

CSES-Limadou data processing at ASI-SSDC

Matteo Merge^{a,b,*} on behalf of the LIMADOU-HEPD Collaboration

(a complete list of authors can be found at the end of the proceedings)

^aASI - Italian Space Agency,
via del politecnico s.n.c., Rome, Italy

^bINFN - Sezione di Roma “Tor Vergata”, V. della Ricerca Scientifica 1, I-00133 Rome, Italy
E-mail: matteo.merge@roma2.infn.it

The CSES space mission, an international collaboration between China and Italy, aims at monitoring the perturbations originated by electromagnetic emissions in the ionosphere, magnetosphere and in the Van Allen radiation belts, and at investigating possible correlations with seismic events. The Italian collaboration, named LIMADOU, contributed to the mission with the realization of the High Energy Particle Detector (HEPD), an instrument realized on the basis of a long experience in developing advanced space detectors for charged and neutral particles and gamma rays – on a wide range of energies – for applications in solar physics as well as in extra-galactic astrophysics and cosmology. The CSES Satellite was launched from the Jiuquan Satellite Launch Center on February 2, 2018 and the expected mission lifetime is 5 years. Satellite data are transferred to the Institute of Crustal Dynamics (ICD) of the China Earthquake Administration (CEA) in Beijing, China. After the donwlink HEPD raw data are transferred to the Italian Ground Segment. In the IGS, HEPD raw data are processed from level0 to level2 after calibration and equalization in a high-availability processing server and stored in a high-resilience storage. In this proceeding we present a schematic of the HEPD detector data structure and the processing pipeline that has been built at the Italian Space Agency – Space Science Data Center

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter

1. Introduction

In the framework of the “Limadou fase D2” project, the design activity of the Italian Limadou Ground Segment at the Space Science Data Center (SSDC) located in the ASI HQ has been carried out.

This facility performs: storage of all CSES scientific payloads data (level2), storage and processing of HEPD raw data, storage of the HEPD Level2 data, bookkeeping of all the transferred files, server to transfer the HEPD level2 to ICD-CEA, access to the CSES data to the Italian Scientific Community (INFN, INAF, INGV etc. . .) via the web page of ASI-SSDC.

2. HEPD Detector

The High-Energy Particle Detector (HEPD2), developed by the Italian Collaboration, detects electrons, protons and light nuclei. The main objective is to measure the increase of the electron and proton fluxes due to short-time perturbations of the radiation belts caused by solar, terrestrial and anthropic phenomena. The energy range explored is 3 - 100 MeV for electrons and 30 - 200 MeV for protons.

The instrument consists of several detectors. Two planes of double-side silicon microstrip sensors placed on the top of the instrument provide the direction of the incident particle. Just below, two groups of plastic scintillators, one thin segmented, gives the trigger; they are followed by a calorimeter, constituted by 16 scintillators planes and a layer of LYSO sensors. A scintillator veto system completes the instrument. The power supply and electronics are inserted in a box placed at one side of the detector. The HEPD is contained in an aluminum-honeycomb box. Four different models have been realized and fully tested: Electrical Model (EM), Structural and Thermal Model (STM), Qualification Model (QM) and Flight Model (FM). More details can be found in [1] and [2].

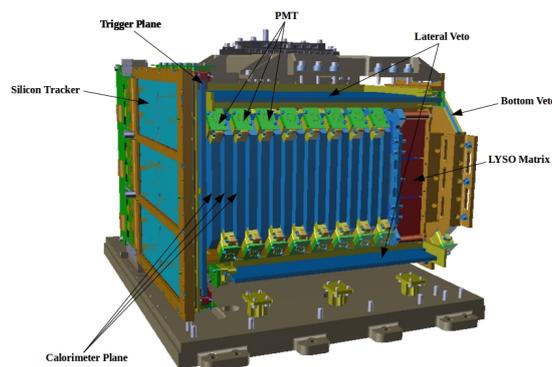


Figure 1: CAD of the HEPD Detector.

3. HEPD Data Scheme

The reconstruction occurs in three phases, which determine three different data formats, namely 0, 1 and 2, with increasing degree of abstraction.

Raw data as down-linked from CSES are converted to ROOT format (Level0) to be integrated with the processing software. Level1 data contain all detector responses after calibration and equalization. Level2 data contain higher level information, such as: estimated energy loss in the tracker, estimated energy loss in the trigger planes, estimated energy loss in the scintillator tower, estimated arrival direction, particle identification estimators.

