1981A&AS...44..371G

ASTRONOMY & ASTROPHYSICS
SUPPLEMENT SERIES

JUNE 1981, PAGE 371

Astron. Astrophys. Suppl. Ser. 44, (1981) 371-374

AN EXTENSION OF FITS FOR GROUPS OF SMALL ARRAYS OF DATA

E. W. GREISEN and R. H. HARTEN (*)

National Radio Astronomy Observatory Charlottesville, VA 22901, U.S.A.
(*) Netherlands Foundation for Radio Astronomy Dwingeloo, the Netherlands

Received June 11, accepted October 29, 1980

Summary.- An extension of the FITS tape format for the interchange of astronomical images is presented. This extension allows a FITS "image" to consist of a collection of "groups", where a group consists of a sub-array of data and one or more parameters describing the non-regular coordinates associated with the sub-array. This extension follows all the rules of basic FITS and only requires six new keywords, a few new values for standard keywords, and a new data stucture. The new format is ideally suited for collections of arrays, such as spectra or radioastronomical visibility data, which occur on a non-regular grid.

Key words : Tape Format - Data Processing - Data Transport.

1. Introduction.- In a recent paper, Wells, Greisen and Harten (1980, paper I) described the FITS ("Flexible Image Transport System") format for the interchange of astronomical images and other digital array data on magnetic tape. FITS provides a simple and flexible mechanism for the unambiguous transmission of n-dimensional, regularly spaced data arrays together with an unlimited number of parameters which are used to describe and document the data arrays. Although the basic FITS discussed in Paper I is very powerful, it is not optimum for the transmission of large numbers of small arrays. In many cases, the small arrays occur in sets characterized by numerous common parameters and one or more regularly spaced coordinates axes. However, the small arrays are distributed irregularly on one or more additional coordinate axes.

A simple example would be radio or optical spectral line data. Each spectrum would be an array of data sampled uniformly in frequency or velocity ; however, the spectra could be distributed in a non-uniform manner on the sky. All arrays might have common rest frequency, spacing, etc., but the spatial coordinates would vary with each array. Another example would be a large photographic plate which has been digitized in only a number of small regions surrounding the objects of interest. A more esoteric, but very important, example is radio astronomy synthesis telescope data. All of these data types share the common characteristic of consisting of a group of small arrays which occur in an essentially random manner in one or more axes. What is needed is a way to organize these arrays and their accompanying descriptive parameters in a simple manner which can easily be written onto, and deciphered from, magnetic tape. This paper presents an extension of the FITS format designed to accomodate such sets of small arrays.

2. Summary of the basic FITS.- The goals, rules, structures and parameters of FITS were described in detail in paper I. The basic format can be summarized as follows. Each "image" (n-dimensional data array) is written as one file or label on a tape using a physical block size or record length of 23040 bits (2880 8-bit bytes). Each file contains "header" records followed by data records. The header records contain the parameter information in the form of 36 80-byte card images per record. The character code is 7-bit ASCII (also known as ISO/R646 and CCITT Alphabet # 5), right justified in 8-bit bytes. The header records contain the information necessary to describe the array (number of axes, points per axis, etc.) as well as additional information as deemed necessary by the user. The first data record occurs immediately after the last header record. The image data are packed into the records with maximum efficiency and no fixed relationship between row number and data record number need exist. The data are stored on tape in one of three data formats : 8-bit unsigned binary integers, and 16-bit and 32-bit twos complement signed binary integers. The header record card images have the format :

Keyword = value / comments

An extensive list of recommended keywords is given in paper I. The extension discussed in this paper involves a few additional keywords, some fixed conventions with the required keywords, and a revised structure in the data records. These changes are discussed in the next section.

3. An extension of FITS for ordered sets of small arrays.- This extension is a slight modification of the existing FITS and retains all of the rules, conventions and keywords of the basic FITS as described in paper I. The extension involves the addition of six new keywords, a few standard keyword values, and a slightly revised data structure.

