Space VLBI Data Recording Format RDF Parsing and Data Decoding Postprint

Authors: Zhang Hao, Zhang Juan, Liu Lei, Zheng Weimin, Rui Ping

Date: 2019-11-13T00:00:00+00:00

Abstract

Space VLBI, owing to its longer baselines, can achieve higher angular resolution than ground-based VLBI at the same observation frequency. China is currently advancing research in space VLBI science and technology. The Russian RadioAstron space VLBI project concluded in 2018, and its successful experience can serve as a valuable reference. The project employed a specialized data format distinct from conventional ground-based VLBI, RDF (RadioAstron Data Format), to record observational data. It utilized equipped onboard or ground uplink hydrogen atomic frequency standards to provide frequency references. After acquisition and quantization, the data were transmitted to ground stations, time-tagged, and recorded on hard disks. To process RadioAstron data, this paper analyzes the RDF format, accomplishes RDF data decoding and VLBI correlation processing, and subsequently conducts a comparative analysis of the characteristics between universal ground VLBI data formats such as Mark5B and VDIF and the RDF format. These efforts will accumulate experience for the design and processing of data formats for China' s future space VLBI initiatives.

Full Text

Analysis and Decoding of Space VLBI Data Recording Format RDF

Hao Zhang[1,2], Juan Zhang[1], , Lei Liu[1], , Weimin Zheng[1,3], #, Ping Rui[1]

[1]Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China

Shanghai Key Laboratory of Navigation and Positioning, Shanghai 200030, China

Corresponding author: Weimin Zheng, zhwm@shao.ac.cn

Abstract

Space VLBI achieves higher angular resolution than ground-based VLBI at the same observation frequency due to its significantly longer baselines. China is actively advancing research in space VLBI science and technology. The Russian RadioAstron space VLBI project, which concluded in 2018, offers valuable successful experience for future missions. RadioAstron employed a specialized data format called RDF (RadioAstron Data Format) that differs from conventional ground VLBI formats. The system utilized either onboard or ground-based uplink hydrogen atomic frequency standards to provide frequency references. After acquisition and quantization, data was transmitted to ground stations, time-tagged, and recorded on hard disks. This paper analyzes the RDF format, completes RDF data decoding and VLBI correlation processing, and compares RDF with common ground VLBI formats such as Mark5B and VDIF. This work will accumulate essential experience for designing and processing space VLBI data formats in China' s future missions.

Keywords: Space VLBI; RDF format; Data decoding; Mark5B

Very Long Baseline Interferometry (VLBI) is currently the astronomical observation technique with the highest angular resolution, widely applied in astronomical observations, geodesy, and deep space exploration [1]. Its resolution depends on both antenna aperture and baseline length between antennas. As VLBI scientific research and engineering missions such as Mars exploration advance, increasingly higher demands are placed on VLBI spatial resolution. To further improve the resolution of VLBI networks, longer baselines are required. However, constrained by Earth' s dimensions, ground-based VLBI baselines cannot exceed Earth' s diameter, making research on Space Very Long Baseline Interferometry (SVLBI) increasingly urgent.

As early as the early 1980s, the European Space Agency (ESA) and NASA jointly proposed the QUASAT (QUASAR SATELLITE) program to ESA. The plan involved placing a 15-meter antenna in space and conducted extensive studies during the pre-research phase on SVLBI technical requirements, observation performance, data analysis and imaging capabilities, and expected scientific objectives, providing valuable experience for subsequent SVLBI development [2]. Between 1986 and 1988, NASA conducted three VLBI observations using the 4.9-meter antenna on the TDRSS (Tracking Data and Relay Satellite System) and two 64-meter radio telescopes on the ground, targeting a sample of compact extragalactic radio sources at frequencies of 2.3 GHz and 15 GHz. Interference fringes were detected for most sources, confirming the feasibility of SVLBI technology [2]. During this period, the former Soviet Union proposed the RadioAstron program. Affected by the dissolution of the Soviet Union, the project was temporarily suspended before resuming in 2010 for official SVLBI launch and observation missions. Japan also proposed the VSOP (VLBI Space Observatory Programme) space interferometry plan in February 1997, which provided the first multi-telescope space VLBI imaging capability for high-resolution mapping of active galactic nuclei and radio stars [2].

