

# Readout architecture for the Pixel-Strip (PS) module of the CMS Outer Tracker Phase-2 upgrade

---

**Alessandro Caratelli\***

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

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

**Davide Ceresa†**

*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)

**Simone Scarfi**

*CERN, Geneva, Switzerland.*

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

The Outer Tracker upgrade of the Compact Muon Solenoid (CMS) experiment at CERN introduces new challenges for the front-end readout electronics. In particular, the capability of identifying particles with high transverse momentum using modules with double sensor layers requires high speed real time interconnects between readout ASICs. The Pixel-Strip module combines a pixelated silicon layer with a silicon-strip layer. Consequently, it needs two different readout ASICs, namely the Short Strip ASIC (SSA) for the strip sensor and the Macro Pixel ASIC (MPA) for the pixelated sensor.

The architecture proposed in this paper allows for a total data flow between readout ASICs of  $\sim 100$  Gbps and reduces the output data flow from 1.3 Tbps to 30 Gbps per module while limiting the total power density to below  $100 \text{ mW/cm}^2$ . In addition a system-level simulation framework of all the front-end readout ASICs is developed in order to verify the data processing algorithm and the hardware implementation allowing multichip verification with performance evaluation. Finally, power consumption and efficiency performance are estimated and reported for the described readout architecture.

*The 25th International workshop on vertex detectors*

*September 26-30, 2016*

*La Biodola, Isola d'Elba, ITALY*

---

\*Corresponding author for the CMS collaboration

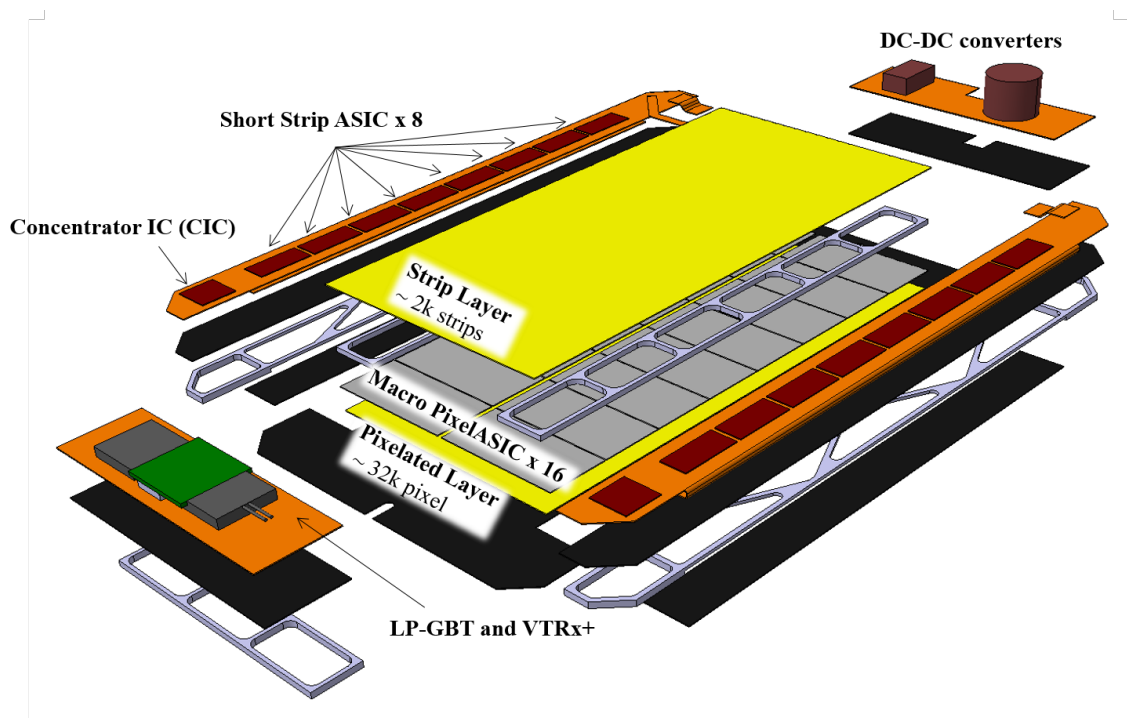
†Speaker and corresponding author for the CMS collaboration