The basic problem is how to handle a number of small arrays of data as well as a limited number of "coordinate" parameters which are associated with each array and which do not occur in a uniform grid. Most of the parameters describing the small arrays or sub-arrays are the same for each sub-array and can be described in the standard FITS header records. The parameters which vary with each sub-array must be associated with it in some manner. The way we chose to solve this problem was to introduce the concept of a "group". A group consists of a number of parameters (written as binary integers on the tape) followed by the associated sub-array of data. The groups of parameters and sub-arrays are packed into the fixed length (2880 byte) data records. A schematic illustration of this type of structure is given in figure 1. Note that all data within a group are written in binary integer form in the data records.

The standard FITS header records are used to describe the data and its organization and to control access to the data records. The additional keywords and standard keyword values necessary for this extension are given in table I. At least 4 keywords are required in the first header record of each file. They are, in order, SIMPLE, BITPIX, NAXIS, NAXIS1, NAXIS2, ... These convey, respectively, whether the file conforms to the normal FITS standard, the number of bits to represents each point or pixel value the number of coordinate axes, and the number of points or pixels along each axis. These keywords are now used to describe each sub-array. It is necessary to indicate that the group format is being used. We have chosen to do this by adopting the convention of setting NAXIS1 to zero and by setting the GROUPS keyword to true. The former indicates a non-standard data array structure and the latter indicates a group structure format. The keywords NAXIS2, ... NAXIS*n* are used to describe the data array in the standard manner. The only change is that the value of NAXIS is now one more than the actual number of axes in the sub-array ; thus, since the first NAXIS*n* must indicate the number of points or pixels along the most rapidly varying axis, that axis is NAXIS2. The keywords CRVAL*n*, CRPIX*n*, etc. have their usual meanings except that CRVAL1, CRPIX1, etc. have no meaning (since NAXIS1=0).

In addition to providing the keyword GROUPS and giving it a value of *T*, two additional keywords must be provided : PCOUNT, which indicates the number of parameters included in each block, and GCOUNT, which indicates the number of groups appearing in this file or "image" on tape. The keywords PTYPE*n*, PSCAL*n*, PZERO*n* apply to the parameters just as the keywords CTYPE*n*, etc. apply to the axes.

The binary parameter values are required to have the same format (8-, 16-, or 32-bit integer) as the binary array data values. This requirement is needed in order to prevent a binary integer from being split between two physical tape records. For cases in which a parameter requires greater precision than the data, the writer of a FITS group tape may split the parameter into two separate parameters. Both of the separate parameters must, by convention, be assigned the same value of PTYPE*n*. The recipient of such a tape may then assume that the true parameter value is given by the numerical sum of the two separate parameter values. Please note that, if this convention is followed, the data format used to record all the data and parameter values must meet the precision and sign requirements specified by the value of BITPIX. One must not write a 16-bit twos-complement number in the place of two 8-bit unsigned integers. An example of this convention is described in the following section and illustrated in figure 2.

4. Examples.- In order to illustrate this extension of FITS, two examples are given. In the first example, the observer has a collection of spectra at various locations around an astronomical object. An appropriate FITS header is shown in figure 2. The tape follows all the basic conventions for number representation and data ordering (SIMPLE=*T*). It uses 16-bit twos-complement binary integers to represent pixel and parameter values (BITPIX=16) and does not contain a standard regularly gridded matrix (NAXIS1=0). The file contains data in a block structure (BLOCKS=*T*). The data sub-arrays consist of *1*-dimensional arrays (NAXIS=2, but NAXIS1=0) with 384 pixel values (NAXIS2=384). Preceding each sub-array are four parameters (PCOUNT=4) ; however, there are really only two parameters (*l*,*b*) since each is divided into two numbers to obtain the desired accuracy.The parameters are the galactic coordinates (PTYPE1='GLON', PTYPE3='GLAT') in that order. Each value is split into two parts. The first parameter contains the integral number of degrees (PSCAL1=PSCAL3=1) while the second parameter contains the fraction in units of ten-thousandths of a degree (PSCAL2=PSCAL4=1.E-04). The file or "image" contains 100 (GCOUNT=100) "groups" of parameters plus their respective sub-arrays. Each group contains 776 (8+768) bytes of information. The header information is terminated with an END keyword.

In this example the header information requires only one tape record (2880 bytes). This is immediately followed by the data records containing the parameters and data of each group. The information for each group is packed sequentially in the records, and the information of a single group may span two or more physical records. In this case, the first record contains the information from groups number 1 (bytes 1 thru 776), number 2 (bytes 777 thru 1552), number 3 (bytes 1553 thru 2328) and the first 552 bytes of group number 4. The information for the remaining groups is packed in a similar manner. In total there are 27 data records (plus a header record). The unused words of the last data record are filled, by convention, with zeros. The last data record is followed by a tape mark.