Currently, only Japan and Russia have completed formal space VLBI projects. Shanghai Astronomical Observatory is advancing China' s space VLBI development. The "Space Low-Frequency Radio Observatory" project plans to launch two large-aperture low-frequency radio telescopes into near-Earth space to form Earth-space and space-space interferometric capabilities. A Moon-Earth VLBI program is also underway. Due to significant differences between space and ground environments, space VLBI research faces several urgent challenges, including maintaining stable signal links between space radio telescopes and ground tracking stations and establishing practical and effective space VLBI delay models. This paper focuses on the space VLBI data terminal recording format issue. By studying RadioAstron raw observation data, we analyze its data recording format, complete data decoding, and enable correlation processing with ground VLBI observation data. Currently, multiple ground-based VLBI formats are used internationally, while space VLBI formats (such as RDF) are only used by Japan and Russia.

The MARK5B data format was designed by the Haystack Observatory at MIT as a VLBI data terminal recording format. It employs hard disk storage, which offers low cost and facilitates data transmission via the Internet, thereby promoting e-VLBI technology development. The MARK5B format strictly complies with the VSI-H (VLBI Standard Interface Hardware Interface Specification) and is compatible with all data recording formats using VSI interfaces, effectively reducing the frequency of VLBI data recording hardware updates [3]. MARK5B data is divided into multiple disk frames, each consisting of a fixed-size header followed by a data block. The header contains synchronization words, CRC checksums, frame numbers within each second, and start UTC time, effectively ensuring data integrity [4]. Due to frame loss and disorder phenomena, MARK5B requires statistical calculation of data efficiency every second as a reference for data reliability and network diagnostics [4]. All five stations in China' s CVN network use the China Data Acquisition System (CDAS) developed by Shanghai Astronomical Observatory [5]. CDAS records raw observation data in MARK5B format and has successfully completed China' s lunar exploration engineering tasks.

VDIF (VLBI Data Interchange Format) is a new standardized VLBI data exchange format that supports real-time and near-real-time e-VLBI as well as disk file storage [6]. Based on variable-length data frames, VDIF frame headers carry critical information such as frame start time, frame length, and thread number. The data frame length is specified in the frame header, allowing users to freely set the frame length to optimally match the chosen data transmission protocol [6]. For example, in real-time network transmission scenarios, the frame length can be appropriately selected so that each online data packet carries exactly one data frame. Additionally, VDIF supports multi-threading, where time-series data frames from the same frequency channel are stored in the same thread, while data from different frequency channels are stored in multiple parallel threads. For serial network data transmission, multiple parallel threads are merged into a single serial "data stream" [6]. VDIF' s advantages lie in data network exchange and compatibility, though currently few devices support VDIF, limiting its application.

The Russian space VLBI project RadioAstron adopted the RDF format for ground data acquisition. Since RadioAstron did not carry large-capacity storage devices, acquired space VLBI data had to be transmitted to the ground in real time. The space VLBI telescope data acquisition actually comprises two parts: the VLBI space antenna and the ground tracking station. The VLBI space antenna receives raw VLBI observation signals, while the ground tracking station locks onto the space antenna' s downlink signal, acquires the VLBI data transmitted from the space antenna in real time, and records it in RDF format.

The space VLBI radio telescope systems of VSOP and RadioAstron employ a structure different from conventional ground VLBI stations (Figure 1 [Figure 1: see original paper]). To achieve all-weather real-time tracking of the space antenna' s elliptical orbit, multiple ground tracking stations must be established worldwide. One primary function of ground tracking stations is to provide a stable frequency reference for the space telescope [7], ensuring precise control of the space antenna' s sampling rate. Additionally, due to instability in the space antenna' s time and frequency system, ground tracking stations record the arrival time of space VLBI data at the tracking station to ensure accurate data reception timing. This timing increases with data field growth, and during correlation processing, the processor calculates the light travel time (distance between space antenna and ground tracking station divided by the speed of light) to retroactively determine the corresponding time at the space antenna for each data field. To minimize the data format while meeting scientific requirements, an RDF file only needs a time tag at the beginning of the file header, with subsequent data reception times calculated sequentially based on the data sampling rate. Since space-ground data transmission does not support complex data formats, RDF has a simple structure and may continue to be used in future space VLBI projects.

