The Flexible Image Transport System (FITS) in Astronomy

Preben Grosbøl
European Southern Observatory
Munich, Germany

ERPANET Seminar
Wien, May 11, 2004

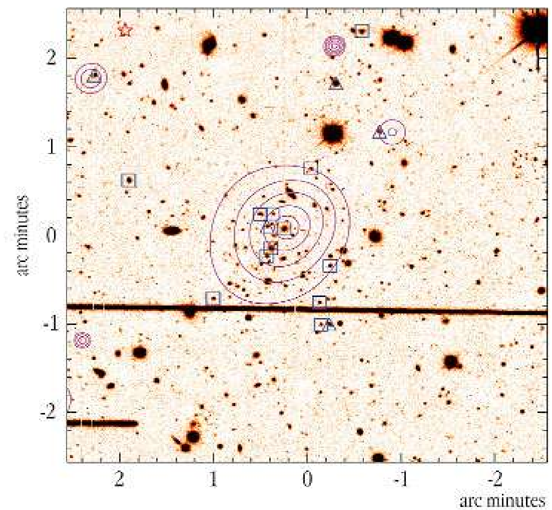
Overview

1. Need for interchange format
2. Control of format
3. Basic structure of FITS
4. FITS data structures
5. Usage in astronomy

Astronomical Images and Spectra

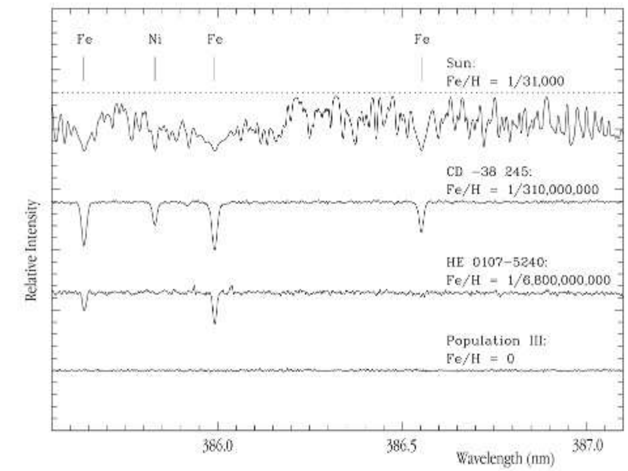
Spiral Galaxy Messier 83 (VLT ANTU + FORS1)

ESO PR Photo 41/99 (29 November 1999)



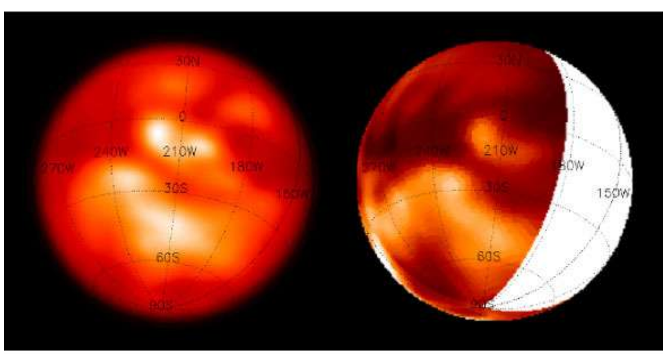
XMM-LSS X-ray Contour Map Overlaid on a CFHT I-band Image

ESO PR Photo 19c/03 (14 July 2003)



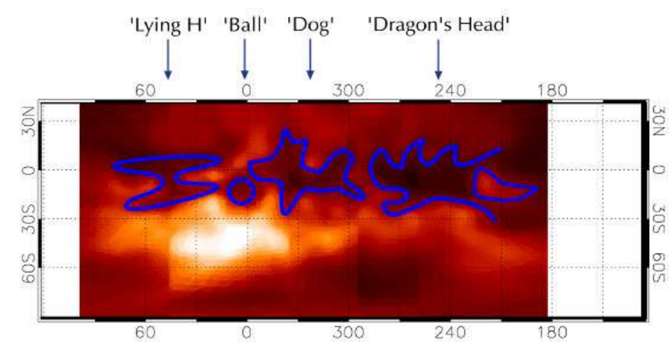
Spectra of Stars with Different Metal Content

