DAMPE data processing and analysis at CNAF

G. Ambrosi[1], D. D'Urso[1,2,*], M. Duranti[1,3], F. Gargano[4,5], S. Zimmer[6]

[1] INFN, Sezione di Perugia, I-06100 Perugia, Italy
[2] ASDC, I-00133 Roma, Italy
[3] Università di Perugia, I-06100 Perugia, Italy
[4] INFN, Sezione di Bari, I-70125 Bari, Italy
[5] Università di Bari, I-70125 Bari, Italy
[6] University of Geneva, Departement de physique nuclaire et corpusculaire (DPNC), CH-1211, Genève 4, Switzerland
DAMPE experiment `http://dpnc.unige.ch/dampe/`, `http://dampe.pg.infn.it`

E-mail: `domenico.durso@pg.infn.it`

Abstract. DAMPE (DArk Matter Particle Explorer) is one of the five satellite missions in the framework of the Strategic Pioneer Research Program in Space Science of the Chinese Academy of Sciences (CAS). DAMPE has been launched the 17 December 2015 at 08:12 Beijing time into a sun-synchronous orbit at the altitude of 500 km. The satellite is equipped with a powerful space telescope for high energy gamma-ray, electron and cosmic ray detection. The main scientific objective of DAMPE is to measure electrons and photons with much higher energy resolution and energy reach than achievable with existing space experiments in order to identify possible Dark Matter signatures. It has also great potential in advancing the understanding of the origin and propagation mechanism of high energy cosmic rays, as well as may enable new discoveries in high energy gamma-ray astronomy.

1. Introduction

DAMPE is a powerful space telescope for high energy gamma-ray, electron and cosmic ray detection. In Fig. 1 a scheme of the DAMPE telescope is shown. The top, the plastic scintillator strip detector (Psd) consists of two layers of scintillating plastic strips that serve as anti-coincidence detector, followed by a silicon-tungsten tracker-converter (STK), which is made of 6 tracking layers. Each tracking layer consists of two layers of single-sided silicon strip detectors measuring the two orthogonal views perpendicular to the pointing direction of the apparatus. Three layers of Tungsten plates with thickness of 1 mm are inserted in front of tracking layer 2, 3 and 4 to promote photon conversion into electron-positron pairs. The STK is followed by an imaging calorimeter of about 31 radiation lengths thickness, made up of 14 layers of Bismuth Germanium Oxide (BGO) bars which are placed in a hodoscopic arrangement. The total thickness of the BGO and the STK correspond to about 33 radiation lengths, making it the deepest calorimeter ever used in space. Finally, in order to detect delayed neutron resulting from hadron showers and to improve the electron/proton separation power, a neutron detector (NUD) is placed just below the calorimeter. The NUD consists of 16, 1 cm thick, boron-doped plastic scintillator plates of 19.5×19.5 cm^2 large, each read out by a photomultiplier.

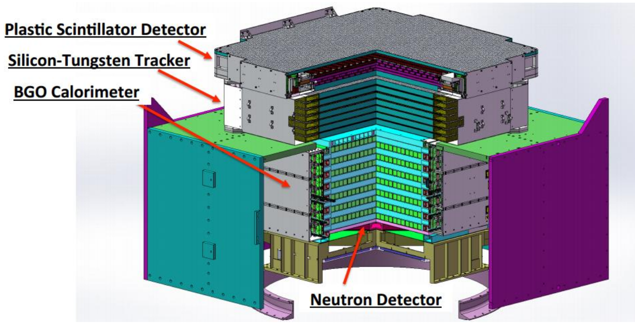

Figure 1. DAMPE telescope scheme:a double layer of the plastic scintillator strip detector (PSD); the silicon-tungsten tracker-converter (STK) made of 6 tracking double layers; the imaging calorimeter with about 31 radiation lengths thickness, made of 14 layers of Bismuth Germanium Oxide (BGO) bars in a hodoscopic arrangement and finally the neutron detector (NUD) placed just below the calorimeter.

The primary scientific goal of DAMPE is to measure electrons and photons with much higher energy resolution and energy reach than achievable with existing space experiments. This will help to identify possible Dark Matter signatures but also may advance our understanding of the origin and propagation mechanisms of high energy cosmic rays and possibly lead to new discoveries in high energy gamma-ray astronomy.

DAMPE was designed to have an unprecedented sensitivity and energy reach for electrons, photons and cosmic rays (proton and heavy ions). For electrons and photons, the detection range is 2 GeV-10 TeV, with an energy resolution of about 1.5% at 100 GeV. For cosmic rays, the detection range is 100 GeV-100 TeV, with an energy resolution better than 40% at 800 GeV. The geometrical factor is about 0.3 m^2 sr for electrons and photons, and about 0.2 m^2 sr for cosmic rays. The expected angular resolution is $0.1°$ at 100 GeV.

2. DAMPE Computing Model and Computing Facilities

As Chinese satellite, DAMPE data are collected via the Chinese space communication system and transmitted to the China National Space Administration (CNSA) center in Beijing. From Beijing data are then transmitted to the Purple Mountain Observatory (PMO) in Nanjing, where they are processed and reconstructed.

2.1. Data production

PMO is the deputed center for DAMPE data production. Data are collected 4 times per day, each time the DAMPE satellite is passing over Chinese ground stations (almost every 6 hours). Once transferred to PMO, binary data, downloaded from the satellite, are processed to produce a stream of raw data in ROOT [1] format (1B data stream, $\sim$ 15 GB/day), and a second stream that include the orbital and slow control information (1F data stream, $\sim$ 15GB/day). The 1B and 1F streams are used to derive calibration files for the different subdetectors ($\sim$ 400MB/day). Finally, data are reconstructed using the DAMPE official reconstruction code, and the so-called 2A data stream (ROOT files, $\sim$ 70 GB/day) is produced. The total amount of data volume produced per day is $\sim$ 100 GB.

