Postprint: Planning of a Downlink Data Processing System for Astronomical Satellites

Authors: Zhao Haisheng, Li Chengkui, Jia Shumei, Song Liming

Date: 2018-07-13T00:00:00+00:00

Abstract

Data acquired by astronomical satellites must undergo a series of processing steps by the astronomical satellite downlink data processing system to generate analyzable data products, which together with their corresponding software are to be released to domestic and international users. Simultaneously, the data processing system must also monitor payload status, data quality, and celestial source burst conditions, among other aspects. Thus, the construction of the downlink data processing system is directly related to the acquisition of physical results, making the planning of this system highly significant. This paper introduces the planning of this system from the perspectives of data product definition, subsystem planning, and data flow, and proposes a modular development approach to coordinate the development of various subsystems holistically, enabling each system and its software to cooperate with one another and jointly promote scientific output.

Full Text

Planning of the Data Processing System for Astronomical Satellites

Zhao Haisheng, Li Chengkui, Jia Shumei, Song Liming
Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract

Data acquired by astronomical satellites must undergo a series of processing steps by the downlink data processing system to generate analyzable data products. These products, along with the corresponding analysis software, are released to users worldwide. Simultaneously, the data processing system monitors payload status, data quality, and celestial source outbursts. Consequently, the construction of the downlink data processing system directly impacts the acquisition of scientific results, making its planning highly significant. This paper introduces the system planning from the perspectives of data product definition, subsystem architecture, and data flow, and proposes a modular development approach to coordinate the development of various subsystems holistically, enabling the systems and their software to work in concert and jointly enhance scientific output.

Keywords: Astronomical Satellite; Data Preprocessing; Data Analysis; Monitor and Quick Look
Classification: P171.5

Data acquired by astronomical satellites must undergo a series of processing steps to generate analyzable data products. During this process, payload status, data quality, and celestial source outbursts must also be monitored. Together with data interfaces, these functions constitute the Astronomical Satellite Downlink Data Processing System (hereinafter referred to as the Downlink Data Processing System). Prior to satellite launch, ground-based data processing capabilities must be established, and post-launch software parameters and algorithms must be promptly adjusted to carry out calibration work. Therefore, planning the downlink data processing system is of great significance. Currently, the Einstein Probe (EP) satellite [1] has been formally approved, and the enhanced X-ray Timing and Polarimetry Observatory (eXTP) is in the preresearch phase, making the construction of downlink data processing systems for these satellites an imminent task.

The main tasks of the downlink data processing system include three aspects: first, releasing scientific data products, which include calibrated and preliminarily screened data (Level 1 products, comprising engineering and scientific data) and scientifically extracted data after complete screening (Level 2 products, including energy spectra, light curves, images, and energy response files), while also releasing the calibration database with real-time updates; second, releasing software for processing data products (Level 1), providing software parameters and user manuals; third, real-time monitoring of payload status and data quality, and identification of outburst and transient sources, particularly discovering new sources. In practice, the downlink data processing system must complete software development, including requirements design, development, and testing, while also conducting research on key algorithms in data processing, especially completing instrument ground calibration. This paper focuses on pre-launch construction of the downlink data processing system, including data definition and flow, functions and interfaces of each subsystem, and system operation procedures. It also introduces major system interfaces, primary work content, and software engineering practices, aiming to enable coordinated system operation, promote scientific output, and improve software development and operational efficiency.

1.1 Data Definition and Data Flow

Satellite downlinked data typically undergo simple deframing, grouping, and deduplication to form a series of data files. Based on the downlink channel, we define these as telemetry data (0T) and payload data (0L, Level 0), which serve as inputs to the downlink data processing system. 0L includes data collected by the satellite platform, engineering data from the payloads, and scientific data. The processing of scientific and engineering data is critical to the system, requiring preprocessing steps such as depacketization, data conversion, file splitting and merging, followed by calibration processing (e.g., electronic gain correction, noise subtraction, event reconstruction, etc.) before release.

We define 1U, 1C, and 1L (Level 1) to represent data products generated during these steps: 1U denotes uncalibrated data products; 1C denotes calibrated data products; and 1L denotes publicly released data products, which are generally a subset of 1C or 1U. 2L (Level 2) represents advanced data products generated from 1L that have undergone complete data screening and calibration, ready for direct scientific analysis—for instance, fitting energy spectra to obtain source parameters. Monitoring and quick-look processing steps are generally simpler; their inputs may be 0T or intermediate data from 0L processing, with outputs defined as 2M.

