

## Space-borne Hard X-ray Compton Polarimeter POLAR Data Center at PSI

---

**Hualin Xiao\*, Wojtek Hajdas, Radoslaw Marcinkowski**

*Paul Scherrer Institute*

*E-mail:* [hualin.xiao@psi.ch](mailto:hualin.xiao@psi.ch)

POLAR is a space-borne hard X-ray Compton polarimeter built by a collaboration of institutes from Switzerland, China and Poland. Precise detection of the polarization can be a powerful tool to unveil emission mechanisms of e.g. Gamma-Ray Bursts (GRB) or Solar Flares (SF). POLAR is equipped with an array of 1600 scintillator bars dimensioned for precise measurements of the polarization of hard X-rays in the energy range from 50 keV to 500 keV. The instrument was launched into space on September 15th, 2016 on-board the Chinese Space Laboratory TG-2 for up to 3 years long observation period. Telemetry data from its operation in space may reach up to 50 GB daily. To store and process such huge amount of data both dedicated hardware and specialized software are required. Moreover, constant data inflow also requires fully automated and safe guarded data processing. For this purpose a dedicated data center was established at PSI. We present its design concept and structure as well as demonstrate main features with respect to data processing, quick look utilities and alerting functions.

*35th International Cosmic Ray Conference — ICRC2017  
10–20 July, 2017  
Bexco, Busan, Korea*

---

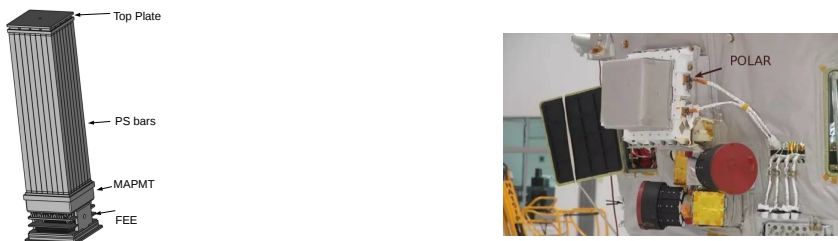
\*Speaker.

## 1. Introduction

POLAR is a space-borne hard X-ray Compton polarimeter built by a collaboration of institutes from Switzerland, China and Poland. It is equipped with an array of 1600 scintillator bars allowing for precise measurements of the polarization of hard X-rays in the energy range from 50 keV to 500 keV. It has a large field of view of about 1/3 of the sky and a sensitive area of about 80 cm<sup>2</sup>. Its minimal detectable polarization (MDP) for stronger Gamma-Ray Burst (GRB) events is lower than 10%. Polarization detection in Solar flares can be determined with even better accuracy. POLAR was launched into space on an orbit around 380 km on September 15th, 2016 on-board the Chinese Space Laboratory TG-2 for up to 3 years long observation period.

The background rates of POLAR in space may account for thousands of events per second. During its normal space operation POLAR is taking data continuously apart of short passages through the South Atlantic Anomaly (SAA). As a result, the daily telemetry data sets are as large as about 50 GB. Both dedicated hardware and specialized software are required to receive, analyse and archive such huge amount of data. Constant data flow also requires fully automated and safe guarded data processing. A dedicated data center at Paul Scherrer Institute (PSI) was established and an elaborated database, for data processing, quick look analysis and alerting purposes was created. In this paper we present design concept and structure of the POLAR data center at PSI as well as its main features.

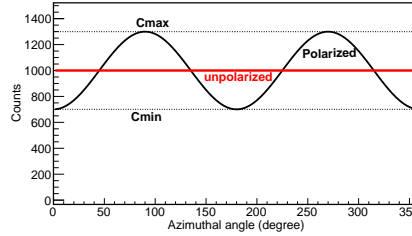
## 2. POLAR instrument



**Fig. 1:** POLAR detector module structure (left) and a photon of POLAR flight model on TG-2 (right). Note that the module carbon fibre socket is not shown in the left panel. POLAR instrument consists of 25 identical modules placed inside the carbon fibre of the enclosure shown in the right panel.

