INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Publication Year	2022
Acceptance in OA@INAF	2023-07-12T15:55:57Z
Title	The Gamma-Flash real-time data pipeline for ground observation of terrestrial gamma-ray flashes
Authors	Addis, A.; Aboudan, A.; BULGARELLI, ANDREA; Baroncelli, L.; CAMPANA, RICCARDO; et al.
DOI	10.1117/12.2627961
Handle	http://hdl.handle.net/20.500.12386/34274
Series	PROCEEDINGS OF SPIE
Number	12189

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

The Gamma-Flash real-time data pipeline for ground observation of terrestrial gamma-ray flashes

A. Addis, A. Aboudan, A. Bulgarelli, L. Baroncelli, R. Campana, et al.

A. Addis, A. Aboudan, A. Bulgarelli, L. Baroncelli, R. Campana, A. De Rosa, A. Di Piano, F. Fuschino, N. Parmiggiani, A. Ursi, E. Virgilli, "The Gamma-Flash real-time data pipeline for ground observation of terrestrial gamma-ray flashes," Proc. SPIE 12189, Software and Cyberinfrastructure for Astronomy VII, 121891G (29 August 2022); doi: 10.1117/12.2627961

Event: SPIE Astronomical Telescopes + Instrumentation, 2022, Montréal, Québec, Canada

SPIE.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The Gamma-Flash real-time data pipeline for ground observation of terrestrial gamma-ray flashes

A. Addis[a], A. Aboudan[b], A. Bulgarelli[a], L. Baroncelli[a], R. Campana[a], A. De Rosa[a], A. Di Piano[a], F. Fuschino[a], N. Parmiggiani[a], A. Ursi[c], and E. Virgilli[a]

[a]INAF/OAS Bologna, Via P. Gobetti 93/3, I-40129 Bologna, Italy.
[b]CISAS G. Colombo University of Padua Italy.
[c]INAF-IAPS, National Institute for Astrophysics, Roma, Italy

ABSTRACT

Gamma-Flash is an Italian project funded by the Italian Space Agency (ASI) and led by the National Institute for Astrophysics (INAF), devoted to the observation and study of high-energy phenomena, such as terrestrial gamma-ray flashes and gamma-ray glows produced in Earth's atmosphere during thunderstorms. The project represents the ground-based supplement to the work of the ASI AGILE satellite in this particular field. This contribution presents the architecture of the Gamma-Flash data pipeline placed at the Osservatorio Climatico "O. Vittori" on the top of Mt. Cimone (2165 m a.s.l., Northern-Central Italy). It consists of RedPitaya ARM-FPGA boards designed for acquiring events at different energies from scintillator crystals coupled to photomultiplier tubes, and a main computer that executes a real-time software pipeline. The software performs several data processing steps, data acquisition, data reduction level, algorithms for waveform selection, and finally it produces the cumulative energy spectrum of the gamma radiation collected by the photomultipliers. Data is stored in different layers, each with a different purpose, and it is available to the scientific community as HDF5 files. The pipeline has a modular architecture to provide good maintenance and flexibility, allowing for easy extensions in the future. A specific subset of data is stored in a database connected to a real-time graphical dashboard for quick-look analysis, showing the acquisition products and the environmental telemetry data.

Keywords: data processing pipeline, Gamma-Flash project, high-energy physics, real time data analysis

1. INTRODUCTION

Terrestrial Gamma-ray Flashes (TGFs) are intense gamma-ray emissions produced by thunderstorm clouds, lasting less than one millisecond, currently observed from space by gamma-ray detectors designed for the observation of cosmic sources. Long-lasting emissions, up to several tens of seconds, have been observed from ground detectors designed for the study of cosmic rays or environmental monitoring. Although these different forms of emission are believed to share the same basic physical mechanism, many points are still obscure: the highest achievable energy, the relationship with lightning, the role of particles and cosmic rays and the relationship with the chemistry of clouds.