## 1. Introduction

One of the objectives of the CMS Outer Tracker upgrade for the High Luminosity LHC (HL-LHC) is the adoption of double layer sensor modules to allow the local identification of high- $p_T$  tracks ( $> 2$  GeV) and their transmission to the Level 1 (L1) trigger system at the 40 MHz bunch crossing rate. For the first time data coming from a silicon tracker will be used in the L1 trigger decision of a high luminosity hadron experiment. Another transmission channel will send triggered events to the experiment back-end at a nominal average trigger rate of 750 kHz. Summarizing the front-end module is required to provide:

- **Trigger data:** primitives of particles with high transverse momentum which are transmitted for every event;
- **L1 data:** with complete pixel and strip events when requested by the Level 1 trigger.

Among the modules for the Outer Tracker described in the CMS Technical Proposal [1], the Pixel-Strip modules, shown in Figure 1, are more technologically challenging because they combine a strip sensor with a pixelated one. The pixel layer is segmented into macro-pixels of  $100 \mu\text{m} \times 1446 \mu\text{m}$ , while strips measure  $100 \mu\text{m} \times 23136 \mu\text{m}$ . The readout ASICs extract hits (binary signal) from the sensor signals. Considering the 32 K channels of the Pixel-Strip module at 40 MHz, they represent roughly an amount of 1.28 Tbps per module. A compression factor of around



**Figure 1:** Pixel-Strip (PS) module exploded view. The stack consists of (bottom to top) a cooling plate (black), a pixel sensor (yellow), a layer of 16 MPAs (grey), two Al-CF sensor spacers (light blue), two front-end hybrids (orange) housing the SSAs (red) and the Concentrator IC (red), two service hybrids (orange) housing the optical link (green) and the DCDC converters (brown) and a short-strip sensor (yellow).

20 in the front-end (FE) ASICs is necessary in order to reach an almost lossless data communication. This compression combines zero-suppression techniques with the capability of recognizing particles with high transverse momentum as proposed in [2].

The different sensor types require two front-end ASICs: the Short Strip ASIC (SSA) and the Macro Pixel ASIC (MPA). The large area of the sensor ( $5 \times 10 \text{ cm}^2$ ) makes necessary 16 MPA as well as 16 SSA chips for the full module readout and consequently a chip for data aggregation called Concentrator IC (CIC). After the aggregation, data are transmitted through an optical link with a speed of 5 or 10 Gbps using the LP-GBT and the VTRx+ ASICs [3]. Power is provided by DC/DC converters [4] which convert the input power supply from 10-14V down to the power supply needed by the ASICs, between 1.0 V and 2.5 V.

This paper focuses on the architecture of the readout ASICs (SSA and MPA), providing the description and the performance of the chosen scheme. Power and bandwidth limitations require an optimization of the readout architecture which must not affect the particle recognition algorithm efficiency and the data readout capabilities. With this purpose, a versatile module-level simulation framework is an essential development tool for the design, optimization and verification of the system architecture. System-level design and verification are currently done in industry through complex environments based on common verification methodologies like Open Verification Methodology (OVM) and Universal Verification Methodology (UVM) [5], built on top of the the hardware description and verification language System Verilog [6]. A similar approach has been used for the development of a dedicated simulation environment for the PS-module, which is described in Section 3. Simulation results from this module framework, shown in Section 4, guided the development and the choice of the final architecture presented in the next Section 2.

