OPEN COSMIC RAY DATA STANDARDS

OCTOBER 2017

AUTHOR(S):
Darko Lukić

SUPERVISOR(S):
Markus Nordberg

CERN
openlab

PROJECT SPECIFICATION

In this project open cosmic data standards should be proposed, as well as an architecture of cosmic ray detector which can fulfil the standard. An open data standard for cosmic rays is important to scientists and researchers in order to collect and process data from multiple cosmic ray detectors all around the world.

Creating an open cosmic data standard requires answers a number of questions, such as who will host open data standard, who will host data about detected cosmic rays and what are the minimum requirements that each cosmic ray detector has to meet.

The focus of a cosmic ray detector architecture should be to provide precise time measurement of cosmic ray detection in order to enable cosmic ray reconstructions, to maximize the usefulness of a distributed cosmic ray detection system. Precise time measurement is a delicate task and it requires a hard real-time device, therefore FPGA should be used. A realisation of such an FPGA system was simulated and studied with VHDL code using software packages from Lattice Semiconductor.

ABSTRACT

Aiming at proposing an open data standard for cosmic rays, a generic architecture for collecting data and a cosmic ray detector that meets the open data standard are presented. The proposed standard is a result of the experience of successfully established open data standards, an analysis of data related to cosmic rays already available online, and discussions with staff from the CLOUD experimental collaboration at CERN. Moreover, required fields (that every contributing detector should provide) and the option to include custom fields specific to a detector architecture or version are described. Possible host organisations for the data standard, as well as for data itself, are proposed.

In this paper, a general architecture of cosmic ray detector, using scintillator plates, is presented. An FPGA based signal processing unit is described, which has two main components: signal processing from fast ADCs and precise time measuring using GPS.

Keywords: FPGA, cosmic ray, open data standard, cosmic ray detector, VHDL, precise time measurement, fast ADC and processing

TABLE OF CONTENTS

Cosmic Rays

Primary cosmic rays are fully ionized atomic nuclei and other particles accelerated at astrophysical sources and reaching the Earth. Sources most common source of cosmic rays is generated during supernova explosions and in supernova remnants in our galaxy.

The discovery of cosmic rays by Victor Hess in 1912 made it possible to verify both the Einstein relation between particles mass and energy and the Dirac theory about the existence of antimatter. [1]

Also, in 1997 it was reported the correlation between cosmic rays and Earth's cloud cover over a solar cycle, first reported by Svensmark and Friis-Christensen. [11]

Cosmic Rays and Creation of Clouds

Two mechanisms by which cosmic rays may affect cloud droplet concentrations or ice particles are described below. We call these two mechanisms the ion-aerosol clear-air mechanism and the ion aerosol near-cloud mechanism.

The ion-aerosol clear-air mechanism is based on the expectation that the presence of ions enhances the birth and early growth of aerosol particles in the atmosphere. A fraction of these may eventually grow into cloud condensation nuclei, leading to the formation of clouds.

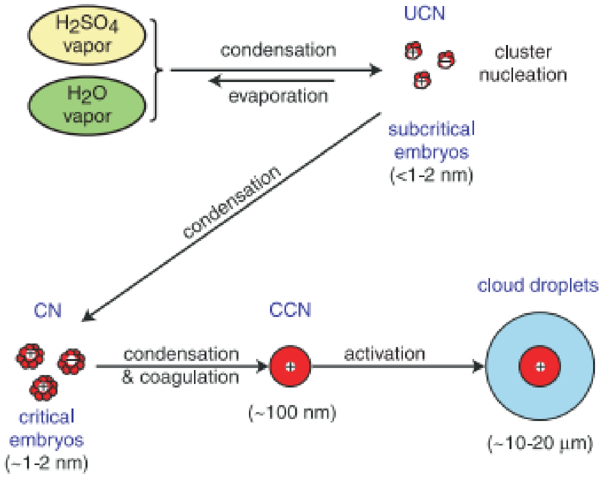

Figure 1. An "ion-aerosol clear-air" mechanism proposed to link variations in cosmic ray intensity with cloudiness. [3]