Gamma-Flash is an Italian program devoted to the study of radiation and particle properties of these phenomena. It is funded by the Italian Space Agency (ASI) and lead by the National Institute for Astrophysics (INAF), with the collaboration of many institutions and universities. The goal of the program is to develop a payload containing gamma-ray and neutron detectors, placed on-ground and on an aircraft, to perform either on-ground and in-flight measurements of the high-energy emission.The main targets of Gamma-Flash is the investigation of high-energy radiation in thunderstorms, which can have substantial impact in many fields, such as local and global climate change and atmospheric plasma physics.[1] The Gamma-Flash suite of detectors is designed to detect both short-duration transients (e.g. TGFs), as well as minute-lasting gamma-ray emissions (e.g. gamma-ray glows) and associated high-energy particle emissions (e.g. neutrons). The Gamma-Flash team

Further author information: (Send correspondence to A. Addis)
A. Addis: E-mail: antonio.addis@inaf.it

Software and Cyberinfrastructure for Astronomy VII, edited by Jorge Ibsen, Gianluca Chiozzi, Proc. of SPIE Vol. 12189, 121891G · © 2022 SPIE
0277-786X · doi: 10.1117/12.2627961

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

is a group specialized in the field of TGF studies, atmospheric physics, high-energy particle and radiation instruments, radiation damage, data analysis and simulations, taking advantage of the experience acquired over more than a decade with the AGILE mission.[4]

This contribution presents the ground-based Gamma-Flash data processing pipeline, which is installed at the Osservatorio Climatico "O. Vittori" on the top of Mt. Cimone (2165 m a.s.l., Northern-Central Italy). Recent studies have demonstrated that the observatory represents an ideal location to investigate thunderstorms and detect associated high-energy emission due to the particularly high rate of lightning.[2]

2. THE GAMMA-FLASH DATA ACQUISITION AND CONTROL SYSTEM

2.1 System Context

The Gamma-Flash Data Acquisition and Control System (DACS) is composed of a plurality of Data Acquisition Modules (DAMs) and one Main Control Computer (MCC). Each DAM can be configured and controlled independently from the others. The events detection, the waveform acquisition, and the data processing steps are performed automatically. All the data acquired is processed online and the results are stored on the MCC. It is possible to monitor them using a quicklook dashboard installed on the MCC.

Fig. 1 shows the main components of the DACS:

- PMT is a scintillator-based photomultiplier tube detector for gamma-ray or neutron events. The Proximity Electronics (PE) contains a high voltage generator, a charge sensitive amplifier, and the hardware needed to power on/off the detector and to monitor its temperature and the supplied high-voltage.
- DAM: each Data Acquisition Module contains a complete data acquisition pipeline made up of an Analog-to-digital converter (ADC), an event triggering logic, and a microcontroller used to handle the data acquisition. Each unit has some local data processing and storage capabilities.
- MCC: the Main Control Computer is a high performance machine where the collected data are stored and processed. The MCC is connected with the external world using an Internet connection for data transfer and remote control. It controls the overall data acquisition process providing data storage, online data processing, quick-look capabilities, and a unique interface for remote control of the experiment.
- A weather station monitors the environmental temperature and triggers alerts when a specific threshold is reached on specific environmental parameters. It is also used at the beginning of the startup session to decide if the system can be turned on.

DACS is designed to perform the following functions:

- Data acquisition
  - provide the interfaces to the remote control to configure and control the experiment;
  - provide on-site monitor and control capabilities on the MCC;
  - control each detector and collect House-keeping (HK) info;
  - send raw waveforms events to the MCC through a TCP-socket;
  - provide alarms and rate counting in real-time or near real-time to the remote control;
  - save the collected data and produce the data layers;
- Control System
  - control each DAM;
  - control the PE of each detector;
  - provide power supply to all the DACS subsystems;
  - show to the user the results of the data acquisition through a quick-look dashboard;

