ASO-S 卫星数据预处理系统设计与关键算法研究

于勤思 1 马福利 1 杨晓艳 2 杨广建 2 黄宇 3 佟继周 1
(1 中国科学院国家空间科学中心 北京 100190)
(2 中国科学院空天信息创新研究院 北京 100094)
(3 中国科学院紫金山天文台 南京 210023)

摘要

先进天基太阳天文台（Advanced Space-based Solar Observatory，ASO-S）是我国太阳物理界自主提出的一颗太阳空间探测卫星。数据预处理系统是 ASO-S 科学卫星地面段的重要组成部分，负责对卫星下行的科学数据进行快速预处理和质量分析，为开展太阳观测数据分析提供重要输入。文章基于 ASO-S 卫星数据预处理系统需求以及科学载荷数据结构、处理特点和难点，设计了数据预处理系统架构、工作流程以及核心处理算法，解决了卫星奇偶双通道下行、多类型数据融合匹配以及高准确度数据质量分析等关键技术问题。各项测试结果表明，ASO-S 卫星数据预处理系统运行稳定，功能和性能达到了预期的设计目标。该系统已成功应用于 ASO-S 卫星地面系统，支持卫星在轨运行。

The Advanced Space-based Solar Observatory (ASO-S) is China's first satellite mission dedicated to comprehensive solar observation from space. To ensure efficient utilization of its scientific payloads, a robust data preprocessing system has been developed as a core element of the satellite's ground segment. This system is responsible for the rapid conversion, organization, and quality evaluation of raw downlinked data, laying a solid foundation for further scientific analysis. This paper introduces the overall design of the preprocessing system, detailing its architecture, operational workflow, and key algorithms. Specific technical challenges—such as handling the satellite's dual-channel (odd-even) downlink structure, synchronizing diverse data streams from multiple payloads, and achieving high-precision quality control—are addressed through targeted solutions. System testing confirms that the implementation meets all functional and performance requirements, and the system has been fully integrated into the ASO-S ground operations, supporting the satellite's in-orbit mission successfully.

1 引言

空间科学卫星数据的预处理是实现科学探索目标的关键步骤。其重要性在于将原始的下行数据转化为标准化的数据产品，以便进行深入的数据处理和进一步的科学数据分析。此外，每颗空间科学卫星的科学任务都是独特的，其所携带的科学载荷在设计和功能上具有高度创新性。这些特征导致了数据格式、处理方法和整个数据处理工作流程的多样性和特殊性。这些特殊要求不仅要求空间科学卫星的地面数据处理系统具备高效的数据处理能力，以满足紧迫的时效性需求，还要确保数据处理的准确性，以保证高质量的科学成果。因此，地面数据处理系统的设计需要针对每个任务的独特要求进行特定考虑。

中国的先进天基太阳天文台（Advanced Space-based Solar Observatory, 简称 ASO-S)是中国首个综合性太阳任务，于 2017 年底获得中国科学院的批准，属于空间科学战略性先导科技专项。ASO-S 的科学目标简称为“一磁两暴”，“一磁”即太阳磁场，“两暴”即指太阳上两类最剧烈的爆发现象—太阳耀斑和日冕物质抛射，即观测和研究太阳磁场、太阳耀斑和日冕物质抛射的起源及三者之间可能存在的因果关系[1-2]。为此，ASO-S 搭载三个科学载荷：全日面矢量磁像仪(Full-disc vector MagnetoGraph，FMG)用来观测太阳光球矢量磁场；硬 X 射线成像仪(Hard X-ray Imager，HXI)用来观测太阳耀斑非热物理过程；莱曼阿尔法太阳望远镜(Lyman-alpha Solar Telescope，LST)主要用来观测日冕物质抛射的形成和早期演化[3-4]。ASO-S 卫星独特的载荷组合首次实现了在一颗卫星上同时观测太阳全日面矢量磁场、太阳耀斑高能辐射成像和日冕物质抛射的近日面传播。

ASO-S 卫星正式运行后，三个科学载荷每日产生海量的科学观测数据，地面数据处理系统需要能够及时的进行下行数据的虚拟信道分离、多站数据融合、源包提取、数据排序去重、异常值检测与去冗、打时空戳、参数解算等处理[5-6]，并根据载荷数据的特征生成分时段的组织良好的 FITS 格式数据产品。相比之下，美国的 SDO 卫星主要采用统一格式和单一数据源，在处理流程上具备相对集中性；而 ASO-S 不仅有着庞大的数据量，还涉及多个载荷、多个数据

格式和多站接收融合，这些因素显著增加了数据预处理系统的设计复杂度和运行压力，对ASO-S 卫星数据预处理系统的设计提出了新的挑战。

本文将针对 ASO-S 卫星载荷数据格式的特征以及数据产品格式要求，设计了一种具有分布式并行处理能力的 ASO-S 卫星数据预处理系统，系统支持载荷数据并行处理以及数据质量控制，能够实现对 FMG、HXI 和 LST 载荷预处理以及 FITS 标准格式数据产品生成，满足科学家开展后续高级数据处理以及科学数据分析的功能和性能需求。

2 系统需求与难点分析

2.1 预处理数据产品设计

为了实现 ASO-S 卫星的科学目标，并促进不同卫星观测数据的比较分析，遵循标准化数据格式至关重要。在太阳物理学领域，已建立了数据产品分级的规范 [7]，以确保能够进行国际间的客观比较。虽然各种科学卫星都遵循这些共同标准，但每颗卫星在这些标准下可能会有特定的定义。ASO-S 卫星数据产品的分类基本原则与太阳物理学界采用的分级原则和方法相一致。详细的预处理产品级别定义如下：

表-1 预处理数据产品分级设计

Table 1. Hierarchical structure of level-0 data products

数据级别	级别描述
RAW	原始数据传输产品包括卫星的原始传输帧数据文件和原始数据传输属性信息文件。通过地面天线接收到多通道原始传输帧数据，经过解调、帧同步、解码和解扰处理生成传输帧二进制文件。属性信息文件提供与原始数据文件相对应的解释信息和数据质量信息。
0A	通过对单次卫星下行传输的原始数据进行解压缩、解码、虚拟信道分离、源包提取、验证、排序、按应用过程标识符分割、添加关键元数据以及格式化输出等处理后生成的载荷科学源包数据产品。
0B	通过对多站接收的载荷 0A 级科学包数据产品进行多站融合、数据去重、合并/分割、添加关键元数据以及格式化输出等操作生成的按小时组织的数据产品。
0C	基于 0B 级科学源包数据产品，进行科学源包提取、CRC 校验、观测模式提取、时间解算、图像匹配等处理后生成的科学数据产品。
Q0	在 0C 级载荷科学数据产品基础上进行解压缩，初步处理后成像或辅助成像的数据产品。

2.2 数据处理需求

在 ASO-S 卫星原始数据落地后，直接由地面接收站生成的 RAW 数据，地面接收站将 RAW 数据发送至 ASO-S 卫星数据预处理系统开展数据处理与产品生成。依据表-1 中的预处理数据产品分级定义，首先需要将多通道下传的帧文件合并，之后按照载荷的数据特征以及格式要求顺序进行预处理数据处理与产品生产。

（1） 载荷 0A/0B 级数据处理需求

载荷 0A 级数据产品对卫星单次过站的数传原始数据进行解压缩、解帧、奇偶帧合并、虚拟信道分离、源包提取、验证排序、按 APID 拆分，格式化（含添加关键字）等处理后的生成的 FMG/LST/HXI 载荷科学数据产品。在 0A 级科学源包数据产品基础上进行排重、拼接/拆分、产品格式化处理（含添加关键字）生成的按小时组织的科学数据产品，载荷 0B 级数据产品在整小时切分的基础上向前合并五分钟的数据，以确保数据产品的完整性。

（2） FMG 载荷 0C 级数据处理需求

FMG 载荷是一台 140mm 口径的太阳矢量磁场望远镜，在 532.4nm 波段对太阳全日面的偏振信号进行测量，最终通过偏振定标得到全日面矢量磁场。FMG 载荷有三种工作模式：常规模式、快速模式和定标模式，常规和快速模式用于生成原始科学观测数据，定标模式用于生成科学数据生产中用到的定标数据[8]。

每种模式的数据可以解调出 Stokes Q、U、V 图像（其中 V 是与视向磁场相关的圆偏振参量图像，Q、U 是与横向磁场相关的参量图像，视向磁场与横向磁场是矢量磁场的两个分量）。在进行 FMG 载荷 0C 级产品生产时，需要提取 FMG 载荷观测模式，对定标模式、常规模式和快速模式进行区分，也需要识别不同类型的图像数据，并将数据分组存入到一个 FITS 文件中，