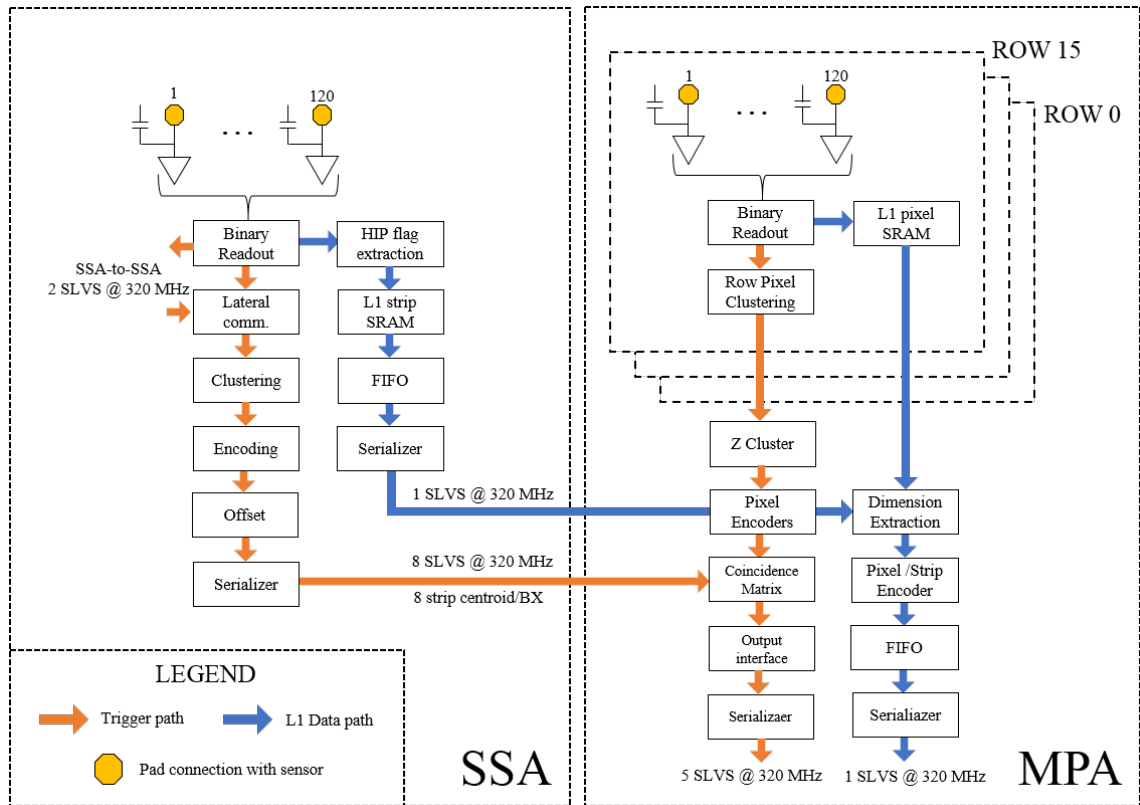
## 2. Readout Architecture

A schematic view of the readout ASICs architecture is shown in Figure 2. The SSA reads out the strip sensor signals, stores the strip L1 data and sends strip trigger data to the MPA. The latter reads out the pixel sensor, stores the pixel L1 data and processes the pixel and strip trigger data: it correlates the pixel sensor hits with the strip sensor hits received from the SSA in order to reject low- $p_T$  particles and provides only high- $p_T$  particle data to the detector back-end electronics. L1 data are encoded and sent to the detector back-end electronics when requested by a L1 trigger signal.

In the following paragraphs, a detailed description of the Readout ASICs is reported.