Ion-aerosol near-cloud mechanism. The ion-aerosol near-cloud mechanism is less well understood. It hinges on the fact that the aerosol electrical charge - and how this charge varies with changes in the ionization rate - is very different near clouds than it is in clear air. [3]

The cylindrical CLOUD chamber, located at CERN, is an electro-polished stainless steel tank volume of 26.1 m3 which aims to simulate creation of clouds. Different inlets and outlets at the chamber wall can be used to connect sampling probes, to introduce trace gases into the chamber, and to evacuate the chamber. [15]

As it is previously described ("Cosmic Rays and Creation of Clouds") cosmic rays have an impact on creating clouds, which means a corresponding impact on cloud cover.

Cosmic Rays and Climatic Changes

Cloud cover variations significantly alter the heat balance of the earth's climate system on an hourly time scale, but their effects are profound on a longer timescale, from seasonal through to decades and centuries. Findings show that variations of cloud cover have significantly contributed to contemporary climatic changes. It is known that the presence of clouds is associated with a cooler surface, but significant uncertainties are involved in assessing the role of cloudiness changes in the climate system, in particular under conditions related to increased greenhouse gases. [6]

Whilst cosmic rays have been studied since 1912, there is currently no global network combining individual experiments into a coherent framework that can be accessed by climate scientists. The development of a low cost, distributed, cosmic ray detection system and open data standards to share data between existing experiments, has the potential to enhance our understanding of the relationship between cosmic rays and cloud cover by providing real time, and long time sequence empirical data for large areas of the globe. This data could then be cross referenced with satellite observations and existing climate models to gain a better understanding of the relationship between cosmic rays and climate, thereby addressing one of the unknowns in our understanding of the wider process of climate change.

Open Data Standards

The definition of an open standard is one developed through an open, consensual process in which all identifiable stakeholders have been invited to participate. [4] However, there are many specific definitions by individual organization, institutions, companies and even governments. The most accepted definition is given by five organizations together (IEEE, ISOC, W3C, IETF and IAB) where they defined five key “OpenStand Principles”:

- Cooperation,
- Adherence to Principles,
- Collective Empowerment,
- Availability and
- Voluntary Adoption.

We will seek to follow these principles in the open data standard for Cosmic rays.

History of Open Data Standard

An instructive example of an open data standard is RSS which stands for RDF Site Summary, also referred to as Really Simple Syndication. It is data format for the web, established in March 1999, with the goal of keeping users updated about the latest news via a simple and lightweight data structure.

RSS has changed owners through its history and remains popular today. It was first introduced by Netscape at Version 0.9 in 1999. This was based on work by Dave Winer of UserLand, who developed the scriptingNews format in 1997. The scriptingNews format was further developed by Winer in 1999 to version 2.0b1, adopting the features of RSS 0.9. A first step to unification of the standards was the release of RSS 0.91 which includes most features from scriptingNews 2.0b1. This lead to the deprecation of the scriptingNews format.

The RSS team at Netscape was then closed, leading to the publication of UserLand’s RSS 0.91 specification in 2000. RSS 1.0 was published as a proposal in the same year, produced by a number of individuals under the leadership of Rael Dornfest. This was a completely new standard with only minimal elements retained from RSS 0.9.

The framework for RSS 2.0 was further developed by Winer after leaving UserLand in 2002. Finally, RSS 2.0 standard copyright is assigned to Harvard's Berkman Center for Internet & Society. [9] The assignment of the copyright to a non-commercial third party was a significant step in aiding the adoption of the standard as competing firms no longer had to worry about changes to the standard for commercial advantage. [12]

Example of Open Data Format in Science

The NASA Ames format is a text-based, self-describing, portable format. Each NASA Ames file (made of printable ASCII character set) contains header section, which describes data, and a data section. It aims to portable, self-described and readable by humans. [10]

Because of its simplicity and readability it is widely used in climate science. The ACTRIS European Research Infrastructure for the observation of Aerosol, Clouds and Trace gasses uses NASA Ames format to share climate measurement data.