最终生成 FMG 载荷 0C 级的定标数据和科学数据，FMG 载荷 0C 级数据产品组织模型如下图所示。

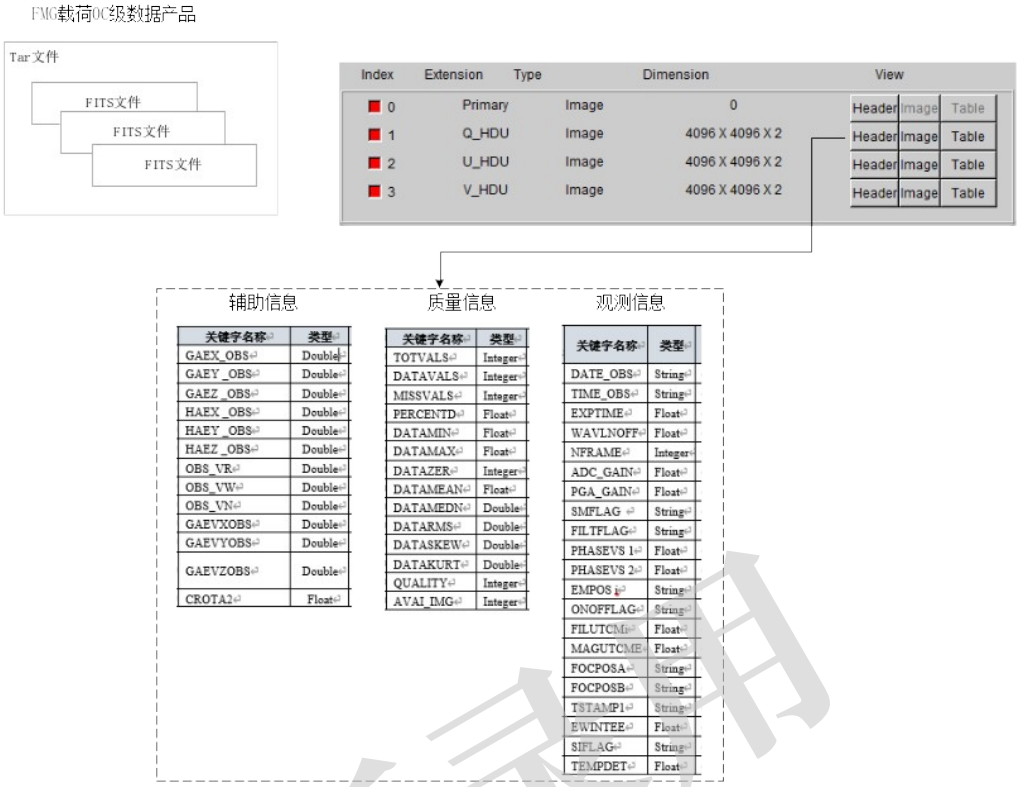


图 1 FMG 载荷 0C 级数据产品组织模型

Figure 1. Data organization model of fmg level-0c products

（3） LST 载荷 0C 级数据处理需求

LST 载荷包含一台口径 60 毫米莱曼阿尔法全日面成像仪（SDI）、一台口径 60 毫米的日冕仪（SCI）、一台口径 130 毫米白光望远镜（WST）和两个 30 毫米口径的导行镜（GT）[9]。地面数据与处理系统需要对接收到的星上传输原始数据进行解压缩、解帧、源包提取、验证排序、拆分/拼接、添加关键字、格式化等处理后生成 FITS 格式的 LST 载荷 0 级数据产品。根据处理程度和数据组织的不同，可分为 0A、0B 和 0C 级产品，如上表-1 内容所示。

在进行 LST 载荷 0C 级处理时，需要提取 LST 载荷观测模式，对定标模式和科学观测模式（常规、爆发、用户模式）进行区分，生成 LST 载荷 0C 级的定标数据和科学数据。LST 载荷的 0C 级科学数据产品是在 0B 级科学数据的基础上，对数传科学数据、定标数据、工程参数及辅助数据进行分析和处理得到的数据产品。LST 载荷的 0C 级载荷科学数据产品按照整小时进行组织，载荷定标数据按次组织。每个产品包括多个 FITS 文件，每个 FITS 文件对应一幅、两幅和四幅等不同数量的图像组合。

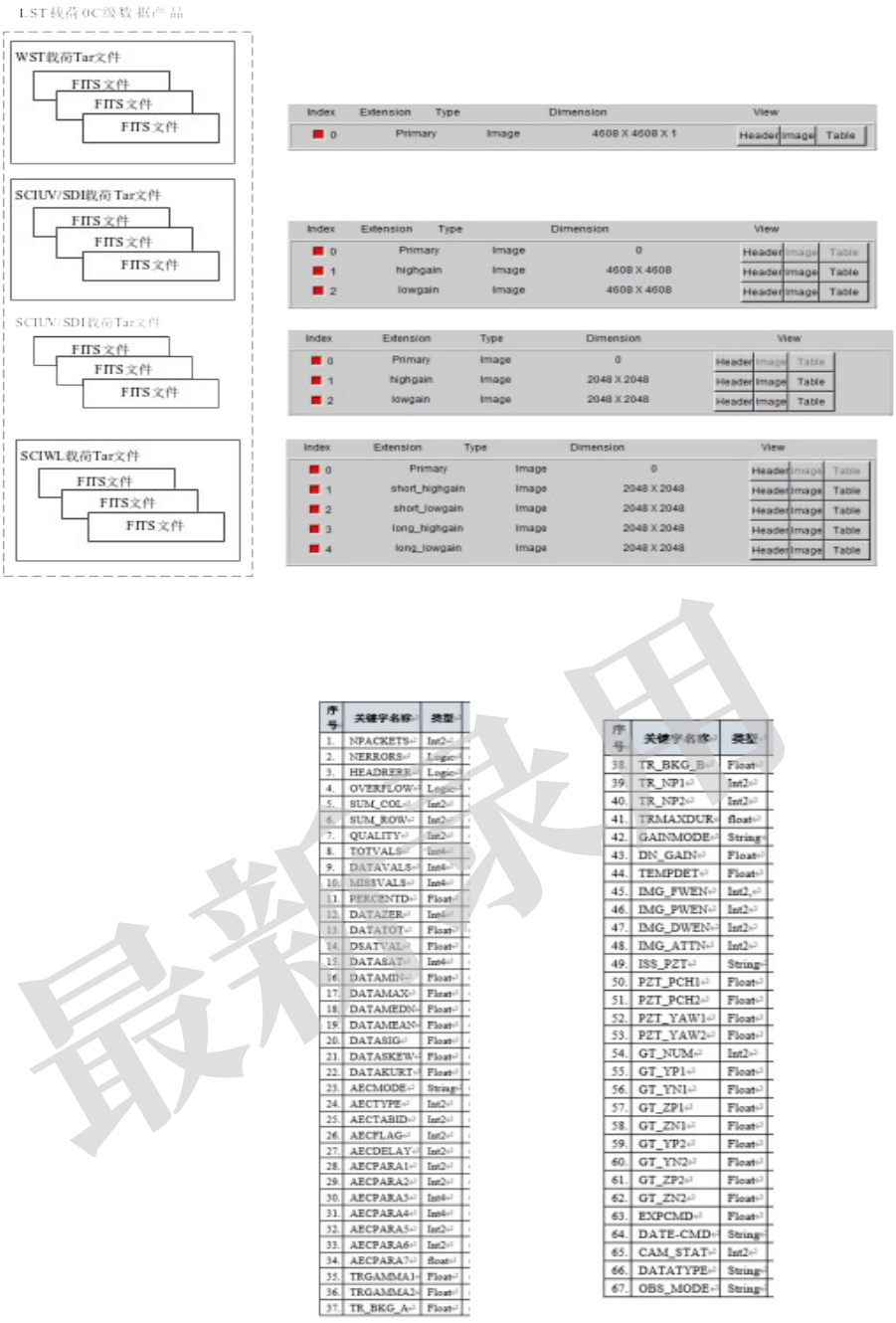


图 2 LST 载荷 0C 级数据产品组织模型

Figure 2. Data organization model of lst level-0c products

（4） HXI 载荷 0C 级数据产品设计

HXI 载荷负责对太阳爆发现象（如太阳耀斑）中的硬 X 射线进行成像与能谱观测，主要覆盖的能段范围为 30–200 keV。其科学目标是研究爆发过程中高能电子的加速、输运与辐射等非热物理过程。HXI 的数据记录模式共有三种，分别为逐事例数据、通道数据和四能段数据。其观测模式分为两种：在无爆发时通常采用 4 秒曝光时间进行常规观测；而在出现耀斑等强硬 X 射线爆发事件时，则切换为更短的曝光时间（0.125 秒、0.25 秒或 0.5 秒）以提升时间分辨率 [10]。

地面数据预处理系统基于数传下行的原始数据进行一系列处理操作，包括数据解压、解码、源数据包提取、验证、排序、拆分/合并及格式化，最终生成 HXI 载荷的 0 级科学数据产品。依据表 1 所示的产品分级定义标准，HXI 载荷的 0 级数据产品可根据处理深度与数据组织方

式分为 0A、0B 和 0C 三级，用以支持不同层次的科学研究需求。

2.3 难点分析

1) 奇偶双通道数据合并处理

ASO-S 卫星搭载的科学载荷每天生成约 500GB 的科学数据。这些数据最终通过 X 波段数据传输下行到地面部分，然而由于卫星可用的频带宽度有限，无法满足科学数据的及时下行需求。为了解决这一问题，卫星采用了极化分频复用技术，以提高数据传输速率。通过同一频带发送两个调制载波信号，并满足特定的极化隔离要求，以便于地面站接收[11]。地面站利用极化的正交特性在同一频率和时间域中分离两个数据流，从而增加下行数据量，改善数据下行的及时性。

传输帧数据在地面被分为两个数据流：一个是奇数帧，另一个是偶数帧。在提取载荷科学源数据包之前，地面系统需要依次合并这两个数据流，以正确提取传输帧中的数据。在处理奇数帧和偶数帧时，可能会出现许多困难和异常。例如，可能在每个数据流中发生数据丢失。如果在卫星与地面之间传输时奇数帧的数据丢失，地面系统仅拥有对应的偶数帧，无法完全提取科学载荷数据帧；两个数据流可能会出现顺序混乱，这使得无法使用简单的合并排序算法来合并数据[12]。此外，某些传输帧可能在星地传输的过程中导致错误，无法通过地面的 RS 校验和 CRC 校验，未能通过验证的数据无法进一步处理，从而会导致部分数据丢失。