### 2.1 Short Strip ASIC

The SSA reads out the strip sensor with a double threshold binary system. The signal from the sensor is amplified by the front-end which incorporates two discriminators with two thresholds: the detection threshold is nominally set around 1/4 of a MIP (Minimum Ionizing Particle energy), while the second threshold is set around 1.5 MIP. The latter is called High Ionizing Particle (HIP) threshold since it allows to distinguish the HIPs. Discriminator pulses are sampled with the 40 MHz bunch crossing (BX) clock and stored in a Static RAM (SRAM) until the arrival of a Level-1 trigger. Detection threshold data, called strip data, are stored without any further compression, while HIP threshold data, called HIP data, are stored with an addition compression technique which limits



**Figure 2:** PS-module readout ASICs architecture. The legend is reported on the bottom-left corner of the figure.

to 24 HIP per BX. L1 trigger commands are sent to the MPA through a single differential link at 320 Mbps. Strip data from the detection threshold are also processed by the trigger data path together with the strip data from neighbor chips. Two differential links, operating at 320 Mbps, provide strip data from two neighboring SSA chips allowing the detection of interesting particles which cross the module between two chips. Large clusters (approximately  $> 400 \mu\text{m}$ ) are discarded, while the centre positions of the remaining clusters are encoded. In addition a programmable offset is applied to centroids depending on their module coordinates. This offset corrects the parallax error generated by approximating a cylindrical geometry with sensors that are actually planar strips. This information is continuously sent to the MPA with 8 differential links operating at 320 Mbps. The total bandwidth between SSA and MPA is 2.88 Gbps.

## 2.2 Macro Pixel ASIC

The MPA reads out the pixel sensor with a single threshold binary system already prototyped and tested [7], where the threshold is about 1/4 of MIP. As in the SSA, the pixel data are stored in SRAMs until the arrival of a Level-1 trigger. Events required by the L1 trigger are processed by the L1 data logic. The same data processing is carried out on strip and pixel data: the cluster information is extracted and encoded with the position of the first pixel/strip in the cluster and its width. A HIP flag is added to the strip cluster to notify the presence of a HIP hit. L1 data are

transmitted over a single 320 Mbps differential link.

Concerning the trigger data path, large pixel clusters are discarded as in the SSA, while the remaining are encoded. The logic gathers also the strip clusters and selects only the pixel and strip cluster pairs which show a position difference below a certain programmable threshold. The position difference limit varies between  $200\ \mu\text{m}$  and  $400\ \mu\text{m}$  depending on the momentum threshold desired and on the module position in the tracker. The selected pairs are encoded as stubs which contain the position of the pixel clusters and the position difference between strip and pixel, called bending. Complete details about the trigger path processing logic can be found in [8]. Trigger data is sent out of the MPA in block synchronous mode with 5 differential links at 320 Mbps, where stubs are aggregated over two consecutive bunch crossings, hence smoothing the chip occupancy fluctuations in time. The total bandwidth towards the data aggregation chip is 1.92 Gbps per MPA.

