

The Informationized Application of On-orbit Data Processing of the Dark Matter Particle Explorer



Jin Chang, Jingjing Zang, and Liang Liu

Abstract DAMPE is the first satellite-based observatory of China targeting on astronomical objects. During its on-orbit observation, over 6 billion of high energy cosmic rays (CRs) have been detected and analyzed, providing several hundred of terabyte (TB) of data for various scientific purposes. Such rich observations have largely benefited the research of electrons (positrons), protons and gammas within the high energy CRs. In this paper, we would take a deep insight into DAMPE. Firstly, the structure and features of its data processing software DAMPESW is introduced. Then constructions of DAMPE infrastructures is explained that are built on the demand of high data processing performance. Finally, in hope of providing reference for other similar projects in the future, we summarize the informationized application of data processing on DAMPE.

Keywords DAMPE · On orbit operation · Cosmic rays · DAMPESW

1 Introduction

1.1 Overview of DAMPE

DARK Matter Particle Explorer (DAMPE) is a space satellite aiming at high energy CRs observations. Benefited by its high resolution and wide energy range of detecting high energy CRs and gamma ray emissions, DAMPE is used in the indirect search of dark matters through measuring or constraining physical properties of dark matter (DM) particles, including mass, annihilation cross section, decay lifetime and so on. Meanwhile, the detection of TeV electrons and heavy nucleuses would provide valuable clues for investigating the origin of CRs, and the observation of high energy gamma rays would both discover and exam the quantum gravity effects with high precision. In conclusion, the scientific goals of DAMPE are:

J. Chang (✉) · J. Zang · L. Liu
Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, China
e-mail: chang@pmo.ac.cn

- (1) The first priority of DAMPE is the search of dark matter particles. The high spatial resolution and wide dynamic energy range allow DAMPE to search and study DM particles indirectly by detecting high energy electrons and gamma rays. These observations could not only provide measurements or precise constraints of the properties of DM particles (such as mass, annihilation cross section, decay lifetime), but can also limit the spatial distribution of them. Therefore, the observation of DAMPE is extremely valuable for making breakthroughs on DM researches.
- (2) Investigating the origin of high energy electrons and nucleuses in CRs by detecting TeV particles.
- (3) Obtaining achievements on gamma ray astronomy by detecting high energy gamma ray emissions.

1.2 Summary of DAMPE Payloads

In order to achieve scientific goals described in last section, 4 detectors are integrated within DAMPE as sketch shown in Fig. 1. From top to bottom they are the plastic scintillator detector (PSD), the silicon tungsten tracker (STK), the BGO calorimeter (BGO), and the neutron detector (NUD) made of plastic scintillator units, respectively.

- (1) The plastic scintillator detector
The main scientific tasks of PSD are: determine the direction of incident particles, identify photons and electrons, and discriminate high energy heavy ions

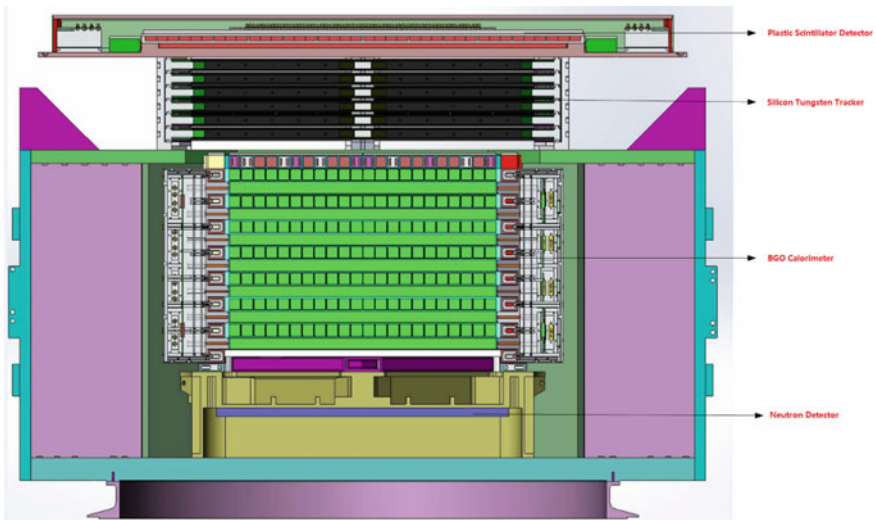


Fig. 1 The sketch of DAMPE payloads

categories ($Z = 1\sim 20$). The PSD, has effective detecting area of $820 \times 820 \text{ mm}^2$, is consisted of two layers, in which a total of 82 plastic scintillator units are placed perpendicularly in X and Y directions. Among them 78 units have identical sizes of $884 \times 28 \times 10 \text{ mm}^2$, and the other 4 have sizes of $884 \times 25 \times 10 \text{ mm}^2$. For next processing steps, photomultiplier tubes (PMTs) are set connecting to both ends of those units to convert optical signals to electrical ones.

(2) The silicon tungsten tracker

The STK is made of silicon micro-strip sub-detectors which are very sensitive to the position of incident particles. The main scientific tasks of STK are measure the direction of incident particles, identify electrons (charged particles) and gamma rays (non-charged particles), as well as measure high energy nuclides ($Z = 1\sim 26$). Specifically, there are 6 layers of silicon micro-strip detectors in STK, each layer is consisted of two sub-layers of XY paralleled silicon micro-strip detectors to locate position (X and Y). The area of STK is $80 \times 80 \text{ cm}^2$. Three tungsten plates are placed as conversion medium of gamma rays to meet the configuration requirement of STK. In order to identify charged and non-charged particles, especially gamma rays and electrons, tungsten plate do not place on the top of the first layer of silicon micro-strip detector. Instead, they are placed between the first and second layers, the second and third layers, and the third and fourth layers. Note that the outputs of STK are charge signals.

(3) The BGO calorimeter

The BGO is a fully absorption electromagnetic calorimeter. Its main scientific task is detecting CRs, especially the energy of high energy electrons and gamma rays (5 GeV~10 TeV). In the meantime, it can identify electron and hadrons based on different horizontal and vertical growths of showers, so that the high energy hadron (mainly protons) background can be excluded. Structurally, the BGO calorimeter is made of 308 BGO crystals and its area is approximately $60 \times 60 \text{ cm}^2$. There are 7 major layers within the it and each of them are mad of two XY paralleled sub-layers, located by X and Y axis. Within each sub-layer, there are 22 identical detector units (BGO crystal) which have sizes of $2.5 \text{ cm} \times 2.5 \text{ cm} \times 60 \text{ cm}$. PMTs are used connecting to the both ends of units to convert optical signals to electrical ones.

(4) The neutron detector

NUD is a comprehensive detector of neutron moderation and detection. Its main scientific task is detecting secondary neutrons generated in the interaction between CR hadrons (mostly protons) and its top materials. Based on energy deposition of these neutrons, NUD could identify incident particles independently, which further supplement the identification of protons and electrons of BGO. Specifically, NUD is consisted of 4 independent squares cut from a $693 \text{ mm} \times 693 \text{ mm}$ plastic scintillator detector, which is 1-cm thick and boron-doped (BC454, produced by Saint Gobain corporation). A corner has been removed for each square in order to couple the PMT, from which a compete detecting surface could be read out.

During the operation period of DAMPE, there are two observation modes for every detector listed above, calibration mode and observation mode. According to the initial design of DAMPE, the calibration mode of detectors is divided into three work settings: pedestal calibration, Digital-to-analog converter (DAC) calibration, and Multi-band Imaging Photometer calibration (MIPs). The purpose of observation mode is detecting CRs, such as high energy electrons, gamma rays, nuclides and so on. Correspond to CRs types chosen to measure, different triggering settings should be applied in order to rule out background events (mostly protons) as many as possible. During flight, detectors would keep repeating following steps: calibration → optimize parameters (high pressure, trigger delay, trigger threshold, etc.) → observation → calibration.

