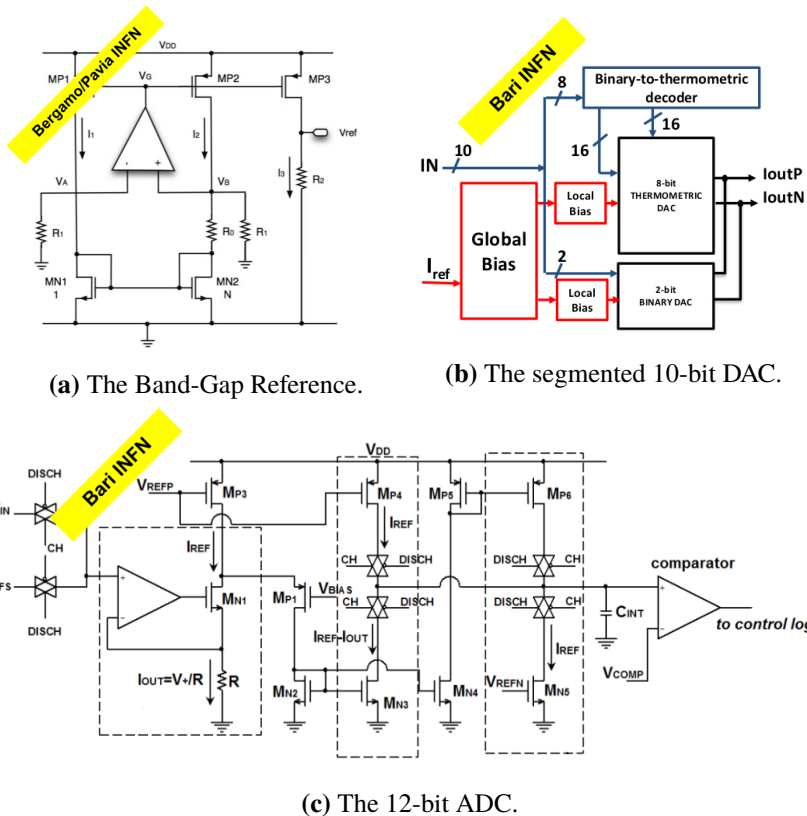
(a) The Band-Gap Reference.

(b) The segmented 10-bit DAC.

(c) The 12-bit ADC.

Figure 2. Block diagrams of the Bandgap Reference, DAC and ADC embedded in the design.

(a) The synchronous Front End.

(b) The asynchronous Front End.
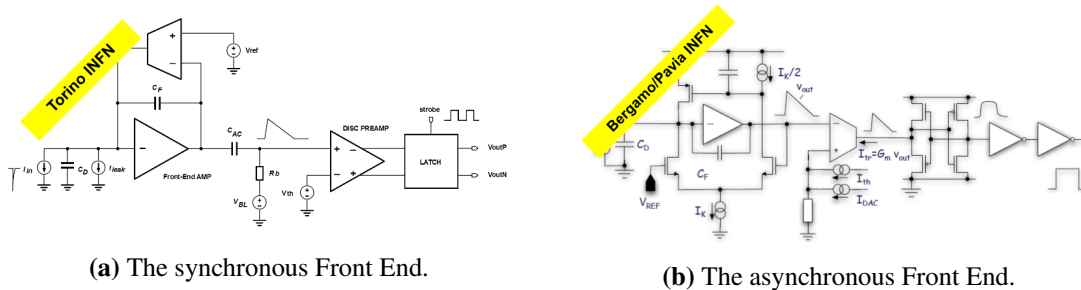
Figure 3. Functional schemes for the 2 Front Ends.

gain shows a very good uniformity with a 2.2% RMS dispersion, the Equivalent Noise Charge (ENC) is equal to $80\,e^-$ at C_{input} $50\,fF$, the offset compensation mechanism and the local oscillation both work consistently with simulation results. Irradiation measurements show that the front-end is still fully working at $600\,Mrad$, with negligible degradation of the analog parameters.