以下为系统运行过程中遇到的一个具体异常处理实例：在某次地面接收过程中，由于偶通道的短时信号衰减，导致连续多个偶数帧未被成功接收，而对应的奇数帧数据则完整下传。按照数据格式设计要求，每个完整的科学数据帧需同时包含奇数帧和偶数帧内容，以支持源包的完整提取。因此，在该场景下，仅凭奇数帧信息无法完成数据拼接，导致该时段的载荷数据包无法正确重构。

进一步分析发现，由于偶通道丢失帧的时间跨度不固定，且奇数帧仍在持续入库，造成合并队列中奇偶帧的“对齐点”无法确定，原有的按帧号顺序简单排序合并策略不再适用。对此，系统采用带时间戳缓冲窗口的队列机制，结合帧头信息中的同步标识，动态调整合并策略，以尽可能保留完整帧，提高数据恢复率。

此类异常虽在整体数据中占比较低，但对特定科学观测时段的数据完整性影响显著，因此数据预处理系统必须具备相应的自动识别与应对机制，以确保科学数据质量和处理连续性。

2) 基于特征数据的载荷图像匹配

根据 FMG 载荷 0 级数据产品的设计要求，在数据预处理过程中需根据 Stokes 参数对图像进行分组整理。处理系统需根据图像的时间戳信息及载荷的工作模式，将同一观测周期内的 Q、U、V 参数图像按既定顺序（Q、U、V）组合成一个完整观测单元，并打包存储为一个单一的 FITS 文件。在实际运行中，由于图像的生成、传输及接收可能发生不同步，可能出现某一参数图像（如 Q、U 或 V）缺失、顺序错乱或延迟到达的情况。为此，地面预处理系统需具备图像类型识别与观测周期还原能力，能够对乱序、异常、缺失等情况进行处理，并输出结构一致、质量标注完整的标准化产品。如图 3 所示，产品中包含数据完整性、配对一致性等关键质量标识。

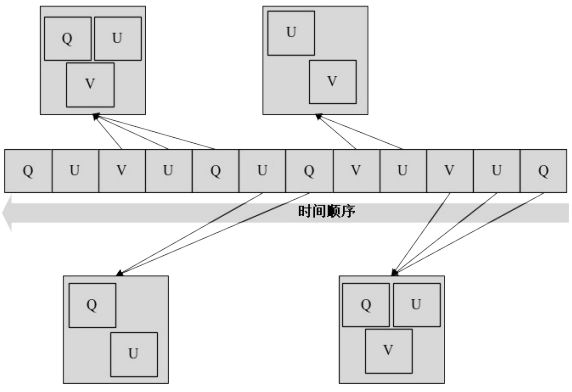


图 3 FMG 载荷数据处理过程示意图

Figure 3. Schematic diagram of the fmg data processing workflow

在 LST 载荷中，SCIWL 探测器图像处理过程涉及更多属性的组合。根据其产品设计，预处理系统需依据图像的曝光时间和增益设置，对短曝光高增益、短曝光低增益、长曝光高增益和长曝光低增益四类图像进行组合整理。每个偏振片角度下，系统首先接收短曝光图像对，再接收长曝光图像对，最终按时间信息与属性标签将四幅图像组织成一个观测单元，并存储为一个 FITS 文件。与 FMG 载荷类似，SCIWL 图像也面临图像乱序、缺失和异常等问题，地面系统需基于图像元信息完成图像归类与组合，确保产品结构完整，并在输出产品中提供相应的质量控制标识，说明该组图像的组织完整性与配对一致性。

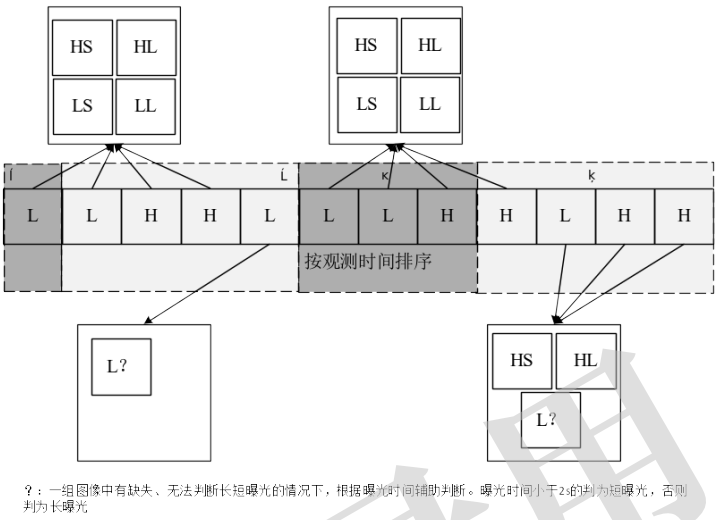


图 4 LST 载荷 SCIWL 探测器长短曝光、高低增益图像匹配算法示意图

Figure 4. Schematic of the image matching algorithm for short/long exposure and high/low gain modes of the sciwl detector on the lst payload

3) 多源多层数据信息识别与匹配

卫星搭载的三个科学载荷数据通过三层数据结构进行组织，首先科学载荷产生以 EB90 为同步码的载荷科学数据帧，之后载荷科学数据帧会组织成同步码为 E225 的源包格式数据存储在传输帧的数据域中，每一层数据结构都会包含有一些头、尾信息，用于进行源包数据类型的识别与数据一致性的校验，每层具有其独特的数据特征和处理要求，如图 5 所示。

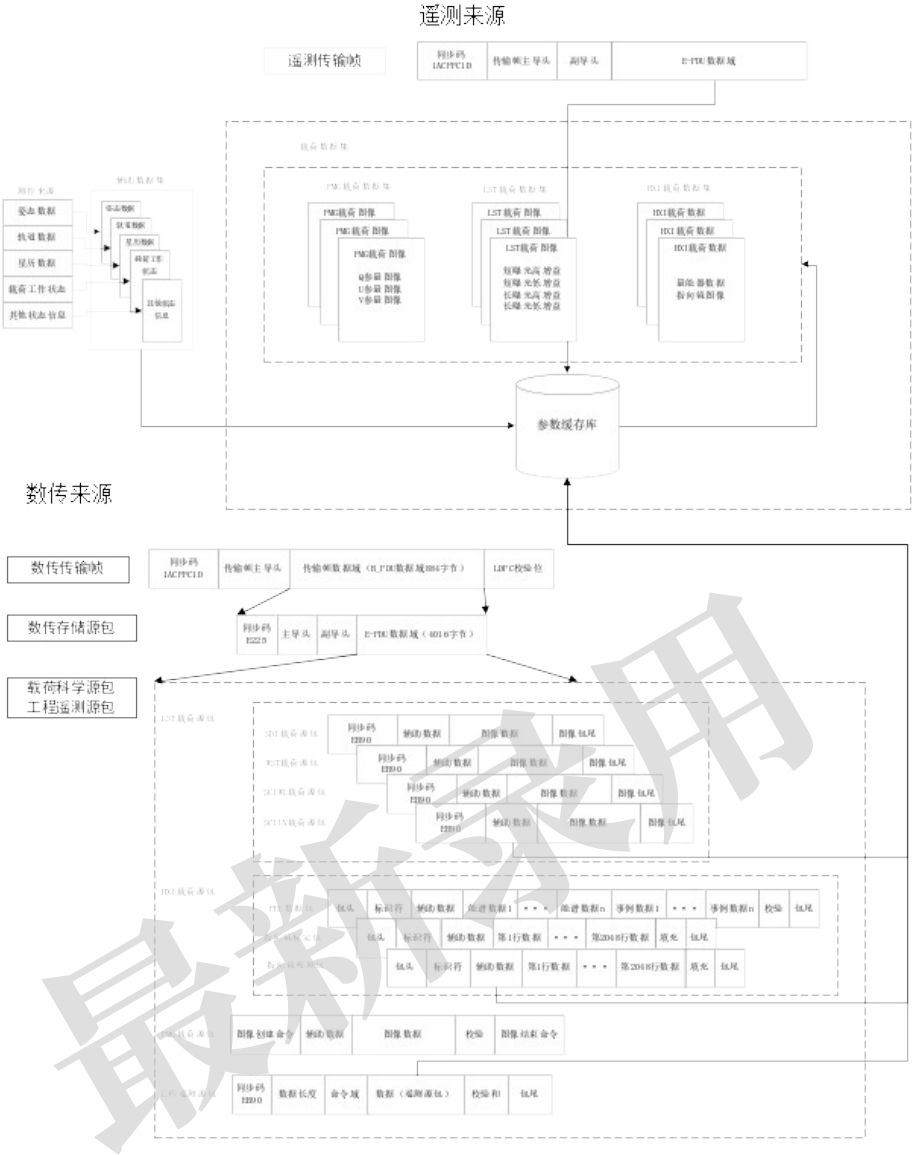


图 5 科学载荷三层数据结构组织模型

Figure 5. Three-layer data structure organization model for scientific payloads

3 系统设计与实现

3.1 系统设计

ASO-S 卫星数据预处理系统需要将卫星原始数据通过一系列处理步骤，生成符合产品格式要求的 0A、0B、0C 级产品。系统的具体功能如图 6 所示，其中包含了传输帧及源包层数据处理、载荷包层数据处理、产品格式化输出和任务与状态管理四个部分。