1.3 On Orbit Operation of DAMPE

At the present, the on-orbit operation of DAMPE is accomplished by the cooperation of ground supporting system from National Space Science Center, Chinese academy of Science, scientific application system from Purple Mountain Observatory, Chinese academy of Science, as well as the satellite telemetry and control of many related units. From 2015-12-20, the very day of receiving the first frame of scientific data, to 2019-08-05, the cumulative data receiving from DAMPE are 20208 orbits, which including 1324 days of on-orbit operation and 7 times of full sky survey. A total of 6.67 billion of high energy particles are detected and processed, generating 1B data 19.92 TB, 1F data 12.53 TB, 2A data 112.43 TB. Shown in Figs. 2 and 3 are the total events and average events per day, respectively. In the past three years of

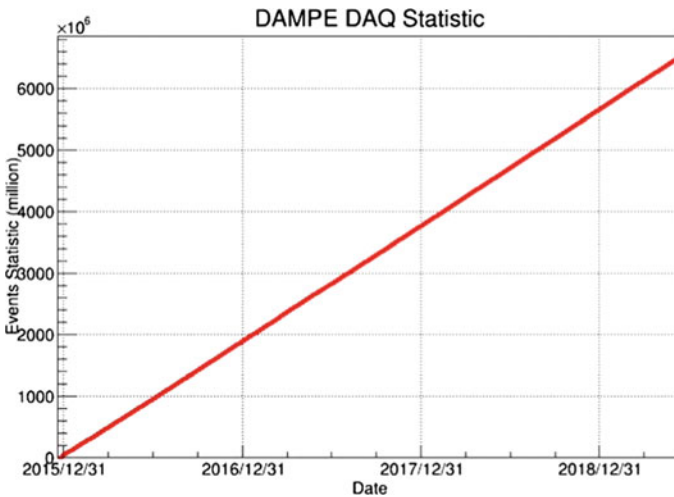


Fig. 2 Accumulated event data of DAMPE

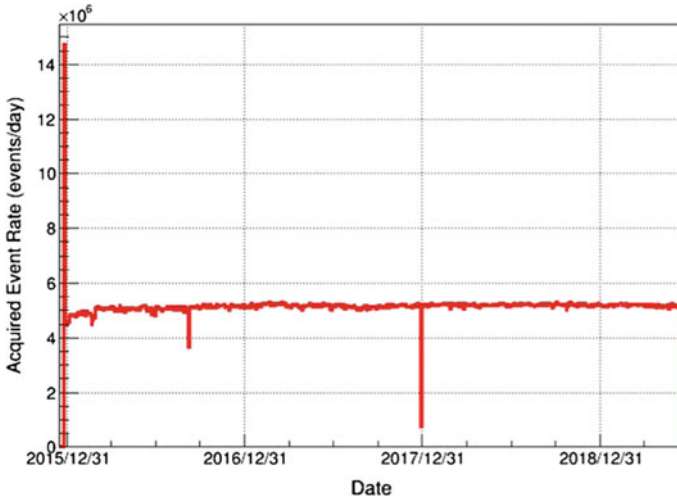


Fig. 3 The event rate of DAMPE

on-orbit operation, multiple issues appeared (the reset of payload data manager, bad cluster in flash storage of data tube, single event effect on high voltage power supply cabinet, etc.) and are adequately disposed. Compared to the period when satellite was just launched, no new bad tracks of electronic channels (70 thousand and more) are founded, and the whole satellite is still maintaining in a good condition, suggesting that design of satellite platform and payloads are very reliable.

At this moment, the most important task for DAMPE is continuously cumulate the observation data and release them to scientific groups for making scientific achievements.

1.4 Present Research Status

Currently, there are a number of international satellites also aiming at CRs detections, such as the Large Area Telescope instruments on board on Fermi Gamma-ray Space Telescope (Fermi-LAT), the Alpha Magnetic Spectrometer mounted on International Space Station (AMS02), as well as the CALorimetric Electron Telescope (CALET). Adopted to different properties and scientific goals of these projects, different data processing pipelines and software are developed.¹ For instance, the FermiTools is developed for the analysis of Fermi-LAT data (<https://github.com/fermi-lat/Fermitools-conda/>) and it has been continuously updated to meet new requirements, the latest release is v11r5p3. AMS02 also has its own offline data analysis software AMSsoft. However, the version of this software is unknown since it is not open to

¹<http://fermi.gsfc.nasa.gov/ssc/data/analysis/software>.

the publics. Therefore, toward DAMPE observation, it is necessary to develop a data analysis software for scientific purposes.

1.5 The Advancement of DAMPESW

Borrowing data analysis experience from other similar satellite projects, multiple concepts, such as Kernel, Algorithm, Service and Event, are introduced into DAMPE Software (DAMPESW) that make it not only functional on the basic data processing, but also accessible, as a unified processing platform, for all users. Therefore, software development can be made between users without interference, making it possible to be continuously update.

The core and mainly body of DAMPESW are coded by C++. Boost.Python is employed linking C++ with Python so that the two languages could identify and coordinate with each other. Python, as a well-known programming language, is used for configuring user-customized algorithms to help them work properly. Meanwhile, the scripting language Bash is also used for the environmental configuration of software.

So far, DAMPESW has integrated three programming languages, C++, Python, and Bash, and it can also cooperate with ROOT, which is a data analysis software commonly used in particle physics. Consequently, DAMPESW is very convenient for new users that can process data easily and quickly.

To summarize, all the analysis threads of DAMPE observation data are unified in DAMPESW, meaning that it is convenient for independent users to compare their outcomes and improve their analytical techniques through communications. In other word, such a software would guarantee that outcomes are reliable, and DAMPESW could improve the efficiency scientific outcomes as well.

2 Processing of Scientific Data

2.1 The Scientific Application System of DAMPE

In engineering, the scientific application system of DAMPE is responsible for its operation and scientific data processing.

The system is also responsible for making observation plan, monitoring the status of payloads on board the satellite, generating advanced data products, calibrating payload, and organizing scientific research.

In order to ensure the achievement of scientific goals and the accomplishment of the overall task, the scientific application system is divided into three sub-systems (considering the function and workflow). They are the scientific operation sub-system, the scientific data & user managing sub-system, and the advanced data products processing sub-system. Each of those sub-systems are built to implement

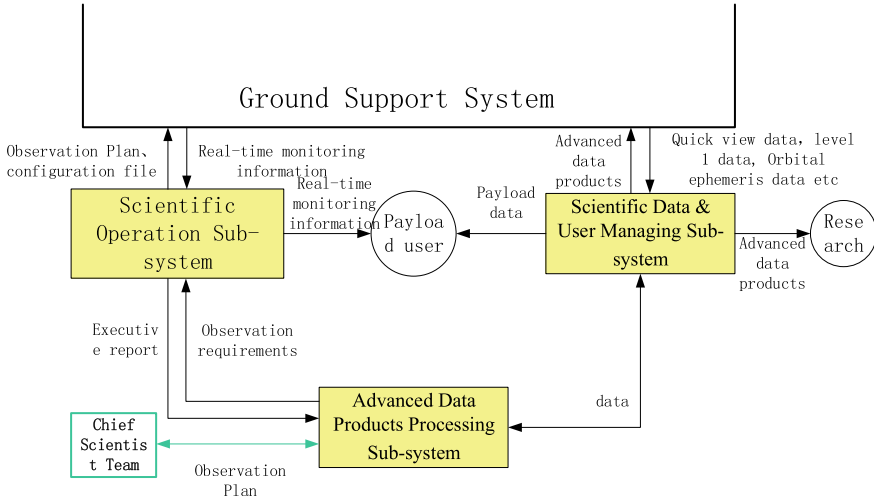


Fig. 4 Composition of the scientific application system

different functions corresponding to different stages. Among them, the scientific operation sub-system is responsible for building an effective mechanism for the operation and management of the satellite, including making short or medium-long term observation plan for payloads, monitoring the status of payloads, and determining the uplink parameters of satellite. The scientific data & user managing sub-system is mainly responsible for the management of data, including data intactness (guarantee data is intact during transmission), data transformation, data classification, storage of huge volume data, data calling and long-term preservation, calibrating and unifying time of scientific data, data archiving, backup and releasing, organizing scientific research and so on. Meanwhile, the advanced data products processing sub-system would take charge for generate advanced data products, analysis the background, identify particles, and build data base.

