

# Design and simulation of a 65 nm Macro-Pixel Readout ASIC (MPA) for the Pixel-Strip (PS) Module of the CMS Outer Tracker Detector at the HL-LHC

---

**Alessandro Caratelli** <sup>†</sup>

*Microelectronic System Laboratory (LSM), École polytechnique fédérale de Lausanne (EPFL), CERN, Geneva, Switzerland.*

*E-mail:* [Alessandro.Caratelli@epfl.ch](mailto:Alessandro.Caratelli@epfl.ch)

**Davide Ceresa**\* <sup>†</sup>

*CERN, Geneva, Switzerland.*

*E-mail:* [Davide.Ceresa@cern.ch](mailto:Davide.Ceresa@cern.ch)

**Jan Kaplon**

*CERN, Geneva, Switzerland.*

*E-mail:* [Jan.Kaplon@cern.ch](mailto:Jan.Kaplon@cern.ch)

**Kostas Kloukinas**

*CERN, Geneva, Switzerland.*

*E-mail:* [Kostas.Kloukinas@cern.ch](mailto:Kostas.Kloukinas@cern.ch)

**Jan Murdzek**

*CERN, Geneva, Switzerland.*

*E-mail:* [Jan.Murdzek@cern.ch](mailto:Jan.Murdzek@cern.ch)

**Simone Scarfi**

*Microelectronic System Laboratory (LSM), École polytechnique fédérale de Lausanne (EPFL), CERN, Geneva, Switzerland.*

*E-mail:* [Simone.Scarfi@epfl.ch](mailto:Simone.Scarfi@epfl.ch)

The MPA is the pixel readout ASIC for the hybrid Pixel-Strip module of the Phase-II CMS Outer Tracker upgrade at the High Luminosity LHC (HL-LHC). It employs a novel technique for identifying high transverse momentum particles and provides this information at 40 MHz rate to the L1-trigger system. The chip also comprises a binary pipeline buffer for the L1-trigger latency, and a data path to support the readout of full events with a maximum trigger rate of 1 MHz and a latency of 12.8  $\mu$ s. The design and implementation in a 65 nm CMOS technology of the first prototype ASIC that integrates all functionalities for system level operation with a power density lower than 90 mW/cm<sup>2</sup> are presented in this contribution.

*Topical Workshop on Electronics for Particle Physics*

*11 - 14 September 2017*

*Santa Cruz, California*

---

\*Speaker.

<sup>†</sup>Main authors

## 1. Introduction

The objective of the CMS Outer Tracker upgrade for the High Luminosity LHC is to adopt the use of double layer sensors to facilitate the quick, on-detector, identification of high-pT tracks ( $>2\text{ GeV}/c$ ) and their transmission to the L1 trigger system at the  $40\text{ MHz}$  Bunch Crossing (BX) rate. For the first time data coming from the Tracker will be used in the L1 trigger decision of a high luminosity hadron experiment. In parallel, a readout channel will transmit triggered events to the Data AcQuisition (DAQ) system at a nominal average trigger rate of  $750\text{ kHz}$ . A complete description of the CMS Outer Tracker upgrade can be found in the CMS Technical Proposal [1].

The Macro Pixel ASIC is the readout chip for the Pixel-Strip module which extracts hits (binary signals) from the pixelated sensor. The chip comprises a fast front-end with leakage current compensation and amplitude discriminator circuits with binary readout for a  $120 \times 16$  pixel sensor array with dimensions of  $12.0\text{ mm} \times 23.16\text{ mm}$ . The readout of 1920 channels per chip at a  $40\text{ MHz}$  BX rate provides a data throughput of  $\sim 80\text{ Gbps}$  per chip. Such an amount of data is combined in real time with the input data ( $2.56\text{ Gbps}$ ) coming from the strip sensor layer. The latter is read out by another readout ASIC, called Short Strip ASIC (SSA). The on-chip data processing combines zero-suppression techniques with the capability of recognizing particles with high transverse momentum. The system achieves an almost lossless data transmission to the back-end with an output bandwidth of  $1.6\text{ Gbps}$  per chip. In parallel, every event is stored for a maximum latency of  $12.8\mu\text{s}$  and it can be acquired with a trigger signal. The chip supports a maximum trigger rate of  $1\text{ MHz}$  and provides the encoded position and dimension of pixel and strip clusters. A complete description of the analog front-end, as well as the details about the data processing logic can be found in [2]. The results, mainly about the front-end performance, obtained with the first prototype are summarized in [3]. This paper focuses on the implementation of the first full size  $65\text{ nm}$  MPA including all functionalities required by the CMS Tracker.