A second example is shown in figure 3. In this case, the sub-arrays consist of weighted complex fringe visibilities at 256 frequencies, 4 Stokes polarizations and a single phase-tracking position. Each of these sub-arrays is a single observation characterized by the non-uniformly gridded values of baseline, hour angle, and the spatial Fourier transform coordinates *uu* , *vv* , and *ww*. The data are organized as cosine/sine pairs plus weight for 4 Stokes polarizations at each of 256 frequencies. All data and parameters are written as 16-bit integers. The data, in units of Jy, are written with an lsb of 0.24 mJy in the example. The second parameter, hour angle, has units of degrees and is written with an lsb of 1.32 seconds of time. The third, fourth, and fifth parameters, the projected baseline coordinates, have units of seconds, and are written with an lsb of 4 ns. There is a total of 94, 631 groups or observed points.

This example illustrates an important concept in the use of FITS : the tape writing institution must consider the capabilities of the tape reading institution. If the data blocks on the tape described by figure 3 are sorted such that the hour angle varies quickly and the baseline varies slowly, then the data order and structure are well suited for the analysis of Fourier transform spectra. However, an institution wishing to make single frequency maps on a small computer would not appreciate the structure outlined in figure 3. They would prefer that the data be separated by frequency and Stockes parameter. This could be done either by writing a separate FITS file for each frequency and Stokes parameter or by handling frequency and Stokes as if they were irregularly gridded parameters like the spatial Fourier transform coordinates. The user should always try to keep the requirements and restrictions of the reading institution in mind.

5. Conclusions.- A format for the interchange of astronomical images on magnetic tape has been developed (Wells, Greisen and Harten, 1980) and is already in use at several major observatories. The FITS format provides a simple, flexible and powerful mechanism for the transmission of *n*-dimensional, regularly spaced data arrays along with extensive documentary and parametric information. The basic FITS format is not well suited for the transmission of small arrays of data which are related but which do not necessarily occur on a uniform grid. This paper presents an extension of the basic FITS format which should remove this difficulty. The extension requires the creation of a new simple data structure plus the addition of 6 new keywords and a few standard keyword values to indicate the data array structure. The extended FITS format has been used to exchange radio aperture synthesis data between the Westerbork Synthesis Radio Telescope and the Very Large Array.

References

WELLS, D.C., GREISEN, E.W., HARTEN, R.H. : 1980, *Astron. Astrophys. Suppl.* in press.

TABLE I.- *New Keywords and Keyword values*

A. Keywords :

GROUPS	logical	if true, specifies that the data structure has a "group format".
PCOUNT	integer	the number of parameter values preceding each data block.
GCOUNT	integer	the number of groups contained in the file or "image".
PTYPEn	character	type of the n^{th} parameter (descriptive name).
PSCALn	floating	true value of parameter n = tape value * PSCALn + PZEROn.
PZEROn	floating	see above.

B. Added values for CTYPE and PTYPE :

'UU '	U coordinate of visibility data, a Fourier transform pair with LL (units : seconds).
'VV '	V coordinate of visibility data, a Fourier transform pair with MM (units : seconds).
'WW '	W coordinate of visibility data, a Fourier transform pair with NN (units : seconds).
'HA '	Hour angle (units : degrees).
'COMPLEX'	complex valued : has values 1=real, 2=imaginary, 3=weight (N.B. this is an extension of the definition given in Paper I).