The internal I/O between these three sub-systems and the ground supporting system are present in Fig. 4.

2.2 Scientific Data Processing Software DAMPESW

2.2.1 The Facing Challenges of DAMPE Scientific Data

DAMPE is a multifunctional high energy particle detector with multiple component. Its main body is made of four sub-detectors and each of them is consist of a large number of sensitive detecting units. Same incident particle interacting with detector have various information that are recorded in every unit. From those units, the read-out information would be integrated, forming into the so-called Event at

data level. Therefore, the data processing of users basically is to combine information and pick up interested events by adopting appropriate computing method. Such method, the so-called the Algorithm, will go through every event and perform identical operation to each information successively. Generally, in order to obtain a specific physical property of an incident particle, multiple algorithms are required. Also, at the different stages of data processing, algorithms should be adjusted based on different level of processing. Hence, to achieve one single goal, different algorithms need to be connected in some particular orders, and the interaction of outcomes between different algorithms would occurs. For instance, in the case of tracing particles, the track could be obtained in the cluster of STK through Kalman filtering, in which shower axis is needed (as the seed of track) that is reconstructed by BGO calorimeter. The reconstruction then requires the knowledge of the shower center of gravity of BGO calorimeters within each layer, in which the center of gravity of shower could only be calculated on the basis of knowing the energy of each BGO crystal. Therefore, various algorithms are needed to reconstruct the track of incident particles and they should be connected in a certain sequence.

The data processing of events demand for many physical algorithms which is developed by numerous of scientific groups based on different requirements. During processing, the execute of algorithms need auxiliaries from various functional modules which may also be demanded in other algorithms. Thus, the functional module, which is called as Service, should be callable and open to algorithms. Taking I/O for instance, data input and output are used in every algorithm and a I/O module (I/O Service) would be useful. Note that I/O service will only take charge for I/O. The only difference between service and algorithm is that service don not analysis and process the actual data of events, it existed only for providing assistant to algorithms in order to reduce pressure of algorithm developments.

DAMPESW is developed exactly for the requirements described above, the detailed functions are as follows:

1. Provide a unified development platform

First of all, interfaces (virtual functions) are offered for developers to access the core of DAMPESW. Therefore, developers only need to focus on the specific functions of those interfaces without concerning other things, like the order of software drivers and the management of data (interaction and storage). They would run by the core.

Secondly, all users' data processing would under the same software framework, which is convenient for communicating and sharing between users. This is good for the division and cooperation of developing software.

2. Provide unified processing mechanism

All algorithms are executed by user defined python scripts (a text file) for flow control. Users need to load their algorithms and services into the core, then the software would be executed by following logics:

Step 0. Configurations (formulating algorithms, inputs and outputs, etc.).

Step 1. Initialization (DmpCore:Initialize(), like customizing histogram, define variables, etc.)

- Step 2. Execution (DmpCore::Run(), go through every events).
 - Step 3. Cleaning and exiting current work (DmpCore::Finalize(), writing data to output files)
3. Provide flexible scalability
Depending on specific requirements, users could create their own algorithms, event classes, and services.

2.2.2 Components of DAMPESW

DAMPESW is consist of 5 main functional modules and 5 auxiliary modules. Those 5 main functional modules are RawDataConversion module, Reconstruction module, Calibration module, Simulation module, and Visualization module. The 5 auxiliary modules are Kernel module, Analysis module, Event module, EventOrbitSimulation & OrbitSimulation module, and Geometry module, shown in Fig. 5.

The functions of main functional modules are as follows:

RawDataConversion module: Converting raw data (binary) to TTree format could be recognized by ROOT, meaning that convert 1A/1B scientific data product to 1E/1F data product.

Reconstruction module: Converting the read-out values of electronic channels of 1E/1F data product to physical objects (energy, direction, particle types, etc.) through complicated calibration algorithm. Meaning of converting scientific data product of 1E/1F to 2Q/2A.

Calibration module: Analyzing calibration 1E/1F data products and generating calibration coefficients.

Simulation module: Mainly used for simulating the response of every sensitive detecting units inside detectors after particles (in the radiational space environment) entered detectors.

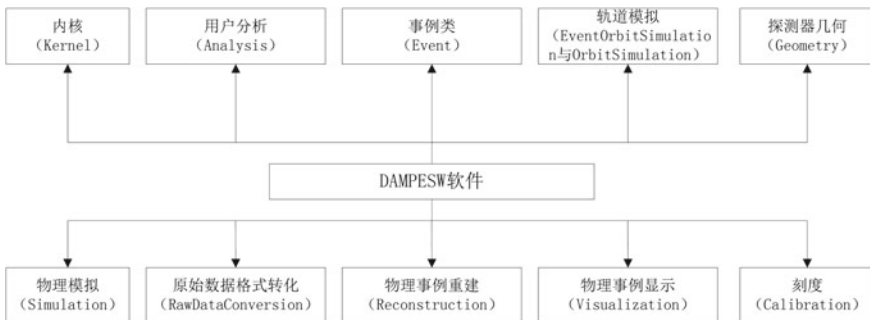


Fig. 5 Software structure of DAMPESW

Visualization module: Read 2Q/2A scientific data product and exhibit energies of detective units on the 3D geometric model intuitively, using colors to represent the energy level.

The functions of auxiliary modules are as follows:

Kernel module: Kernel module is the core of DAMPESW. It is responsible for executing and driving of all other modules, and provide unified I/O format and data interaction mechanism (between modules for all modules in the software). It would also provide base class for the extension of services and algorithms.

Analysis module: Integrated module of user-defined algorithms.

Event module: Definitive module for event data format.

EventOrbitSimulation & OrbitSimulation module: Module that simulate the environment of satellite orbit.

Geometry module: Module that storage the geometry data of detectors.

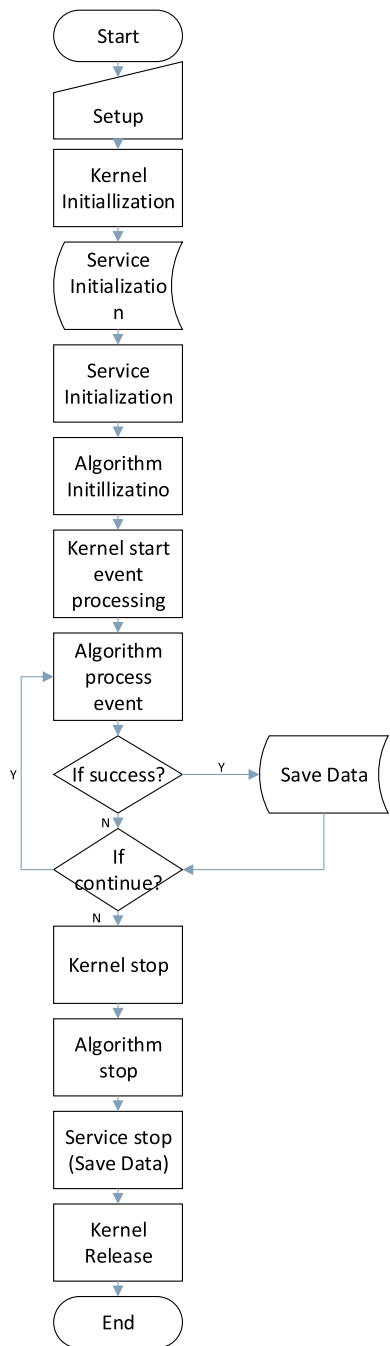
2.2.3 DAMPESW Control Flow

The software control flow is presented in Fig. 6. Before actual running of the software, the configuration of services and algorithms need be conducted at first. Then the initializing shall start on the sequence of core to services to algorithms. After completing services and algorithms initializing, the event processing procedure will be started by the core. Algorithms will analysis an event and determine whether the event is processed successfully. If so, the flow will continue after this event been saved. If not, the flow will continue without saving the event. Next, if the conditions (already finished all event processing or achieve the manual termination conditions) are satisfied, the event processing procedure will be ended and the flow will go to next stage. Otherwise the flow will go into a loop and prepare for processing the next event. The core will be terminated after the end of event processing procedure. Then services will be ended after algorithms terminated. By the time of services stopped, all data will be writing into hard disk. And after all algorithms and services are stopped successfully, memory will be freed up and flow will be finished.