最新录用

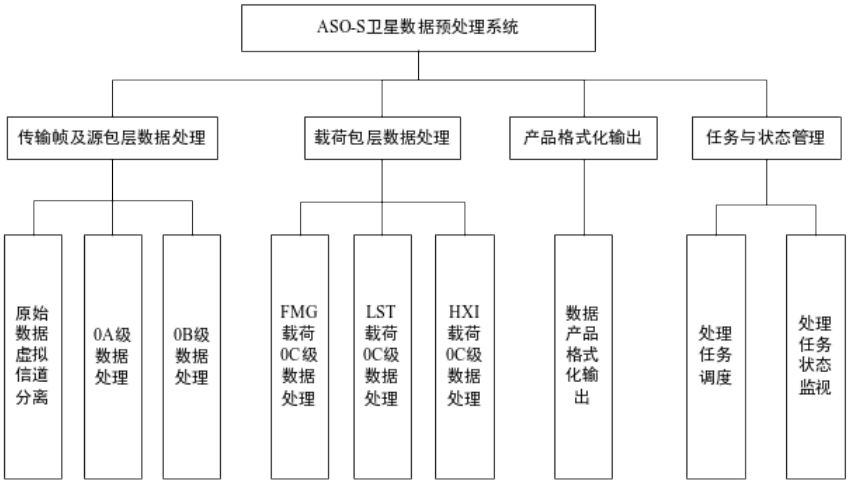


图 6 ASO-S 卫星数据处理系统设计

Figure 6. Design of the aso-s satellite data processing system

传输帧及源包层数据处理，负责将卫星下行的原始数据通过虚拟信道分离、奇偶帧数据合并，生成虚拟信道分离后的数传帧数据；通过源包提取与校验、源包排序、按应用过程标识符分割，生成 0A 级产品；对多站接收的 0A 级产品进行多站融合、数据去重、合并/分割，生成 0B 级产品。

载荷包层数据处理是系统的一项核心内容，需要基于 FMG、LST、HXI 三个载荷数据的特点，通过科学源包提取、CRC 校验，得到载荷包数据；然后根据各载荷具体处理需求，进行图像解压缩、观测模式提取、时间解算、图像匹配等处理后，分别生成 FMG 载荷 0C 级产品数据、LST 载荷 0C 级产品数据、HXI 载荷 0C 级产品数据。

产品格式化输出，依据 ASO-S 卫星产品格式要求，对以上生成的 FMG、LST、HXI 三个载荷 0C 级产品数据添加关键元数据信息、FITS 格式化处理等操作，输出标准的 FITS 格式 0C 级产品。

任务与状态管理，负责接收、解析数据处理任务订单，验证订单信息的正确性；根据任务订单生成处理任务，调用相应的数据处理单元进行处理；任务执行过程中及任务结束后，对外发送任务执行状态。

3.2 处理流程设计

在已有任务数据处理系统设计的基础上[13]，本文针对 ASO-S 卫星数传数据处理特点，设计了一套能够处理 ASO-S 卫星实时平台遥测、回放工程数据、回放科学数据等多虚拟信道数据，以及 FMG、LST、HXI 多载荷数据的综合处理流程，具体如图 7 所示。系统接收到任务订单后，自动启动处理流程，对订单指定的数传原始数据进行数传帧提取与校验、虚拟信道分离，生成按照虚拟信道分类后的数传帧数据；根据传输帧计数及其周期信息，匹配对应的奇/偶帧数据，并完成奇偶帧双通道数据合并处理。根据虚拟信道标识符类型，进入相应的分支流程。针对实时平台遥测数据、回放工程数据，启动对应的工程数据处理流程，进行 0A、0B、0C 级工程数据产品生产；针对回放科学数据，首先进行源包提取与校验、按照应用过程标识符对源包分类、排序排重、合并/分割等处理过程，生成 0A、0B 级产品，然后根据应用过程标识符识别不同载荷，进入相应载荷的 0C 级产品处理分支流程。

不同虚拟信道、不同载荷数据处理分支流程可以并行处理。特殊地，在科学数据 0C 级产品生产过程中，部分步骤需要读取 0C 级工程数据产品的信息，通常情况下工程数据处理速度较快，不影响并行处理；异常情况下，当科学数据 0C 级产品生产所需的工程数据不存在时，处理任务按照预设时长（比如 30 分钟）进入等待状态，超时后采用默认值填充相应信息并输出产品，同时给出错误提示。

其中，奇偶双通道数据合并处理算法、FMG 载荷图像匹配算法、LST 载荷 SCIWL 仪器图像匹配算法是流程中的关键算法。

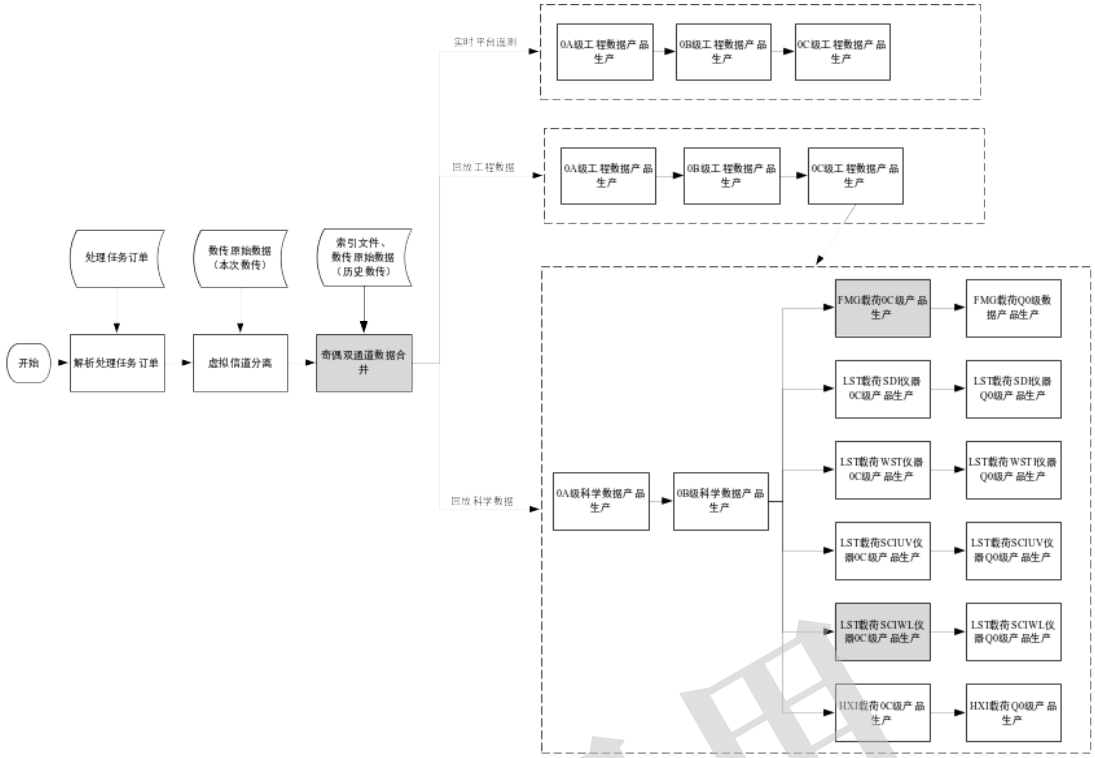


图 7 ASO-S 卫星数据处理流程

Figure 7. Data processing workflow of the aso-s satellite

4 关键算法

4.1 基于帧索引的奇偶双通道数据处理算法

数据预处理系统在接收到地面站发送的奇偶双通道下行的数据后，首先需要依据数据格式要求对传输帧数据进行格式检查以及一致性校验，剔除异常传输帧数据，依据虚拟信道标识符对传输帧的数据进行分类提取。由于卫星依据虚拟信道标识符单独给各虚拟信道设计传输帧计数规则，因此需要在进行虚拟信道分离后的数据基础上进行奇偶双通道下行数据合并处理，同时考虑到当前数据与历史数据衔接的需求，在双通道数据处理时需要引入缓存数据。针对上述情景，设计了一种基于帧索引的奇偶双通道数据处理算法，如图 8 所示，算法主要步骤如下：

（1） 数据校验与筛选

首先对接收到的事后数传数据帧执行格式校验。假设单帧数据可表示为F_i，其对应的校验函数为$\left(\check{F}_i\right)$，即:

$$\left(\check{F}_i\right)=\begin{cases}1, if\ the\ data\ format\ correct.\\0, if\ the\ data\ format\ incorrect.\end{cases}$$

当$\left(\check{F}_i\right)=1$时，帧$F_i$被保留，继续后续处理；否则，帧被丢弃，不参与后续计算。

（2） 虚拟信道分离与索引记录

每个数传帧 F_i带有虚拟信道标识符 VC_i信息，将根据该标识符将数据帧分配到不同的虚拟信道中。设虚拟信道集为$\{VC_1,VC_2,\ldots,VC_n\}$，定义虚拟信道分离函数将不同$F_i$将帧按信道信息进行分类，同时提取帧计数周期$C_i$、虚拟信道帧计数$V_i$，并记录其在原始文件中的偏移量 $OFFSET_i$。这些信息将保存在帧索引文件中，即$Index\left(F_i\right)=\{C_i,V_i,OFFSET_i,VC_i\}$，用于后续排序和合并处理。

（3） 帧索引合并与排序

从历史缓存中加载该数据的奇/偶配对文件，并提取其索引信息，将历史索引 $Index_{history}$与当前索引 $Index_{current}$合并。合并后的帧索引按帧计数周期C_i和虚拟信道帧计数V_i进行排序。