ESO PR Photo 25b/02 (30 October 2002)



NACO Images of Titan's Surface
(VLT YEPUN + NACO/SDI)

ESO PR Photo 11e/04 (14 April 2004)



Provisional Names of Titan Dark Surface Features
(VLT YEPUN + NACO/SDI)

ESO PR Photo 11f/04 (14 April 2004)

Data in Astronomy

- Catalogs
  - *Enuma Anu Enlil* (Babylonian < 900 BC), ~70 tables
  - Almagest (Ptolemy ~141 AD), 1028 stars, ~10 arcmin
  - HST Guide Star Catalog (NASA, 1990), 15 Mstars, ~0.1 arcsec
- Spectra and Images
  - Photographic
    - Carte du Ciel (1890-1950)
    - Schmidt telescope sky surveys (Palomar, ESO/S) (1950-1980)
  - Digital detectors
    - Optical CCD mosaics (e.g. ESO VST)
      - 16kx16k array of 2kx4k CCD's, ~ 0.5Gb per frame or ~30 Gb per day
    - Radio interferometers (e.g. ALMA)
      - Complex visibilities, ~6 Mb/s or ~500 Gb per day
  - Astronomical archives
    - ESO archive currently ~10 Tb, exponential growth, 1.5 year to double

Basic needs for interchange

- Comparing astronomical data
  - From different observatories
  - At different wavelengths e.g. X-ray, optical, infrared, radio
    - Associated to different physical properties
    - Different registration and resolutions
- Single, standard format
  - Open standard controlled by standards body
  - Independent of computer architecture
  - Possible to extend (once FITS, always FITS)
  - Multi-dimensional data cubes and tables
  - Self-documenting

Evolution of FITS

- 1979: Initial FITS Agreement and first interchange of files
- 1981: Published original (single HDU) definition [FITS]
- 1982: Formally endorsed by the International Astronomical Union (IAU)
- 1988: Defined rules for multiple HDUs [XTENSION]
- 1988: FITS Working Group established by IAU [IAUFWG]
- 1988: Extended to include ASCII tables [TABLE]
- 1990: Extended to include IEEE floating-point data
- 1994: Extended to multiple image arrays [IMAGE]
- 1995: Extended to binary tables [BINTABLE]
- 1997: Adopted a Y2K-compliant date format
- 2001: Reiterated existing standard in one paper [NOST]
- 2002: Approved conventions for world coordinates [WCS1/2]

Control bodies

- IAU FITS Working Group
  - Controls to FITS standard
    - Any change must be approved by IAUFWG
    - Under IAU Division 12, Commission 5
    - IAUFWG executive committee
      - Chair, vice-chair and chairs of Regional FITS Committees
    - ~20 members representing astronomical community
      - Major data producing organizations (e.g. NASA, ESA, ESO, HST)
      - Major data centers and archives
  - Three Regional FITS Committees
    - European, North American and Japanese