2.2.4 Release of DAMPESW

The development of DAMPESW is independent. The speed of software version updates will be proportional to the number of developers involved. From the practical perspective, engineering management is not an appropriate approach for scientific data analysis. However, in order to guarantee the quality and progress of development, we have taken a series of measures. Firstly, the development is undertaken by various of international cooperation groups. Online group meeting is arranged weekly for task assignment, progress discussion, and schedule making. Secondly, version control repository is built on the computer platform of Purple Mountain Observatory, who is responsible for the construction of scientific application system. As demonstrate

Fig. 6 Software control flow



in Fig. 7, development codes from every cooperation groups would be submitted to DAMPESWSVN version control repository and SVN would manages version info of codes automatically. So that the version control issue can be solved and development quality can be ensured.

| Name | Size | Rev | Age | Author | Last Change |
|--------------------|------|------|-----------|--------|--|
| ./ | | | | | |
| DmpSoftware-1-0-0 | | 417 | 6 years | andrii | |
| DmpSoftware-1-1-0 | | 684 | 6 years | andrii | Last release of non-GDML non-EventClass? DAMPE software. |
| DmpSoftware-2-0-0 | | 945 | 5 years | andrii | First stable release of GDML & EventClass? based DAMPE offline software |
| DmpSoftware-2-1-0 | | 999 | 5 years | andrii | Last release of old-framework DAMPE offline software. Next releases will ... |
| DmpSoftware-3-0-0 | | 1107 | 5 years | andrii | First release of the new framework |
| DmpSoftware-3-0-1 | | 1134 | 5 years | andrii | STK raw data conversion based on trbread2.c script |
| DmpSoftware-3-1-0 | | 1261 | 5 years | andrii | Last stable release before introducing the combined stable kernel (by Chi ... |
| DmpSoftware-3-2-0 | | 1264 | 5 years | andrii | Combined stable Kernel from Chi and Andrii. |
| DmpSoftware-3-2-1 | | 1271 | 5 years | andrii | First stable version of global Raw Data Conversion (provided by Jingjing? ... |
| DmpSoftware-3-2-2 | | 1302 | 5 years | andrii | RDC for shanghai cosmos from Jingjing?, only STK part is modified by ... |
| DmpSoftware-3-2-3 | | 1400 | 5 years | andrii | Frozen release between moving to combined RDC for the beam test |
| DmpSoftware-4-0-0 | | 1471 | 5 years | andrii | Frozen release of software for v1 beam-test data production |
| DmpSoftware-4-0-1 | | 1488 | 5 years | andrii | bug fixes for the v1 production. This is the last release before moving to ... |
| DmpSoftware-4-1-0 | | 1508 | 5 years | andrii | v2 beam test data production |
| DmpSoftware-4-1-1 | | 1564 | 5 years | andrii | v2 beam test data production (bugfix for the STK raw data conversion) |
| DmpSoftware-4-1-2 | | 1669 | 4 years | andrii | Last release before getting to 2015 March beam test processing |
| DmpSoftware-4-2-0 | | 1803 | 4 years | andrii | v0 reprocessing of March 2015 beam-test data |
| DmpSoftware-4-3-0 | | 1856 | 4 years | andrii | stk-calb, stk-recon, stk-compare -- tools for quick STK monitoring, aimed ... |
| DmpSoftware-4-3-1 | | 1859 | 4 years | andrii | Fully functional stk-compare, stk-calb and stk-recon tools |
| DmpSoftware-4-3-2 | | 1868 | 4 years | andrii | First reprocessing of April 2015 cosmos |
| DmpSoftware-4-3-3 | | 1865 | 4 years | andrii | v0 reprocessing of March 2015 beam-test data (STK RDC bug fixed): request ... |
| DmpSoftware-4-3-4 | | 1882 | 4 years | andrii | stk-* commands tested by Valentina |
| DmpSoftware-4-4-0 | | 1989 | 4 years | andrii | v0 reprocessing of May 2015 Shanghai cosmos data of DAMPE |
| DmpSoftware-4-4-1 | | 2004 | 4 years | andrii | v0 reprocessing of June 2015 beam-test data |
| DmpSoftware-4-4-2 | | 2016 | 4 years | andrii | V1 reprocessing of March 2015 beam-test data: fixed PSD reconstruction ... |
| DmpSoftware-4-4-3 | | 2022 | 4 years | vgallo | REC3 reprocessing of cosmos April 2015 clusterseed 4 |
| DmpSoftware-4-4-4 | | 2042 | 4 years | andrii | v0 reprocessing of June 2015 beam-test data with DAMPE+AMS merging |
| DmpSoftware-4-4-5 | | 2048 | 4 years | andrii | reprocessing of DAMPE integration-test data, June 25, 2015 |
| DmpSoftware-4-4-6 | | 2152 | 4 years | andrii | Calibration of magnet for beam test, June 2015 |
| DmpSoftware-4-4-7 | | 2154 | 4 years | andrii | Production: satellite inflight simulation data, August 25-28, 2015 |
| DmpSoftware-4-4-8 | | 2206 | 4 years | andrii | bug fix in the STK raw mode decoding reported by Xin |
| DmpSoftware-4-4-9 | | 2231 | 4 years | andrii | Last release before moving to the Kernel update: enable multiple input ... |
| DmpSoftware-4-5-0 | | 2251 | 4 years | andrii | bug fix in STK DLD package unpacking (problem of duplication in .txt ... |
| DmpSoftware-4-5-1 | | 2254 | 4 years | andrii | STK package (frame) size information added to STK metadata |
| DmpSoftware-4-5-2 | | 2299 | 4 years | andrii | processing1 of aging test data in Geneva, November 2015 |
| DmpSoftware-4-5-3 | | 2316 | 4 years | andrii | processing1 of aging test data in Geneva, November 2015 (bug fixes: ... |
| DmpSoftware-4-5-4 | | 2383 | 4 years | andrii | v0 processing, beam test November 2015 |
| DmpSoftware-4-5-5 | | 2712 | 4 years | andrii | Last stable release before merging rep1 and rep in a single Orbit ... |
| DmpSoftware-4-5-6 | | 2788 | 4 years | andrii | Last release before moving to the new STK calo-seeded tracking: 1D+1D ... |
| DmpSoftware-5-0-0 | | 2870 | 4 years | pmo | svn r2869, DMPESW v5.0.0 |
| DmpSoftware-5-1-0 | | 3111 | 3 years | pmo | release r3110 as gamma version 5.1.0 |
| DmpSoftware-5-1-1 | | 3152 | 3 years | pmo | Tag: put software version number to 5.1.1 |
| DmpSoftware-5-1-2 | | 3220 | 3 years | pmo | Tag: 5.1.2 |
| DmpSoftware-5-1-3 | | 3419 | 3 years | pmo | Tag: version 5.1.3 |
| DmpSoftware-5-1-4 | | 3448 | 3 years | pmo | Tag: v5.1.4 |
| DmpSoftware-5-1-5 | | 3537 | 3 years | pmo | tag: v5.1.5 |
| DmpSoftware-5-2-0 | | 3835 | 3 years | pmo | Tag: 5.2.0 |
| DmpSoftware-5-3-0 | | 4425 | 3 years | pmo | Tag: 5.3.0 |
| DmpSoftware-5-3-1 | | 4543 | 3 years | pmo | Tag: 5.3.1 |
| DmpSoftware-5-3-2 | | 4735 | 3 years | pmo | Tag: 5.3.2 |
| DmpSoftware-5-3-3 | | 5084 | 3 years | pmo | Tag: 5.3.3 |
| DmpSoftware-5-4-0 | | 5141 | 3 years | pmo | Tag: 5.4.0 |
| DmpSoftware-5-4-1 | | 5804 | 2 years | pmo | Tag: 5.4.1 |
| DmpSoftware-5-4-2 | | 6047 | 2 years | pmo | Cherry-pick : DmpSoftware?-BugFix/5-4@c6046 : Modify minor bug in ... |
| DmpSoftware-6-0-0 | | 6417 | 2 years | pmo | Tag: 6.0.0 |
| DmpSoftware-6-0-1 | | 7674 | 16 months | pmo | Tag: 6.0.1 |
| DmpSoftware-6-0-2 | | 7866 | 14 months | pmo | Tag: 6.0.2 |
| DmpSoftware-6-0-3 | | 7957 | 13 months | pmo | Tag: 6.0.3 |
| DmpSoftware-6-0-4 | | 8214 | 10 months | pmo | Tag: 6.0.4 |
| DmpSoftware-6-0-10 | | 8517 | 3 months | pmo | Tag : 6.0.10 |
| ReleaseNote | | 8519 | 3 months | pmo | Add : Release note 6.0.10 |