2.2. Monte Carlo Production

Analysis of DAMPE data requires large amounts of Monte Carlo simulation, to fully understand detector capabilities, measurement limits and systematics. In order to facilitate easy workflow handling and management and also enable efficient monitoring of a large number of batch jobs in various states, a NoSQL metadata database using MongoDB [2] is being developed with a prototype currently running at the Physics Department of Geneva University. Database access is provided through a web-frontend and command tools based on the flask-web toolkit [3] with a client-backend of cron scripts that will run on the selected computing farm. The design and implementation of this workflow system is heavily influenced by the implementation of the Fermi-LAT data processing pipeline [4] and the DIRAC computing framework [5].

Fundamentally, each production consists of repeated batch submissions with minimal differences across different batch jobs. To this end, a MC shifter creates a *Job* which contains the basic information regarding the type of task (MC generation, digitization, reconstruction, etc.) and each *Job* contains a number of *JobInstances*. Web-frontend and command line tools provide means to manage and access these Jobs and JobInstances. On the computing farm a set of cron jobs establish connections with the web-frontend (and by extension the database) and request JobInstances of type *New*. Once a new instance is pulled from the database, a batch job is created and submitted. Upon submission the batch ID is reported back to the database and used for further tracking. This cron job is supplemented by a second job which acts as a watchdog that monitors running jobs in the batch farm and requests jobs to be terminated if they exceed memory or cpu requirements and otherwise reports running jobs back to the database. The termination of jobs by the watchdog ensures that failed job status are reported back with maximum transparency to the exact reason of their failure. To this extent, the database utilizes a combination of major and minor status. The major status is mapped to one or more of the status that are provided by the batch system while the minor status is set at application level.

Once submitted, each batch job continuously reports its status to the database through outgoing http requests. To that end, computing nodes need to allow for outgoing internet access. Each batch job implements a workflow where in- and output data transfers are being performed (and their return codes are reported) as well as the actual running of the payload of a job (which is defined in the metadata description of the job). Dependencies on productions are implemented at the framework level and jobs are only submitted once dependencies are satisfied. While log files are stored on the computing farm, an external rsync script ensures permanent storage on a server in Geneva.

This framework is currently under test on the Geneva computing farm and at CNAF.

3. CNAF contribution

The CNAF computing center is the mirror of DAMPE data outside China and will be the main data center for Monte Carlo production. To keep available DAMPE data to the European DAMPE Collaboration, DAMPE data are transferred from PMO to CNAF using the gridFTP [6] protocol. Every time a new 1B, 1F or 2A data files are available at PMO, they are copied to a server at CNAF, gridftp-plain-virgo.cr.cnaf.infn.it, to the DAMPE storage area. The connection to China is passing through the Orientplus [7] link of the Géant Consortium [8]. The data transfer rate is currently limited by the connection of the PMO to the China Education and Research Network (CERNET), that has a bandwidth of 100 Mb/s.

On the user interface at CNAF, every hour a copy of each stream is triggered to the Geneva computing farm by means of lsf jobs. Dedicated lsf jobs are submitted once per day to asynchronously verify the checksum of new transferred data from PMO to CNAF and from CNAF to Geneva. In addition we foresee to have a set of cron jobs running on the user interface which will be a trusted host to connect to the web server in Geneva.

4. Activities in 2015

DAMPE has been launched on December 17$^{\mathrm{th}}$ 2015, so only some implementation test have been performed in the last days of 2015. The implementation of all the data transferring pipeline has been then completed in the first months of 2016. Currently, data are copied daily at CNAF from PMO and a test of the Monte Carlo production framework is on-going. From 2016, DAMPE is officially supported by the CNAF center with 380 HS06 and 24 TB of storage disk.

5. Outlook for 2016

Due to the late launch of DAMPE, data transfer, analysis and MC production activities started only in Q1/2016. During the first three months critical tools have been implemented and it is now possible to fully exploit the CNAF in order to significantly contribute to DAMPE science. Over the next few months we will have a clearer picture on the need for resources (CPU and storage) and file an updated request accordingly.

6. Acknowledgments

We thank the DAMPE groups of the Purple Mountain Observatory and of the Geneva University for their collaboration in the definition and the implementation of the DAMPE Computing Model.

References

[1] Antcheva I. *et al.* 2009 *Computer Physics Communications* **180** 12, 2499 - 2512, https://root.cern.ch/guides/reference-guide.
[2] https://www.mongodb.org
[3] http://flask.pocoo.org
[4] Dubois R. 2009 *ASP Conference Series* **411** 189
[5] Tsaregorodtsev A. et al. 2008 *Journal of Physics: Conference Series* **119** 062048
[6] Allcock, W.; Bresnahan, J.; Kettimuthu, R.; Link, M. (2005). "The Globus Striped GridFTP Framework and Server". ACM/IEEE SC 2005 Conference (SC'05). p. 54. doi:10.1109/SC.2005.72. ISBN 1-59593-061-2. http://www.globus.org/toolkit/docs/latest-stable/gridftp/
[7] http://www.orientplus.eu
[8] http://www.geant.org
[9] http://www.cernet.edu.cn/HomePage/english/index.shtml