Example of Open Data Standard

The OMG Standard seeks to promote the free flow of information between government agencies and citizens by establishing a common set of technical standards for organizing and sharing public data. [5]

This is an example of sharing spatial data with an open standard. Whilst this covers one of the key requirements for an open cosmic ray data format, further critical elements are required which will be covered later in this document.

Data Format and Data Standard

To standardize open data OMG provides required and extra fields, but does not require a specific data format. Users are free to choose the most convenient data format, however OMG recommends users store their data in CSV, JSON, or XML format since they have the most granular fields. It discourages the use of KML and RSS 2.0 due to their limitations. OMG Standard already has many users, including the City of Louisville USA, Jefferson County USA and Your Mapper. The use cases include provision of local municipal data, real estate sale information and crime statistics.

Management of Open Data Standards

Following best practice from other Open Data standards, the standard for collecting and distributing data about cosmic rays should be stored by a trusted organization and made publicly available.

The RSS example has an educational history as it almost died when Netscape was sold to AOL. It is therefore prudent that an open data specification should be hosted by the independent organization, with the copyright owned by this organization. A suggested host for an open cosmic ray data format could be the International Particle Physics Outreach Group (IPPOG), or an international organization such as CERN.

The chosen organization should have a good understanding of cosmic rays. Indeed the CERN convention makes specific reference to the study of cosmic rays, so it could be a good fit with the aims of the organisation.

Software Implementation

Software and precisely defined data standard has to enable to various cosmic ray detectors to store data in same database and uniform API for exporting data.

The open cosmic ray data standard has to provide:

- Support for multiple data formats and
- The definition of required and optional parameters.

Cosmic Ray Details

Every detected cosmic ray has to contain a minimal number of details to provide enough information to be scientifically useful.. Some cosmic ray detectors are able to describe cosmic rays with additional parameters. These additional parameters could be very useful to scientists and researcher depending on the experiment and hypothesis being tested.

Required Cosmic Ray Details

- **Longitude** (real number, character) - Geographic coordinate that specifies the east-west position of a point expressed in degrees. Dynamic.

  Example 4916.46,N Latitude 49 deg. 16.45 min. North

- **Latitude** (real number, character) - Geographic coordinate that specifies the north–south position of a point expressed in degrees. Dynamic.

  Example 12311.12,W Longitude 123 deg. 11.12 min. West

- **Altitude** (real number) - Height above sea level of a location in meters. Dynamic.

  Example 545.4,M Altitude, Meters, above mean sea level

- **Location precision** (real number) - Location precision presents maximal deviation of location in meters. This value should be the maximum of the latitude and longitude, representing the radius of a sphere within which the detector may be located.

  Dynamic.

- **Time** (date and time) - Time have to be formatted in ISO8601 standard (“%Y-%m-%dT%H:%M:%S.%f”, eg. 2016-06-10T21:42:24.7607389988).

  Dynamic.

- **Time precision** (number) - Time precision presents the maximum deviations in nanoseconds for given event.

 Dynamic.

- **Standard Operating Procedure** - A document produced by the designer/manufacturer of the equipment detailing the technical performance of the detector and how it should be operated in order to produce reproducible results. Fixed technical characteristics of the detector, such as the maximum resolution, theoretical limits for parameter accuracy and any relevant points for data analysis should be included in this document.

These two items should be defined in the Standard Operating Procedure:

- **Detector Surface** (real number) - Size of surface for cosmic rays detection in square meters. Static, for a given unit, software version.
- **Error Rate** (number) - Number of missed and false positive detections per 1 million cosmic rays. Static, for a given unit, software version.

