Packaging Radio/Sub-millimeter Spectral Data in FITS

Zhong Wang

Smithsonian Astrophysical Observatory, Cambridge, MA 02138 U.S.A.
Email: zwang@cfa.harvard.edu

Abstract. We report on an experiment to incorporate the multiple extension FITS file format in building data processing software for spectroscopic observations in radio and millimeter-submillimeter wavelengths. By packaging spectral header information into table extensions and the actual spectral data into either table or image extensions, this approach provides a convenient and effective way to organize large amount of spectroscopic data, especially for pipeline-like processing tasks. Tests conducted based on a software package developed for the SWAS mission have demonstrated this to be a viable alternative to the conventional ways of organizing spectral data at these wavelengths.

1. Introduction: Characteristics of Submillimeter Spectral Data

Historically, spectral data were taken roughly one spectrum at a time, and most existing data formats reflect such a methodology. The traditional ways in which FITS can be used for spectral data are less than adequate in meeting the needs of a new generation of instruments that produce large amount of relatively homogeneous data, which nevertheless need to be reduced individually. A set of more efficient, better streamlined reduction procedures are necessary, which in turn require more careful considerations in data packaging. This is particularly important for projects that adopt non-interactive, pipeline-style data processing as their primary mode of operation.

The SWAS[1] (Submillimeter Wave Astronomy Satellite) mission is essentially a small orbiting radio telescope working at relatively high (about 500 GHz) frequencies (Wang 1995). It is designed to make simultaneous spectroscopic measurements at several sub-mm line wavelengths that are inaccessible from the ground. The spectral data have some special characteristics:

- Integration time for each individual "scan" is set to be short (on the order of a few seconds). High signal-to-noise ratios are achieved by co-adding a large number of such scans at the data processing stage;

---

[1]Details of the SWAS mission can be found on the internet in SAO's SWAS homepage at http://cfa-www.harvard.edu/cfa/oir/Research/swas.html

- Because of the noise properties and frequency shifts with time (due to the motion of the spacecraft as well as instrumental drifts), individual scans need to be calibrated separately before co-adding;

- More than one spectral band is recorded at any given time, thus a complete dataset consists of spectra of several different wavelength ranges from independent detectors. Yet they are of the same source and are taken under very similar observing conditions;

- Observations are broken up into individual "segments" each lasting typically 30–40 minutes. During each segment, many spectroscopic scans (aside from calibration data) are simple repeats under nearly identical conditions, and therefore share values of many header parameters with only slight variations (drifts).

The data reduction plan for the mission calls for a highly efficient pipeline processing scheme in which a large amount of spectroscopic data can be sorted, selected, calibrated and co-added based on some user-changeable criteria. All processing operations have to be recorded and traceable, and are preferably done with minimal human intervention.

2. Software Approach: Packaging Data in FITS

We have selected IRAF as our primary programming platform for the SWAS pipeline software, and use FITS as the main format for our data product. However, to store individual scan data in the conventional one- (single spectrum) or two-dimensional ("multispec" or echelle) format would be very inefficient and even confusing.

Our approach is to make use of the BINTABLE and IMAGE extensions of FITS and store spectral header and pixel data in these two types of extensions in a single FITS data file. The basic rules we adopted are:

- Each FITS file corresponds to a specific time interval of data taking (an observing segment). It contains a single BINTABLE extension and (optionally) several IMAGE extensions. It has a conventional header section containing some none-time-variable, shared parameters of the spectra;

- The table extension has one row per spectral scan taken, and stores all of the *time-variable* spectral header information corresponding to that scan (i.e., time, coordinates, frequency, noise statistics, etc);

- The pixel data of the spectrograph output (the spectrum itself) can be stored in one of two ways: they can either be individual "array elements" of a table row, or an image row similar to the two-dimensional arrays used to save Echelle and multi-fiber spectra in the optical.

  In case of using variable-length, vector (array) columns to store pixel data, a table row can have more than one array element, each representing data of a separate spectral band. The length of an array element is the number of pixels in a spectrum of that band.

In case of images, each image row corresponds to a row in the header parameter table. The number of columns of the image is the same as the number of pixels in a spectrum. There can be multiple image extensions in each data file, with different extensions representing data of different spectral bands.

The choice between using the table extension of FITS alone versus using table plus image extensions depends on the actual application. In principle, saving everything in a single table is structurally more appealing and efficient, but not many existing software tools can accept variable-length vector column FITS tables — which means more development work for basic tools. On the other hand, making use of 2-D images can facilitate data inspection procedures with various tools that examine FITS images. Its shortcoming is that information about a single spectrum is saved in two or more extensions, requiring care in managing them.

3.   Advantages and Efficiency Gains

Given the lack of a widely-accepted standard in formatting a largely homogeneous spectroscopic dataset, we find the packaging using FITS extensions to be a viable alternative to the conventional ways in which spectral data are saved:

- It provides a convenient way to organize, sort, edit, monitor and visualize the large spectral dataset (much of it is done with existing tools within IRAF or those directly working with FITS), while preserving all the essential information of each individual spectrum;

- The programming of calibration tasks to process a relatively large, nearly-homogeneous dataset is very much simplified. In particular, the spectral header parameters can be processed either independently or as table columns. It also means that, for example, a subset of the header parameters of a large number of spectra can be easily manipulated without even accessing the actual pixel data;

- The efficiency of data processing is clearly enhanced. Since spectral pixel data are always processed as arrays, the fractional time taken by file I/Os and data buffering is substantially reduced when, for example, those data are accessed as part of an image.

Many of the existing software tools that deal with FITS table and image data are readily available for use with these new data files, sometimes with minor modifications. This can significantly save the programming effort and shorten the software development cycle.

4.   Comments and Future Work

Despite the considerable merits of this approach, some problems remain to be addressed, and we are working on future enhancement of our software.

One of the main problems is the interface with existing software tools (such as visualization tools from certain existing tools packages). In practice, we were able to partly circumvent this shortcoming by writing customized interfaces, but the solutions are not always satisfactory. It appears that if a tabular style file format is deemed sufficiently important, a more fundamental (kernel-level) interface for spectral data in FITS should be developed.

The approach described in this paper is an explorative step in enhancing the use of FITS, especially in taking advantage of the multiple extension development and the use of vector columns in FITS tables. With the rapid advancement in database and data warehousing technology, one would ideally like to have a more rigorous database-like framework to deal with the management of the observational data. This, however, needs to be balanced with the requirement of cost-effective and quick prototyping of many projects' practical applications. The latter often means maximizing the use of existing software tools and adopting conventional approaches. The multiple extension FITS adopted by the community and being incorporated by NOAO and STScI's IRAF teams (e.g., Zarate & Greenfield 1996, Zarate 1998) and other software groups is an important and innovative development, which preserves the usefulness of many old tools while allowing new methods to be tested for astronomical data applications. We believe that more efforts are needed in exploring ways to take full advantage of this very useful data format.

References

Wang, Z. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne & J. J. E. Hayes (San Francisco: ASP), 402

Zarate, N. & Greenfield, P. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. George H. Jacoby & Jeannette Barnes (San Francisco: ASP), 331

Zarate, N. 1998, this volume