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

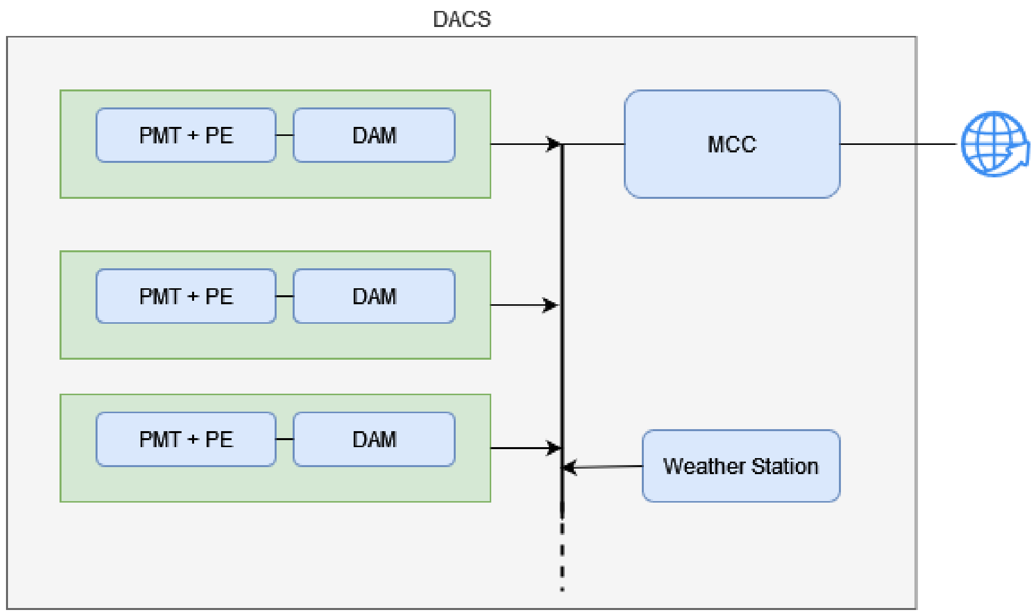

Figure 1. DACS Overview: every block PMT-PE-DAM is a complete acquisition system that shares the same power and ethernet channel with the others. MCC coordinates the acquisition process and collect all the data sent by DAMs.

2.2 Physical description

The main electrical connections between DACS subsystems are shown in Fig. 2 and belong to three different classes: power, digital (e.g. TTL, RS232, or Ethernet connections), and analog signals used to monitor the detectors and the PE.

3. DAM FIRMWARE AND SOFTWARE

3.1 DAM overview

The DAM software is made up of two different parts, the FPGA firmware and the application software both running on a RedPitaya board*. RedPitaya is an open source computer-board equipped with a 125 MS/s fast ADC connected with a programmable FPGA. The FPGA is connected with a dual ARM Microcontroller (MC) running Ubuntu Linux 18.04.

3.2 Firmware description

The FPGA firmware completely defines the functions implemented in the FPGA side of the board. Two different versions exist:

- Default firmware: provided by the supplier of the board, contains the DAC control logic and the event triggering logic, the logic used to handle UART and GPIOs plus several other general purpose functions (signal generation, logic analyzer). This is the version of the firmware currently used for the Gamma-Flash project.
- Custom firmware: it contains all the features of default firmware, and adds a Direct Memory Access (DMA) data transfer logic to move the waveforms to the MC, and in the future it may include also waveform processing algorithms.

The FPGA is connected with the MC by an internal bus. The communication between the MC and the FPGA is based on a custom driver developed using C language and designed to maximize the data transfer speed.