Optional Cosmic Ray Details

- **Temperature** (real number) - Temperature of the detector in degree Celsius (°C)
- **Pressure** (real number) - Pressure presented in bars at the detector.
- **Magnetism** (real number, real number, real number) - Magnetic field presented in teslas (T) at the detector. Representing the field in the X, Y and Z axes respectively.

 Example: 1.1, 1.0003, 0.003 Tesla

- **Acceleration** (real number, real number, real number) - Gravitational field, represent the force exerted by gravity on the detector in order to deduce the orientation.

 Example: 0.05, 0.05, 9.71 M/S/S where the detector is oriented with the Z axis parallel to the earth's gravitational centre.

- **Angle** (real number, real number, real number) - Unit vector that describes an incidence angle of cosmic ray. X coordinate is parallel to north.

Custom Data Fields

Custom data fields, or reserved fields, are fields dedicated to specific datasets may be provided as optional details. They can be used to describe thunderstorms, rainfall or more technical details like, CPU and RAM usage. All custom data fields for a given detector type and version should be explained in the Standard Operating Procedure.

All keys should be lowercase and without special characters in order to be compatible with different data formats.

Data Format Examples

JSON

```
[
  {

"Longitude":"4916.46,N",

"Latitude":"12311.12,W",
     "Altitude":100,
     "LocationPrecision":5,
     "Time":"2016-06-
10T21:42:24.7607389988",

"TimePrecision":"1000",

"StandardOperatingProcedure
":"C20312032F34",
     "Temperature":23.2,
     "Pressure":1.023
  },
  {

"Longitude":"4916.46,N",

"Latitude":"12311.12,W",
     "Altitude":100,
     "LocationPrecision":3,
     "Time":"2016-06-
10T21:42:24.76075942845",

"TimePrecision":"1000",

"StandardOperatingProcedure
":"C20312032F34",
     "Temperature":23.2,
     "Pressure":1.023
  }
]
```

XML

```
<?xml version="1.0" encoding="UTF-8"?>
<Events>
  <Event>
     <Longitude>4916.46,N</Longitude>
     <Latitude>12311.12,W</Latitude>
     <Altitude>100</Altitude>

<LocationPrecision>5</LocationPrecision>
     <Time>2016-06-
10T21:42:24.7607389988</Time>
     <TimePrecision>1000</TimePrecision>

<StandardOperatingProcedure>C20312032F34</St
andardOperatingProcedure>
     <Temperature>23.2</Temperature>
     <Pressure>1.023</Pressure>
  </Event>
  <Event>
     <Longitude>4916.46,N</Longitude>
     <Latitude>12311.12,W</Latitude>
     <Altitude>100</Altitude>

<LocationPrecision>3</LocationPrecision>
     <Time>2016-06-
10T21:42:24.76075942845</Time>
     <TimePrecision>1000</TimePrecision>

<StandardOperatingProcedure>C20312032F34</St
andardOperatingProcedure>
     <Temperature>23.2</Temperature>
     <Pressure>1.023</Pressure>
  </Event>
</Events>
```

Data Standard Validation and Security

Submitted data has to be validated during the input and to provide status in response. If errors occur, validation should identify this and a response should be made. This process should be part of the SOP for each detector type.

- Data validation should be done in layers and to check if:
- there is a syntax error in data format,
- all required fields are provided,
- custom fields do not contain special characters and

values are inside of the expected ranges (eg. number of events per second, governed by known physical parameters).

Data Organization and Accessibility

After collected data is stored in database additional derived parameters could be provided to users. This is useful for scientists and researchers to avoid unnecessary data manipulation.

Date and time are difficult to parse because of nanosecond precision. API should support different output format for date and time. HiSPARC organization from Netherlands, which collects data about cosmic rays, uses three fields to describe time, date, time and nanoseconds from last second. [13] In order to easily manipulate time in various software packets, time should be exported and presented, in a few formats.

A useful example of pre-calculated value is a surface area which depends on the orientation of cosmic ray detector. That value could help scientist to filter data by surface area more quickly, in order to exclude either high or low rate detectors.