设排序函数为$Sort()$，按照C_i进行升序排列，当C_i相同时，再依据V_i排序。则排序后的索引$Index_{sorted}$可以表示为：

$$Index_{sorted} = Sort\left(Index_i \middle| C_i, V_i\right)$$

（4） 重复帧剔除

遍历排序后的帧索引文件，对于帧计数周期C_i和虚拟信道帧计数V_i相同的帧，只保留一个，去除其余重复帧。

（5） 奇偶配对与数据输出

读取去重后的帧索引文件，检测是否存在奇偶配对。如果F_i和F_{i+1}的索引满足$C_i = C_{i+1}, V_i = V_{i+1} - 1$）则符合匹配条件，进行成对匹配输出。若不存在奇偶配对，则生成空的输出文件清单。

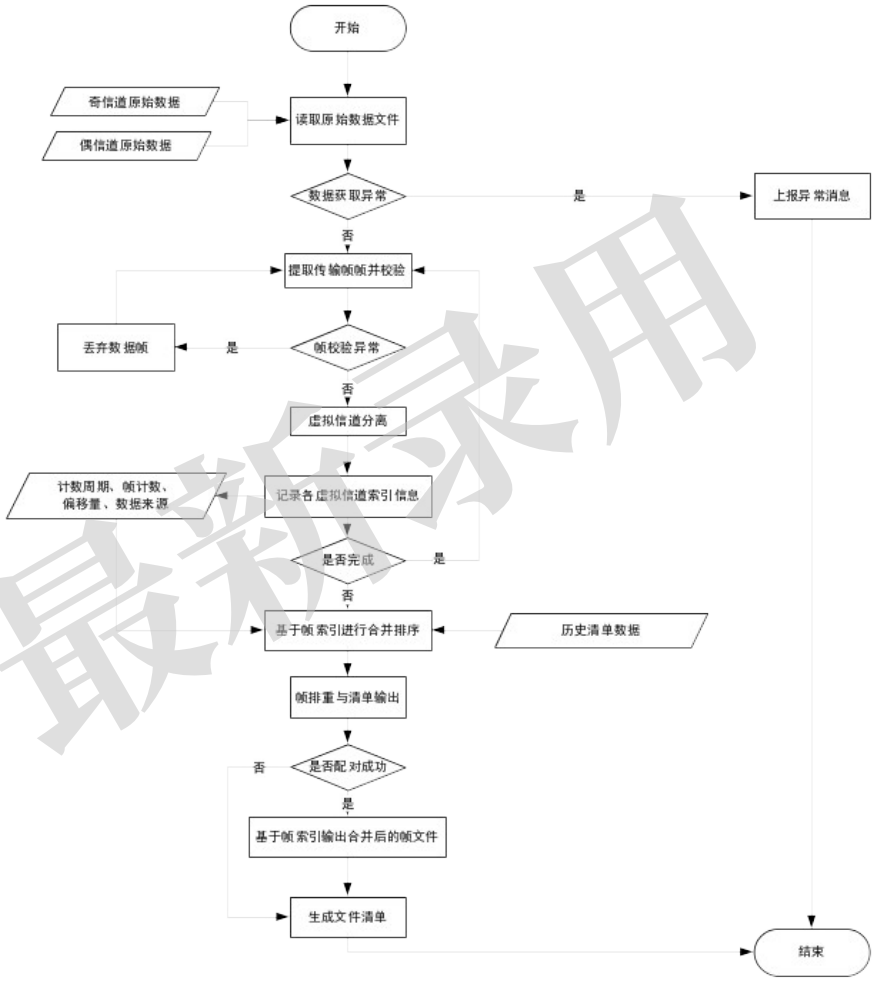


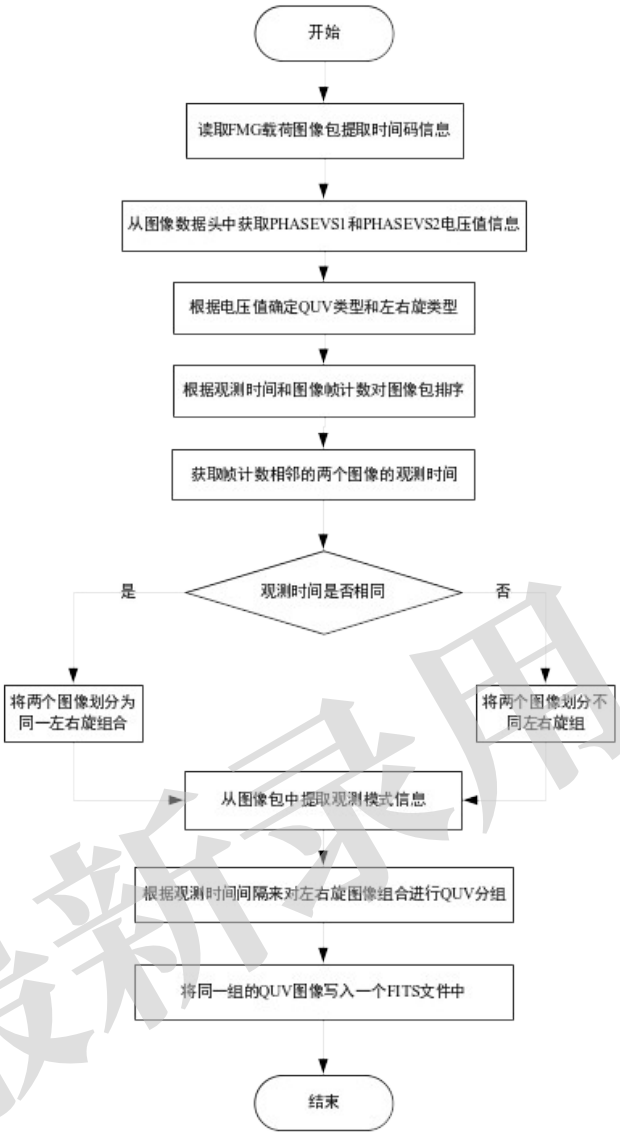
图 8 ASO-S 卫星数据处理流程 AOS-S 奇偶帧双通道数据处理流程图

Figure 8. Dual-channel processing workflow for even and odd aos-s frames in the aso-s satellite data pipeline

4.2 FMG 载荷图像提取与匹配处理

FMG 载荷在轨常规模式和快视观测模式工作时会进行 Stokes Q、U、V 图像数据的循环观测，每次观测会同时对太阳进行拍摄两张左旋和右旋图像。因此在进行 0C 级产品生产时，需要按照 Q、U、V 的顺序进行匹配，将匹配得到一组 Q、U、V 图像数据写到同一个 FITS 文件中。FMG 载荷图像匹配的算法流程如下所示：

最新录用



最新录用

图 9 FMG 载荷图像匹配算法流程图

Figure 9. Workflow of the image matching algorithm for the fmg payload

（1） 提取时间码并匹配电压值

从 FMG 载荷图像中提取时间码信息，随后依据该时间码从工程参数中匹配提取对应时刻的 PHASEVS1 和 PHASEVS2 电压值信息。设图像I_i的时间码为T_i，工程参数集合为E，对于图像I_i，若存在$e_j \in E$使得e_j的时间与T_i对应，则提取e_j中的 PHASEVS1 和 PHASEVS2 电压值并与I_i关联。这一步骤确保每张图像对应的电压数据能够与图像时间准确匹配。

（2） 确定 QUV 和旋转类型

通过电压值 PHASEVS1 和 PHASEVS2 的组合来确定图像的 Stokes 参数（QUV）及其左右旋类型。

（3） 基于观测时间与帧计数进行排序

根据图像的观测开始时间t_i和图像帧计数f_i，对所有图像进行排序。以观测开始时间作为首要排序依据，对于任意两幅图像I_j和I_k，若$t_j < t_k$则I_j排在I_k前面；若$t_j = t_k$且$f_j < f_k$，则I_j排在I_k前面。

（4） 左右旋图像分组

对于排序后的图像，定义左旋图像为Q_L，右旋图像为Q_R。通常情况下，观测时间相同、

帧计数相邻的有两幅图像，按照观测顺序判定前一个为Q_L，后一个为Q_R。异常情况下，如果存在超过两幅观测开始时间相同的图像，则联合帧计数信息，按照Q_L在前、Q_R在后的顺序将帧计数相邻的两幅图像匹配为一组；一个观测时间仅有一幅图像，默认其为Q_L。

（5） QUV 图像分组

A. 常规观测模式处理

当观测模式为常规观测时，如果相邻两组 QU 或 UV 图像之间的观测时间间隔小于等于 40 秒，则它们视为同组。设两组图像的观测时间分别为t_a和t_b，若$|t_a-t_b|\le 40$，则这两组图像视为同组。

对于缺少 U 图像的相邻 QV 图像组，如果观测时间间隔小于等于 80 秒，则这些 QV 图像视为同组。若$|t_c-t_d|\le 80$，其中t_c和t_d为相邻 QV 图像组的观测时间，则视为同组

B. 快速观测模式处理

在快速观测模式下，相邻的 QU 或 UV 图像组如果观测时间间隔小于等于 13 秒，则它们视为同组。若$|t_a-t_b|\le 13$，其中t_a和t_b为同组。

缺少 U 图像的 QV 图像组如果观测时间间隔小于等于 26 秒，则它们视为同组。即若$|t_c-t_d|\le 26$，其中t_c和t_d为缺少 U 图像的 QV 图像组的观测时间，则视为同组。