### 3. Module-level simulation framework

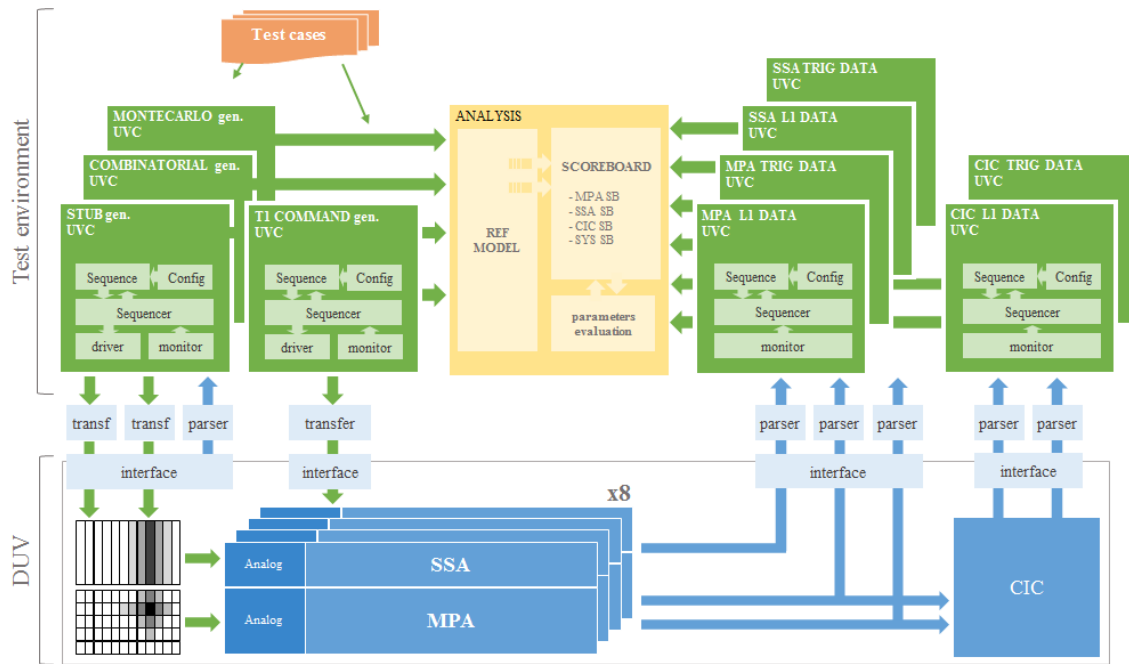
The PS-Module simulation framework is based on the SystemVerilog hardware description and verification language and on the Universal Verification Methodology (UVM) from which inherit the base classes. The main functionalities of the tool are:

- Verify and assist the circuits implementation at register-transfer Level (RTL). Different parts of the design can benefit from a single versatile simulation environment without the need of developing multiple test-benches. The modular implementation and the configurable test scenarios allow to focus the simulation on the functionalities of a specific subsystem and to verify its effect at PS-module level. A PS-module level simulation allows moreover to verify, at clock-cycle level precision, the sub-system integration, the communication between modules and the communication protocols between chips.
- Provide accurate performance evaluation. Bandwidth and power limitations require to optimize the architecture of the system without affecting the overall efficiency. For this reason it becomes necessary to evaluate the efficiency of the particle recognition algorithm and of the data readout. The tool allows the comparison with an ideal reference model and to extract and report efficiency parameters.

Two different types of stimuli can be provided to the Design Under Verification (DUV). For formal verification, random generated stimuli allow to stress the design and reach high test coverage. For performance evaluation the stimuli are extracted from Monte Carlo simulations of the CMS Outer Tracker detector, in order to evaluate parameters based on the physics events.

The DUV includes the actual implementation of the module composed by 8 Macro-Pixel ASICs (MPA), 8 Short-Strip ASICs (SSA) and the Concentrator (CIC), as described in Section 2. It can be either the Register Transfer Level (RTL) description of the PS-Module or the gate-level netlist after synthesis and place and route with back-annotated delays. The analog front-ends are modeled by their accurate behavioral description.

The test environment includes three main types of UVM verification components (UVC), related to the stimuli generation, the output monitors and the analysis components. Those components are implemented at the Transaction Level Modeling (TLM) level. At this level of abstraction,



**Figure 3:** Architecture of the PS-Module verification environment.

which is higher than RTL, the details of the communication between different components are decoupled from the details of their implementation. Channels hide the complexity of the protocols and the communication is implemented in the form of function calls instead of signals.

Four different components provide the stimuli, as described in Figure 3. Accordingly to the UVM methodology [5], each stimuli UVC is composed as follows:

- A Sequence class creates the series of transactions at TLM level.
- A Sequencer randomizes the sequence items, transmits the transactions to the driver and returns the responses to the sequence class.
- A Driver converts TLM transactions into RTL signals that can be given as input to the Design Under Verification through interfaces.
- A Configuration class whose elements configure the stimuli generation and can be controlled from the test cases through the UVM factory mechanism, an object-oriented design pattern that provides the ability to configure the type of objects from a test class or anywhere else in the code.