0T data are primarily used for monitoring and quick-look, offering strong timeliness but limited information, which aids ground-based assessment of payload performance and timely discovery of new sources for observation planning. 0L data have poorer timeliness but are comprehensive, forming the foundation for scientific research. 1L may be either uncalibrated or calibrated data; for example, HXMT [2][3] releases uncalibrated data, whereas Swift [4] releases calibrated data. The advantage of releasing uncalibrated data lies in the clear separation between preprocessing and analysis stages, whereas releasing calibrated data requires processing through certain data analysis software modules beyond preprocessing. This paper recommends the latter approach, where 1L is a subset of 1C, for three reasons: (1) Released calibrated data can be directly used for monitoring and quick-look, as calibration helps identify noise and background, benefiting new source discovery, while monitoring also requires evaluating calibration parameters to provide a basis for recalibration; (2) Users can directly extract energy spectra and light curves without recalibration, though it should be noted that analyzing archival data generally requires recalibration due to frequent updates to the calibration database; (3) Calibration processing reduces data volume, facilitating transmission and analysis.

During the 0L to 1L processing, data files must be reorganized and segmented according to observation plans or monitoring results to form a complete observation, with data from different operating modes (data acquisition strategies, operating frequencies, etc.) separated within each observation. It should be noted that normal observations may be interrupted by certain circumstances (e.g., discovery of new outbursts), requiring timely updates to observation plans, while monitoring results (particularly attitude data) are prone to gaps. Attitude adjustment data before and after observations may either be retained within the observation (potentially causing overlapping data between adjacent observations) or treated as separate observations.

0L data may suffer from data duplication and file duplication during transmission, necessitating deduplication strategies in subsequent processing. Deduplication can be performed during data segmentation and reorganization. For each data downlink, new observation data are generated by splitting and reorganizing according to observation plans (file collections are generally defined as an exposure). During this process, the completeness and validity of each new file must be recorded, and version numbers assigned. If subsequent downlinks contain duplicate data or files, the completeness and validity of both versions are compared, with the superior version overwriting the existing file. Completeness indicates whether the start-to-end time is fully covered (i.e., whether data timestamps span the entire exposure time), while validity refers to data missingness (packet loss and CRC verification [5] failures and their severity) and correctness (interface with the downlink agency, indicating file regeneration due to issues such as software bug fixes).

1.2 Subsystem Planning

We have improved upon the downlink data processing system of the HXMT satellite, successfully launched in June 2017, by restructuring it into four subsystems. Figure 1 [Figure 1: see original paper] illustrates the planning of these four subsystems within the overall system, with their relationships manifested through system operations (implemented via software). The Data Reception and Distribution Subsystem serves as the scheduler in the architecture; once data are downlinked (either payload or telemetry), it orchestrates software from other subsystems to complete data processing and stores the results. The Data Preprocessing Subsystem forms the foundation of data processing, primarily responsible for extracting original payload packets (depacketization), which serves as the basis for both the Data Analysis and Data Monitoring/Quick-look Subsystems while generating complete observation data files. The Data Analysis Subsystem performs calibration and other tasks based on preprocessed data, releasing the resulting 1L and 2L data to the Data Reception and Distribution Subsystem for storage and distribution. The Monitoring and Quick-look Subsystem processes telemetry data, depacketized payload status data, and calibrated data, also delivering results to the Data Reception and Distribution Subsystem. The entire operation is file-driven.

Each subsystem will undertake software development and related auxiliary work, with primary planning as follows:

(1) **Data Reception and Distribution Subsystem**: Responsible for developing data reception and distribution software, primarily completing data reception (0T and 0L), invoking software from other subsystems for data

processing and monitoring/quick-look, and calling preprocessing modules to segment, reorganize, and deduplicate generated data for archival, while providing release interfaces for 1L, 2L, and 2M. Key interfaces include input data transmission strategies and data release interfaces. This subsystem serves as the scheduler for the entire downlink data processing system.