Fig. 7 SVN version control repository of DAMPESW releasing

2.3 Satellite Data Processing Flow

The flow diagram of DAMPE data is presented in Fig. 8.

After DAMPESW receiving the order of data processing, the RawDataConversion module will read-in file from the order which contains 1A/1B data product, and unpack it to verify the content. The verification including remove head and tail of FITS data packet, separate different types of data (observation data, pedestal calibration data, Digital-to-analog converter calibration (DAC) data, MIPs calibration data and pedestal update data), and convert all of them to TTree format (1E/1F scientific data product) that could be recognized by ROOT. Observation data will be reconstructed to physical objects by Reconstruction module, others will be converted to calibration coefficient files through a series of complicated calculating by Calibration module.

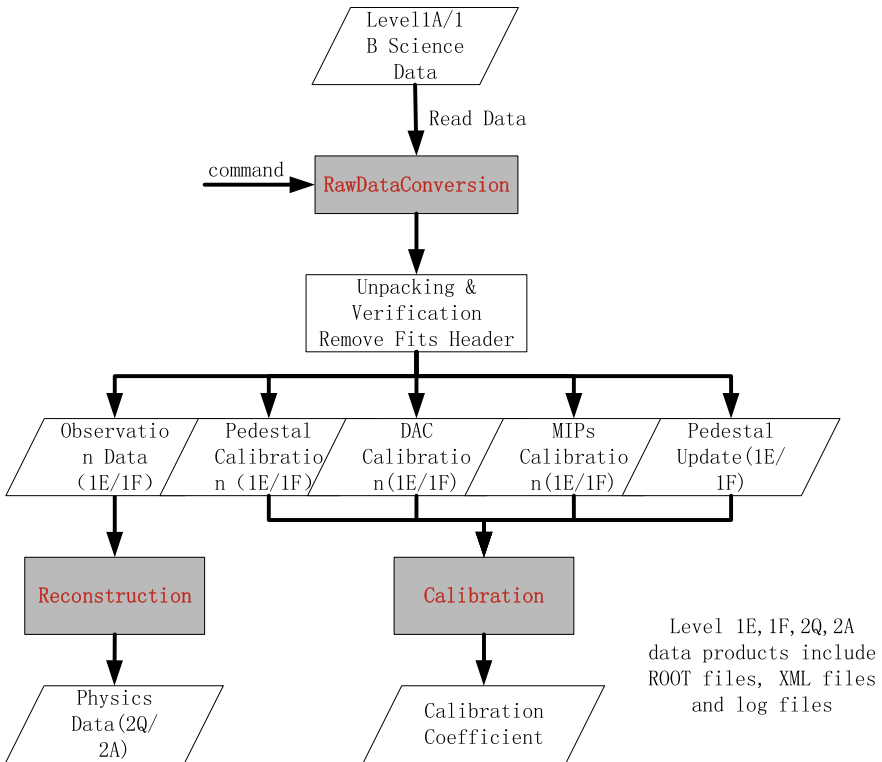


Fig. 8 Workflow of satellite data processing

2.3.1 Workflow of RawDataConversion

As shown in Fig. 9, the first step of flow is read-in 1A/1B data product. Then head and tail of FITs data packet will be removed. After that the inputting data will be split. Data packet of one single event will be read-out into a series of caches (BGO, PSD, NUD, STK, and TRG), and module will judge whether event in STK cache is intact. If intact, event will be filled in TTree. Otherwise module will continue to read next event. In the case of event been successfully filled in TTree, module will

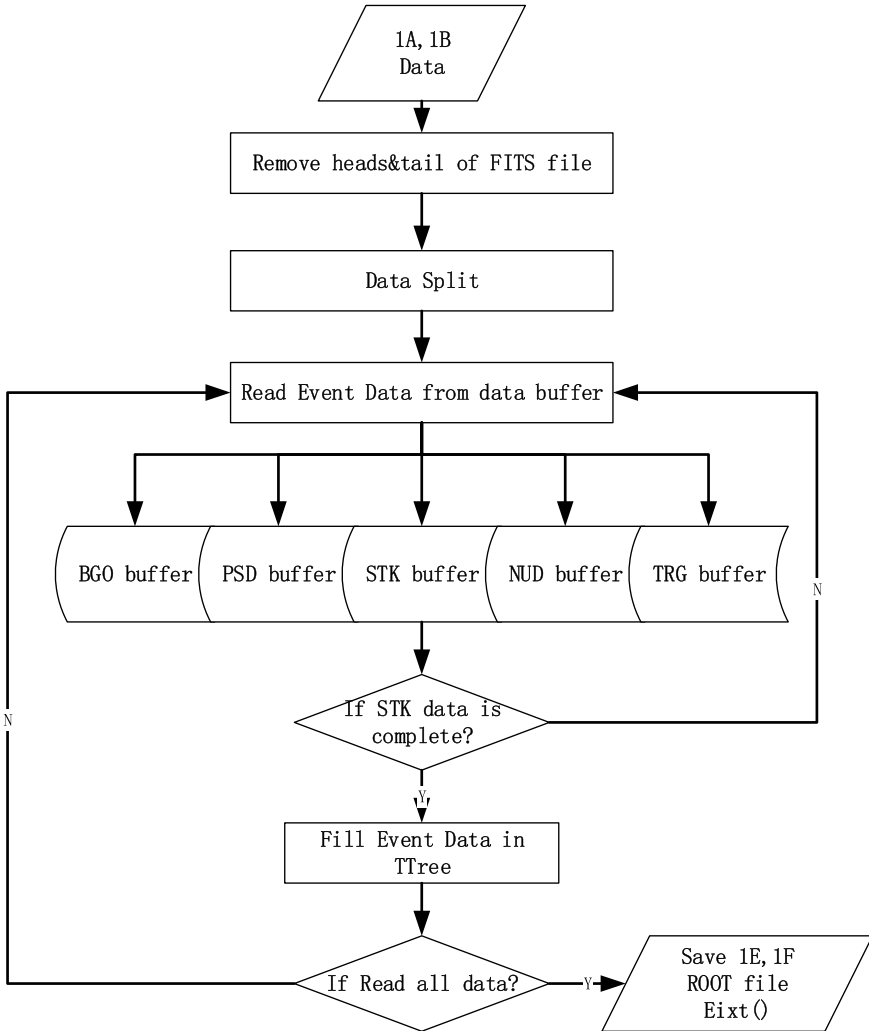


Fig. 9 Workflow of raw data conversion

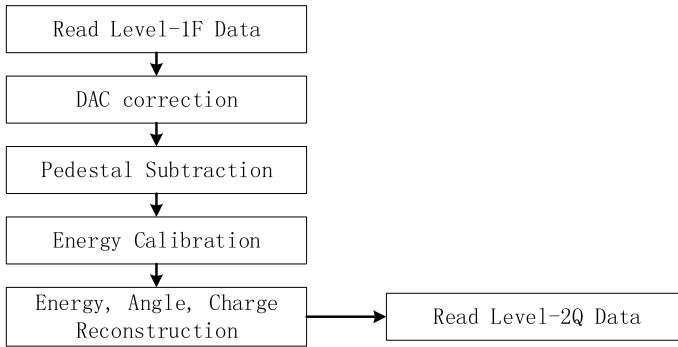


Fig. 10 Workflow of reconstruction

judge whether the generated ROOT file should be terminated. If not, module will keep processing the next event, otherwise TTree file will be saved and exited.