## 2. Design Implementation

The strict requirement of a total power density lower than  $90\text{ mW}/\text{cm}^2$  has driven the choice of a  $65\text{ nm}$  CMOS technology featuring 7-metal with aluminum ReDistribution Layer (RDL). This technology provides 5 thin metals which are used for signal routing, while the remaining 3 thicker metals are used for power routing. Voltage drop across the die is of particular importance because the power can be supplied only from the bottom of the chip since the other edges are covered by the sensor.

The very low power requirement makes also necessary to adopt several low power design techniques. The MPA exploits a Multi-Supply Voltage (MSV) technique to strongly reduce the digital power consumption without degrading the analog front-end and data transmission performance. Therefore, the digital core is powered at  $1\text{ V}$ , while the Analog Front-End and the custom-sLVS [4] drivers and receivers are powered at  $1.25\text{ V}$ . The sLVS differential interface utilizes a programmable current to optimize the power consumption while maintaining good signal integrity. The digital core is implemented using standard cell libraries of different threshold voltage devices (Multi- $V_T$  design) to locally improve performance or reduce power consumption. In particular, the design exploits low- $V_T$  devices for the serializer and deserializer logic running at  $320\text{ MHz}$ , while

the rest of the design uses normal- $V_T$  devices. Three of the main challenges presented by the MPA design are the optimizations of the clock distribution, the reduction of the memory power consumption and the radiation hardening. The solutions adopted in the MPA are presented in details in the next paragraphs.

## 2.1 Clock Distribution

The 40 MHz sampling clock tree is based on a clock trunk for the distribution to the pixel rows, shown in Figure 1B, which allows to achieve the very low skew ( $\ll 1\text{ ns}$ ) [5] required for the digitalization of the analog output. The clock trunk is routed manually with a thick metal in order to minimize the resistance. Consequently the skew among the pixel rows is reduced, but the larger line capacitance requires a multi-cell buffer placed in the periphery to be driven. The clock distribution inside the row is done with the automatic clock tree synthesis to achieve the same delay for all the pixels. This clock tree starts from a manually placed row input buffer.

Minimizing the skew has an important drawback: the entire pixel matrix switches within a very short period ( $\ll 1\text{ ns}$ ) causing high spike of current and consequently a large IR-drop. In order to avoid this effect, the sampling clock is used only for the first sampling flip-flop after the analog discriminator, while the remaining logic exploits a second clock distribution. The latter is based on a re-buffering stage per pixel row which increases the skew ( $> 1\text{ ns}$ ) and reduces IR-drop at the clock transition to less than  $20\text{ mV}$  (see figure 1A).

## 2.2 Memory Gating

The MPA is required to store the entire pixel matrix hits for a maximum latency of  $12.8\ \mu\text{s}$  with a BX period of  $25\text{ ns}$ , which corresponds roughly to  $1\text{ Mb}$  of data. This data is readout when an external trigger is received and a lossless communication is required. The maximum trigger rate is  $1\text{ MHz}$ , which makes more convenient to store data without compression. Indeed, any data compression step before the memory needs to operate at  $40\text{ MHz}$ . Given the occupancy and the lossless requirements, every technique investigated exceeded the gain in power consumption achieved by the reduced memory size. Our approach was to store uncompressed data locally on every pixel row thus limiting the data transport across the large die only for the read operations of hits. A custom made radiation hardened SRAM block [6] of  $128\text{ bits} \times 512\text{ words}$  is employed per pixel row.