The asynchronous architecture, instead, features a folded cascode preamplifier, followed by a fast comparator composed of a transconductor stage and a transimpedance amplifier. The fast comparator receives the signal from the preamplifier and turns it into a Time over Threshold. The threshold discriminator architecture is based on a low power transimpedance amplifier for fast switching operation. Given the triangular shape of the preamplifier response, featuring a very fast

leading edge and a constant slope return to baseline, a linear relationship between amplitude (or input charge) and ToT is expected. The threshold dispersion issue is addressed by means of a local, in-pixel circuit for threshold adjustment, based on a 4-bit current steering DAC. Measurements show that the noise response is fully compliant with the specifications, and that leakage currents up to 15 nA do not affect the preamplifier performance. The architecture was designed to comply with a maximum input signal of 30000 electrons and features an output dynamic range around 450 mV, a charge sensitivity of about 90 mV/fC and an ENC of 114 electrons RMS for a sensor capacitance $CD = 100$ fF. A time walk not exceeding 25 ns is achieved in circuit simulations with a threshold of 700 electrons and signals 1000 electron in amplitude or larger. TID irradiation has been performed up to 800 Mrad, showing no significant degradation in the preamplifier signal shape and a 20% increase in noise with a $Cdet = 50$ fF at 800 Mrad. [4]

2.3 Digital Architecture

The pixels are grouped together to form 4×4 submatrices: these are called Pixel Regions. The Pixel Regions contain 16 analog Front Ends, themselves grouped in 4 analog islands in the layout, and a shared digital architecture which stores the pixels' data, handles the configuration, performs trigger matching, and communicates with the readout block at the chip periphery.

Focusing on the goal of having a very low event loss, the shared logic was devised so as to have a very large shared buffer. Theoretical predictions show, in fact, that a 4×4 Pixel Region would need 16 buffer rows in order to achieve an event loss lower than 0.1%, given the design hit rate and trigger latency. Further studies also showed that events, both in the center of barrel and the edge, rarely would have more than 5 pixels hit in a region. TOT compression could then be performed in order to optimize the area and power. These estimations have been verified with realistic simulations, performed with the RD53 Verification Environment: VEPIX53. [5]

(a) Functional scheme of the architecture.

(b) Placed view of the assembled design, showing the 4 analog islands.

Figure 4. The Pixel Region contains and manages 16 pixels.
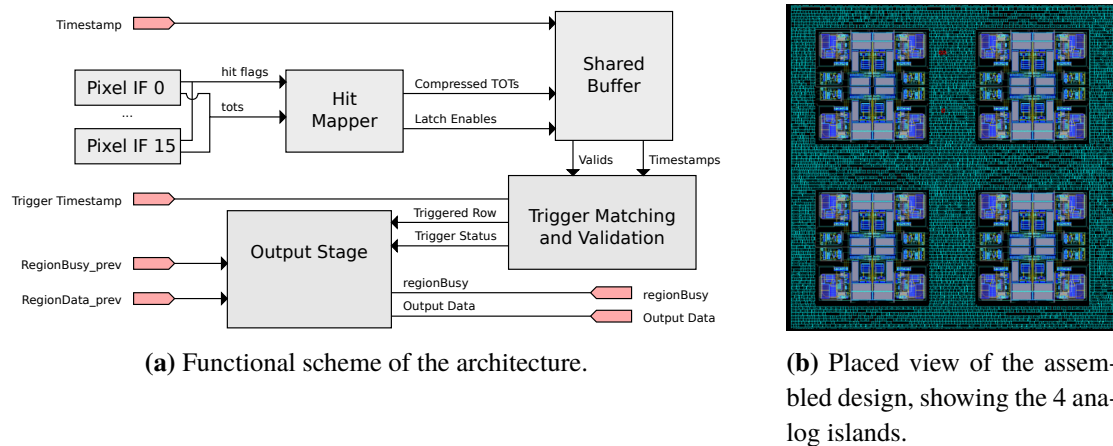
In order for the compression to be implemented, every pixel must be synchronized with each other: pixels hit together must be processed at the same time, regardless of their TOTs. This synchronization is ensured by using deadtime counters. They enforce a fixed deadtime on the pixels, during which they are processing the hits and are, thus, incapable of serving new ones. It

should be noted that this freezing is per-pixel, and not region-wide. Once the deadtime has elapsed, dedicated logic selects the first 6 pixels according to a priority queue. Their TOTs, along with a hit map and timestamp information, is then saved into the shared buffer. The hit map is the binary output for the event, and is used to rebuild the event off-chip by reversing the priority queue.

