Data Transfer for the AT

R.P.Norris

7 January 1987

Introduction

The data obtained by the AT must be transferred to an analysis computer which may be at Culgoora, Epping, or some other institution, and some format must be chosen for this transfer which is (a) efficient, and (b) easily read on any probable data analysis computer. When this was considered this some while ago in AT/25.1.1/008, the conclusion was reached that whilst the FITS format (an internationally recognised data transfer format) satisfied (b), it did not satisfy (a). In particular, the criticism was made that some of the features of FITS rendered it unsuitable for real-time data collection. Thus a variant of FITS (called RPFITS), which overcame most of the disadvantages of FITS, was proposed for AT data.

Since then, partly in response to such criticisms, the FITS standard has been modified so that it now incorporates many of the features of RPFITS. It is therefore appropriate to re-examine the question of whether we should adopt FITS or RPFITS. In this document, which is intended as a basis for discussion rather than a final definition, I argue that we should adopt the new FITS standard for data transfer.

FITS

FITS (Flexible Image Transport System) was originally proposed (Wells et al., 1981, Astr. Astrophys. Suppl., **44**, 363) as a format for mag tape transfer of astronomical images. The first enhancement (Greisen & Harten, 1981, Astr. Astrophys. Suppl., **44**, 371) extended the 'images' to include irregularly gridded arrays of data, such as u-v data from radio synthesis arrays. Since then a number of extensions have been introduced, such as the introduction of tables and other extension files. FITS is now accepted throughout the astronomical community as the way to exchange data, and it was always envisaged that external users of the AT would take home their data in the standard FITS format.

However, FITS was not suitable for real-time recording of data for reasons to be discussed shortly, and so a modification of FITS, called RPFITS, was devised. It was intended that AT data should be recorded in RPFITS, that all data transfer between AT sites would be via RPFITS, and that the data reduction computer would accept RPFITS tapes. In those cases where a visiting astronomer wished to take the data back to his home institution, any AT computer could translate the RPFITS tapes to FITS tapes.

The need for RPFITS

FITS was not suitable for real-time data collection for the following reasons, which are detailed in AT/25.1.1/028:

(1) The FITS data had to be written in blocks of 2880 bytes, which is a very inefficient block size for a VAX (which prefers multiples of 512 bytes) and for high speed tape drives (which prefer longer blocks). The RPFITS format changed the block size to 2560 bytes to maximise speed on a VAX. Longer blocks (integral mutiples of 2560 bytes) could be used for mag tape transfers.

(2) FITS required each source to occupy one file. Thus a long observation of multi-source switching would have to be broken up into many short FITS files. RPFITS overcame this problem by allowing additional headers within the data, so that source changes (or other parameter changes) could be marked by new headers.

(3) FITS insisted that the data be written in integer format, with a scaling factor written in the header. This is clearly unsuitable for real-time data, since (a) the maximum value of the data must be known before the observation starts, and (b) converting from REAL*4 (as used by the correlator) to integer is wasteful of real-time processor time. RPFITS overcame this by allowing the data to be written as IEEE REAL*4.

(4) FITS relied heavily upon tables written at the end of the data. Thus a computer crash a few minutes before the end of an observing run would result in the loss of all the data from that run. Similarly, the integrity of the data on a magnetic tape depended on the integrity of a few critical blocks at the end of the tape. RPFITS overcame this by banning tables at the end of the data and instead writing all the relevant information into the headers within the data.

New Extensions to FITS

The FITS standard has now been modified to answer some of these criticisms. Specifically, in answer to each of the four points above:

(1) FITS now allows block sizes which are integral multiples of 2880 bytes. It thus allows long blocks, enabling efficient writing to mag tapes, and, by permitting block lengths of 8*2880 = 23040 = 9*2560 bytes, allows efficient writing to disk by VAXes. Tests done on the 8250 at Epping shows that writing with this block size (and with judicious use of buffer parameters) is as efficient as with PTR's BLK_WRITE routines using RMS calls (but see the note at the end of this document on i/o rates).

(2) FITS now allows multi-source files. It does so by assigning source numbers which refer to a source table at the end of the data. By doing so, however, it worsens problem (4).

(3) FITS now allows data to be written as IEEE REAL*4 numbers.

(4) The problem of fragility to computer errors and media errors has not been solved. If we adopt FITS, we can circumvent this problem to some extent by keeping a copy of all table information within a database on the disc. This database may need to be preserved indefinitely (possibly as a backup tape) so that media errors can be rectified.

FITS vs RPFITS

The debate of FITS vs RPFITS boils down to the following arguments:

(1) **Pro-FITS:** FITS is universally acceptable. Thus if we wrote our data in FITS, any visiting astronomer would be able to take home his raw data from Culgoora without any further processing. Similarly, the data could be read straight into AIPS at Epping or Culgoora without needing any locally-written tasks to read an RPFITS tape.

(2) **Pro-RPFITS:** FITS is very fragile to computer or media errors. Thus we would need to write all the calibration and source information into a separate database and preserve that information (as a backup) along with the data. This would not present too much of a problem on-line, as the database would need to be saved for the duration of the observation anyway (a) for calibration, and (b) to write out the tables at the end. The real penalty is that a backup copy of this database must be made and kept along with the data. As an alternative, depending on the reliability of the medium used (high for optical disks, for example), it may be decided that once the data had been read and checked the database backup could be disposed of.

CONCLUSION

The choice between FITS and RPFITS depends on factors which are more philosophical than technical. My personal inclination is to adopt FITS, but I would be grateful for any other comments.

A Note on I/O Rates

The maximum data rate of the AT is

8192 channels * 15 baselines /5 second integration = 24576 complex visibilities/s.

If each visibility is written as a 2*REAL*4 complex pair, then this gives a data rate of 200 kbyte/s. If an additional REAL*4 weight is attached to each visibility, this rises to 300 kbyte/s. This weight is current practice in AIPS, but is not required by FITS. In view of its inefficiency, I do not recommend that we flag data in this way, but instead flag faulty data using an extension table or by setting amplitudes to zero.

Tests on a moderately busy VAX 8250 show that writing FITS blocks to disk (with a maximum rate of 1.8 Mbyte/s (Eagle) or 2.2 Mbyte/s (RA81)) can take place at a typical rate of 300 kbyte/s. Laser discs are likely to have a maximum rate around 200 kbyte/s, and VHS tape devices have maximum rates around 100 kbyte/s. Thus the AT working flat out will certainly need more than one drive, and even then will be on the edge of feasibility. It may be that if an astronomer really does want 8192 channels then his integration time will need to be increased. On the other hand, it should be noted that the vast majority of experiments will come nowhere near this data rate, and can be handled comfortably by the slowest drive.