(2) **Data Preprocessing Subsystem** [6]: Responsible for developing data preprocessing software, primarily completing parsing of engineering and scientific data (0L) to form scientific events and records, reconstructing complete scientific events and correcting certain information, and generating complete observation data files (1U). The main interface is the software-hardware interface, which requires clear definition of hardware data acquisition strategies and data structures. Key tasks also include data level definition and content format design, as well as research on data processing and correction algorithms. Data preprocessing must also compute space environment variables, which together with satellite voltage, temperature, and other parameters constitute screening conditions for physical analysis.

(3) **Data Analysis Subsystem** [7]: Responsible for developing data analysis software, primarily starting from 1U-level products to complete data calibration, screening, and extraction of advanced data products (1C/2L), while providing the Calibration Database (CALDB). Main interfaces include the software-hardware interface and the scientific user interface—the former focuses on the data analysis process, while the latter emphasizes selection of physical criteria. Key tasks also include ground-based data analysis and calibration. Ground analysis must address event grading, noise subtraction, bad pixel identification, etc., while calibration focuses on gain correction, energy-to-channel relationship (E-C), energy response, and detection efficiency, as well as CALDB design and I/O operations. The data analysis software must be released to scientific users; following the HXMT model (Reference [7]), it is developed modularly, including modules for data calibration, good time calculation, data screening, energy spectrum and light curve extraction, energy response generation, and background generation, among others. The software must also provide a pipeline module that completes all functions of the aforementioned modules, with its output being 2L data. The CALDB, as part of the analysis software, must also be released to users.

(4) **Data Monitoring and Quick-look Subsystem**: Responsible for developing data monitoring, quick-look, and display modules, primarily completing payload status monitoring, data quality monitoring, scientific quick-look, and result display. Key tasks include studying data characteristics, searching for new sources, and extracting source positions. The top priority in its construction is rapid identification of physical phenomena (quick-look operations), utilizing intelligent algorithms to replace manual

monitoring.

1.3 System Flow Planning

The scheduling function in the system is performed by the Data Reception and Distribution Software, which orchestrates various software components for each data downlink to complete data processing. Figure 2 [Figure 2: see original paper] illustrates the scheduling of this software. Upon completion of a data downlink, the Data Preprocessing Software is first invoked to complete the 0L to 1U process, which involves numerous files and is generally divided into multiple steps (Reference [6]). These files may have dependencies; for example, when parsing timestamps in event files, the UTC-to-oscillator correspondence [8] resides in engineering files. The processing strategy is as follows: record information for each downlink, including file types and start/end times; process independent files directly, while suspending dependent files pending the next downlink (or utilizing previous downlink results). Data file completeness also significantly impacts this process—generally, 1U should be complete (especially for files required in analysis), as analysis software uses many auxiliary files (e.g., temperature files for gain-temperature dependence) during event calibration and preliminary screening. However, a single 0L downlink may not satisfy completeness requirements, commonly exhibiting missing data files or inconsistent timestamps across files. The processing strategy is that during file merging and splitting, if these files cannot cover an entire exposure time, that exposure cannot be processed until file count and temporal requirements are met. If subsequent downlink files overlap in time with this exposure and have version updates, the exposure file is regenerated. The 1U data are then processed through certain analysis software modules—primarily calibration and event screening modules (with looser criteria than physical analysis)—to produce 1C. A subset of 1C generates 1L, which is then processed through the analysis software pipeline to produce 2L.

0T downlinks occur frequently and are crucial for monitoring and quick-look; the scheduler must generate 2M results for each telemetry downlink. Both the 0L-to-1U process and portions of 1C data from payload downlinks are used for monitoring and quick-look: the 0L-to-1U process primarily for new source discovery, and 1C primarily for monitoring calibration information (e.g., whether recalibration is needed). Thus, monitoring and quick-look have three main checkpoints: telemetry data, 0L-to-1U intermediate data, and 1C data. For each payload or telemetry downlink, the Data Reception and Distribution Software must invoke the monitoring and quick-look software, though not all three checkpoints are used—selection depends on file availability. Specifically, during payload downlink, if 1C cannot be generated, only 0L-to-1U data are used. The three checkpoints have different emphases: telemetry focuses on monitoring payload operational status and discovering new sources; 0L-to-1U emphasizes new source discovery; and 1C emphasizes calibration monitoring.

1L, 2L, and 2M are all releasable products. The scheduler controls read/write permissions for released files. Table 1 shows the data inputs and outputs for each subsystem.