In addition, a memory gating technique, shown in Figure 1C, has been developed to limit the power consumption exploiting the low hit occupancy in the Outer Tracker. Since the memory is read with a fixed latency, the SRAM is managed as a circular memory: every location corresponds to a determined cycle. When the memory pointer reaches the last location, it restarts from the first location, overwriting the previous data. Consequently, if an event is not required within the defined latency, it is deleted. Considering an average occupancy of 1-2 cluster per bunch crossing, only two memories are written every cycle. By using an OR-tree, if a hit is detected a write operation is executed, otherwise the clock to the memory is gated. Doing so, the power consumption from the memory is reduced by  $\sim 8$  times, but the SRAM is not refreshed every cycle. Consequently a tag is saved with the data in order to identify if the data which are being read have been updated. The tag is provided by an 8-bit counter, called Latency counter, which increments every time the circular

memory completes a full cycle. When the memory is read, the counter tag is compared with the current Latency counter, and the data are processed only if the counter tag value comes from the previous latency cycle. This method requires a refresh of the memory before the Latency counter overflows, i.e. around every 3 ms, which is obtained by disabling the OR logic. The final power consumption for the memory with the additional gating logic is  $\sim 5$  times lower than the power consumption without any gating technique.

### 2.3 Radiation Hardening

Radiation effects are one of the main challenges in the MPA design. The maximum Total Ionizing Dose (TID) foreseen in the CMS Outer Tracker is 100 MRad. As demonstrated in [7], the degradation induced by high TID in 65 nm MOS transistors is strongly gate-length dependent. For this reason, minimum length transistors are not used in the analog design. On the other hand, the digital design is based on standard cell libraries from the foundry with which the designer cannot size freely the transistor dimension. Based on the results from [8] and considering the low operating temperature of the tracker (cooling system should work between  $-30^{\circ}\text{C}$  and  $-40^{\circ}\text{C}$ ), we have implemented the digital logic utilizing a 9-tracks standard cell library from the foundry. Simulations employing radiation device models predict a frequency degradation around 20 % at 200 MRad and  $-20^{\circ}\text{C}$ .

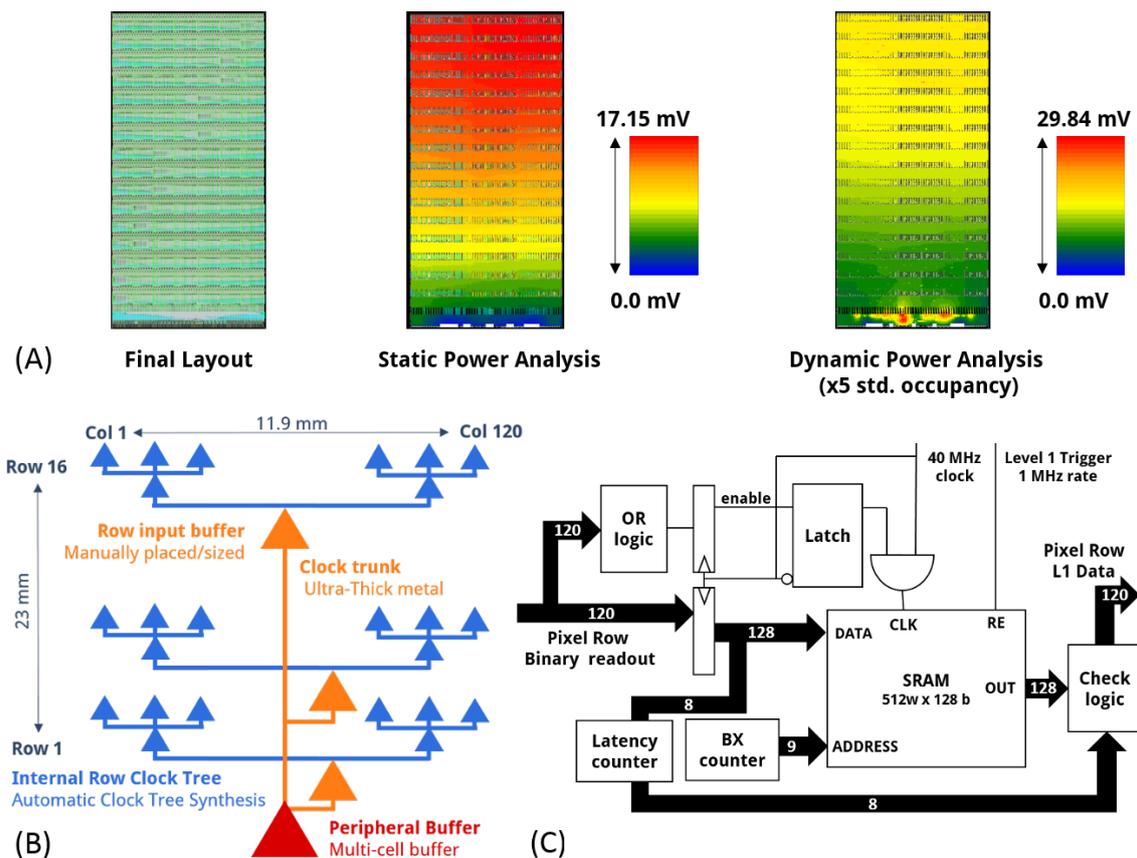
Concerning Single Event Effects, the MPA implements a partial Triple Module Redundancy (TMR) methodology. The control logic, including the system clock (but not the sampling clock), and the configuration logic have been triplicated, while a full TMR has been excluded due to the restricted power budget. The large number of configuration bits is an important contribution to the power consumption, which can be mitigated by clock gating: the registers are clocked only when addressed and, in order to keep the Single Event Upset immunity, an error detection and correction logic is added. This additional logic sends a clock pulse to the registers when it detects a discrepancy among the output values of the registers. The voted output is then written in all the registers, correcting the bit upset.