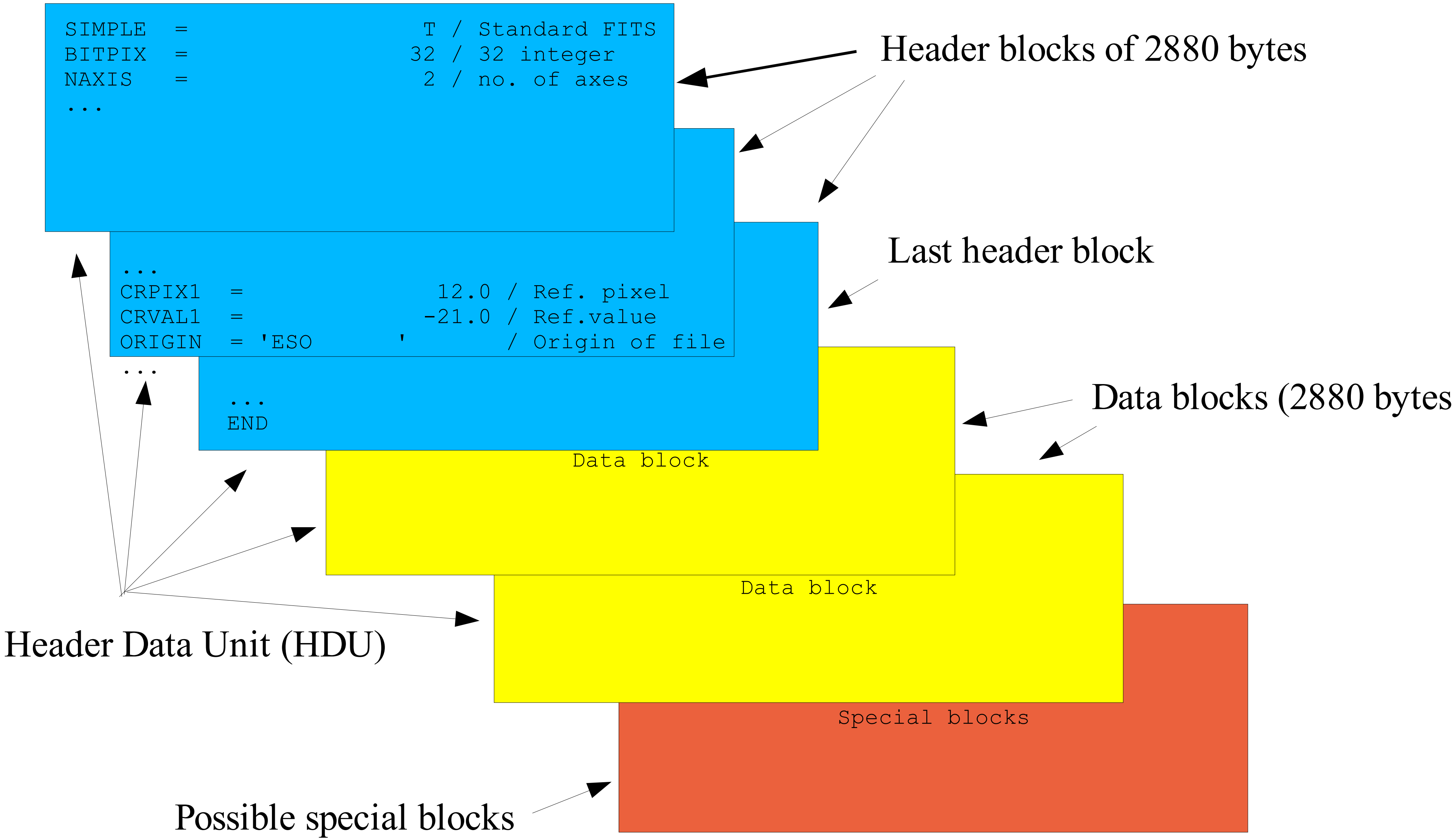
Change procedures

- Proposal forwarded to Regional FITS Committee (RFC)
  - Regional discussion and vote
- If approved by RFC, forwarded to other RFC's
  - Discussion and vote by other RFC's
- If approved by all RFC's, forwarded to IAU FITS WG
  - Final discussion and proposal
  - Final vote
    - Requires ¾ majority and no votes against
    - If no-vote, 6-month moratorium for discussion and possible revision
    - Second vote require ¾ majority
- Voting rules currently under revision
  - To better accommodate Internet discussions
  - To avoid possibly conflicts between RFC's

Basic structure

- Logical 2880 8-bit blocks
  - evenly divisible by byte/word lengths of all known computers
- Header and Data Unit (HDU)
  - Header
    - Integer number of logical records
    - Defines size and type data unit
    - Card images of 80 printable ACSII characters (36 per record)
    - Keyword-values structure
  - Data
    - Big-endian binary data
    - Multi-dimensions arrays in FORTRAN order
    - Null values (Not-A-Number) can be specified
    - Patted with zero to integer number of logical records
  - 'special' blocks may exist after HDU