POLAR detector consists of 1600 plastic scintillator (PS) bars grouped in 25 identical detector modules. Each module consists of 8×8 PS bars, a 64 pixel multi-anode PMT (Hamamatsu MAPMT H8500) and a readout front-end electronics (FEE) as shown in the left panel of Fig. 1. Each bar has a dimension of 5.8×5.8×176 mm<sup>3</sup>. The 64 bars are fixed together with two 8 × 8 plastic grids and coupled to the MAPMT via an 1 mm thick optical pad. The FEE provides divided high voltage to all MAPMT dynodes, processes signals from the MAPMT and communicates with the POLAR Central Task Processing Unit (CT). FEE also packs together signals from event timestamps, digitized energy depositions and trigger hit pattern (as well as some auxiliary data such as FEE health status) and forms a module science packet. Module science packets are sent to the POLAR CT. The CT is responsible for control and monitoring of all 25 modules. It also make

trigger decisions from incoming events, manages both low and high power supplies and handles communications with the space-lab. All 25 modules, the CT, the power supplies and interfacing electronics are enclosed in a Carbon box shown on the right panel of Fig. 1. The whole instrument is mounted on the outside panel of the space-lab.



**Fig. 2:** Theoretical modulation curve measured by a Compton polarimeter for a polarized and an unpolarized beam.

Gamma-rays with energies from 50 keV to 500 keV interact with plastic scintillators mainly via Compton effect. The Compton scattering cross-section is given by Klein-Nishina formula

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2}{2} \left( \frac{E'}{E} \right)^2 \left( \frac{E'}{E} + \frac{E}{E'} - 2 \sin^2 \theta \cos^2 \eta \right), \quad (2.1)$$

where  $r_0$  is the classical radius of the electron,  $E$  and  $E'$  are the energy of the incident photon and the scattered photon respectively,  $\theta$  is the angle between incident photon direction and scattered photon, and  $\eta$  is the azimuthal scattering angle between the initial electric vector (polarization vector) and the direction of the final photon. Polarization measurements of the Compton polarimeter are accomplished by measurements of azimuthal scattering angles  $\eta$  and building their distribution called a modulation curve. In the case of unpolarized incident beam, the distribution of  $\eta$  is flat. For a polarized incident beam it follows a  $\cos(2\eta)$  pattern as shown in Fig. 2. The level of the polarization in the measured signal is quantified by a modulation factor  $\mu_p$  given by

$$\mu_p = \frac{C_{\max} - C_{\min}}{C_{\max} + C_{\min}}, \quad (2.2)$$

where  $C_{\max}$  and  $C_{\min}$  are the maximum and minimal number of counts in the modulation curve.  $\mu_p$  is proportional to the polarization degree of the incident beam. The polarization degree  $p$  is given by  $p = \mu_p / \mu_{100}$ , where  $\mu_{100}$  is obtained either by simulations or experimental calibration. In practice a measured modulation curve has to be corrected for background signal and geometry effects. More details on POLAR detection principle and polarization reconstruction method can be found in Refs. e.g. [3, 4, 5].

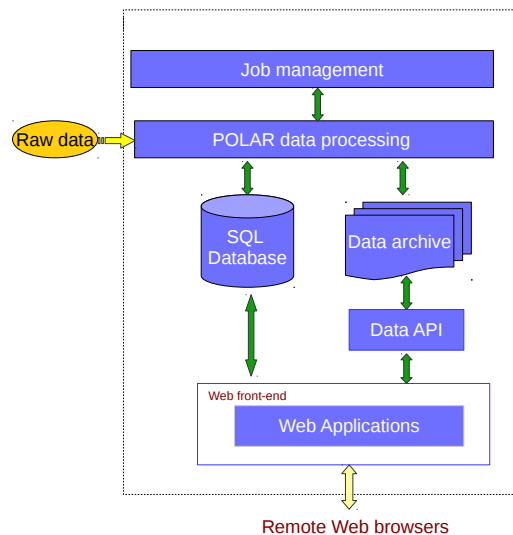
### 3. PSI POLAR data center hardware

PSI POLAR data center (PPDC) relies on high performance computing resources installed at the PSI IT premises. It consists of two powerful servers dedicated for data storage and data