1 RDF Data Format

The RDF format is a disk-based VLBI data format in which the space radio telescope RadioAstron acquires VLBI data and transmits it to the ground at a specific frequency, where ground tracking stations record and output the data.

Ground tracking stations record time-continuous radio astronomy data in a set of files, with each file containing approximately 15 minutes of data, except for the last file. When a file reaches the specified data volume, the system creates a new file to record subsequent data. This process repeats until data recording terminates.

Files are named according to the "station name + UTC" convention, using the UTC time corresponding to the first VLBI data record in the file, which must be an integer second. The specific format is: "station_YYYYDDDHHMMSS.rdf" , where "station" is the station name, "YYYY" is the Gregorian year, "DDD" is the day of year, and "HH" , "MM" , and "SS" are the hour, minute, and second in UTC, respectively.

Unlike MARK5B, which divides data files into multiple equal-length disk frames with fixed-format headers containing frame information, a single RDF file contains only one file header, with all subsequent data being raw VLBI observation data, as shown in Figure 2 [Figure 2: see original paper].

1.1 File Header Format

The RDF file header occupies a fixed 256 bytes of disk space, with VLBI raw data recorded continuously starting from byte 257 until data sampling stops. The header records critical information including data start UTC time, data transmission rate, and reception mode. The structure is as follows:

Figure 3 [Figure 3: see original paper] marks the starting byte number and size of each data item in the header. The file type is a synchronization word consisting of 4 ASCII characters used to identify radio astronomy data files and must be "RDF1" . The header size is always 256 bytes, represented by the binary number "00000001 00000000" . UTC represents the date and time when data recording began, stored as an ASCII string in the form "YYYY DDD-hh:mm:ss" and occupying a fixed 18 bytes. The data rate is stored as a 2-byte ASCII string; RadioAstron' s data rate is 16 Mb/s, so it is stored as "16" . The reception mode is indicated by a single character with four options: "I" , "Q" , "A" , and "R" , with RadioAstron using "R" mode. The data source is stored as a 3-byte string, with ground telescopes marked as "DAS" and space telescopes as "DEC" . Bytes 63-160 are filled by the correlator, which obtains these parameters from the VEX file, where telescope coordinates are calculated from orbital data. Bytes 161-256 are divided into 16 channels, with each channel occupying 6 bytes to record frequency, sideband, and polarization information for that channel.

1.2 Data Word Format

The data acquisition terminal selects the required observation frequency channels from the intermediate frequency (IF) signals sent by the receiver, then converts them to baseband signals through frequency conversion, followed by digital sampling and encoding before recording on disk [8]. In practice, multiple frequency channels often need to be selected simultaneously [8].

The recording format of the data stream in the data block is determined by the number of channels and quantization method, with each channel corresponding to one bitstream. The MARK5B disk recorder can only record bitstreams in quantities of 1, 2, 4, 8, 16, or 32 [9]. Figure 4 [Figure 4: see original paper] illustrates the data unit recording format for 8 or 16 channels with 1-bit or 2-bit quantization.

For example, with 16 channels and 1-bit quantization, the data at bits 0, 16, 32, 48, 64, 80, etc., in the data block portion represent the bitstream from channel 0 recorded sequentially, while bits 1, 17, 33, 49, 65, 81, etc., represent the bitstream from channel 1. This pattern continues similarly for other channels. Figure 4 is only a schematic diagram; in reality, the channels are not arranged so regularly, and within multi-bit quantization modes, the bit positions within a channel are not necessarily adjacent. The number of channels and the bit arrangement within data units are specified by the VEX file provided by the ground tracking station.

Figure 5 [Figure 5: see original paper] shows a sample field from the VEX file specifying the bit arrangement rules. Figure 6 [Figure 6: see original paper] provides a schematic diagram of bit arrangement for 4-channel, 2-bit quantization.

2 VLBI Data Decoding

Data decoding is a preprocessing step before correlation processing, and different decoding algorithms are employed for different data formats. The key to designing a decoding algorithm lies in understanding the data storage structure. Through international space VLBI cooperation projects, we obtained partial RDF source files and developed a practical and feasible decoding algorithm specifically for RDF.