Basic layout of FITS file



Header Format

- Arbitrary number of fixed format 80-char. card images
  - Card format: **`keyword = value / comment`**
  - Keywords: 8 char.
  - Values: logical, integer, real (complex),string and date
- File must start with '**`SIMPLE = T`**' card
- Header ends with '**`END`**' card

```
         1         2         3         4         5         6
1234567890123456789012345678901234567890123456789012345678901234567..
SIMPLE  =                    T / file does conform to FITS standard
BITPIX  =                   16 / 16-bit twos-complement pixels
NAXIS   =                    2 / 2-dimensional image
NAXIS1  =                  512 / first axis length
NAXIS2  =                  512 / second axis length
COMMENT  -----------------------------------------------------
COMMENT   FITS (Flexible Image Transport System) format is defined
COMMENT   in 'Astronomy and Astrophysics', volume 376, page 359;
COMMENT   bibcode: 2001A&A...376..359H
COMMENT  -----------------------------------------------------
ORIGIN  = 'observatory-name   ' /
DATE    = '2003-11-22T05:23:45' / when this file was written
END
```

Hierarchical keywords

- Extend 8 char. flat name space
  - Define hierarchical name space
  - Use comment card i.e. no '=' in column 9
    - **HIERARCH DOMAIN LEVEL1 ... LEVELn KEY = VALUE / Comment**
  - Domain name space define in Data Dictionary

```
Parameter Name:     TEL FOCU LEN
Class:              header
Context:            Telescope
Type:               double
Value Format:       %.3f
Unit:               m
Comment Field:      Focal length (m)
Description:        Effective focus length of telescope
```

```
HIERARCH ESO TEL FOCU LEN   =      8.009 / Focal length (m)
HIERARCH ESO TEL FOCU SCALE =     11.650 / Focal scale (arcsec/mm)
HIERARCH ESO TEL FOCU VALUE =  25484.000 / M2 setting (eu of 0.00125 mm)
```

Data Unit

- Multi-dimensional arrays: up to 999 axes
- Data types: unsigned 8-bit integer, signed 16/32 integers and 32/64 IEEE floating point numbers
- Linear world coordinates: scale, offset, and rotation
- Scaling of data values

```
BITPIX  =                  -32 / 32-bit IEEE float
NAXIS   =                    2 / 2-dimensional image
NAXIS1  =                  512 / first axis length
NAXIS2  =                  512 / second axis length
CRPIX1  =                14.30 / RA ref. pixel
CRVAL1  =                143.3 / RA of ref. pixel
CDELT1  =             -0.00232 / RA step size (deg)
CTYPE1  = 'RA      '           / Right ascension axis
CRPIX2  =               -321.2 / DEC ref. pixel
CRVAL2  =            -30.00032 / DEC of ref. pixel
CDELT2  =              0.00232 / DEC size in x (deg)
CTYPE2  = 'DEC     '           / Declination axis
OBS-DATE= '2003-03-14T12:20:21.3' / Start of observation
EQUINOX =               2000.0 / Equinox of coordinates
BZERO   =               0.543  / Value = BZERO + BSCALE*data
BSCALE  =               0.300  / Data scale
BUNIT   = 'mJy     '           / milli-Jansky
```

General Extensions

- Use 'special' records to append extra HDU's
- Must start with **`XTENSION= 'type'`** card
- Bit size of data matrix defined as:
  - Bites = `GCOUNT*(PCOUNT+NAXIS1*...*NAXISn)*abs(BITPIX)`
  - Enables to skip over unknown extensions
- A new extension HDU may follow

```
XTENSION= 'BINTABLE'           / Binary table extension
BITPIX  =                    8 / 8 bit bytes
NAXIS   =                    2 / 2-dimensional image
NAXIS1  =                 1202 / No. of bytes in row
NAXIS2  =             12321232 / No. of rows in table
PCOUNT  =                 1232 / Size of heap
GCOUNT  =                    1 / One group of data
...


ORIGIN  = 'ESO   ' /
DATE    = '2003-11-22T05:23:45' / when this file was written
END
```