Subsystem Name	Data Input	Software Developed	Data Output
Data Reception and Distribution	Payload/Telemetry Data	Data Reception and Distribution Software	All
Data Preprocessing	Data Reception and Distribution Software	Data Preprocessing Software	1U/1L
Data Analysis	Data Preprocessing Software	Data Analysis Software	1C/2L
Data Monitoring and Quick-look	1U/1L	Data Monitoring and Quick-look Software	2M

The Data Reception and Distribution Subsystem requires a database (e.g., MySQL [9]) to store important information: (1) details of each payload and telemetry downlink, particularly files generated at intermediate processing stages; (2) key processing parameters, such as time correspondences, observation information, and detector unit classification (e.g., bad detectors); (3) calibration-related information, especially parameters reflecting calibration drift. This information is best presented through monitoring displays.

2 System Engineering

The establishment of the entire system depends on defining data interfaces, data flows, and calibration and background methods. The downlink data processing system must determine various interfaces and implementation methods. The most critical interface is the software-hardware interface, which includes data acquisition workflows and data structure definitions—both primarily related to the satellite payload and platform systems. Platform-collected data, such as orbit, attitude, and some payload operational parameters, are typically designed by specialized organizations (e.g., China Academy of Space Technology) with relatively simple and fixed formats. However, the interface with the payload system is more complex: the data acquisition workflow determines data file organization and formats, while data structure definitions directly dictate depacketization algorithms for data products. Since this system is generally developed concurrently with the payload system, this interface undergoes frequent changes. Traditionally, interfaces have been defined and modified through documentation, which suffers from significant lag. Implementing software-hardware interfaces through a real-time read-write system (e.g., database + display or similar to ARP) would be more efficient. This paper recommends using configuration files to define payload data (see Reference [6]), with these configuration files serving as software-hardware interfaces that can be presented through a dedicated system.

Research on data processing workflows, data calibration, background, space environment, and instrument performance is crucial to this system and forms the fundamental guarantee for completing data preprocessing, analysis software, and monitoring/quick-look functions. Such research is generally based on two types of data: simulated data and ground test data. Simulated detector data acquisition aids calibration and data processing studies; space environment simulation assists background research; ground-based generation of payload event data based on data definitions can be used for developing and testing data processing workflows. Ground test data are typically the responsibility of the payload system, but this system must also participate in data analysis. These data closely resemble satellite downlinked data, and their study is extremely helpful for researching data processing algorithms and calibrating instrument performance, particularly for determining event noise and event grade reconstruction, which play key roles.

Planning for satellite downlink systems involves many uncertainties, particularly regarding data-influencing factors and calibration uncertainties. These primarily stem from the difficulty in estimating the intense space radiation environment (background), potential deformation of payload shape and position, and possible changes in detection characteristics. This necessitates strong maintainability and extensibility in the system. This paper recommends adopting modular development, separation of data and code, and object-oriented development patterns, with holistic software development coordination.

Astronomical satellite data processing typically employs a modular approach, where each module has its own input parameters and can run independently, with data files passed between modules. This design reduces coupling between steps, enhancing system maintainability—particularly for nonlinear operations (mainly those involving correlations between adjacent events), which can be isolated as separate modules for maintenance without affecting other components. Additionally, modularity facilitates invocation by the Data Reception and Distribution Software. Data processing involves numerous parameters: data selection criteria, detector configuration, calibration, and background parameters, among which calibration and background parameters are updated frequently. Software development should separate code from these parameters as much as possible to allow independent maintenance. For monitoring and quick-look, data and display interfaces should also be separated.

Object-oriented development offers excellent encapsulation and inheritance characteristics, greatly facilitating systematic development. Figure 3 [Figure 3: see original paper] shows the software framework of the current HXMT user analysis software. The top layer (BaseLIB library) comprises a collection of basic functionalities, including data I/O classes, information output classes, energy spectrum and light curve classes, CALDB reader classes, basic data structure classes, algorithm classes, etc. BaseLIB should be prioritized in development as it can serve all software or payloads; we recommend using the HXMT user analysis software as a reference library for its development. The middle layer is payload-specific (or software-specific, LIB(A), LIB(B), etc.), including file system composition classes (implementing pipelines, etc.) and various payload-related (or software-related) data structure classes; this layer may be omitted for simpler software or payloads. The lower layers of libs and modules store payload-related algorithms and special function classes, respectively, while tasks implement individual modules, primarily completing module workflow operations. This design is hierarchical, separating interfaces, algorithms, and functional implementations, with workflow functions connecting them. Changes to one part do not affect others. Another advantage of this design is reduced software development and testing costs—if software components were isolated, both development and testing costs would be high. In this design, module testing can repeatedly utilize libraries like BaseLIB, thereby enhancing reliability.