### 3. Conclusions

A module-level verification environment, described in [9], allowed extensive simulation of the MPA with back-annotated delays extracted from the final layout (in Figure 1A). Moreover, activity-based power verification provided the IR-drop maps shown in Figure 1A: static and dynamic power analysis provide a detailed study of the voltage drop for the chip. In conclusion, full chip simulation and power verification run with EDA tools from Cadence (Innovus - Digital IC design, Voltus IC Power Integrity Solution for the digital domain and Virtuoso Analog Design Environment for the analog domain) show the achievement of the expected performance with a total power density lower than  $90\text{ mW}/\text{cm}^2$ .

### Acknowledgements

We thank Prof. Datao Gong, Ye, Jingbo, Yang Dongxum and Jian Wang from SMU for the design of the Phase Aligner used for the sampling clock de-skewing.



**Figure 1:** A) Layout and IR drop analysis; B) Sampling Clock Distribution; C) Memory gating technique.

## References

- [1] CMS collaboration, *Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid*, CERN-LHCC-2015-010 / LHCC-P-008.
- [2] D. Ceresa et al., *Macro Pixel ASIC (MPA): The readout ASIC for the pixel-strip (PS) module of the CMS outer tracker at HL-LHC*, 2014 JINST 9 C11012
- [3] D.Ceresa et Al., *A 65 nm pixel readout ASIC with quick transverse momentum discrimination capabilities for the CMS Tracker at HL-LHC*, 2016 JINST 11 C01054.
- [4] G. Traversi et al., *Design of low-power, low-voltage, differential I/O links for High Energy Physics applications*, 2015 JINST 10 C01055.
- [5] L. Gaioni et al., *Low-power clock distribution circuits for the MPA*, 2015 JINST 10 C01051
- [6] R. Brouns et Al., *Development of a radiation tolerant low power SRAM compiler*, AMICSA 2014.
- [7] F. Faccio et Al., *Radiation-induced short channel (RISCE) and narrow channel (RINCE) effects in 65 and 130 nm MOSFETs*, *IEEE Trans. Nucl. Sci.*, vol. 62, no. 6, pp. 2933-2940, Dec. 2015.
- [8] L. M. Casas et al., *Characterization of radiation effects in 65 nm digital circuits with the DRAD Digital radiation test chip*, 2017 JINST 12 C02039
- [9] D. Ceresa et al., *Readout architecture for the Pixel-Strip (PS) module of the CMS Outer Tracker Phase-2 upgrade*, PoS Vertex 2016 066