Generic Backend Architecture

Generic architecture for collecting, storage and presenting data is given in Figure 2. Cosmic ray detector or data about cosmic rays is sent to a provider using HTTP POST request. Each POST request should contain at least two parameters, format and actual data presented in the given format.

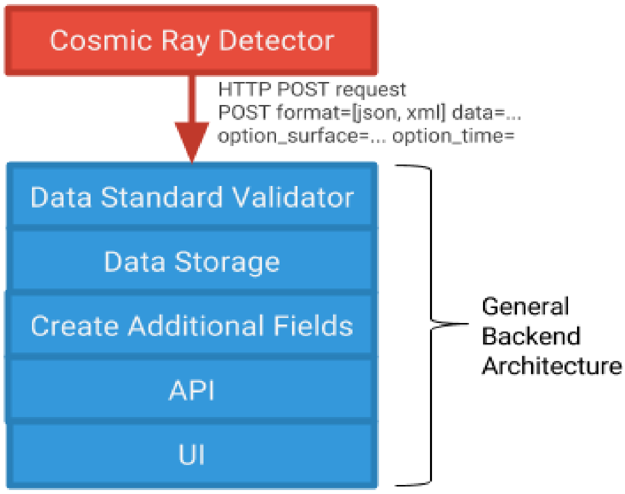

Figure 2. Generic Backend Architecture

The server application is responsible for data validation and data storage when data is posted to the server. Also, the server application should provide a consistent API to filter and export data.

Hardware Implementation

Example hardware for a typical cosmic ray detector is described in this chapter.

Scintillator

A scintillator is a material that can be stimulated by traversing charged particles. The electrons in the molecules are raised to a higher energy level. A part of this excitation energy is released in form of optical photons when the molecules de-excite to their normal states.

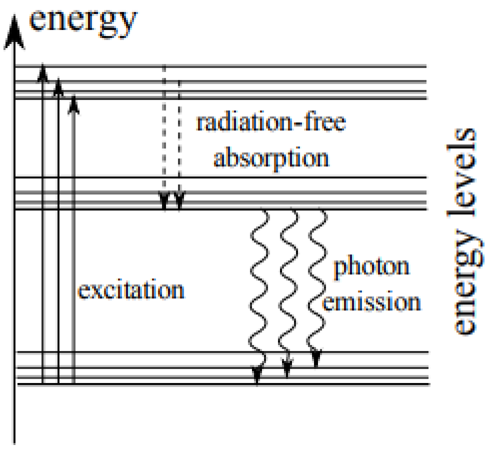

Figure 3. Absorption and emission in a scintillator material

The amount of emitted light is dependent on the energy deposited. A highly sensitive photodetector, such as a PhotoMultiplier Tube (PMT) or Silicon PhotoMultipler (SiPM) can be used to measure the number of photons produced inside a scintillator. [7]

Cosmic Ray Detection

Generally, cosmic ray detectors use either a scintillator or direct conversion silicon detector to transfer energy of muons to an electrical signal.