（6） 写入 FITS 文件

将经过分组处理后的 QUV 图像组合写入 FITS 文件中，确保数据以标准的天文数据格式保存，便于后续分析。

4.3 LST 载荷 SCIWL 探测器图像匹配算法

LST 载荷 SCIWL 探测器图像需要按照观测开始时间和高低增益类型进行分组，将观测开始时间间隔小于等于 2s 的短曝光高增益、短曝光低增益、长曝光高增益、长曝光低增益 4 幅图像分在一组，并将一组图像数据打包成一个 fits 文件。匹配算法如下所示：

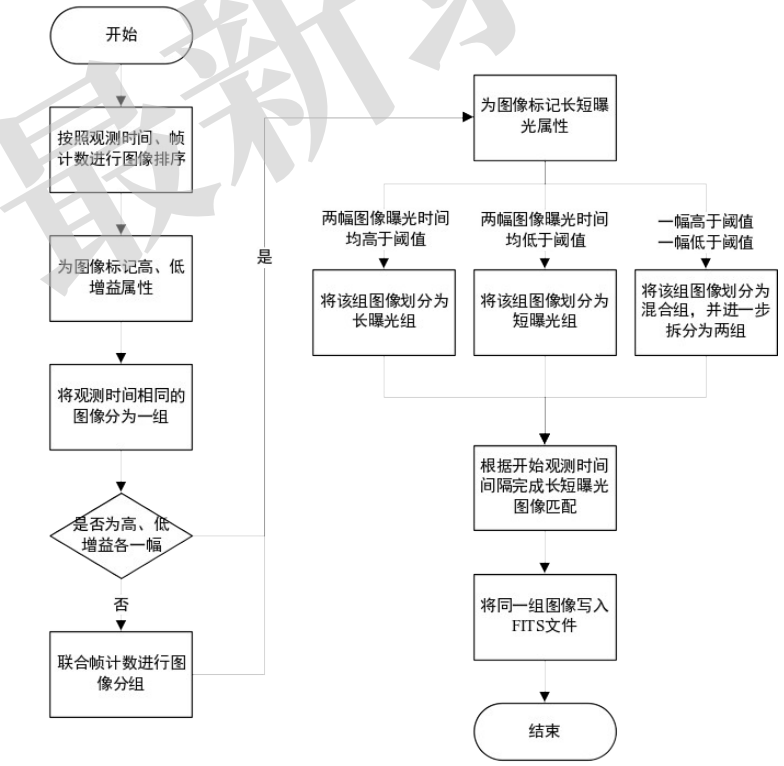


图 10 LST 载荷 SCIWL 探测器长短曝光、高低增益图像匹配算法流程图

Figure 10. Workflow of the image matching algorithm for short/long exposure and high/low gain modes of the sciwl detector on the lst payload

最新录用

（1） 基于观测开始时间与帧计数进行排序

在对 SCIWL 图像进行处理时，首先从图像头中提取图像的观测开始时间t_i和图像帧计数f_i。以观测开始时间作为首要排序依据，对于任意两幅图像I_j和I_k，若$t_j<t_k$则I_j排在I_k前面。当观测开始时间相同时，以帧计数为次要依据进行进一步排序，即若$t_j=t_k$且$f_j<f_k$，则I_j排在I_k前面。通过这种方式实现对图像的有序整理。

（2） 高低增益图像匹配

A. 图像类型标志位提取与分类

从 SCIWL 图像头中提取图像类型标志位，依据此标志位将图像标记为高增益图像或低增益图像。设图像I_i的增益类型标志为g_i，若$g_i=1$表示高增益图像，$g_i=0$表示低增益图像。

B. 正常情况匹配

通常情况下，观测开始时间相同的图像中会有高、低增益图像各一幅。此时，对于图像I_j和I_k，若$t_j=t_k$且$g_j=1$、$g_k=0$，则将这两幅图像匹配为一组，以便后续处理。

C. 异常情况处理

若存在超过一幅观测开始时间相同的高/低增益图像，则结合帧计数信息进行匹配。按照高增益在前、低增益在后的顺序，对于图像I_j和I_k，若$t_j=t_k$且$g_j=1$、$g_k=0$，同时帧计数满足相邻关系，即$|f_j-f_k|=1$，则将这两幅图像匹配为一组。若出现匹配不到配对图像的情况，则将其单独分组。

（3） 基于高低增益组合的长短曝光图像匹配

A. 高低增益图像整体标记

将匹配分组得到的一组高低增益图像视为一个整体，根据图像头中的曝光持续时长e_i对该整体进行标记。

B. 长曝光组、短曝光组和长短曝光混合组判定

设阈值为 T，对于一组图像中的两幅图像I_j和I_k，若$e_j>T$且$e_k>T$，则该组为长曝光组；若$e_j<T$且$e_k<T$，则该组为短曝光组；若$e_j>T$且$e_k<T$或者若$e_j<T$且$e_k>T$，则该组为混合组，此时需要将该组图像拆分为长曝光、短曝光两组。

C. 最终匹配规则

按照短曝光组在前、长曝光组在后，且两组图像观测开始时间间隔小于设定阈值的规则，将其匹配为一组图像。即对于短曝光组图像集合S和长曝光组图像集合L，若$|t_s-t_l|<T$（其中为t_s短曝光组中某图像的观测开始时间，t_l为长曝光组中某图像的观测开始时间，T 为时间间隔阈值），则将这两组图像匹配为一组。

（4） 写入 FITS 文件

将匹配得到的一组图像按照短曝光高增益、短曝光低增益、长曝光高增益、长曝光低增益的顺序写入 FITS 文件，并添加图像关键元数据信息，按照产品格式输出为 SCIWL 载荷 0C 级产品文件。

除 SCIWL 探测器外，LST 载荷的其余三个探测器（SCIUV、SDI 和 WST）虽不涉及偏振片和多组图像匹配，但其图像数据的预处理流程在整体结构上与 SCIWL 类似，亦包括基于观测时间和图像属性的信息进行图像组合与质量控制。相比之下，这些通道的图像匹配逻辑更为简单，主要聚焦于高低增益图像的配对与异常检测，因而处理流程更加直接，可由相同的处理框架中以模块化方式实现。

5 系统部署运行

5.1 系统的部署与架构

ASO-S 卫星数据预处理系统运行于国家空间科学数据中心的业务集群环境中。业务集群由三个关键部分共同构建。首先是高性能计算集群，它由调度管理节点和处理计算节点组成。调度管理节点采用双路机架式服务器，以主备运行模式（Active/Standby）确保系统的稳定可靠。处理计算节点运用四路机架式服务器，依据业务具体需求，通过调度管理算法实现硬件资源的科学分配。其次是超融合计算环境，同样配备四路高性能服务器，借助 FusionSphere 虚拟化软件对物理服务器的各类硬件资源进行解耦处理，使得单个物理服务器能够并行运行多个独立的虚拟机，大幅提高资源利用效率。最后是分布式文件存储服务器。这三个部分之间依靠高速互

联交换机实现数据的高效传输与交换，有力保障系统的顺畅运作。

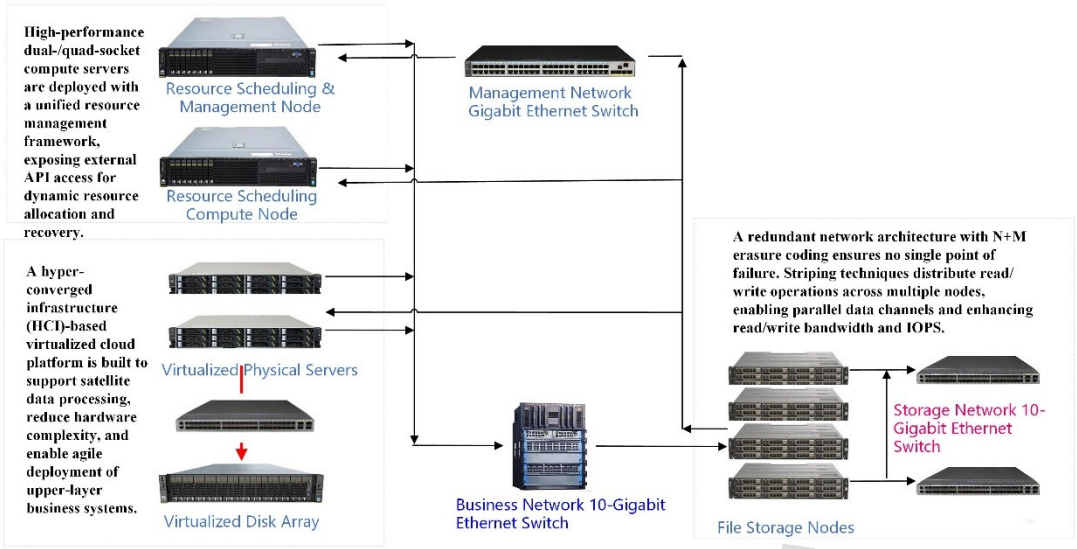


图 11 ASO-S 卫星数据预处理系统部署

Figure 11. Deployment of the aso-s satellite data preprocessing system

计算集群由 8 台物理计算节点组成，整体配置包括 544 个 CPU 和 2TB 的内存，单节点最大内存为 256GB。该集群不仅支持 ASO-S 卫星，还同时承载怀柔一号（Gravitational Wave High-energy Electromagnetic Counterpart All-sky Monitor，GECAM）、悟空号（Dark Matter Particle Explorer，DAMPE）及慧眼（Hard X-ray Modulation Telescope，HXMT）等多颗空间科学卫星任务。为了优化计算资源的分配，系统设计了“共用动态分配+专用固定占用”的调度方法，其中专用资源是基于各卫星数据的数据量、数据时效性要求进行固定分配，为该卫星数据处理任务独占使用。共用动态分配的计算资源则根据各卫星任务的实时需求进行动态调整。当前为 ASO-S 卫星数据预处理系统分配了 2 台专用计算服务器，逻辑 CPU 总数为 256 个，内存大小达到 500GB，单机最大内存约为 250GB，具体如表-2 所示。