processing. The processing server fulfils heavy duty and intense computing requirements and is equipped with 16 CPU cores and 64 GB RAM. The storage sever currently uses an expandable disk array with a total space of 64 TB. The disks have a Redundant Array of Independent Disks level 5 (RAID5) configuration and use the Network File System (NFS). The RAID-5 configuration of the storage system provides protection of the data even in a case of the whole physical drive failure. Both servers chose as their operation system. They are linked via a gigabit ethernet. A smaller server is also installed as a back-up of both data processing and storage units. It allows for tests of the software solutions and limited space temporary data storage. The data center has been running in a continuous manner for more than six months without any human intervention.

## 4. Software at PSI POLAR data center

### 4.1 Software structure



**Fig. 3:** PSI POLAR data center software structure.

A powerful and versatile software suite was developed at PPDC to facilitate more efficient data analysis and serve science communities involved in the GRBs. As shown in Fig. 3, the suite consists of following packages: data processing software, database, job management software and web-based applications. The data processing software package converts raw data files into products of different complexity levels, conducts calibration data analysis, monitors data quality and performs quick-look tasks and analysis as shown in Fig. 4. It has more than 60 modules mainly developed using C++ and ROOT [8]. The relational database contains 39 tables keeping various information collections. It allows for sharing information between different modules and sub-systems of the PPDC. The job management system is responsible for initialization of the data processing as well as tracking and logging services. The web interfaces consists of more than 50

web-based applications. It provides user-friendly interfaces for data monitors, visualizations of the quick-look analysis results and presentations of database tables and preliminary results and outputs from scientific analysis. More details about each system are presented in the following sections.

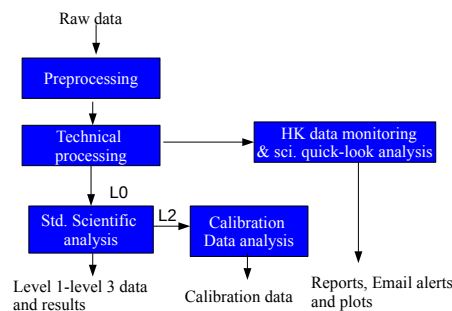
## 4.2 Database

Every day dozens of raw data files arrive to PPDC followed by hundreds of higher level data files and thousands of plots generated by the data processing software. In this phase a critical information pre-processing generates vital inputs to be further used by other software. In order to manage all data files and preserve collected information, a database based on MySQL was developed at PPDC. The database consists of 39 tables with more than 500 columns in total. Different tables contain different information collections depending either on e.g. raw data structures or needs of the processing packages. For example, the raw data table contains raw data file names, directories, data acquisition start and stop times; the run number table has such columns as run numbers, corresponding raw data file names, detector settings, data directories and quick-look results; the detected GRB table stores POLAR run numbers, names, start and stop times, durations (full and T90), locations and orientations and GRBs incident angles. All information that needs to be exchanged among different system branches or software packages is kept in the database. Columns in different tables are associated with each other with ‘foreign keys’. Majority of columns in the database tables are filled automatically by the data processing software.

## 4.3 POLAR data processing software and methods

### 4.3.1 Data flow at PPDC

Raw data arrived at PPDC is processed in a pipeline as shown in in Fig. 4. All raw binary data undergoes preprocessing and technical processing and is converted to the level 0 data. All the level 0 data is processed by housekeeping monitoring or quick-look analysis software. For data with GRB candidates or in-flight calibration data, the level 0 science data is processed with the standard scientific analysis software to higher levels for calibration or further polarization studies. More details about the data processing are presented in the following sections.