*https://redpitaya.com

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

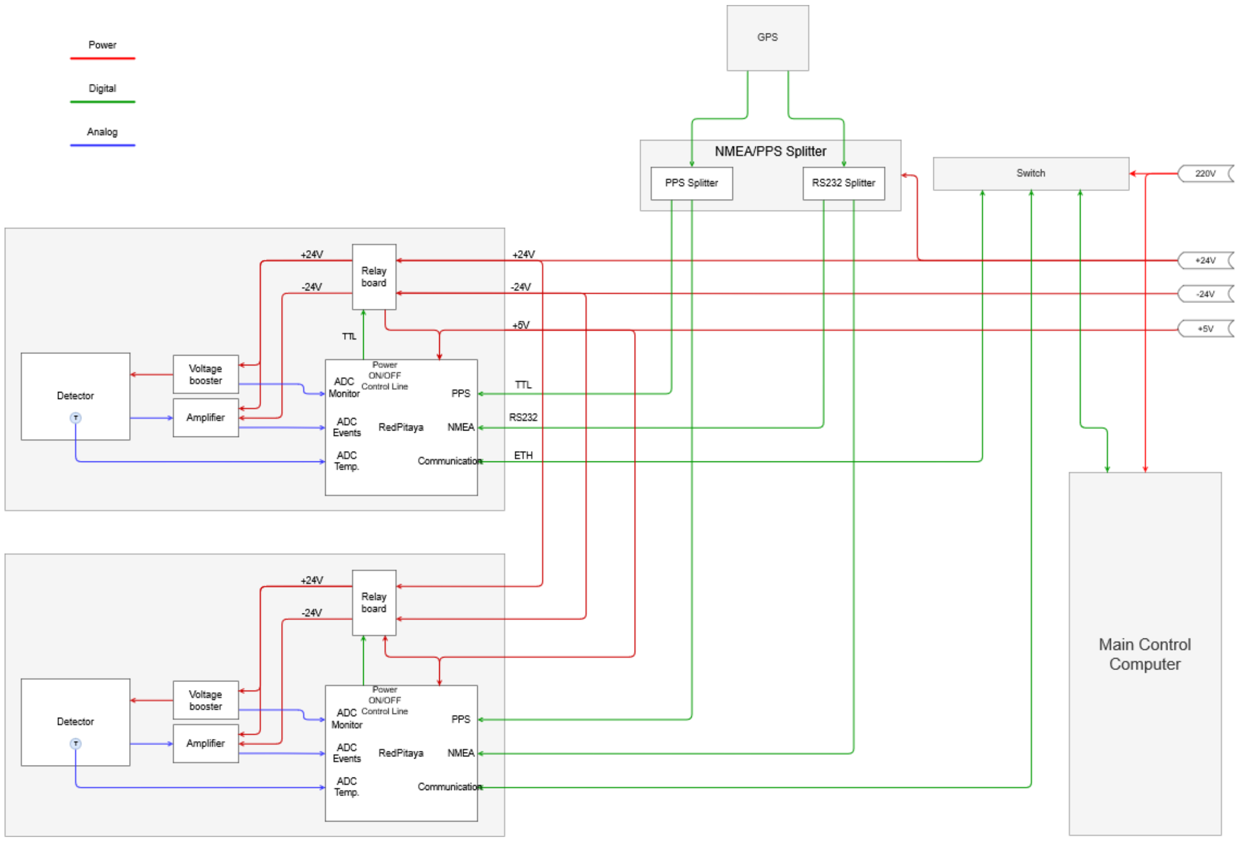

Figure 2. In the figure above is shown the detail of the electrical connections. Voltage boosters require 24V to supply 1KV output for powering detectors. Redpitaya board is connected to the detectors using a SMA cable and send data packets to the MCC through ethernet interface. GPS is connected to a NMEA/PPS splitter that transmits timestags to Redpitayas.

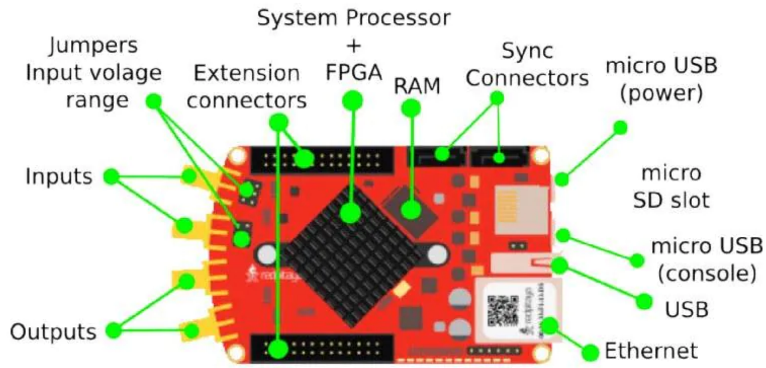

Figure 3. Redpitaya: the system processor and FPGA are included in the same chip, the extension connectors are used as I/O channels for GPS and relays.

3.3 Application Software description (ASW)

The DAM controls the detector, the waveform acquisition process and the interface with the MCC. It is a single multithread process implemented in C/C++ using POSIX API on Linux OS. It implements the data flow shown in Fig. 4: waveform acquisition on the FPGA side, data transfer from FPGA to the ARM CPU, time tagging, data processing and transfer to the MCC.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

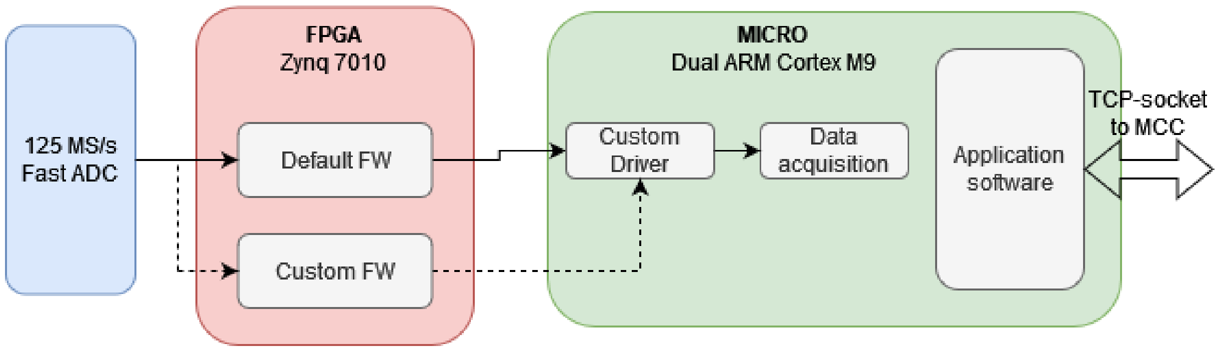

Figure 4. DAM internal components and data flow, the ADC supports up to 125 MS/s, default firmware is currently supported for acquiring data, a custom firmware is under study.