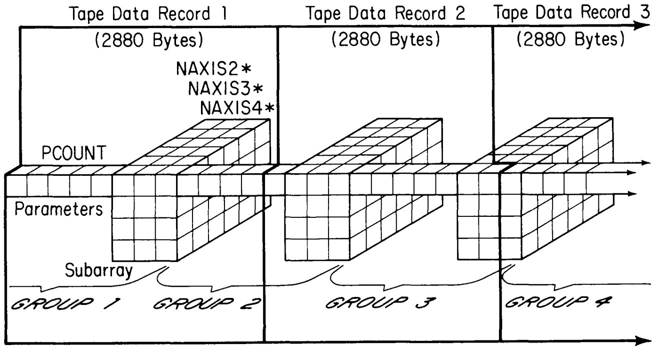

FIGURE 1.- A schematic illustration of the group structure used for binary data. In this example, a three-dimensional subarray containing NAXIS2 x NAXIS3 x NAXIS4 data elements is preceded by PCOUNT parameter elements. The non-commensurability of the group length and the fixed record length is indicated. The data records shown are preceded in the tape file by one or more header records.

```
      0       0 1                     3
CARD  1       9 1                     0

1     SIMPLE  =                       T   / Proper FITS conventions used
2     BITPIX  =                      16   / 16 bits per data point
3     NAXIS   =                       2   /
4     NAXIS1  =                       0   / Non-standard array structure
5     NAXIS2  =                     384   / Points along array
6
7     GROUPS  =                       T   / Group data structure
8     PCOUNT  =                       4   / There are 4 parameters
9     GCOUNT  =                     100   / There are 100 'groups'
10
11    PTYPE1  = 'GLON    '                / Integer Deg. Gal. longitude
12    PTYPE2  = 'GLON    '                / Fraction
13    PTYPE3  = 'GLAT    '                / Integer Deg. Gal. latitude
14    PTYPE4  = 'GLAT    '                / Fraction
15    PSCAL1  =                     1.0   /
16    PSCAL2  =                 1.0E-04   / Fraction in units .0001
17    PSCAL3  =                     1.0   /
18    PSCAL4  =                 1.0E-04   / Fraction in units .0001
19    PZERO1  =                     0.0   / No offset
20    PZERO3  =                     0.0   /
21    CRPIX2  =                   192.0   / Reference point in the array
22    CRVAL2  =               -5.0E+01    / Value at the reference point
23    CRDELT2 =               0.874E+00   / Increment between points
24    CRTYPE2 = 'VELO-LSR'                / LSR velocities
25    BUNIT   = 'K       '                /
26    BSCALE  =               3.333E-03   / Real=tape*BSCALE + BZERO
27    BZERO   =                      0.   /
28    BLANK   =                  -32768   / Undefined data point value
29    OBJECT  = 'W51-21cm'
30    TELESCOP= 'NRAO-140'
31    INSTRUME= '21C-COOL'
32    OBSERVER= "AENEWMAN'
33    DATE-OBS= '31/03/79'                / DD/MM/YY
34    DATE    = '25/12/79'                / DD/MM/YY tape written
35    HISTORY TPOWER BASELINE NFIT=3 BDROP=15 EDROP=12 /
36    END
```

FIGURE 2.- A sample FITS header illustrating the use of the group data structure for spectra observed around a source.

```
         00000000011111111112222222222333333333344444444445555555555666666666677777777778
CARD#    12345678901234567890123456789012345678901234567890123456789012345678901234567890
-----    --------------------------------------------------------------------------------

1/ 1:    SIMPLE  =                    T         / UV FORMAT FROM NRAO(CV) IS T
1/ 2:    BITPIX  =                   16         / 2-BYTE TWOS COMPL INTEGERS
1/ 3:    NAXIS   =                    6         / NUMBER OF AXES
1/ 4:    NAXIS1  =                    0         / NO NORMAL IMAGE IN GROUPS
1/ 5:    NAXIS2  =                    3         / # COMPLEX: REAL,IMAGINARY,WEIGHT
1/ 6:    NAXIS3  =                    4         / # STOKES PARS: I,Q,U,V
1/ 7:    NAXIS4  =                  256         / # FREQUENCIES
1/ 8:    NAXIS5  =                    1         / # RA(1950)
1/ 9:    NAXIS6  =                    1         / # DEC(1950)
1/10:
1/11:    GROUPS  =                    T         / GROUPS DATA STRUCTURE
1/12:    PCOUNT  =                    5         / # PARAMETERS/GROUP
1/13:    GCOUNT  =                94631         / # DATA GROUPS
1/14:
1/15:    BSCALE  =         2.441406E-04         / REAL = TAPE*BSCALE + BZERO
1/16:    BZERO   =                  0.0         / NO BIAS ADDED
1/17:    BUNIT   = 'JY      '                   / UNITS OF FLUX
1/18:    BLANK   =               -32768         / TAPE VALUE THAT => BLANK PIXEL
1/19:    OBJECT  = '0742+103'                   / SOURCE NAME
1/20:    INSTRUME= 'VLA     '                   / NRAO(CV) VLA MAPPING PROGRAMS
1/21:    TELESCOP= 'VLA     '                   / NRAO(CV) VLA MAPPING PROGRAMS
1/22:    EPOCH   =               1950.0         / EPOCH OF RA, DEC (YEARS)
1/23:    OBSERVER= 'BRID    '                   / ID CODE
1/24:    DATE-OBS= '05/01/80'                   / OBSERVATION START DATE DD/MM/YY
1/25:    DATE    = '28/03/80'                   / TAPE WRITING DATE: DD/MM/YY
1/26:    ORIGIN  = 'NRAO(CV) PGM=DUV2FITS(V1)'  /
1/27:
1/28:    HISTORY VLACV SORT   ORDER='BH'        / SECOND VARIES FASTER
1/29:             WHERE B MEANS ASCENDING  BASELINE NUMBER
1/30:             WHERE H MEANS ASCENDING  HOUR ANGLE
1/31:
1/32:
1/33:
1/34:
1/35:
1/36:
```