General layout of FITS file

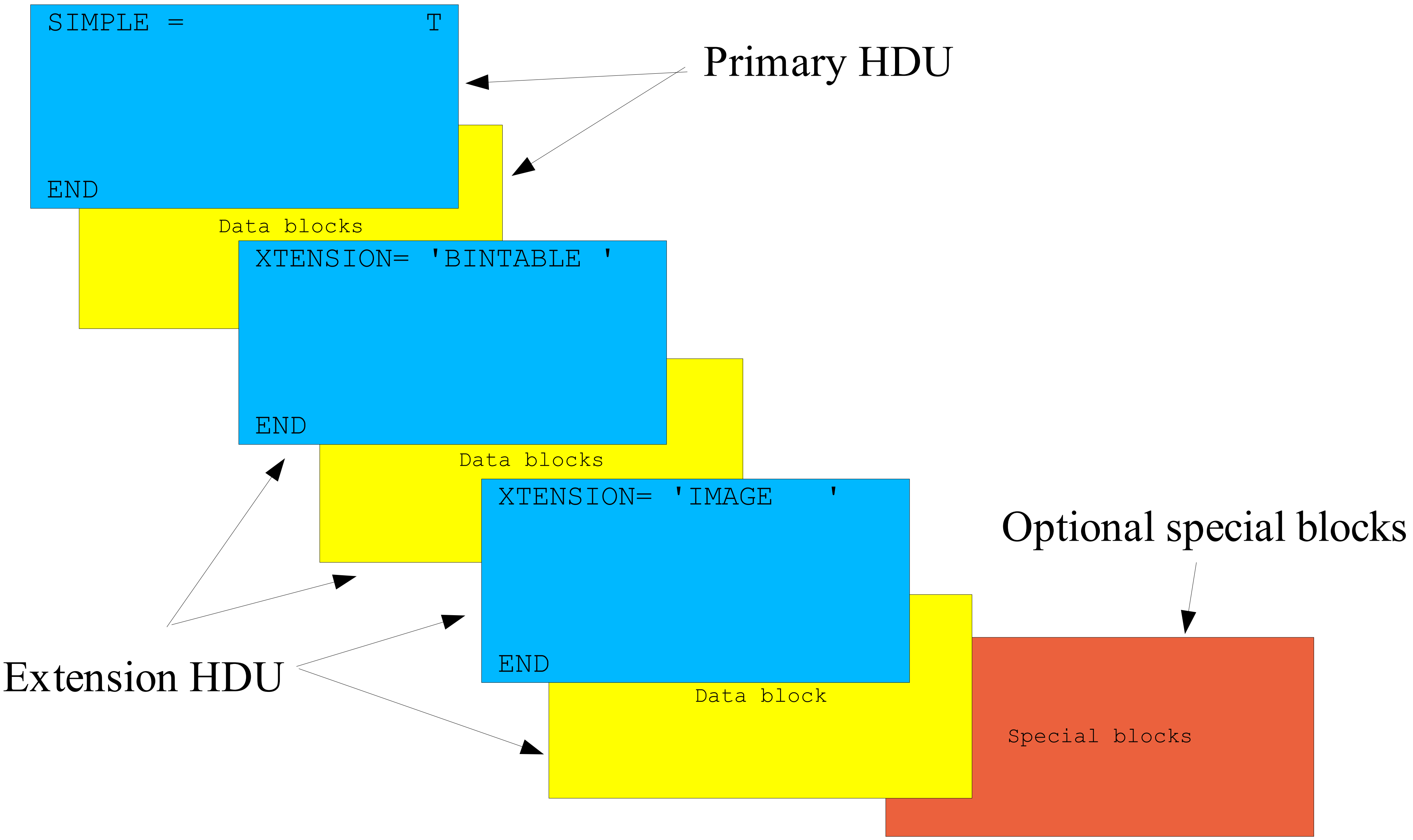


IMAGE extension

- Storage of additional image maps in a FITS file
  - Conform to extension rules (e.g. `PCOUNT,GCOUNT`)
  - Identical to primary data matrix
  - Typical usage:
    - Storage of data observed by multiple CCD's
    - Storage of Point Spread Function for images

```
XTENSION= 'IMAGE   '           / Image extension
BITPIX  =                  -64 / 64 IEEE floating point
NAXIS   =                    1 / 1-dimensional image
NAXIS1  =                12334 / No. of pixels in first axis
PCOUNT  =                    0 / No parameters allowed
GCOUNT  =                    1 / One group of data
...

END
```