The stub generation UVC produces randomized transactions that emulate detectors hits from high- $p_T$  particles. This set of hits is sent to the DUV and any missing stub at the output of the PS-Module under test represents an exception that is handled by the test environment. The stub generation is randomized in position, cluster size, bending and energy. The density of stubs follows a Poisson distribution and is configurable per test case. In order to reach high error coverage, the combinatorial generation UVC allows to generate totally randomized stimuli. It can be activated



separately or in addition to the stub generation UVC. In the latter case, it allows to emulate detector hits that do not represent valid stubs such as noise, machine background or simply not interesting particles. The Monte Carlo generation UVC imports particle hits from the Monte-Carlo (MC) programs for computer simulation of complex interactions in high-energy particle collisions, which provide event samples for the entire CMS Tracker. This technique allows to extract the readout electronics efficiency values according to detector statistics. The T1 generation UVC allows to randomize the PS-Module fast commands for time synchronization and L1 triggers.

In order to analyze the signals of the DUV the test environment implements several monitor classes connected at critical points such as at the output of the PS-Module and at the output of each ASIC. The monitors have the purpose to convert the RTL signals into TLM transactions that can be handled by the analysis components. The monitors related to the trigger data path allow to compare the signals on every clock cycle and decode the information. The monitors related to the L1 data path instead implement an event-driven behavior which generates transactions only when triggered by the stimuli generation UVCs or when it detects activity on the monitored signals. A reference model, implemented at TLM level, reproduces the ideal functionalities of the PS module. It receives an input stimulus from the generation UVCs and generates the expected transactions for each component of the PS-Module. The monitor outputs and the reference model output transactions are stored in queues and are transmitted through the UVM factory to the analysis UVC. In this component, several scoreboards allow to perform conformity checks between predicted and actual DUV outputs. The results of the comparison are reported in the simulation log files and further analyzed: every mismatch can either represent an error in the design implementation or a non-ideality of the architecture.

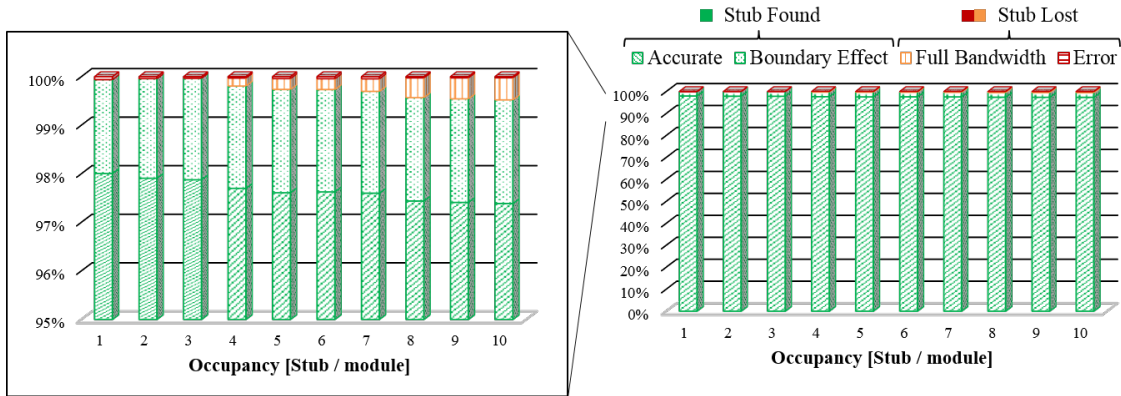
## 4. Results

As explained in Section 3, the module level simulation environment provides a test-bench for code and data protocol verification. In the PS-module development, once a full verification of the DUV is achieved, the same tool has been used to evaluate the efficiency of the algorithm implemented in hardware.