2.3.2 Workflow of Reconstruction

Reconstruct is started with the input of 1E/1F data product, which should be saved in ROOT format produced in other modules. Generally, reconstruct is a series of procedures that convert observational signals to physical signals on the help of calibration coefficients. Figure 10 is the flow diagram of the process.

The first step is DAC correction. Next, pedestal should be deducted and appropriate dynodes will be selected and read out. Through dynode ration transform, measurement of one single detective unit should be converted to the readings of ADC channels of the eighth dynode. Hence, energy of a single detective unit can be obtained by normalizing peak value of MIP. Based on that, advanced energy, direction, and charge are reconstructed.

2.3.3 Workflow of Calibration

Calorimeter is a type of device that absorb energy of incident particle by making it interact with the device. So that part of its energy would be released in a recordable way. To fully understand the performance of this device, the conversion factor between detected signal and incident particle energy is crucial. And the process of obtaining it is the so-called calibration. There are four major parts of calibration, pedestal calibration, MIPs calibration, dynode relative coefficient calibration,

and Digital-to-analog converter calibration. All of them have inputs of 1E/1F data product in ROOT format, and the output should be calibration coefficients.

Pedestal calibration:

In the case of no inputting signals from electronic front-end chip, the output should be the readings of ADC channels, which would yield a normal distribution. Hence, the calibration coefficients should be the arithmetic mean and variance of the normal distribution.

MIPs calibration:

If detector is working under sampling mode, the trigger shall be set at MIPs mode that ADC spectrum of MIPs response can be obtained. In which case calibration coefficients should be the peak position and width of the most probable distribution derived from fitting the spectrum with Gaussian-convoluted Landau function.

Dynode relative coefficient calibration:

The detected signal would be output through the second, fifth, and eighth dynodes of PMTs. After electronic processing, the readings of ADC channels should yield a linear correlation. Therefore, the calibration coefficient should be the gradient and intercept derived from linear fitting.

Digital-to-analog converter calibration:

After digital-to-analog (DA) conversion, orders, which is digital, send from ground base are converted to analog signals which would act as electronic front-end input. Between the output, readings of ADC channels, and the amplitudes of the input, an approximate linear relation existed. Its deviation from linear reflect the non-linear feature of electronic.

2.3.4 Workflow of Data Simulation

By assuming various circumstances that detector has, physical simulation should refer to the process of computing emitted signals from the interaction between incident particles and the detector. The purpose of physical simulation is to create simulated data that are identical to observed data, so that it can be analyzed by DAMPESW following the same method used in observed data.

Figure 11 is the flow diagram of generating simulated data.

Multiple DAMPESW modules are responsible for generating simulated data, which are Geometry, Simulation, OrbitSimulation, EventOrbitSimulation modules.

The physical simulation and electronic simulation are integrated in Simulation module. These two simulations are performing simultaneously for BGO, PSD, and NUD (electronic simulation of STK would be introduced latterly). Meaning that energy deposit would be convert to readings of ADC channels by electronic response function as soon as it was produced by simulating interactions of incident particles

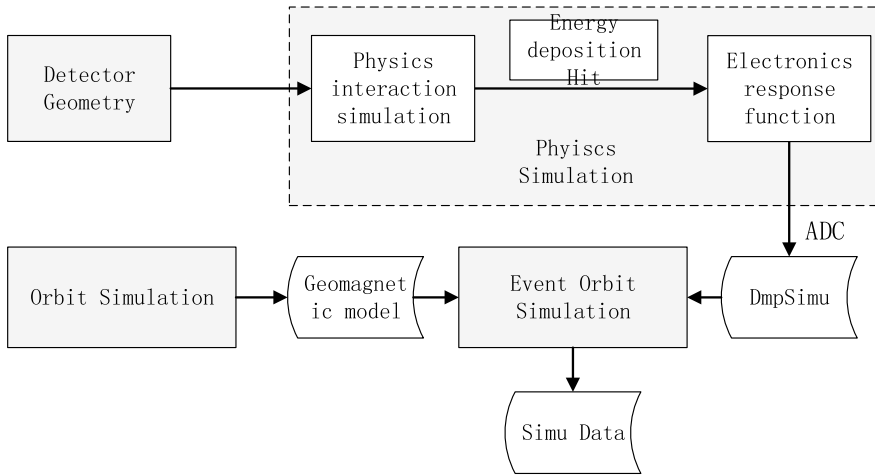


Fig. 11 Workflow of data simulation

and detector. Module will produce root data file (hereafter called DmpSimu) that have same format of data level 2.

The simulation of satellite orbiting environment is divided into two steps: Orbit-Simulation module is responsible for computing the satellite orbit and magnetic field distribution at corresponding altitude based on ephemeris. Root data file (OrbitSimu) will be produced in this step.

Figure 12 shows maps of magnetic fields and vertical geomagnetism cutoff at 500 km altitude.

Figure 13 shows the satellite orbit and experienced geomagnetism during one day EventOrbitSimulation module takes DmpSimu and OrbitSimu as input. It would decide whether the incident particles are originated from outer space using a method called “backward tracing”. If true, a time will be given for each particle and new root data file (EventOrbitSimu) will be generated, which can be used for estimating event rate and trigger rate (with respect to various observation mode) at any satellite positions. The so-called “backward tracing” is the method of judging incident particle position, that is to numerically solve the equation of motion of particle inside magnetic field by inverse its charge and direction. In fact, we only interested in particles came from outer space so that particles from earth are excluded during event rate calculating.

As mentioned before, STK working different from BGO, PSD, and NUD so that its electronic simulation needs to be done separately. Firstly, during physical simulation, every energy deposit on silicon micro-strip should be recorded. After finished one event, whether a silicon micro-strip should be read-out or suspended is depending on the previous record. Then energy will be reassigned to other neighboring read-out silicon micro-strips in order to calculate the number of electron-hole pairs can be produce in such an energy. Finally, depending on the monitoring of respond curve

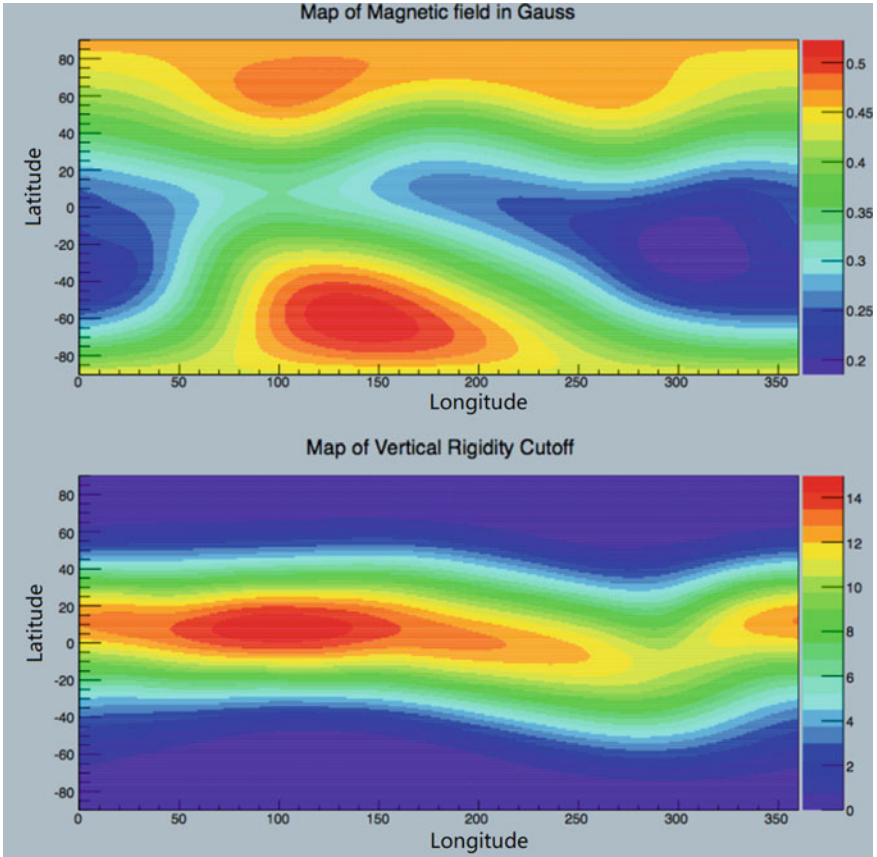


Fig. 12 The map of magnetic fields and vertical geomagnetism cutoff at 500 km altitude

of VA chip within silicon micro-strip (shown in Fig. 14), energy would be converted to ADC readings.