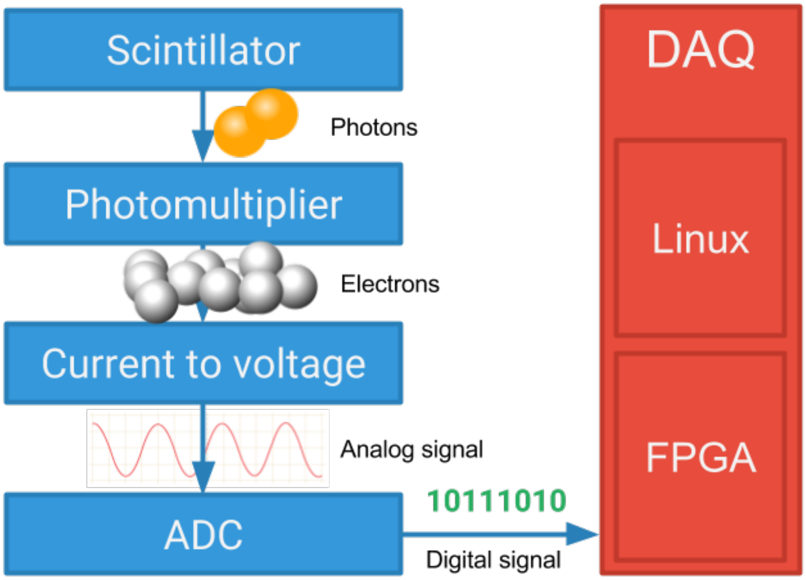

Figure 4. Generic architecture of cosmic ray detector

Photons from scintillator to digital signal

In order to get an analogue electrical signal from scintillation light, a photomultiplier is used. When a photon strikes the mica window surface of the photomultiplier tube (PMT), there is a finite probability of a single electron being released and entering the evacuated portion of the tube. [8]

The photomultiplier tube outputs electrons that correspond number of photons, therefore an additional analogue block is used to convert current to voltage. An ADC is then used to digitise the signal. A silicon photomultiplier (SiPM) operates on similar principles, however a semiconductor is used in place of an evacuated valve amplifier.

Detecting a possible event on single channel

One simple way of detecting events is to compare every ADC value to a predefined threshold. If the sampling rate is fast enough it is not reliable to detect an event by comparing an only single value to a threshold, as it is probably noise. However, if multiple ADC values are above the threshold it is most likely that cosmic ray is detected.

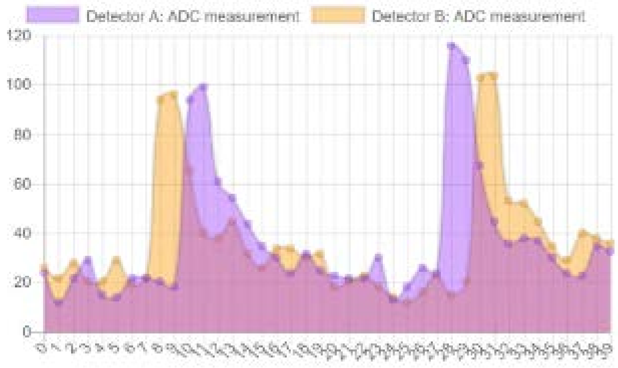

Figure 5. Example of detected two events

Using two scintillators for detecting cosmic rays

Unlike others particles, a muon has enough energy and straight trajectory to go through two scintillators. Considering this behaviour of muons, a reliable way of detecting muons is the detection of events on two or more scintillators at the nearly same time.

Precise Time Measurement

For scientists, it is useful to know when a cosmic ray is detected therefore precise time measuring is very important. Gaining precise time measuring is cheaper and generally easier to implement than higher-resolution quantization within an ADC.