表-2 服务器节点配置信息

Table 2. Hardware configuration of server nodes

	CPU	内存	硬盘	操作系统
共用动态分配服务器 1	128	256GB	1TB	CentOS Linux 7 (Core)
共用动态分配服务器 2	32	256GB	1TB	CentOS Linux 7 (Core)
共用动态分配服务器 3	32	256GB	1TB	CentOS Linux 7 (Core)
DAMPE 专用服务器	32	256GB	1TB	CentOS Linux 7 (Core)
HXMT 专用服务器	32	256GB	1TB	CentOS Linux 7 (Core)
GECAM 固定专用服务器	32	256GB	1TB	CentOS Linux 7 (Core)
ASO-S 固定专用服务器 1	128	256GB	1TB	CentOS Linux 7 (Core)
ASO-S 固定专用服务器 2	128	256GB	1TB	CentOS Linux 7 (Core)

5.2 系统运行

ASO-S 卫星每天下行约 6 轨数传数据，每次下行原始数据量约 50GB，每次下行数据包括卫星平台数据、科学载荷工程数据和科学探测数据。当卫星下行原始数据到达数据预处理系统后，该自动会自动依据文件传输完成通知触发各类科学数据处理，该系统每天可生成 ASO-S 卫星各级科学、工程和辅助数据产品约 230 余种，总体量可达 3TB。

ASO-S 卫星根据数传地面站过境安排，每天约有 6 次~8 次的科学数据下行机会，由于每次下行的数据中包含的数据类型以及数据量都会根据载荷的工作模式调整而有区别，因此数据处理与产品生成的时间也会逐渐变化。根据 ASO-S 科学卫星工程任务要求，卫星单次过境的

数据从落地到向科学应用系统分发不超过 8 小时。经过对该流程中各个环节步骤进行性能指标分解，数据预处理系统需要能够在 4 小时内完成数据处理。根据卫星在轨运行统计结果，如图 12 所示，本文设计并实现的数据预处理系统能够满足 ASO-S 科学卫星工程任务要求。

为验证数据预处理系统的实际运行性能，本文对 ASO-S 卫星多日在轨数据的处理情况进行了统计与分析。如图 12 所示，ASO-S 卫星每日下行数据量在 50GB 至 180GB 之间波动，处理系统在保持稳定运行的前提下，数据处理时间基本控制在 75 至 150 分钟之间，体现出良好的可扩展性与并行处理能力。尽管个别轨道段数据量较大，但系统处理时间未出现显著延迟，表明数据预处理系统在高负载条件下仍具有较强的稳定性。

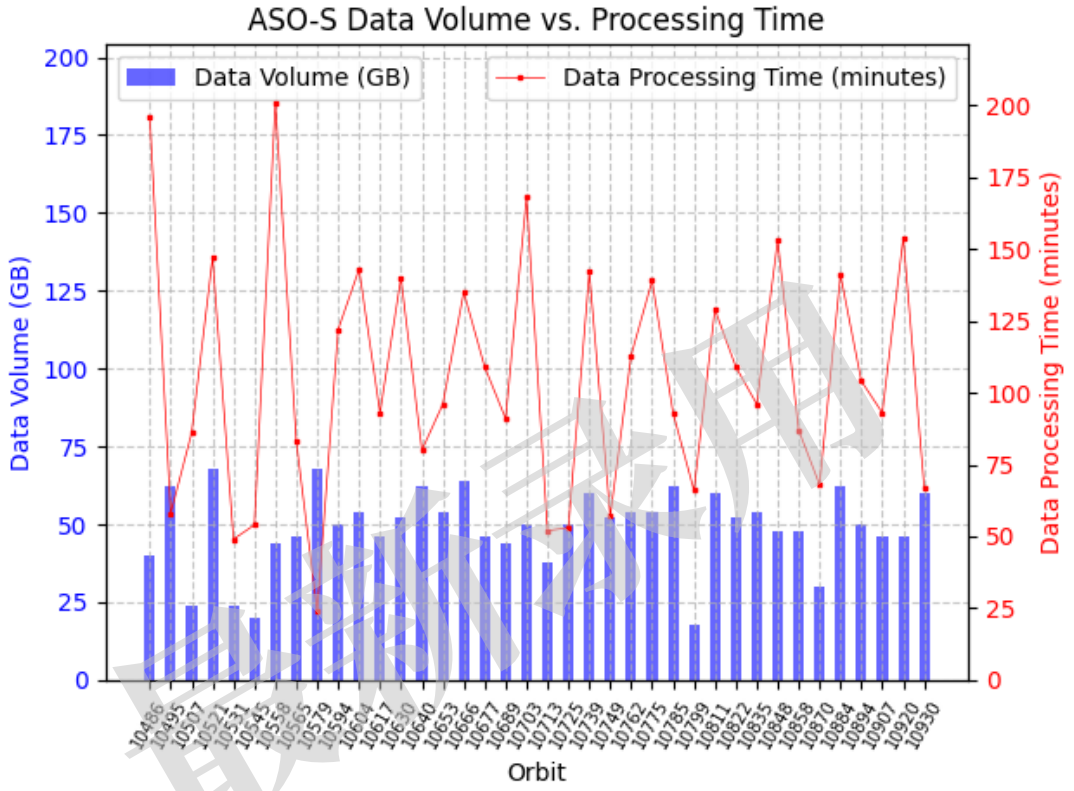


图 12 ASO-S 卫星数据量与数据处理时间对比图

Figure 12. Comparison between data volume and processing time for the aso-s satellite

进一步地，为评估不同载荷处理流程的效率，图 13 展示了 FMG 载荷图像数据的处理性能情况。每轨图像数量在 800 至 2000 幅之间，处理时间多数保持在 40 分钟以内。FMG 图像格式规范、结构相对单一，预处理过程主要涉及图像解码、时间排序和数据质量标识等操作，因此整体处理流程较为稳定。

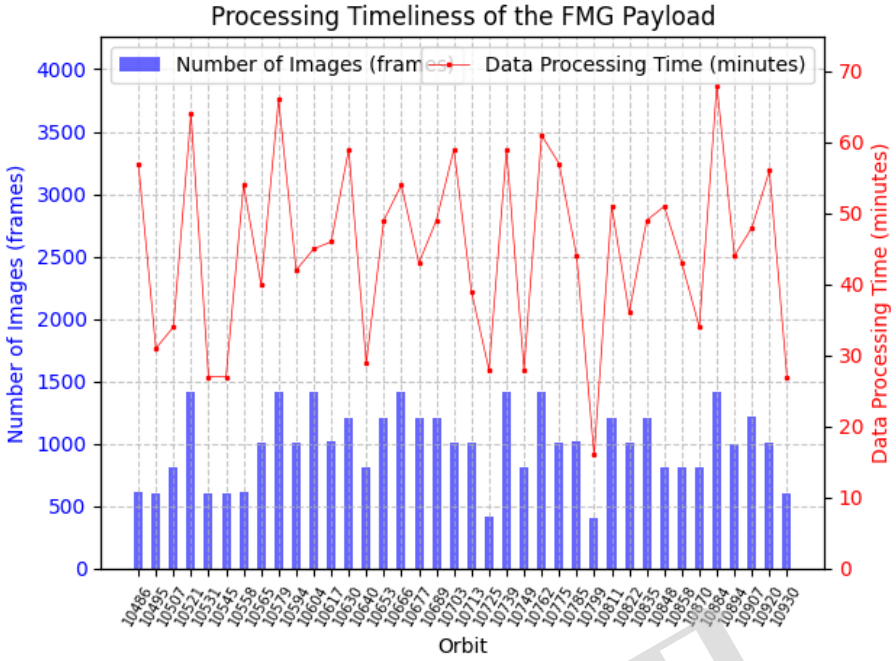


图 13 ASO-S 卫星 FMG 载荷图像数据处理时效性

Figure 13. Timeliness of image data processing for the fmg payload on the aso-s satellite

相比之下，LST 载荷的数据处理复杂度明显更高。图 14 展示了 LST 图像数据处理过程中的图像分组与合并算法的时效性评估结果。LST 每轨图像数量可达 2 万幅以上，且存在不同曝光、增益设置下的图像组合需求，需进行四图像一组的精确匹配。处理时间因此波动范围更大，但大多数情况下仍控制在 60 分钟以内，显示出本文提出的分组合并算法在高复杂度输入场景下的可行性与高效性。

综合上述结果，数据预处理系统在面对大体量、多源异构数据的情况下，仍能保障良好的处理效率和运行稳定性，验证了系统架构与关键算法设计的有效性，能够满足 ASO-S 任务长期在轨运行期间的实际需求。