### 4.1 Efficiency results

The simulation environment evaluates the discrepancies with respect to the ideal model. It is able to distinguish among the sources of inefficiency and report the results accordingly. Discrepancies are categorized as:

- **Boundary effect:** counts the number of stubs which are generated with an offset of  $\pm 50 \mu m$  in the position or bending value. This effect is related to the dead area between MPAs which affects the position of the pixel centroids.
- **Full Bandwidth:** shows the percentage of stubs lost because of the limited bandwidth between the ASICs in the module.
- **Error:** shows the percentage of stubs lost because of artefacts introduced by the stub finding algorithm implemented in the hardware.



**Figure 4:** Efficiency of the stub finding hardware implementation for different stub occupancy values. On the left the details of the losses.

Figure 4 reports the efficiency of the presented architecture for different occupancy values. The maximum occupancy foreseen in the Outer Tracker is close to 4 stubs/module where the inefficiency due to the bandwidth limitation is  $< 0.2\%$ . The boundary effect ranges between 2-3% as expected from the geometry of the module, while the errors are limited below to  $0.02\%$ .

The stubs can be classified as real stubs coming from high- $p_T$  particles, which are useful for event reconstruction, and fake stubs coming from artefacts generated by low- $p_T$  or secondary particles and combinatorics, which will be filtered out in the back-end [9]. The percentage of real stubs respect the total generated stubs is between 5 and 10%. Consequently, the losses which affect event reconstruction due to non-ideality of the hardware implementation of the stub finding algorithm and bandwidth limitations are extremely low:  $< 0.02\%$  of the total number of stubs.

#### 4.2 Power consumption results

Besides the efficiency, the power consumption is another parameter of importance in the design of the PS-module. The differential communication interconnects needed for the continuous flow of data between the readout ASICs are important contributors to the power consumption of the readout ASICs. Given a power consumption of 2.5 mW per differential driver and receiver [10] which operates at a frequency of 320 MHz, the consumption per Mbps is  $\sim 8 \mu\text{W}$ . Consequently, without data reduction the communication from SSA to MPA (approx. 6 Gbps) would cost almost 50 mW. The presented architecture halves the needed bandwidth showing a power consumption for data communication from SSA to MPA of 23 mW. The total data flow among readout ASICs amounts to 5.44 Gbps which corresponds to a consumed power of  $\sim 43 \text{ mW}$ , between 17-18% of the available power budget.

### 5. Conclusions

The readout architecture for the PS module has been defined and evaluated. Data reduction is applied directly by the readout ASICs based on the particle momentum discrimination. These techniques limit the bandwidth required fulfilling the transmission bandwidth requirements and providing also a power consumption reduction. A dedicated test system environment has been



developed to verify the single chip models as well as the communication protocols. Full verification has been achieved and the performance evaluation shows the high- $p_T$  particle primitives lost are limited to  $< 0.02\%$  of the total number of stubs.

## 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. Abbaneo and A. Marchioro, *A hybrid module architecture for a prompt momentum discriminating tracker at HL-LHC*, 2012 JINST 7 C09001.
- [3] P. Moreira et al., *High Speed Optical and Electrical Links for LHC and HL-LHC*, 2014, ECFA High Luminosity LHC Experiments Workshop, Aix-les-Bains, France
- [4] A. Affolder et al., *DC-DC converters with reduced mass for trackers at the HL-LHC*, 2011 JINST 6 C11035.
- [5] Mentor Graphics, *OVM/UVM user's guide*, 2014, <https://verificationacademy.com/topics/verification-methodology>.
- [6] IEEE Std 1800-2012 (Revision of IEEE Std 1800-2009), *IEEE Standard for SystemVerilog-Unified hardware design, specification, and verification language*, 2013 <http://dx.doi.org/10.1109/IEEESTD.2013.6469140>.
- [7] 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.
- [8] 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.
- [9] D. Abbaneo, *Upgrade of the CMS Tracker with tracking trigger*, 2011 JINST 6 C12065.
- [10] G. Traversi et al. *Design of low-power, low-voltage, differential I/O links for High Energy Physics applications*, 2015 JINST 10 C01055.