FIGURE 3a.-Example 2, header record 1 of 3.

```
         00000000011111111112222222222333333333344444444445555555555666666666677777777778
CARD#    12345678901234567890123456789012345678901234567890123456789012345678901234567890
-----    --------------------------------------------------------------------------------

2/ 1:    CTYPE2  = 'COMPLEX '                   / FRINGE VISIBILITIES
2/ 2:    CRPIX2  =                  1.0         / REF PIXEL
2/ 3:    CRVAL2  =                  1.0         / REAL FIRST
2/ 4:    CDELT2  =                  1.0         / THEN IMAGINARY
2/ 5:    CTYPE3  = 'STOKES  '                   / COORD TYPE: #=> I,Q,U,V
2/ 6:    CRPIX3  =                  1.0         / REF PIXEL LOCATION
2/ 7:    CRVAL3  =                  1.0         / REF PIXEL VALUE => I
2/ 8:    CDELT3  =                  1.0         / VALUE INCREMENT
2/ 9:    CTYPE4  = 'FREQ    '                   / VALUE IS HZ
2/10:    CRPIX4  =                  1.0         / REF PIXEL
2/11:    CRVAL4  =        4.8851000E+09         / COORDINATE VALUE
2/12:    CDELT4  =               1000.0         / VALUE INCREMENT
2/13:    CTYPE5  = 'RA      '                   / COORD TYPE: IN DEGREES
2/14:    CRPIX5  =                  1.0         / REF PIXEL LOCATION
2/15:    CRVAL5  =          115.7019375         / REF PIXEL VALUE
2/16:    CDELT5  =                  0.0         / VALUE INCREMENT
2/17:    CTYPE6  = 'DEC     '                   / COORD TYPE: IN DEGREES
2/18:    CRPIX6  =                  1.0         / REF PIXEL LOCATION
2/19:    CRVAL6  =           10.3090667         / REF PIXEL VALUE
2/20:    CDELT6  =                  0.0         / VALUE INCREMENT
2/21:
2/22:    PTYPE1  = 'BASELINE'                   / PAR TYPE: VALUE IS NUMBER
2/23:    PSCAL1  =                  1.0         / NO MEANING TO SCALING
2/24:    PZERO1  =                  0.0         / NO OFFSET EITHER
2/25:    PTYPE2  = 'HA      '                   / PAR TYPE: VALUE IS DEGREES
2/26:    PSCAL2  =       5.49316406E-03         / REAL = TAPE*PSCAL + PZERO
2/27:    PZERO2  =                  0.0         / NO OFFSET
2/28:    PTYPE3  = 'UU      '                   / PAR TYPE: U VIS COORD IN SEC.
2/29:    PSCAL3  =         3.999997E-09         / REAL = TAPE*PSCAL + PZERO
2/30:    PZERO3  =                  0.0         / NO BIAS ADDED
2/31:    PTYPE4  = 'VV                          / PAR TYPE: V VIS COORD IN SEC.
2/32:    PSCAL4  =         3.999997E-09         / REAL = TAPE*PSCAL + PZERO
2/33:    PZERO4  =                  0.0         / NO BIAS ADDED
2/34:    PTYPE5  = 'WW      '                   / PAR TYPE: W VIS COORD IN SEC.
2/35:    PSCAL5  =         3.999997E-09         / REAL = TAPE*PSCAL + PZERO
2/36:    PZERO5  =                  0.0         / NO BIAS ADDED
```