Once the data has been saved, the shared logic resets the pixels, which are then ready to process new hits. This mechanism allows to save area, while maintaining a very low event inefficiency ($\sim 0.1\%$, due to shared buffer overflow) and low charge information loss ($\sim 0.6\%$, due to the limited number of TOTs saved). Details on the inefficiencies are shown in figure 5.
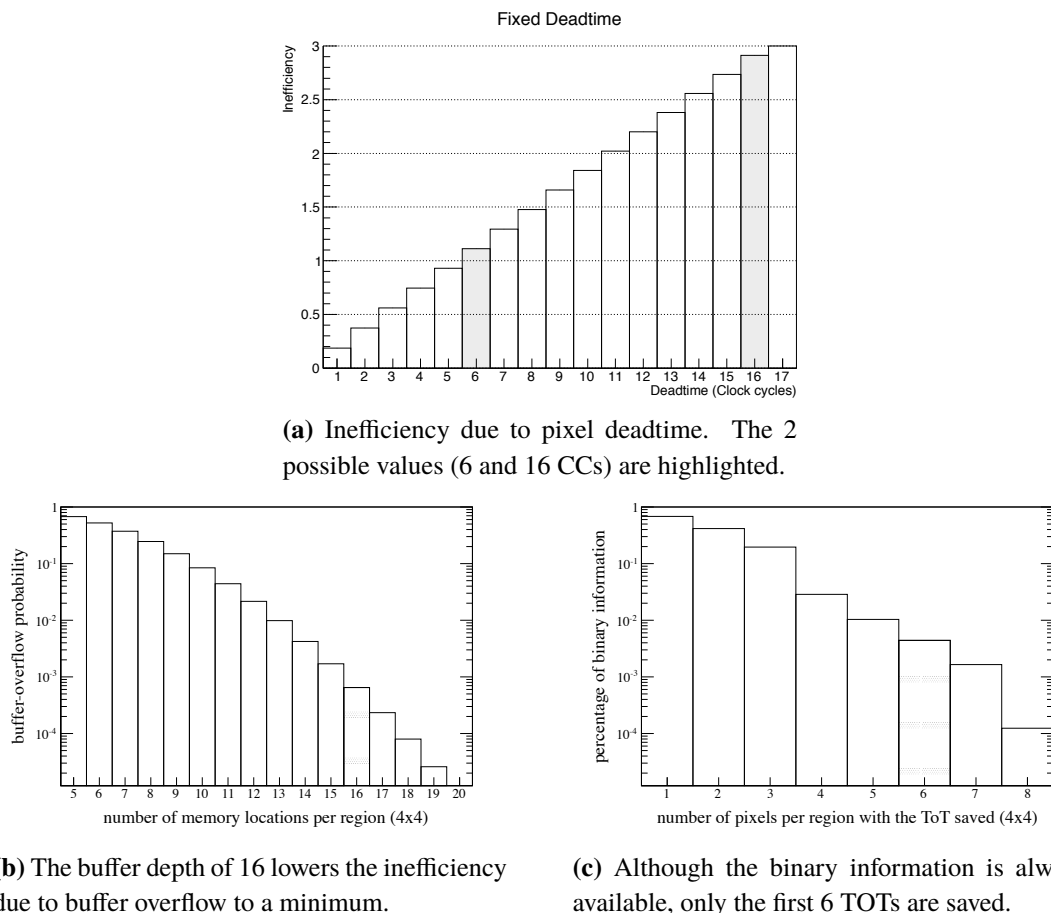

(a) Inefficiency due to pixel deadtime. The 2 possible values (6 and 16 CCs) are highlighted.

(b) The buffer depth of 16 lowers the inefficiency due to buffer overflow to a minimum.

(c) Although the binary information is always available, only the first 6 TOTs are saved.

Figure 5. Inefficiency sources in the design, expressed in percentage.