The detailed decoding algorithm flow is shown in Figure 7 [Figure 7: see original paper]. Since RDF has only one file header, we first read the header to extract the two most critical pieces of information for data decoding: the start UTC time and data rate. We then calculate the data volume per second (T) based on the data rate and cyclically read data of size T from the data block into newly created files named with integer-second UTC times until all data is read.

3 Data Validation

The correctness of the decoding process can be verified through correlation processing of the decoded data. The RDF data used in this study came from a 2014 observation experiment of the extragalactic radio source 0823+033 by the Russian space VLBI network. In addition to RadioAstron, the participating telescopes included the Arecibo telescope in the United States and the Westerbork telescope in the Netherlands, with data recorded in Mark5A and Mark5B formats, respectively. Before correlation processing at the Shanghai Astronomical Observatory VLBI Center, data from each station must be decoded and converted to a specific format for preprocessing input. After processing with the self-developed software correlator SCORR, clear interference fringes were successfully obtained (Figure 8 [Figure 8: see original paper]).

Figure 8 shows the space VLBI data correlation processing results, where the red portion represents the phase spectrum and the blue portion represents the amplitude spectrum. The Ar-Ra baseline shows obvious interference fringes (Ar denotes Arecibo, Ra denotes RadioAstron). These results demonstrate that the preprocessed data obtained through decoding is valid, verifying the correctness of the RDF format data decoding. The Wb-Ra baseline has a relatively low signal-to-noise ratio, likely because both telescopes have relatively small apertures (Westerbork is a synthesis radio telescope composed of 14 25-meter mesh antennas with an equivalent aperture of 66 meters, and its overall sensitivity is not as high as Arecibo; the RadioAstron space radio telescope has a 10-meter aperture), combined with the very long baseline (approximately 200,000 kilometers), resulting in weak fringes. Arecibo has a relatively large aperture (300 meters) and thus shows clearer fringes.

Our signal-to-noise ratio calculation method is consistent with HOPS (Haystack Observatory Postprocessing System) (see Equation (1)). The fringe signal-to-noise ratio for the Wb-Ar baseline is 654.32, for the Wb-Ra baseline is 10.27, and for the Ar-Ra baseline is 102.22.

$$SNR = \frac{Amp\sqrt{2BT_{ap}/n_{ap}}}{n_{lags} - 1}$$

Where *Amp* represents the amplitude obtained by summing all frequency points across all integration periods and channels (with the DC component from the first frequency point removed), B is the channel bandwidth, *T_ap* is the integration period, *n_lags* is the number of frequency points per channel, and *n_ap* is the sum of weights across all integration periods and channels. Weights are read from the correlator output files.

4 Comparison of Three VLBI Data Formats

Table 1 summarizes the key characteristics of the three data formats commonly used in ground and space VLBI: VDIF, MARK5B, and RDF.

The RDF format records only a single file header in addition to raw VLBI data, giving it higher recording efficiency compared to other formats. To ensure reliable data recording, the RadioAstron project employed an additional decoder. When the decoder detects data loss or offset, it creates a new RDF file to begin recording. Since data loss or offset occurs relatively infrequently and RDF files avoid recording frame header data used in the other two formats, this approach achieves both good data recording reliability and high efficiency.

The MARK5B format divides data files into multiple disk frames, each with a fixed length of 10,000 bytes, preceded by an additional 16-byte header. Although frame loss, out-of-order frame numbers, and insufficient frame length are inevitable, the UTC time and frame number within each second in the header can determine each frame' s position in the time series, and appropriate data error correction methods can improve the overall data efficiency of the file [4]. Therefore, the MARK5B format offers high data reliability. Additionally, MARK5B is fully compliant with the VSI standard, maintaining broad compatibility with existing VLBI systems [10]. Currently, the most widely used VLBI terminal systems internationally are the MK series [11], making MARK5B the most broadly applied format.

The VDIF format also uses multiple data frames, but unlike MARK5B, VDIF employs variable frame lengths (variable frame length means that frame length can be set according to different transmission protocols to match the capacity of individual data packets; the frame length is fixed within the same observation file) to ensure that each data packet of the adopted transmission protocol can exactly accommodate one VDIF data frame [6]. In this way, if packet loss occurs or packets are not delivered in order during transmission, the receiver can still recover the data timing and fill in missing data (with zeros) using header information within each packet. If traditional VLBI data recording formats such as MARK5B were used for network transmission, packet loss could cause bit shifts, and VLBI correlation processing is extremely sensitive to data bit shifts [4], potentially making the received data file unreliable. In contrast, VDIF offers higher network transmission security. VDIF was designed to address the inconvenience of data exchange caused by the diversity of data terminal recording formats currently used internationally. In addition to facilitating data exchange, VDIF also supports disk storage [6]. As data terminal equipment continues to be updated, VDIF will be more widely applied and become a broadband VLBI universal data format combining both data recording and network exchange functions.