ASCII Table Extension

- Designed to describe a printed table
  - Two-dimensional array of ACSII characters
  - Mainly used for published tables
  - Provides string and numeric column elements
  - Standard description of each column

```
XTENSION= 'TABLE   '           / ASCII table extension
BITPIX  =                    8 / 8-bit ASCII characters
NAXIS   =                    2 / 2-dimensional image
NAXIS1  =                 1202 / No. of bytes in row
NAXIS2  =               321232 / No. of rows in table
PCOUNT  =                    0 / No parameters allowed
GCOUNT  =                    1 / One group of data
TFIELDS =                   21 / No. of columns
...
TBCOL1  =                   12 / Start of field 1
TFORM1  = 'F12.4   '           / Format of field 1
TTYPE1  = 'K_mag   '           / Label of field 1
...
END
```

Elements and their Format

- Supported column formats
  - Strings
    - printable ASCII characters
  - Numeric formats
    - Support of FORTRAN-77 input formats
    - Scaling done by **TZEROn** and **TSCALn** keywords
    - Undefined values supported through **TNULLn** keyword
  - Support for labels and data units provided

```
TBCOL8  =                         32  / Starting character in row
TFORM8  = 'F6.2     '                 / Format of column
TTYPE8  = 'K_0      '                 / Label of column
TUNIT8  - 'mJy      '                 / Unit of column
TNULL8  = '         '                 / Null value
...
TFORM10 = 'A8       '                 / Character column
TFORM11 = 'I10      '                 / Integer column
```

```
0        1         2         3         4         5         6
123456789012345678901234567890123456789012345678901234567890
NGC 3321   1.023  SB  +33.2342  32.32  galaxy  332  2.0123
```

Binary Table Extension

- Define complex table with binary array elements
  - Similar to ACSII tables but extended with
    - Storage of binary data
    - Elements may be arrays
      - Optional interpreted as multiple dimensional matrices
    - Optional heap area for storage of common data

```
XTENSION= 'BINTABLE'           / Binary table extension
BITPIX  =                    8 / 8-bit byte
NAXIS   =                    2 / 2-dimensional byte map
NAXIS1  =                 1202 / No. of bytes in row
NAXIS2  =             12321232 / No. of rows in table
PCOUNT  =                12400 / Size of heap
GCOUNT  =                    1 / One group of data
TFIELDS =                   24 / No. of columns in table
...
TFORM1  = '8A      '           / Format of column 1
...
END
```

Column Types

- Both ASCII and binary elements
  - Strings with optional interpretation as arrays of strings
  - Integer binary types
    - Bit, unsigned 8-bit, signed 16- and 32 integer
    - Scaling (with **`TSCALn`** and **`TZEROn`**) and null definition (**`TNULLn`**)
  - Floating point values
    - 32-/64-bit IEEE-754 real and complex values
    - Full support of Inf, underflow and Not-a-Number values