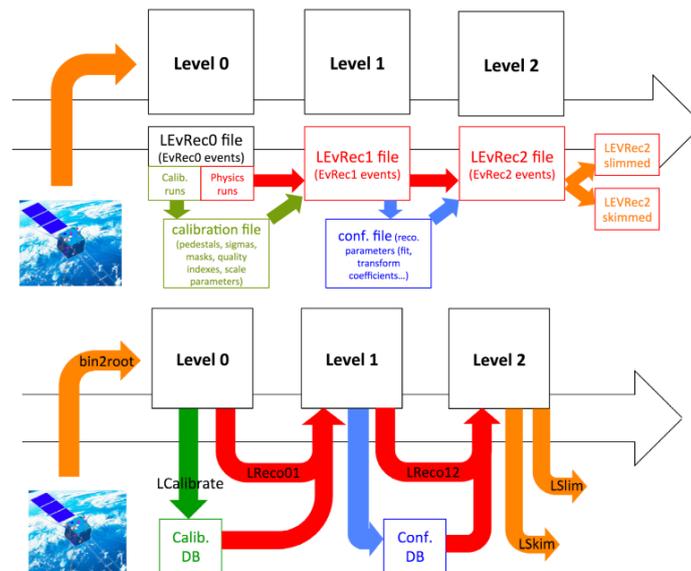


Figure 2: Scheme of the data processing levels and correlations between them.

Further abstraction from level2 is very user-dependent and it must be considered as data analysis rather than event reconstruction. The same activities of slimming the level2 structure, elaborating it or skimming the event list after reconstruction can be easily performed with custom-made software, some example of which are provided. The CSES-HEPD event reconstruction aims at providing the scientist with high-abstraction-level data, immediately usable stand-alone or in correlation with observables from other detectors or experiments. This objective has shapen the event reconstruction software, which is built on criteria of flexibility, scalability and output adaptability.

4. Downlink and Pre-processing

The CSES-Limadou mission's ground segment consists of two elements: the Chinese Ground Segment at ICD-CEA and the Italian Ground Segment at SSDC (ASI, Italy). Major elements of the China Ground Segment are: the Data Receiving System, the Data Processing System and the Data Application System. Quasi real-time data are transferred to the Italian processing ASI-SSDC. As soon as new HEPD raw data are available they are automatically processed by a fully automated pipeline.

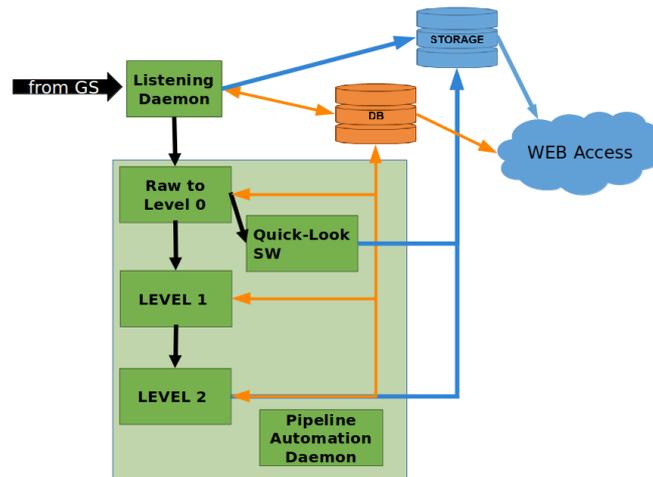


Figure 3: Scheme of the processing pipeline daemons

5. Processing pipeline

The processing pipeline has been developed as several layers of software with an interface to a processing database and to the storage of the infrastructure. A MySQL Database, with JSON capability, is used to keep track of the different stages of processing and for data bookkeeping.

The pipeline before the start of the commissioning phase has been extensively tested with HEPD flight model ground data (cosmic muons). After flight when real data became available the software and database were adjusted to fit with data size and peculiarity of the flight data. The first part up to level0 has not changed much from the commissioning phase, while the level0-to-level1 and level1-to-level2 are constantly adapted to the development of the processing software and fitted to the payload data acquisition configurations. The high modularity of the pipeline allow to achieve asynchronously single steps of the processing.

A Database is used keep track of the different stages of processing, to store calibration data. The main task is the bookkeeping of downloaded and transferred data, most of the software of the pipeline has a direct interaction with the database either to insert a record for a new file but also during the processing to check for new files at each intermediate step. The database is also used for the timing of processing scripts, at each step in the pipeline after the raw file processing, the scripts wait for new entries in the database to start the processing of any given file. MySQL 5.7.22 has been chosen as the database framework to ensure compatibility with JSON data format, for the well know behavior and for availability of large end exhaustive documentation.