FIGURE 3b.- Example 2, header record 2 of 3.

```
         00000000011111111112222222222333333333344444444445555555555666666666677777777778
CARD#    12345678901234567890123456789012345678901234567890123456789012345678901234567890
-----    --------------------------------------------------------------------------------

3/ 1:    COMMENT  FORMULA FOR BASELINE # FROM ANTENNA PAIR I < J
3/ 2:    COMMENT  B = (256 * I)  +  J
3/ 3:
3/ 4:    COMMENT  ANTENNA LOCATIONS IN NANOSECONDS:
3/ 5:    HISTORY VLACV ANT   N= 2 X=  5470.525 Y=-14443.276 Z= -8061.210 ST='AW4'
3/ 6:    HISTORY VLACV ANT   N= 4 X=  1667.280 Y= -4396.334 Z= -2452.399 ST='CW8'
3/ 7:    HISTORY VLACV ANT   N= 5 X=    37.719 Y=   135.627 Z=   -50.585 ST='DE2'
3/ 8:    HISTORY VLACV ANT   N= 6 X=  3353.710 Y= -8816.123 Z= -4910.700 ST='BW6'
3/ 9:    HISTORY VLACV ANT   N= 7 X=   118.761 Y=   445.786 Z=  -170.397 ST='DE4'
3/10:    HISTORY VLACV ANT   N= 9 X= 10924.708 Y=-28961.684 Z=-16194.042 ST='AW6'
3/11:    HISTORY VLACV ANT   N=10 X=    73.382 Y=   271.952 Z=  -103.200 ST='DE3'
3/12:    HISTORY VLACV ANT   N=12 X=  8324.926 Y= 31661.636 Z=-12190.700 ST='AE6'
3/13:    HISTORY VLACV ANT   N=14 X= 14206.476 Y=-37731.068 Z=-21114.612 ST='AW7'
3/14:    HISTORY VLACV ANT   N=15 X= 17842.852 Y=-47447.285 Z=-26566.649 ST='AW8'
3/15:    HISTORY VLACV ANT   N=16 X=  1548.048 Y=  5883.154 Z= -2264.541 ST='CE9'
3/16:    HISTORY VLACV ANT   N=17 X=   509.527 Y= -1338.539 Z=  -745.231 ST='DW8'
3/17:    HISTORY VLACV ANT   N=18 X=  2552.452 Y=  9638.185 Z= -3698.873 ST='BE6'
3/18:    HISTORY VLACV ANT   N=20 X=  -100.221 Y=   -15.904 Z=   152.474 ST='DW2'
3/19:    HISTORY VLACV ANT   N=21 X=  -812.570 Y=  -126.899 Z=  1200.964 ST='DN8'
3/20:    HISTORY VLACV ANT   N=22 X=  1021.275 Y= -2683.726 Z= -1494.627 ST='CW6'
3/21:
3/22:    COMMENT   FORMULA FOR BASELINE BETWEEN ANTENNA I AND J (I < J):
3/23:    COMMENT   BASELINE(IJ) = LOCATION(I) - LOCATION(J)
3/24:
3/25:    COMMENT   FORMULAE FOR UU, VV, WW :
3/26:    COMMENT   UU = BX * SIN(HA)  +  BY * COS(HA)
3/27:    COMMENT   VV = BZ * COS(DEC)  +  SIN(DEC) * (BY * SIN(HA) - BX * COS(HA))
3/28:    COMMENT   WW = BZ * SIN(DEC)  +  COS(DEC) * (BX * COS(HA) - BY * SIN(HA))
3/29:                 WHERE UU AND VV ARE THEN ROTATED TO THE EPOCH GIVEN ABOVE
3/30:
3/31:    END
3/32:
3/33:
3/34:
3/35:
3/36:
```

FIGURE 3c.- Example 2, header record 3 of 3.

FIGURE 3.- A sample FITS header illustrating the use of the group data structure for observations made with a synthesis telescope.