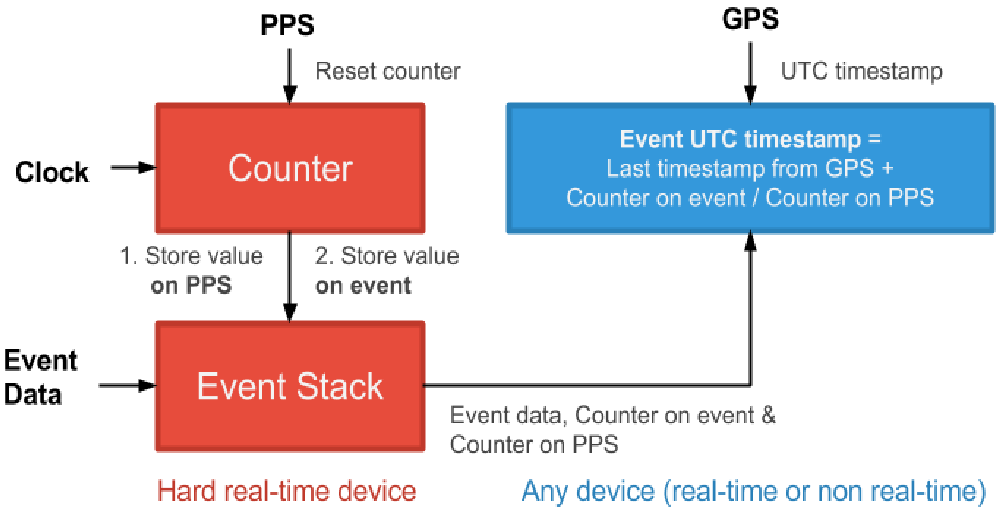

Figure 6. Combining timestamp and PPS from GPS and clock counter to get precise time

Precise time measuring depends on clock speed and jitter on interrupt pin that is connected to GPS (PPS, Pulse Per Second). In Figure 4 a hard-real time device is used to save counter value on PPS because interrupt jitter is lower and saving data from the counter is a time constrained operation.

Data Pipeline

To increase the throughput of ADC data samples a deep data pipeline is required. Figure 5 gives an example of pipeline in the FPGA for the purposes of data acquisition.

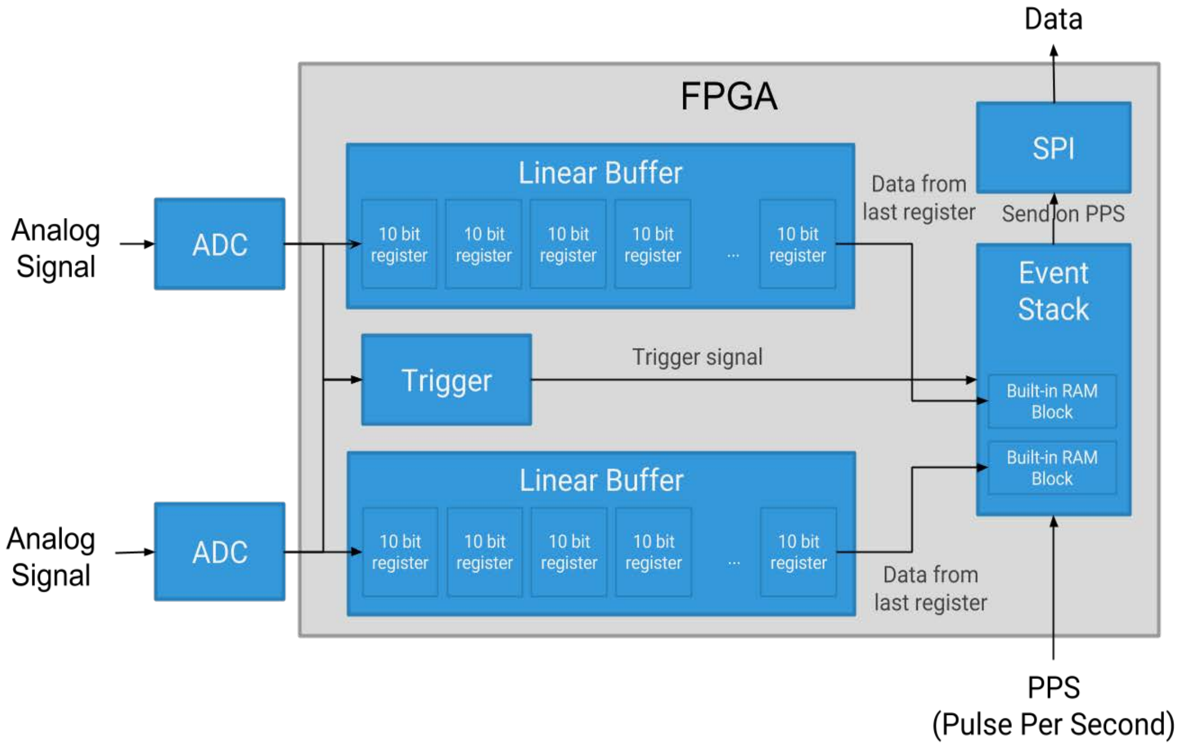

Figure 7. Data Pipeline in FPGA

The FPGA reads data from ADC in every clock and stores that data into a linear buffer. A Linear buffer is used to store values important for signal reconstruction. When an event is detected all relevant data from the linear buffer are stored to the event stack. All data can be transferred over a two or three wire bus for further processing, or processed locally via memory mapping within the FPGA.