As soon as a new file is added to a specified folder in the processing server, it is immediately registered in the database and a copy of the raw file is sent to the processing chain while the original file is transferred to a permanent backup area in the storage.

5.1 Data Ingestion: hepd_rawd listening daemon

This software, written in python using an iNotify wrapper, is responsible to watch the upload folder where the raw data are copied, manage the bookkeeping of the raw files and control the storage.

As soon as a new file is added to the watched folder, the software fills an entry in the `raw_table` inside the database with basic information of the files such as arrival time, filename, filesize etc.. A copy of the file is put in the storage folder and a copy of the raw file is pushed to a working folder where the other software of the pipeline is listening for new files, as soon as the copy is made in the working folder the original file is removed. If any anomaly is encountered in the process the corrupted file is moved to a given folder for further inspection.

5.2 Level0 Processing: Dataparser, Bin2root

The level0 processing is handled with a python script ran in daemon mode, the process waits in background and monitors the database for new raw files entries. As soon as a raw data file, containing data of one half-orbit, is added in the working folder, the `dataparser` executable is run. This program performs checks on data transmission goodness between the satellite and the ground stations. HEPD output data are sent to the satellite via the RS-422 bus in frames of fixed length (4124 bytes), each one with an header at the beginning of the frame and a checksum at the end. The `dataparser` script is responsible to check the sequence and the integrity of each frame, calculating the checksum and comparing it with the one calculated onboard the satellite and is written at the end of the frame. It also checks if the first frame of the file has index equal to 1 and if the file ends correctly. See [3] for more details on the on-board software.

The output will result in a binary file containing only the HEPD scientific information where all the headers and checksums of the frames have been removed. After the execution of the `dataparser`, the `raw_file` database table is filled with the information of the total number of RS422 frames and the number of valid RS422 frames (with checksum ok) in the processed file. A more verbose version, that saves a detailed log file, can be executed in case of particularly corrupted raw files in order to try to recover some portion of the data. The binary file produced with the `dataparser` is then the input of the `bin2root` script. This software, written in C++, performs minimal operation on the raw data and is just used to produce the level0 files in the root format.

A root file is produced for each run; a raw file relative to an half-orbit typically contains 5-15 runs depending on the run duration. Calibration runs, as well as special configuration runs, are saved with a different name respect to the standard acquisition runs. After the `bin2root` software is run, the fields `processed` and `processing_time` of the database `raw_file` table are updated and the `level0_file` table is filled. The produced root files are saved in the storage folder and all the temporary file created in the process are removed.

5.3 Level1 Processing: Reco01

Level1 processing is handled with a python script `hepd_l0to11`, ran in daemon mode. which checks the `level1_files` table for new files. The daemon `hepd_l0to11` looks for not processed L0 files in the database every few seconds and if a not processed entry is found the processing starts. First, with a system of queries to the database, the script pairs each acquisition run with its calibration run, namely the last calibration file before that run (among the not processed ones). The number of events in the acquisition run is checked and required to be at least 1.

Then the script produces a configuration file, called steering file. Afterwards the steering file is produced and it is given to the `Reco01` executable. The software is responsible for the reconstruction of the acquisition runs from Level0 to Level1. This procedure consist in the following

steps: clustering the silicon tracker signal, correcting the silicon tracker signal for the settling time and subtracting pedestals from the photo-multiplier's responses and equalizing them. After the processing the `hepd_l0to11` script updates the fields in the database table of the L0 files, flagging as “processed” the runs used during the processing. Furthermore it adds entries in the L1 table, such as: filename, path, `processing_time`, ids of the files used in the reconstruction of the new reconstructed L1s.

5.4 Level2 Processing

Level2 files contain calibrated and equalized data for each detector and projected orbital and asset information for each event. Also for analysis purposes several on-line cuts are stored and several booleans for the event selection are also included in the level2 on event-by-event base. The level2 production script, `hepd_l1to12` runs a parallelized production of the level2 files using a batch system to handle several files processed at the same time. The production of a level2 runs the following step:

- Detector Calibration and Equalization
- Machine Learning Algorithms for event reconstruction
- Propagation of TLE to the particle events
- Propagation of Geomagnetic Coordinates to the particle events
- Production of the finalized Level2 ROOT file

Further details on the Machine Learning techniques can be found in [4]. The level2 production software is constantly updated to include new variables according to the needs of the data analysis group.