Software development for the data flow component is the top priority of this system, encompassing data preprocessing and data analysis. The development (or design) sequence should be preprocessing first, followed by analysis, with the pipeline module developed later, even post-launch if necessary. Data preprocessing, analysis, and monitoring/quick-look all require deep understanding of the payload and data, whereas data reception and distribution can be developed independently of the other three.

3 Summary and Outlook

The satellite downlink data processing system is a critical component of the satellite ground system, responsible for hierarchically processing downlinked data, releasing calibrated 1L data, and distributing analysis software and CALDB. It also monitors detector status and data quality, and undertakes important tasks such as detecting outbursts and discovering new sources. These functions support and promote satellite scientific output, particularly in new source discovery, directly impacting scientific productivity. This paper divides the satellite downlink data processing system into four subsystems: Data Reception and Distribution, Data Preprocessing, Data Analysis, and Monitoring/Quick-look. These subsystems work in coordination to develop and deploy the data reception and distribution software, data preprocessing software, analysis software, and monitoring/quick-look software, while providing CALDB. Development of the satellite downlink data processing system should follow the principle of using logical design to reduce code volume and intellectual effort to replace manual labor, thereby completing system design and development. Work should also focus on data processing algorithm research and calibration. This paper does not address hardware infrastructure; in fact, data storage and computing environments are extremely important and significantly impact software deployment and operation.

References

[1] 袁为民, 张臣, 陈勇, 等. 爱因斯坦探针：探索变幻多姿的 X 射线宇宙 [J]. 中国科学：物理学, 力学, 天文学, 2018, 48: 039502. Yuan W M, Zhang C, Chen Y, et al. Einstein Probe: Exploring the ever-changing X-ray Universe (in Chinese). Sci Sin-PhysMechAstron, 2018, 48: 039502.

[2] Li Ti-Pei, Wu Mei. The hard X-ray modulation telescope mission [J]. Physics, 2008, 37(09):648-651.

[3] Li Ti-Pei. HXMT:A Chinese High-Energy Astrophysics Mission [J], Nuclear Physics B, 2007, 166:131-139.

[4] Burrows, D. N. et al.: 2000, The Swift X-ray Telescope[R], Proc. SPIE Vol. No. 4140,pp. 64–75.

[5] Han Jiawei, Zhang Hongqun.Error correction technology based on CCSDS standard for remote sensing satellite. Chin. J. Space Sci., 2009, 29(1): 112-

[6] 赵海升, 葛明玉, 李正恒, 聂建胤, 宋黎明. 一种天文卫星数据预处理方法 [J]. 天文研究与技术, 2017, 14(3): 376-381. Zhao Haisheng, GeMingyu, Li Zhengheng, NieJianyin, Song Liming. A Method of Data Preprocessing for Astronomical Satellite[J]. Astronomical Research and Technology, 2017, 14(3): 376-381.

[7] ZHAO Haisheng, LI Chengkui, LI Xiaobo, NIE Jianyin, GE Mingyu, PAN Yuanyue, SONG Liming. High Energy X-ray Telescope data analysis method[J]. Chin. J. Space Sci., 2016, 36(6): 938-944.

[8] 蔡凡, 陈丽. 卫星双模授时设备的关键技术实现 [J]. 天文研究与技术, 2018, 15(1):46-51. Cai Fan, Chen Li. Realization of Key Technology of Dual Mode Satellite Timing Equipment.Astronomical Research & Technology, 2018, 15(1): 46-51.

[9] 钟守波, 韩波, 张彦霞, 等. 天文大数据管理工具的设计与实现 [J]. 天文研究与技术, 2015, 12(4): 510–517. ZhongShoubo, Han Bo, Zhang Yanxia, et al. Design and implementation of s software tool package for managing massive astronomical data[J]. Astronomical Research & Technology, 2015, 12(4): 510–517.

---

[1] Funded by the National Natural Science Foundation of China (Youth Science Fund Project, Grant No. 11503029)

Note: Figure translations are in progress. See original paper for figures.