Conclusion

Scientists are very keen to get a large amount of data about cosmic rays which can be detected across the surface of the globe. The problem is that this data is not easily accessible and it is hard to combine data from different sources, but those problems can be solved by introducing open data standard for cosmic rays. In particular, the use of precise timestamps and the GPS network as a reference will be crucial to this endeavour.

Guidelines for building cosmic ray detector hardware, software, and a backend architecture for collecting data, are presented in this paper. Both software and hardware are capable of meeting the proposed open data standards for collecting and presenting data about cosmic rays.

The open data standard will enable to community to build software libraries to access the backend API, filter data, extract data in various formats and in that way enable easier manipulation over data.

In order to keep the open data standard for cosmic rays alive, and to encourage use by the scientific community, it is important to be hosted by the independent organization. Clear rules about use of the standard, ownership and a roadmap for future changes are also critical in ensuring long term viability.

References

[1] Maurizio Spurio, *"The Cosmic Rays and Our Galaxy"*, Switzerland, 2015

[2] L. I. Dorman, *"Cosmic rays and space weather: effects on global climate change"*, Russian Academy of Science, Russia, 2012

[3] K. S. Carslaw, R. G. Harrison, J. Kirkby, *"Cosmic Rays, Clouds, and Climate"*, 2002

[4] Priscilla Caplan, *"Patents and Open Standards"*, Maryland, 2003

[5] *"OMG Standard - Technical Specificatio"*, http://www.omgstandard.com/techspec.php

[6] Pavel YA. Groisman, Raymond S. Bradley, Bomin Sun, *"The Relationship of Cloud Cover to Near-Surface Temperature and Humidity"*, Massachusetts, 1999

[7] Janina Struth, *"Muon detection with scintillation detectors using indirect SiPM readout"*, Aachen, 2010

[8] Eric Harpell, Willy Langeveld, Don McShurley, Steve Sharipo, John Venuti, *"The CCRT: an inexpensive cosmic ray muon detector"*, Standford, 1995

[9] Andrew King , *"The Evolution of RSS",* 2003

[10] *"NASA Ames Format for Data Exchange"*, https://badc.nerc.ac.uk/help/formats/NASA-Ames/

[11] Henrik Svensmark, Eigil Friis-Christensen, *"Reply to comments on "Variation of cosmic ray flux and global cloud coverage - a missing link in solar-climate relationships""*, Denmark, 1999

[12] *"RSS Specifications, History of RSS"*, http://www.rss-specifications.com/history-rss.htm

[13] *"About HiSPARC"*, http://www.hisparc.nl/en/about-hisparc/

[14] J. Voigtlander, J. Duplissy, L. Rondo, A. Kurten and F. Stratmann, *"Numerical simulations of mixing conditions and aerosol dynamics in the CERN CLOUD chamber"*, Geneva, 2012

Acknowledgements

I wish to express my sincere thanks to my supervisor Markus Nordberg and James Devine. James has been always available and willing to help whenever I ran into trouble or had questions about the project.

I take this opportunity to express gratitude to all openlab, SDG and Cosmic Pi team members for their attention and support.

I would also like to express my gratitude towards Francois Grey and Rosy Mondardini from the SDG summer school at University of Geneva and Hanna Elina Manninen from the CLOUD experiment for their support and advice.

I would also like to place on record my sincere thanks to Hendrik Borras for being a great teammate. I also thank my family and Jelena Banjac for ecouragement and support.