2.4 DAMPESW Workflow

The whole workflow of processing 1A/1B data products (by the advanced data product managing subsystem) into advanced 1E/1F and 2A/2Q data products is complete on data processing server. During that, all data processing procedures are driven by python scripts. Thus, the whole process requires a continuously use of different python scripts for different stages, presented in Fig. 15.

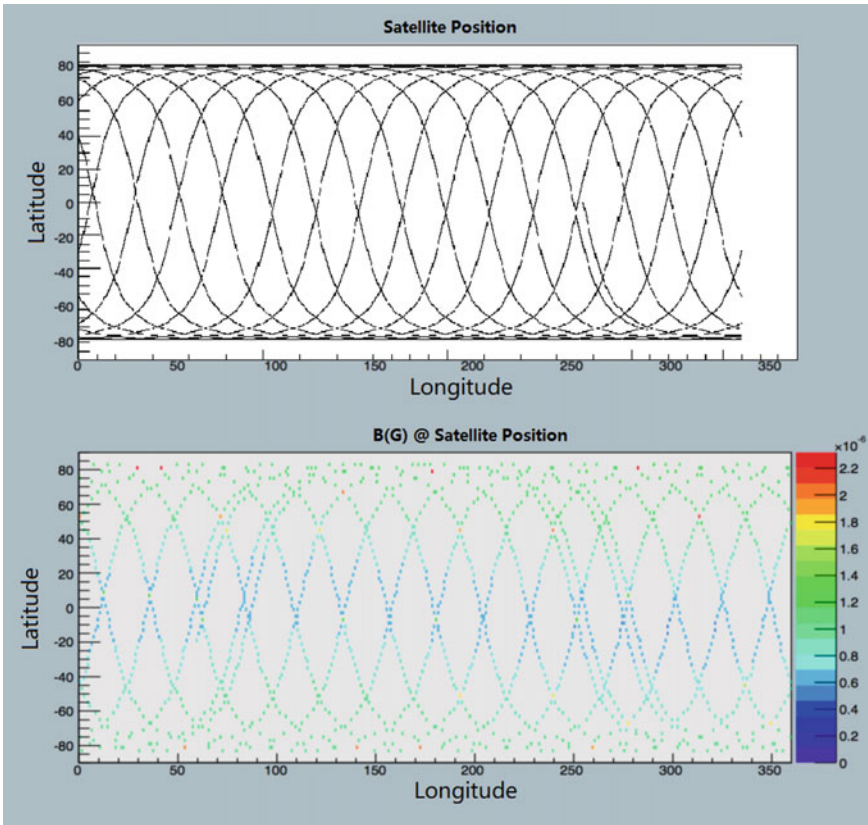


Fig. 13 The satellite orbit and experienced geomagnetism during one day

- (1) Converting FITS data file into another format: Script JobOpt_Fits2Frd.py take file name as input. Its function is to read the head information of FITS file and convert file to frd format. The new file will be saved in a temporary directory.
- (2) RawDataSplit. The original data file of remote observation contains data under many observation modes. Processing of those data requires different procedures that correspond to observation modes. Frd file created in the last step need to be split to multiple files. Each of them should contain data of a single mode with respect to each time period, and they should be marked accordingly (OBS for observation data packet, PED for pedestal calibration data packet, DAC for DAC calibration data packet, MIP for MIPs calibration data packet, UPD for pedestal update calibration packet).
- (3) RDC. Decoding and converting the original data packet of remote observation for each observation modes and produce 1E/1F data product. During its operation, the first step is checking the original data, including CRC check, length check, data packet serial number continuity check, data packet trigger number

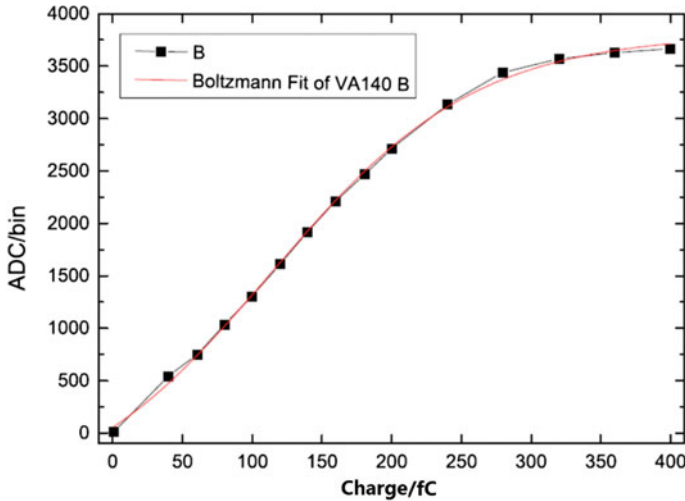


Fig. 14 The response curve of VA chip

continuity check, etc. And the error info will be saved in a file having same name and directory of the data packet. Then trigger data and sub-detectors data will be read-out and written into corresponding EventClass, which will be saved as ROOT file with name consistent with the naming conventions. The key information during data generation will be automatically preserved in a description of advanced data product for scientific data management and user sub-system storage and analysis.

- (4) Calibration. Obtaining calibration coefficient for calibration files (MIP, DAC, PED) through a series of complicated computation. The results will be saved in corresponding directory with a specific format.
- (5) Reconstruction. For data after RDC, it is necessary to reconstruct all events and carry on follow-up scientific analysis. For 1E/1F data product marked as OBS, the most recently used calibration coefficients will be loaded to reconstruct incident particle energies and tracks in every detective unit. The result will be saved as ROOT file in the corresponding directory, named under 2A/2Q naming conventions.

The flows described above would be executed by shell script automatically, including load global variables required for system operation, generate file names based on corresponding naming conventions, manage directories that storage various of temporary files and calibration coefficient files, automatic judge observational mode of data, output error info generating during operation, etc. Again, all those procedures will be automatically executed that all needed is to run shell scripts in data processing server with appropriate input parameters.