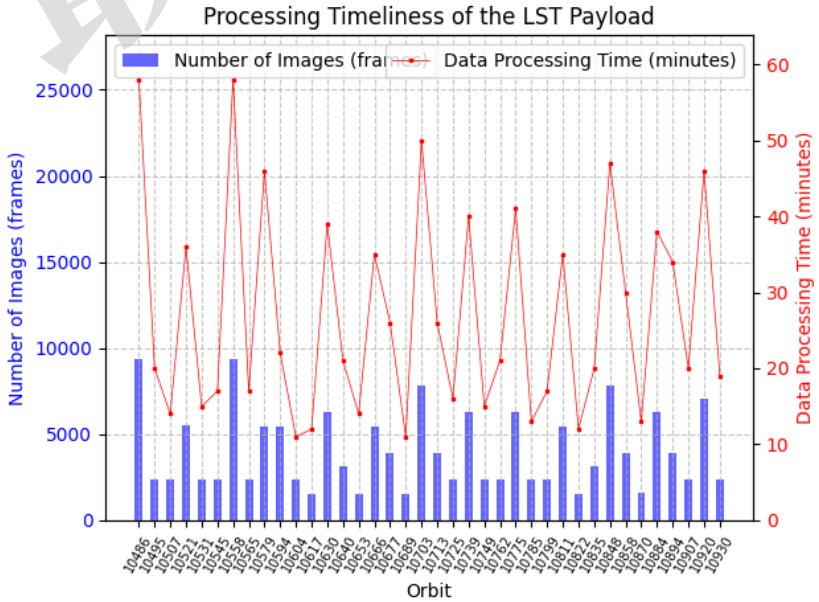


图 14 ASO-S 卫星 LST 载荷图像分组算法处理时效性

Figure 14. Timeliness of the image grouping algorithm for the lst payload on the aso-s satellite

6 结语

本论文围绕 ASO-S 卫星数据预处理系统展开了深入的研究与设计实现。通过对 ASO-S 卫星数据预处理系统需求以及科学载荷数据结构、处理特点和难点的详细分析，设计并实现了一套高效、稳定的数据预处理系统。在系统设计方面，提出了包含传输帧及源包层数据处理、载荷包层数据处理、产品格式化输出和任务与状态管理四个部分的系统架构，明确了各部分的功能与职责。同时，设计了一套综合处理流程，能够处理多种虚拟信道数据以及多载荷数据，其中奇偶双通道数据合并处理算法、FMG 载荷图像匹配算法、LST 载荷 SCIWL 仪器图像匹配算法等关键算法的应用，有效解决了数据处理过程中的核心问题。

经过测试验证，ASO-S 卫星数据预处理系统运行稳定，能够准确地对卫星下行的科学数据进行快速预处理和质量分析，生成符合要求的各级数据产品，功能和性能达到了预期的设计目标。该系统已成功应用于 ASO-S 卫星地面系统，有力地支持了卫星在轨运行，为后续的太阳观测数据分析提供了重要保障。然而，我们也认识到本系统可能存在一些潜在的改进空间。例如考虑进一步提升数据处理的效率；针对下行数据缺失，下行数据质量较差的情况下，如何进行数据产品质量的分析与控制，提升输出科学数据质量。期待在后续的工作中能够不断改进和完善该系统，以更好地适应科学研究需求。

综上所述，本研究为 ASO-S 卫星数据预处理提供了一种可行的解决方案，对我国太阳物理领域的空间探测研究具有重要的意义。同时，希望本研究能够为相关领域的数据预处理系统设计与实现提供有益的参考和借鉴。

参考文献

[1] GAN W Q, FENG L, SU Y. A Chinese solar observatory in space[J]. *Nature Astronomy*, 2022, **6**: 165. DOI: 10.1038/s41550-021-01593-9

[2] GAN W Q. Progress report on ASO-S[J]. *Chinese Journal of Space Science*, 2022, **42**(4): 565-567. DOI: 10.11728/cjss2022.04.yg07

[3] GAN Weiqun, ZHU Cheng, CHEN Bin, *et al*. China's first comprehensive space solar observatory[J]. *Acta Astronomica Sinica*, 2022, **63**(6): 60-1-60-3. DOI: 10.15940/j.cnki.0001-5245.2022.06.001 (甘为群, 诸成, 陈斌, 等. 中国首颗综合性太阳探测卫星[J]. 天文学报, 2022, **63**(6): 60-1-60-3. DOI: 10.15940/j.cnki.0001-5245.2022.06.001)

[4] HUANG Yu, DEGN Lei, XIONG Weiming, *et al*. Introduction of ASO-S satellite mission[J]. *Space International*, 2022(12): 12-16. DOI: 10.3969/j.issn.1009-2366.2022.12.004 (黄宇, 邓雷, 熊蔚明, 等. "先进天基太阳天文台"卫星任务概况[J]. 国际太空, 2022(12): 12-16. DOI: 10.3969/j.issn.1009-2366.2022.12.004)

[5] MA Fuli, SHI Tao, CHEN Ling, *et al*. Design of big data processing system supporting multi-satellites and multi-tasks[J]. *Big Data Research*, 2021, **7**(5): 2021045. DOI: 10.11959/j.issn.2096-0271.2021045 (马福利, 石涛, 陈玲, 等. 面向多星多任务的大数据处理系统设计[J]. 大数据, 2021, **7**(5): 2021045. DOI: 10.11959/j.issn.2096-0271.2021045)

[6] GAO Lijing, CHEN Zhimin, WANG Chunmei, *et al*. Real-time processing method of space science satellite downlink data based on streaming computing[J]. *Computer Simulation*, 2023, **40**(7): 26-31, 290. DOI: 10.3969/j.issn.1006-9348.2023.07.005 (高立京, 陈志敏, 王春梅, 等. 基于 Flink 的空间科学卫星数据实时处理方法[J]. 计算机仿真, 2023, **40**(7): 26-31, 290. DOI: 10.3969/j.issn.1006-9348.2023.07.005)

[7] JI Zhen, TONG Jizhou, HU Xiaoyan, *et al*. Application of data organization framework of space science[J]. *China Science & Technology Resources Review*, 2022, **54**(1): 83-88, 96. DOI: 10.3772/j.issn.1674-1544.2022.01.010 (纪珍, 佟继周, 胡晓彦, 等. 空间科学数据产品组织模型的应用研究[J]. 中国科技资源导刊, 2022, **54**(1): 83-88, 96. DOI: 10.3772/j.issn.1674-1544.2022.01.010)

[8] CHEN Jizhe, ZHANG Haiying, ZHEGN Zhaoying, *et al*. Design and simulation of a flexible-support tip/tilt mirror in the ASO-S/FMG image stabilization system[J]. *Acta Astronomica Sinica*, 2020, **61**(4): 35-1-35-11. DOI: 10.15940/j.cnki.0001-5245.2020.04.004 (陈继喆, 章海鹰, 郑兆瑛, 等. ASO-S/FMG 稳像系统的柔性支撑摆镜设计与仿真[J]. 天文学报, 2020, **61**(4): 35-1-35-11. DOI: 10.15940/j.cnki.0001-5245.2020.04.004)

[9] FENG L, LI H, CHEN B, *et al*. The lyman-alpha solar telescope (LST) for the ASO-S mission – III. data and potential diagnostics[J]. *Research in Astronomy and Astrophysics*, 2019, **19**(11):

162. DOI: 10.1088/1674-4527/19/11/162
[10] CHEN Dengyi, ZHANG Zhe, JIANG Xiankai, *et al*. Development of an X-ray modulation characterization system for HXI payload onboard ASO-S mission[J]. *Chinese Journal of Space Science*, 2023, **43**(4): 747-757. DOI: 10.11728/cjss2023.04.2022-0013 (陈灯意, 张哲, 江贤恺, 等. ASO-S 卫星 HXI 调制定标装置设计与验证[J]. 空间科学学报, 2023, **43**(4): 747-757. DOI: 10.11728/cjss2023.04.2022-0013)
[11] ZHANG Xinyu, YANG Jiasen, ZHANG Xuefeng, *et al*. High performance real-time odd-even channels combination algorithm for transmission baseband data using dual-polarized technology[J]. *Chinese Journal of Space Science*, 2022, **42**(6): 1222-1229. DOI: 10.11728/cjss2022.06.211125125 (张鑫宇, 杨甲森, 张雪丰, 等. 极化复用方式下高性能数传基带奇偶合路实时处理方法[J]. 空间科学学报, 2022, **42**(6): 1222-1229. DOI: 10.11728/cjss2022.06.211125125)
[12] PEI Xin, TONG Jizhou, MA Fuli, *et al*. Research on source packet data splicing and sorting strategy for ASO-satellite[J]. *Computer Engineering*, 2025, **51**(5): 166-176. DOI: 10.19678/j.issn.1000-3428.0068497 (裴欣, 佟继周, 马福利, 等. ASO-S 卫星源包数据拼接与排序策略研究[J]. 计算机工程, 2025, **51**(5): 166-176. DOI: 10.19678/j.issn.1000-3428.0068497)
[13] MA Fuli, CHEN Ling, LI Bing, *et al*. Design and implementation of GECAM preprocessing pipeline[J]. *Astronomical Research & Technology*, 2022, **19**(3): 274-282. DOI: 10.14005/j.cnki.issn1672-7673.20210611.001 (马福利, 陈玲, 李冰, 等. GECAM 卫星快速预处理流程设计与实现[J]. 天文研究与技术, 2022, **19**(3): 274-282. DOI: 10.14005/j.cnki.issn1672-7673.20210611.001)

第一作者：于勤思 男，1992 年 5 月出生，中国科学院国家空间科学中心副高级工程师，主要从事卫星地面数据处理技术方向的研究。



通信作者：马福利 男，1986 年 10 月出生，中国科学院国家空间科学中心高级工程师，主要从事卫星地面数据处理技术和大数据计算与存储技术的研究等。



最新录用