References

- [1] S. Bartocci, R. Battiston, W.J. Burger, D. Campana, L. Carfora, G. Castellini et al., *Galactic cosmic-ray hydrogen spectra in the 40–250 MeV range measured by the high-energy particle detector (HEPD) on board the CSES-01 satellite between 2018 and 2020*, *The Astrophysical Journal* **901** (2020) 8.
- [2] P. Picozza, R. Battiston, G. Ambrosi, S. Bartocci, L. Basara, W.J. Burger et al., *Scientific Goals and In-orbit Performance of the High-energy Particle Detector on Board the CSES*, *Astrophys. J. Suppl.* **243** (2019) 16.
- [3] A. Sotgiu, C. De Donato, C. Fornaro, S. Tassa, M. Scannavini, D. Iannaccio et al., *Control and data acquisition software of the high-energy particle detector on board the china seismo-electromagnetic satellite space mission*, *Software: Practice and Experience* (2020) 1 [<https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.2947>].
- [4] F.M. Follega, R. Iuppa and M.C. on behalf of the LIMADOU-HEPD Collaboration, *Deep learning based event reconstruction for limadou HEPD*, in *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*, Sissa Medialab, July, 2021, DOI.

Full Authors List: LIMADOU-HEPD Collaboration

S. Bartocci¹, R. Battiston^{2,3}, F. Benotto⁴, S. Beolè^{4,5}, W.J. Burger^{3,6}, D. Campana⁷, G. Castellini⁸, P. Cipollone¹, S. Coli⁴, L. Conti^{1,9}, A. Contin^{10,11}, M. Cristoforetti⁶, L. De Cilladi^{4,5}, C. De Donato¹, C. De Santis¹, F.M. Follega^{2,3}, G. Gebbia^{2,3}, R. Iuppa^{2,3}, M. Lolli¹¹, N. Marcelli^{1,12}, M. Martucci^{1,12}, G. Masciantonio¹, M. Mergè^{1,†}, C. Neubuser³, F. Nozzoli³, A. Oliva¹¹, G. Osteria⁷, L. Pacini¹³, F. Palma^{1,†}, F. Palmonari^{10,11}, A. Parmentier¹, F. Perfetto⁷, P. Picozza^{1,12}, M. Piersanti¹⁴, M. Pozzato¹¹, E. Ricci^{2,3}, M. Ricci¹⁵, S.B. Ricciarini⁸, Z. Sahnoun¹¹, V. Scotti^{7,16}, A. Sotgiu^{1,12}, R. Sparvoli^{1,12}, V. Vitale¹, S. Zoffoli¹⁷ and P. Zuccon^{2,3}

¹ INFN-Sezione di Roma “Tor Vergata”, V. della Ricerca Scientifica 1, I-00133 Rome, Italy;

² University of Trento, V. Sommarive 14, I-38123 Povo (Trento), Italy;

³ INFN-TIFPA, V. Sommarive 14, I-38123 Povo (Trento), Italy;

⁴ INFN-Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy;

⁵ University of Torino, Via P. Giuria 1, I-10125 Torino, Italy;

⁶ Fondazione Bruno Kessler, V. Sommarive 18, I-38123 Povo (Trento), Italy;

⁷ INFN-Sezione di Napoli, V. Cintia, I-80126 Naples, Italy;

⁸ IFAC-CNR, V. Madonna del Piano 10, I-50019 Sesto Fiorentino (Florence), Italy;

⁹ Uninettuno University, C.so V. Emanuele II 39, I-00186 Rome, Italy;

¹⁰ University of Bologna, V.le C. Berti Pichat 6/2, I-40127 Bologna, Italy;

¹¹ INFN-Sezione di Bologna, V.le C. Berti Pichat 6/2, I-40127 Bologna, Italy;

¹² University of Rome “Tor Vergata”, V. della Ricerca Scientifica 1, I-00133 Rome, Italy;

¹³ INFN-Sezione di Firenze, V. Sansone 1, I-50019 Sesto Fiorentino (Florence), Italy;

¹⁴ INAF-IAPS, V. Fosso del Cavaliere 100, I-00133 Rome, Italy;

¹⁵ INFN-LNF, V. E. Fermi 54, I-00044 Frascati (Rome), Italy;

¹⁶ University of Naples “Federico II”, V. Cintia 21, I-80126 Naples, Italy;

¹⁷ Italian Space Agency, V. del Politecnico, I-00133 Rome, Italy;

† At ASI Space Science Data Center (SSDC) also, V. del Politecnico, I-00133 Rome, Italy.