Many operation modes are supported for the Pixel Region. The events stored in the buffer can be read out in a triggered or triggerless fashion: if the triggered mode is enabled, a dedicated set of comparators is activated and stands by waiting for a trigger signal. It is driven by the Chip Periphery, and is propagated to the Pixel Regions along with a Trigger Timestamp. When the trigger signal is received by a Pixel Region, the comparators will look in the shared buffer for a timestamp matching the Trigger Timestamp, and mark that line for output. If any line is marked, a busy flag is raised and propagated to the end of column circuitry. The busy flag is passed from a Pixel Region to the next via a chain of OR gates: the triggered data from a Pixel Region is propagated only if the Pixel Region itself has the busy flag set, but all the Pixel Regions preceding it do not. If, instead, the busy

flag coming from the preceding Pixel Region is set, then a multiplexer propagates the incoming data from the preceding Pixel Region. This achieves a full implementation of a Column Drain readout architecture.

In triggerless mode, instead, the events are output as soon as they are saved into the shared buffer. Other operation modes include a binary only mode, in which only the binary information is recorded and the deadtime is minimal; and a debug mode, where the Front Ends are bypassed and the inputs are injected directly into the digital architecture.

2.4 Readout

For the purpose of this demonstrator, it was decided to employ a simple readout architecture. Communication with each macro-column is performed via ad-hoc modules called Macro Column Drainers. These are connected to another component, called Dispatcher, which, in turn, feeds its outputs to the output serializer.

The Macro Column Drainers (MCDs) handle communication to and from the Pixel Regions, handling entire macro columns. This implies that the readout of a column is independent of the others. In triggered mode, the MCDs will buffer the incoming trigger signals along with the corresponding trigger timestamps. A Finite State Machine (FSM) continuously polls the trigger buffer, selects the first one, and propagates it to the Pixel Regions. It would then check the busy flag coming from the Pixel Regions. If the triggerless mode is active, instead, the FSM will be locked in this listening state.

Readout from the Pixel Regions is instantaneous: if any Pixel Region has triggered data, their busy flag would rise after a Clock Cycle and asynchronously propagated to the MCD. There, a data buffer would store all the outputs from the Pixel Regions, for as long as the busy flag is high.

The Dispatcher collects all the outputs from the Macro Column Drainer data buffer, and stores them, in turn, in an internal buffer. This is the final data which is sent to the serializer: another Finite State Machine reads the data from the internal buffer, splits it into 16-bit chunks, performs 8b10b encoding (or simple filling) on these words, and sends them to the 20-bit serializer. A special Start of Packet word precedes the data, and is used by the output reconstruction algorithm off-chip. If no data is available in the shared buffer, then a special IDLE character is sent to the serializer. An output packet is composed of several fields:

Bits	Field
63:60	Macro Column Address
59:50	Timestamp
49:46	Pixel Region Address
45:30	Hit Map
29: 0	6 TOTs

2.5 Configuration

The configuration in the chip is SPI-based. The chip acts as a SPI slave, while an off-chip master sends 20-bit serialized data packets. There are four main types of configuration registers: the Global Configuration Registers (GCRs), the End of Column Configuration Registers (ECCRs), the Pixel Configuration Registers (PCRs) and a some direct-programming registers. The GCRs are a batch of

224 registers, configuring the autozeroing cycle for the Torino FE, the 15 10-bit Global Bias DACs, and the ADC reference currents and multiplexer selection. The ECCRs, instead, are used to set various readout options: the trigger latency (in clock cycles), the triggerless mode, the deadtime (low or high modes), the binary only mode, the 8b10b encoding bypass, the gray encoding bypass and the debug mode. There is also the possibility to mask any of the 16 macro columns from the readout.

The PCRs are distributed in the Pixel Matrix. Each Pixel Region has 8 PCRs, grouped in two modules: 4 register batches for the left half, and 4 for the right one. Each register batch, in turn, contains 16 bits: which configure two adjacent pixels. When a pixel pair has to be configured, the address received from the chip is translated by the configuration module at the Chip Periphery into a Pixel Region address, a Macro Column address, a Left/Right flag, and, finally, a PCR index. In order to speed up the chip configuration, a special auto-increment mode was devised, where needs to be submitted, but only the PCR data.