**Fig. 4:** Data flow during data processing at PSI POLAR data center.

### 4.3.2 Preprocessing

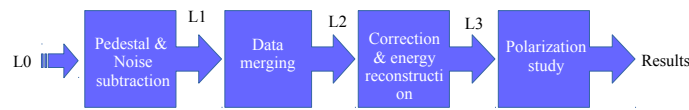
The data received from the server located in the Institute of High Energy Physics IHEP CAS in Beijing, China is considered as the raw data. Before starting automatic processing of the data an additional pre-processing of data files is needed. The size of each file and its MD5 code are calculated and compared with the size and md5 code in a file which is created by the IHEP data sever. If the data file size or its md5 code is not consistent such the file will be synchronized with the one in the IHEP server again. After such consistency checks, the information of the file (e.g. file name, file size, data acquisition start time and stop time, MD5 code, raw data directory) will be pushed into the corresponding mysql database table.

### 4.3.3 Technical processing

When both file i.e. the housekeeping data and the science data of a particular run are ready, the job management system will start their technical processing. The technical processing software will perform several automated tasks. They include e.g. checking CRC code, decoding the housekeeping, science and auxiliary data, synchronizing the event timing (time stamps), calculating counting rates, adding housekeeping data (e.g. temperatures, threshold values and high voltage values) to each event and filling the database (e.g. processing logs, input and output file names, housekeeping data summary, science data summary). The technical processing is done by a single program developed with C++ and ROOT [8]. The data decoded as described above are called the level 0 data and have the same information as the raw data. The ROOT file format is adopted to store the level 0 data due to its clever compression capability and high computing efficiency. FITS format, which is the standard data format in astronomy, is not used as it is quite inefficient in treatment of the data on the event per event basis with involvement of many obsolete structures. Each new and preprocessed run is assigned with an incremental run number. The mapping between the raw data file names and the run number of each run is also managed with the database.

### 4.3.4 Standard scientific analysis

The standard scientific analysis includes a number of predefined steps. The data sources for it are the level 0 data. As shown in Fig. 5, events in the level 0 science data are first checked to separate pedestal events and normal physical events. Pedestal events are accumulated for pedestal position calculation while for physical events both the pedestals and common noise subtractions are performed. The data with subtracted pedestals and common noise values is defined as the level 1 data. It has the same structure as the level 0 data. In the next step all hits belonging (within timestamp resolution) to the same event recorded by different modules are merged and are written to a new data files making the level 2 data set. In the next step the crosstalk corrections and energy calibrations are applied to the level 2 data with the help of the experimentally determined detector response matrices. After above processing the level 2 data set is converted to the level 3 data set. Details of the methods applied during the standard scientific analysis processing are described in Refs. [4, 5]. It should be noticed that most of the events recorded by POLAR constitutes a background that is not useful for either calibration of the instrument or polarization studies. Thus the job management system starts the standard scientific analysis processing only for candidate GRB events or the in-flight calibration data.



**Fig. 5:** POLAR standard scientific data analysis procedure.

#### 4.3.5 Calibration data analysis

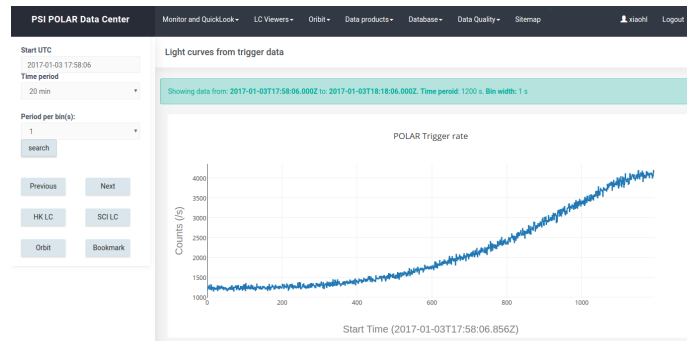
For the in-flight energy calibration of POLAR one uses four weak  $^{22}\text{Na}$  radioactive sources placed in-between instrument modules. As they emit two back-to-back annihilation photons the event selection criteria allow for proper calibration of the whole instrument. Data taken during in-flight runs are used for identification of proper calibration hits and screening of  $^{22}\text{Na}$  events. The selected events are stored in separate files in order to perform fits of the Compton edges. It results in extraction of the energy conversion factors applied to the energy spectra expressed originally in ADC channels. Moreover, selected events are also used for determination and calibration of the crosstalk. The whole information about calibration data processing is recorded in a dedicated database table.