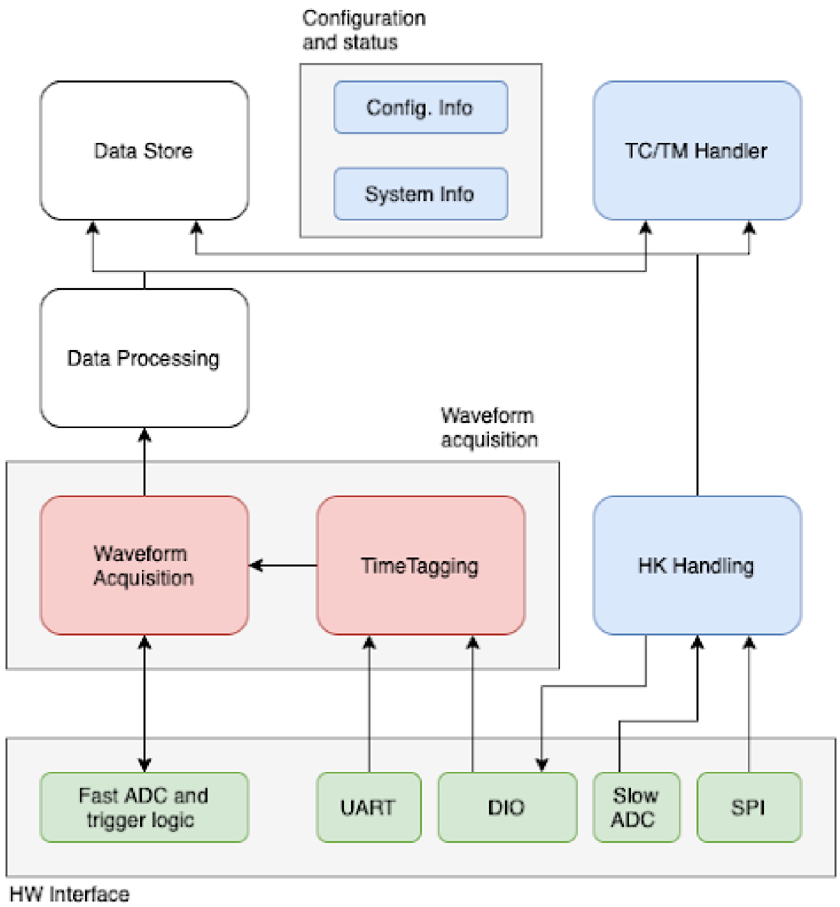

Figure 5. Application software

The main modules of the Application software are shown in Fig. 5. The hardware interface subsystem is in charge of controlling the interface between the ARM CPU and the FPGA, in particular for the configuration and control of the following resources:

- The trigger logic and the fast ADC used to acquire the waveforms.
- The UART used to read NMEA strings from the GPS.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

- The DIO lines. One digital input is used to catch the PPS signal from the GPS and one digital output is used to control a relay to power on/off the detector voltage booster.
- The slow ADC used to monitor the output of the voltage booster.
- The SPI used to read the detector temperature.

The waveform acquisition subsystem is made up of two critical modules with the corresponding threads executed with high priority. The Time Tagging module is in charge of receiving both the PPS signal and the NMEA sentences from the GPS and to correlate them with the internal clock of the board. The waveform acquisition module configures the trigger logic and arm the trigger waiting for an event (waveform), when an event is detected the acquired waveform is transferred from the FPGA to the microprocessor and properly time tagged.

The House-Keeping handling module is a low priority thread in charge of collecting the house-keeping data (voltage booster reference output and detector temperature) and the status of the DAM. The House-keeping data are acquired with a fixed period and sent to the TM/TC Handler.

The TC/TM Handler receives HK data and waveforms from the other modules, creates one or more TM packets and sends them to the MCC. TC received from the MCC are decoded by this module and executed according to the logical status of the DAM. For each TC two TMs are generated to indicate that the TC has been received and the results of the TC execution. The TC/TM Handler implements a TCP-IP server.

The configuration and status subsystem keeps track of the DAM configuration parameters and all the runtime parameters. The Configuration info module loads the configuration parameters from an external XML file. The System Info module implements the DAM state machine keeping all the runtime information such as the status of the power on/off relays, the last values of the HK info.