2.6 Padframe and SLVS drivers

Although the LVDS standard was originally considered for the chip I/O, the harsh radiation environment forced the design to be based on thin gate oxide transistors and a voltage supply of 1.2 V, incompatible with the 1.25V of common mode voltage of the LVDS standard. It was therefore decided to adopt the SLVS standard, which provides link bandwidth up to 1.2 Gbps with 200 mV common-mode voltage and 200 mV differential swing. The driver is composed of four MOS switches in a bridge configuration with two current generators. One of these is controlled by the common-mode feedback circuit which controls the output common-mode voltage. Layout of the I/O pads and SLVS drivers is shown in figure 6.
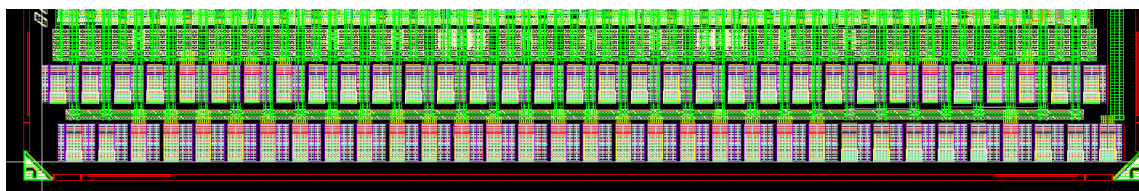

Figure 6. The SLVS pads and their drivers.

3 Conclusions

This demonstrator, measuring $3.5 \times 5.1 \, \text{mm}^2$, and containing a matrix of 64×64 pixels, has been submitted to the foundry in July 2016. Most of the building blocks, developed in the context of the collaboration, have already been tested under radiation, showing promising results in the 500 Mrad range. This chip represents the first attempt to integrate this many of such blocks into a working hybrid pixel detector readout ASIC. In addition, the chip contains a novel digital architecture featuring a digital inefficiency $< 0.1\%$ at the HL-LHC rate ($3 \, \text{GHz/cm}^2$) and providing 5-bit ToT information for 99.6% of hits. For the remaining hits, binary information is provided. The chips were received in late September 2016, and initial testing reveals that all the IP Blocks, Front Ends, and digital circuitry work as expected.

References

[1] N. Demaria et al., *CHIPIX65: Developments on a new generation pixel readout ASIC in CMOS 65 nm for HEP experiments*, in proceedings of the *6th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*, Gallipoli, 2015, pp. 49–54.

[2] E. Monteil, N. Demaria, L. Pacher, A. Rivetti, M. Da Rocha Rolo, F. Rotondo et al., *Pixel front-end with synchronous discriminator and fast charge measurement for the upgrades of HL-LHC experiments*, 2016 *JINST* **11** C03013.

[3] L. Pacher, E. Monteil, A. Rivetti, N. Demaria and M. Da Rocha Rolo, *A low-power low-noise synchronous pixel front-end chain in 65 nm CMOS technology with local fast ToT encoding and autozeroing for extreme rate and radiation at HL-LHC*, *IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, 2015.

[4] L. Gaioni, F. De Canio, M. Manghisoni, L. Ratti, V. Re and G. Traversi, *65 nm CMOS analog front-end for pixel detectors at the HL-LHC*, 2016 *JINST* **11** C02049.

[5] S. Marconi, E. Conti, P. Placidi, J. Christiansen and T. Hemperek, *The RD53 Collaboration's SystemVerilog-UVM Simulation Framework and its General Applicability to Design of Advanced Pixel Readout Chips*, 2014 *JINST* **9** P10005 [`arXiv:1408.3232`].

[6] G.D. Robertis, F. Loddo, S. Mattiazzo, L. Pacher, D. Pantano and C. Tamma, *Design of a 10-bit segmented current-steering digital-to-analog converter in CMOS 65 nm technology for the bias of new generation readout chips in high radiation environment*, 2016 *JINST* **11** C01027.

[7] G. Traversi, F. De Canio, L. Gaioni, M. Manghisoni, S. Mattiazzo, L. Ratti et al., *Characterization of bandgap reference circuits designed for high energy physics applications*, *Nucl. Instrum. Meth.* **A 824** (2016) 371.

[8] E. Monteil, N. Demaria, A. Paterno, A. Rivetti, C. Leng, F. Rotondo et al., *A synchronous analog very front-end in 65 nm CMOS with local fast ToT encoding for pixel detectors at HL-LHC*, in proceedings of *Topical Workshop on Electronics for Particle Physics*, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, 26–30 September 2016.