#### 4.3.6 Monitoring of the housekeeping data and quick-look analysis of science data

Programs for the housekeeping data monitoring and a science events analysis with quick-look were developed at PPDC. They can be started by the job management system assuming that the level 0 data of a desired run are ready. The housekeeping data monitoring program checks all housekeeping variables (e.g. voltage values, temperature readouts etc.) for their values and corresponding ranges extracted from the database table. This allows for determination of the POLAR health status. Email notices are sent to all registered POLAR collaborators if any failure or malfunctioning of the instrument were detected. Moreover, plots of all housekeeping variables are created for each run. The science quick-look analysis program literary produces thousands of plots for each run showing maps of hit pattern, counting rates, pedestal and event spectra, hit non-uniformity maps and so on. All these data are necessary to properly assess the performance status of the detector. Moreover, the science quick-look analysis program also screens light-curves for GRB candidates and performs quick analysis of the confirmed GRB data. The program also allows for automatic sending of the GRB detection notices to all POLAR collaborators by the program. More than 20 GRBs were screened out automatically by the quick-look analysis program within first six months of POLAR operation.

#### 4.4 Web applications

Web applications are used as graphical interfaces to the data center. Their advantage is in a clear cross-platform usability and wide accessibility with an internet browsers. More than 50 web-based applications were developed at PPDC with modern web development techniques such as html5, AJAX, Javascript/jQuery, JSON, php and FastCGI. The web applications provide graphical



**Fig. 6:** Interactive light curve viewer developed at PSI.

interfaces to retrieve and view POLAR light curves, housekeeping monitoring and science quick analysis plots, GRB data quick-look results, space-lab orbit data and so on. They also provide interfaces to access the database tables and the data products stored on the server. As an example of the web applications, Fig. 6 shows one of the light curve viewers developed at PSI. The web applications are accessible through <http://polar.psi.ch>.

## 5. Conclusion

POLAR is a space-borne hard X-ray Compton polarimeter dedicated for precise measurements of the polarization of hard X-rays emitted by transient sources in the energy range from 50 keV to 500 keV. It was launched into space on September 15th, 2016 on-board the Chinese Space Laboratory TG-2 for up to 3 years long observation period. A dedicated POLAR data center was established at PSI (PSI PDC) in order to store and process the huge amount of data from its operation in space. The software packages and files consist of a database, data processing software, web applications and job management system. The data center has been running in a continuous manner for more than six months without any human intervention. POLAR has detected more than 50 GRBs within the first six months. About 20 of them were screened out automatically by the quick-look analysis program developed at PPDC. Polarization studies of the detected GRBs are ongoing.

## References

- [1] D. Lazzati, *New J. Phys.* 8 (8) (2006) 131.
- [2] K. Toma, T. Sakamoto, B. Zhang, et al., *Astrophys. J.* 698 (2009) 1042.
- [3] N. Produit ,et al., *Nucl. Instr. Meth. Phys. Res. A*, 550 (2005) 616 – 625.
- [4] S. Orsi,et al., *Nucl. Instr. Meth. Phys. Res. A*, 648 (1) (2011) 139 – 154.
- [5] H. Xiao,W. Hajdas, B. Wu, et al., *Astroparticle Physics* 83 (2016) 6 – 12.
- [6] H. L. Xiao, W. Hajdas, et al., 2015 IEEE NSS/MIC, 2015,1 – 5. doi:10.1109/NSSMIC.2015.7581896.
- [7] <http://polar.psi.ch>
- [8] <http://root.cern.ch>