4. MAIN CONTROL COMPUTER SOFTWARE DESCRIPTION

4.1 Main Control Computer overview

The Main Control Computer (MCC) is a x86 multi-core machine designed to receive, process and archive raw waveforms from DAMs. Fig. 6 shows the components that have been designed and developed: the DAM servers, the Data Layer archives, the Gamma-Flash data processing pipeline and the Gamma-Flash quick-look GUI.

4.2 Data Layer archives

Four different data layers are designed to store Gamma-Flash data. All data layers except DL3 use the Hierarchical Data Format to store data in a HDF file[†]. HDF is a high-performance data format designed to store and organize large amounts of data, it supports several features as the gzip compression, metadata and query filtering. The latest version is called HDF5 and is currently implemented in all DL except DL3, which uses the Python Pickle[‡] format to serialize Python objects produced at the last stage of the Gamma-Flash pipeline.

Table 1. Data layers

Name	Content	Data	Produced by	Format
DL0	Raw waveforms	y values	DAM Server	HDF5
DL1	Waveforms selected for further analysis	y values	DAM Server	HDF5
DL2	Event list	[time, value]	Gamma-Flash pipeline	HDF5
DL3	Cumulative spectrum histogram	int	Gamma-Flash pipeline	Pickle

In order to decrease the size of the DL0 and DL1 HDF5 files it has been decided to discard the x values, since it is possible to obtain for each point the correct timestamp, starting from the initial time value contained in the HDF5 metadata.

[†]HDF5 hierarchical data format. https://www.hdfgroup.org/solutions/hdf5/.
[‡]Pickle — Python object serialization https://docs.python.org/3/library/pickle.html

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

4.3 DAM Server

The DAM Server is a Python software that receives waveforms from DAMs. It implements a TCP socket for collecting the packets the DAMs send and creates a waveform objects that is pushed on a FIFO queue. Then, a parallel process performs the following steps:

1. Checks if a pre-selected number of events is received
2. If the limit threshold is reached, it consumes the queue
3. Creates DL0 file containing all the waveforms received
4. Saves the file in the DL0 Archive

Some waveforms are selected according to specified criteria (e.g., waveforms containing multiple or saturated peaks), and saved in a different archive, DL1. Within the DL1 archive is possible to identify a subset of waveform for further custom analysis.