In summary, RDF is a data recording format applied at space VLBI ground tracking stations, MARK5B is currently the most widely used ground-based VLBI data recording format, and VDIF, with its flexible design and high network transmission security, may become a universal format in the coming period. It should be noted that the MARK5B format was designed by Haystack Observatory around 2005 [3], while the MARK5A disk-based VLBI data recording format was designed in 2002 [10], and the RDF format was developed during the RadioAstron pre-research phase (1980s), earlier than both MARK5B and VDIF. Over time, RDF may also be replaced by more advanced data recording formats in the future.

This paper presents the first domestic analysis of the RDF format specific to the Russian space radio telescope RadioAstron, completes decoding of actual space VLBI data in RDF format, and successfully performs space VLBI correlation processing and validation using the SCORR software correlator. We have mastered the RDF format decoding method. Finally, we compare RDF with existing ground-based VLBI formats and analyze their respective characteristics. This research will facilitate subsequent in-depth analysis and processing of more RadioAstron actual measurement data and advance China' s space VLBI data processing technology research.

During this research, Dr. Andrianov Andrey from the Astro Space Center of the Lebedev Physical Institute of the Russian Academy of Sciences kindly responded via email regarding detailed questions about the RDF format. Professor Tetsuro Kondo, an international visiting scholar at Shanghai Astronomical Observatory, and Dr. Fengxian Tong provided substantial support for the data correlation processing work, and we express our gratitude to them.

References

[1] Xinying Zhu, Chunlai Li, Hongbo Zhang. Review of VLBI Technology for Deep Space Exploration and Current Status and Development in China [J]. Journal of Astronautics, 2010, 31(8): 1894-1897.
[2] Zhiqiang Shen. Research Status of Space VLBI [J]. Progress in Astronomy, 1998(2): 117-134.
[3] MIT Haystack Observatory. Mark5B System User' s Manual [EB/OL]. https://www.haystack.mit.edu/tech/vlbi/mark5/docs/Mark%205B%20users%20manual.pdf, 2006.
[4] Ying Guo, Weimin Zheng. A High-Performance VLBI Data Preprocessing Method and Its Software Implementation [J]. Annals of Shanghai Astronomical Observatory, Chinese Academy of Sciences, 2013: 114-122.
[5] Wenjun Yang, Longfei Hao. Development History and Future Prospects of VLBI Terminal Systems [J]. Astronomical Research & Technology, 2012, 09(4): 374-381.
[6] Whitney A, Kettenis M, Phillips C, et al. VLBI Data Interchange Format (VDIF) [J]. European VLBI for Geodesy & Astrometry Working Meeting, 2010, 194(2-5): 156-160.
[7] Cornwell T E, Fomalont E B. Synthesis Imaging in Radio Astronomy II [J]. 1989.
[8] Zhihan Qian. Application of Very Long Baseline Interferometry Technology in Deep Space Exploration [C]. Proceedings of the 7th Academic Annual Conference of the Deep Space Exploration Technology Professional Committee of Chinese Society of Astronautics. 2010.
[9] Xinying Zhu, Hongbo Zhang, Chunlai Li. Design and Implementation of VLBI MK5B Channel Detection Software [J]. Astronomical Research & Technology, 2011, 08(3): 236-241.
[10] Wenren Wei, Zhuhe Xue. A New Disk-Based VLBI Terminal System–

MK5A Terminal System [J]. Progress in Astronomy, 2004, 22(3): 269-274.
[11] Zhong Chen, Weimin Zheng, Xiao Chen. Establishment and Application of China' s e-VLBI Network [J]. Annals of Shanghai Astronomical Observatory, Chinese Academy of Sciences, 2015, 36: 136-147.

[1] Reference: Haystack Official Website: https://www.haystack.mit.edu/tech/vlbi/hops.html

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