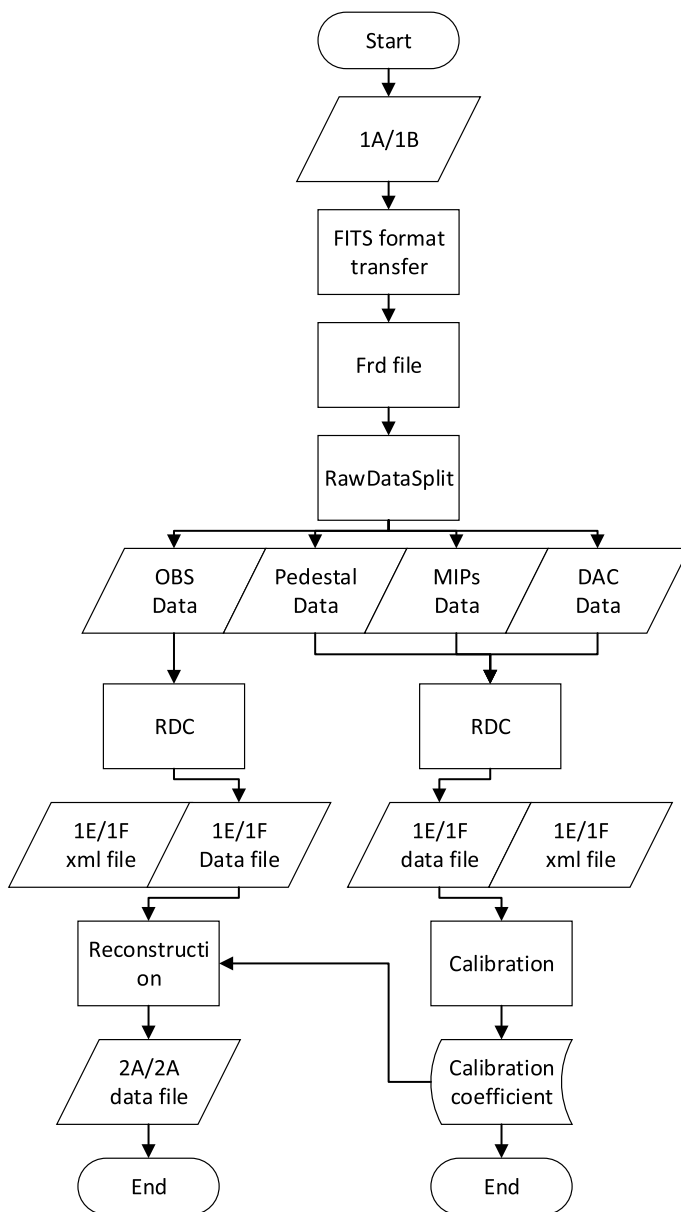


Fig. 15 Workflow of advanced data product sub-system

3 Hardware Components and Network Structure

In order to meet the needs of processing satellite scientific data, a platform is built in DAMPE scientific application system. There several sections of the platform, calculation section, main storage section, hot data storage section, user data storage section, data base section, data backup section, data interaction and network security section, and monitoring hall section, showing on Fig. 16. To be more specific, the calculation section is made of 128 high density blade servers, which have more than 3000 CPUs. The main storage section adopts EMCX410 of high-performance

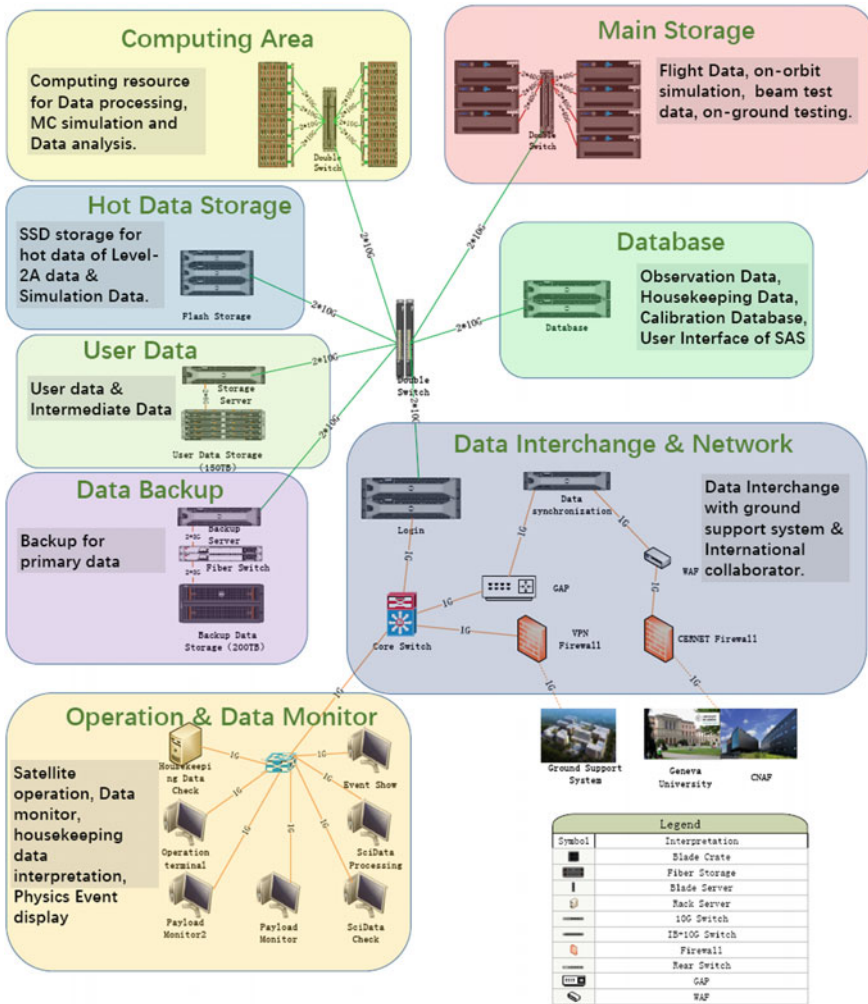


Fig. 16 Sketch of DAMPE scientific data processing platform

parallel storage that have in total 7 storage nodes with a volume of 450 TB, of which the highest data throughput reaches 40 Gbps. It would be used for storage all scientific data products relative to the satellite. The hot data storage section uses flash memory for data storage which has a volume of 175 TB that used for store 2A data product which frequently accessed by users. Large capacity disk storage is used in user data storage section, which store user's personal data generated during data processing. Its volume is 130 TB. Similar to user data storage section, the data backup section (used for backing up primary data advanced data, and simulation data) also adopts disk storage which has 200 TB. The data interaction and network security section are connected to HuaiRou space science center and European cooperation groups through 100 Mbps of private VPN network and 100 Mbps of science and technology network, respectively. While the network security and isolation are guaranteed by firewall, WAF, and gateways. Finally, the monitoring hall is consisted of a series of minor terminals, taking charge of the operation and monitoring data processing procedures.

4 Conclusion

DAMPE is the first Chinese satellite-based observatory built for astronomic purpose. The way to handle its observation data is unprecedented. Internationally, the raw data processing of other same type of satellites are exclusive and hard to be used on processing DAMPE data because of the different instrument structure. In this paper, therefore, we systematically introduced DAMPE data processing flow, methods, software, and hardware based on the perspective of software requirements. We hope that this paper would provide technical reference for future projects.



Jin Chang was born in July, 1966. He graduated from University of Science and Technology of China with master's degree in June, 1992, and his doctor's degree is received in July, 2006. At the present, he is the director of Purple Mountain Observatory, the deputy director of the National Astronomical Observatories (concurrently), chairman of the Key Laboratory of Dark Matter and Space Astronomy, director of the School of Astronomy and Space Science of the University of Science and Technology of China, chairman of research department of dark matter and space astronomy. He obtained the title of academicians of Chinese Academy of Sciences in 2019.

Jin Chang have long been engaged in the detections and researches of gamma ray emission and high energy charged particles that come from outer space, and is one of the pioneers for studying space astronomy in China. He created and developed a new technique of detecting high energy cosmic ray electrons. This technique was used in a long-term balloon experiment (ATIC experiment) of America aiming at detecting cosmic rays in Antarctica. Based on the technique, he led the

DAMPE (also called as “WuKong”) project as the chair scientist. The “WuKong” was successfully launched in December 17, 2019 and it is the very first satellite of China used for astronomical research. Also, this launch marked that China is leading the world at several key techniques and breakthrough on detecting the spectrum of cosmic ray electrons and protons. Furthermore, in order to response for the needs of the major national strategic, Jin Chang and his team successfully developed gamma ray spectrometers for ShenZhou II, Chang’e I and II. His major awards are: the second prize of National Science and technology progress in 2004 (4th to accomplish), the second prize of National Natural Science Award in 2012 (accomplished independent), special prize of National Science and Technology Progress Award in 2012(40th to accomplish), National innovation award in 2017, Ho Leung Ho Lee Foundation Award for Scientific and Technological Progress Award in 2018 (astronomy), Zhang Yuzhe award of China Astronomical Association in 2018, outstanding scientific and technological achievement award of Chinese Academy of Sciences in 2018, ten advances in Chinese Science in 2018, science and technology award of China Academy of Space Science (first), etc.