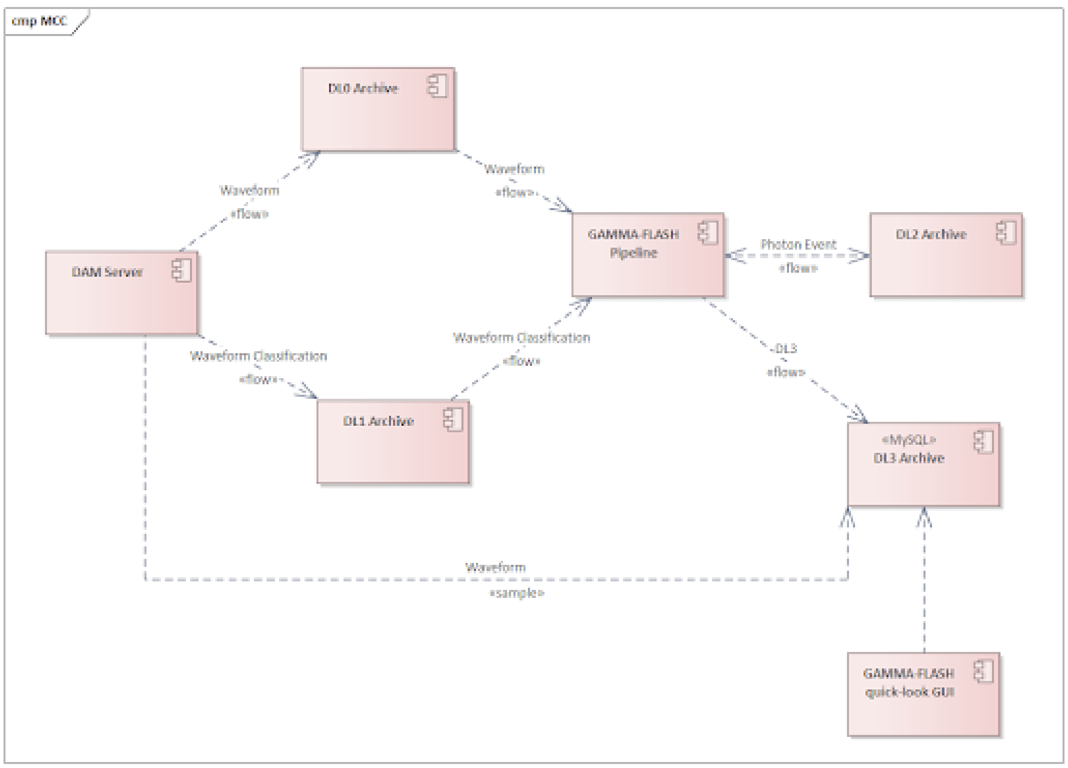

Figure 6. MCC components overview: the image shows the components installed in MCC. DAM Server generates DL0 and DL1 archives, the Gamma-Flash pipeline listens on those archives for new file events and generates DL2, then it listens on DL2 archive and creates the cumulative energy spectrum (DL3). All the process are parallelised to improve the global performances. DL3 is available in a binary file format (Pickle) and written in a MySQL database. Gamma-Flash GUI reads from the database and plots the data in a local webpage.

4.4 Gamma-Flash data processing pipeline

The Gamma-Flash pipeline is a real-time data processing pipeline written in Python that uses high-performance libraries such as Numpy, pytable and rta-dq[3] to generate the Gamma-Flash data products. The software requires writing a XML configuration file for describing the pipeline architecture: the input/output directories or database tables, MySQL database parameter connections, supported file formats and different analysis methods. With

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

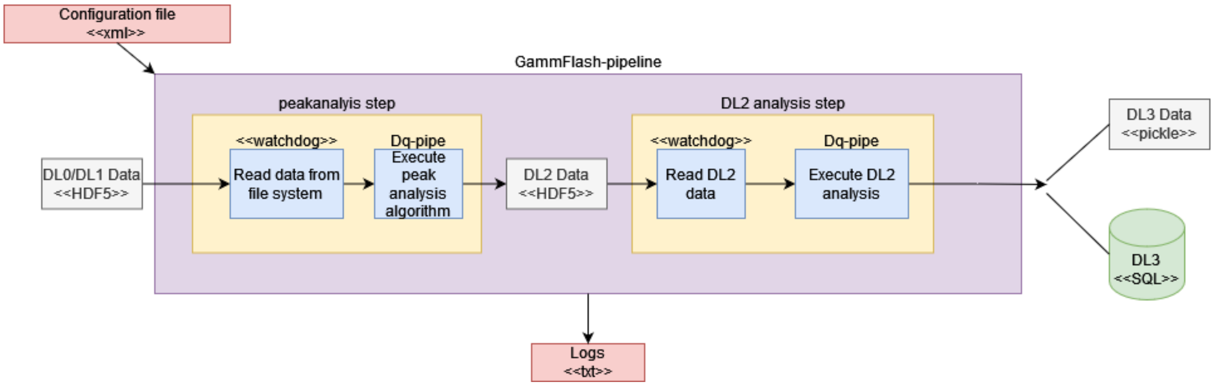

Figure 7. Gamma-Flash pipeline: the yellow blocks implement a watch-dog and a parallel analysis step.

configuration files is possible to customize the entire behavior of the software, allowing for future extensions or to use it in other projects. In the context of Gamma-Flash project, several configuration files are available, each of them implementing a different analysis approach. Fig. 7 shows the current architecture of the data processing pipeline:

1. When a DL0 has been written in the input directory, a watch-dog triggers a new-file event
2. The pipeline reads the content of DL0 file and performs the analysis required on the configuration file
3. A DL2 event list file is produced in the output directory where another watch-dog is listening
4. DL2 is read by the pipeline and processed to finally generate DL3 on a MySQL table or pickle file

The software generates a volume of 40 GB per day for each DAM, since DL0 contains the raw waveforms acquired by the latter. The supported algorithm for waveform processing are the maximum value, the integral of the curve and the Scipy find_peaks algorithm[§]. Given its multiprocess architecture log files are required for identifying issues or bad behavior, and they are available in text format for each step of the pipeline.

4.5 Gamma-Flash quick-look GUI

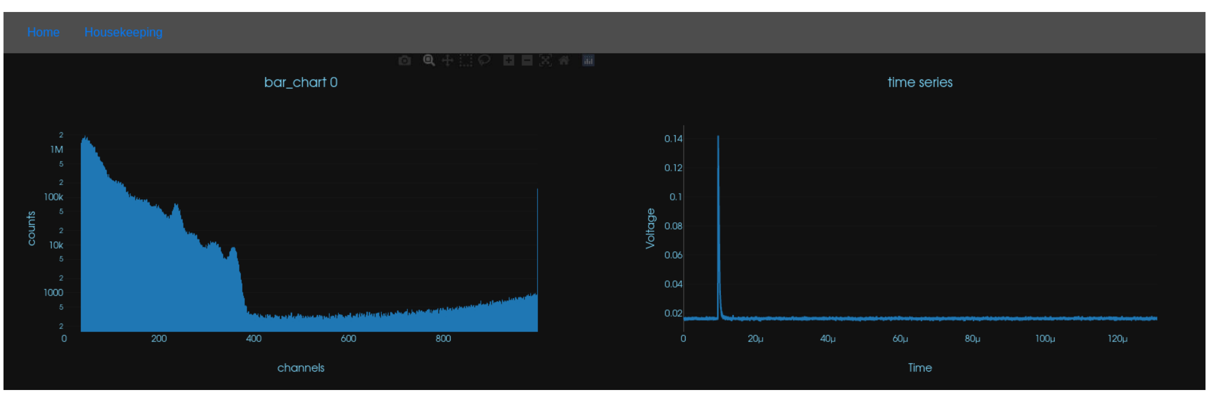

Figure 8. Gamma-Flash real-time dashboard: the figure shows the cumulative energy spectrum and a raw waveform collected by a DAM. The dashboard is deployed on a local webserver and is available using a standard browser.

A quick look dashboard is used for checking the system health and monitoring the real time processing. It is connected to a MySQL database where the data is stored and it shows the cumulative energy spectrum of

[§]https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

gamma radiations collected from DAMs, DL0 waveforms randomly chosen in a selected time window and the temperature curve produced by the weather station. For each DAM two plots are available, showing DL3 and DL0 to the left and to the right, respectively. An XML configuration file is used to organize the views and to set values as the refresh interval value. Plotly[¶] uses a nginx web server to allow the access via browser, and can be remotely deployed on the cloud or in an external local machine.

5. CONCLUSION

This work presented the hardware and software architecture of the Gamma-Flash data processing pipeline for the ground section of the project. It is a real-time software pipeline composed of RedPitaya ARM-FPGA boards for acquiring events from gamma and neutron detectors, and a main computer that receives the data using a low-level TCP socket. The data acquisition module is written in C and Python and generates almost 40 GB of raw HDF5 products per day. It produces several data layers, supporting data reduction level, algorithms for waveform selection, and finally it produces the cumulative energy spectrum of gamma-rays collected by the PMTs. A quick-look dashboard shows to the user the results of the data acquisition and the health of the entire system. The software has been designed to allow flexibility and high customization, and it will be part of the payload installed on an aircraft for the Gamma-Flash flight campaign planned for late 2022.

ACKNOWLEDGMENTS

Gamma-Flash is a mission of the Italian Space Agency (ASI), with programmatic and scientific coordination of the National Institute for Astrophysics (INAF) and partecipation of several research institutes and universities, supported by the ASI-INAF agreement N. 2020-5-HH.0.

REFERENCES

[1] Ursi, A., Rodriguez Fernandez, G., Tiberia, A., Virgilli, E., Arnone, E., Preziosi, E., Campana, R., and Tavani, M., "A study on tgf detectability at 2165 m altitude: Estimates for the mountain-based gamma-flash experiment," *Remote Sensing* **14**(13) (2022).

[2] Ursi, A., Rodriguez, G., Tiberia, A., Virgilli, E., and Tavani, M., "A study of tgf detectability at 2165 m altitude: Simulations for the mountain-based gamma-flash program," *AGU Fall Meeting 2021* (December 2021).

[3] Baroncelli, L., Bulgarelli, A., Parmiggiani, N., Fioretti, V., Addis, A., De Cesare, G., Di Piano, A., Conforti, V., Gianotti, F., Russo, F., Maurin, G., Vuillaume, T., Aubert, P., Garcia, E., and Zoccoli, A., "rta-dq-lib: a software library to perform online data quality analysis of scientific data," *ADASS 2021* (2021).

[4] Tavani, M., Barbiellini, G., Argan, A., Boffelli, F., Bulgarelli, A., Caraveo, P., Cattaneo, P., Chen, A., Cocco, V., Costa, E., et al., "The AGILE mission," *Astronomy & Astrophysics* **502**(3), 995–1013 (2009).

[¶]Plotly python graphing library. https://plotly.com/

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 07 Feb 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use