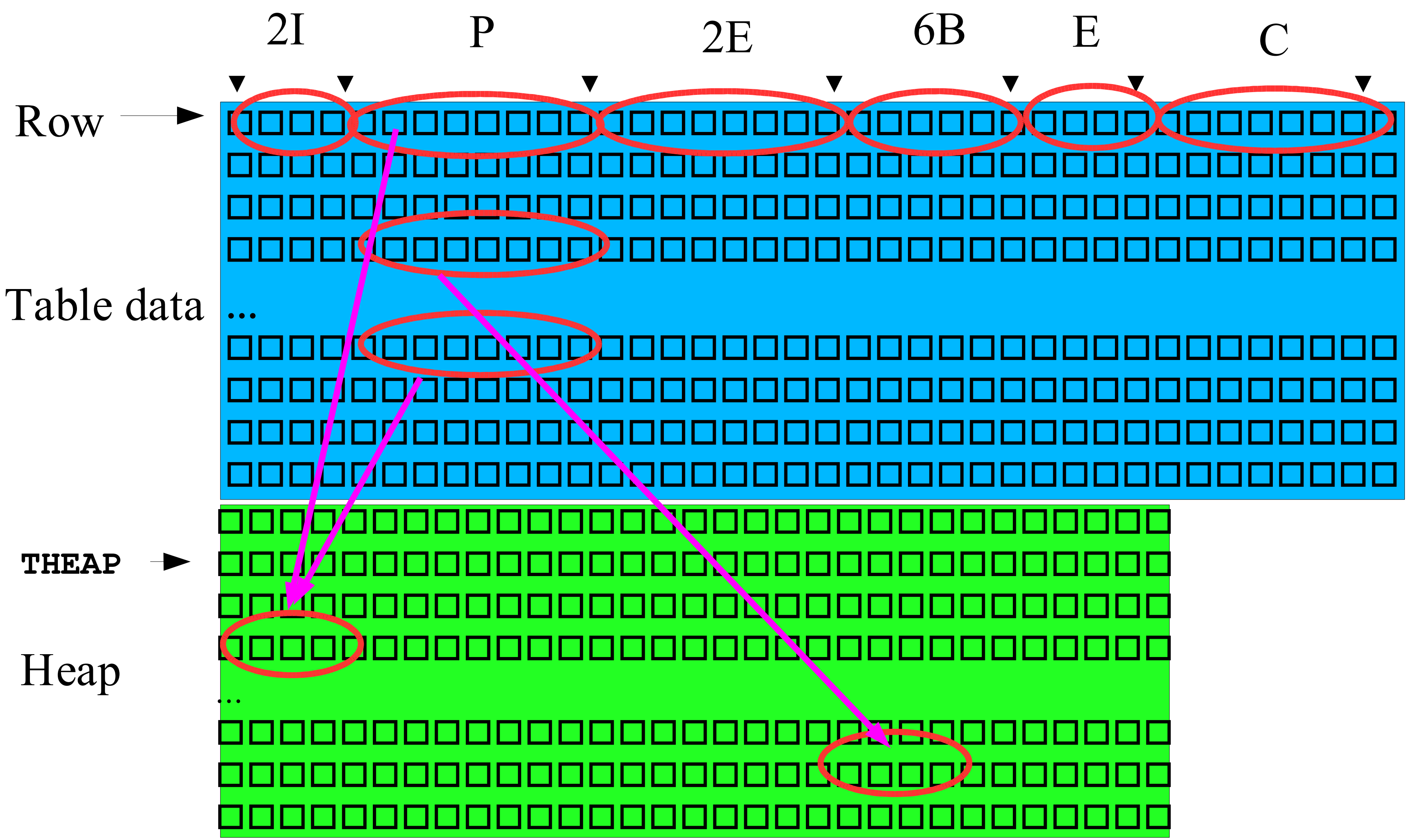
```
TFORM5  = '1024E'                 / Single precision array
TTYPE5  = 'Spectrum'              / Label of column 5
TUNIT5  = 'mJy     '              / Data unit is milli-Jansky
TDISP5  = 'F8.3    '              / Display format
...
TFORM6  = '32A     '              / ASCII string
TFORM7  = '8L      '              / Logical aray
TFORM8  = '1024X   '              / 1024-bit array
TFORM9  = '9J      '              / 32-bit integer array
TFORM10 = 'D       '              / Double precision
TFORM11 = 'M       '              / Double precision complex
```

Array and Heap elements

- Single elements may contain arrays
  - Size defined by **`r`** in **`TFORM`** keyword e.g. **`'rE'`**
  - Option to interpret array as multi-dimensional matrix by '**`TDIMn`**'
- Element may point to separate *heap* area
  - Size of heap defined by '**`PCOUNT'`**
  - Optional offset specified by '**`THEAP'`**
  - Efficient for sparse matrices and repeating data

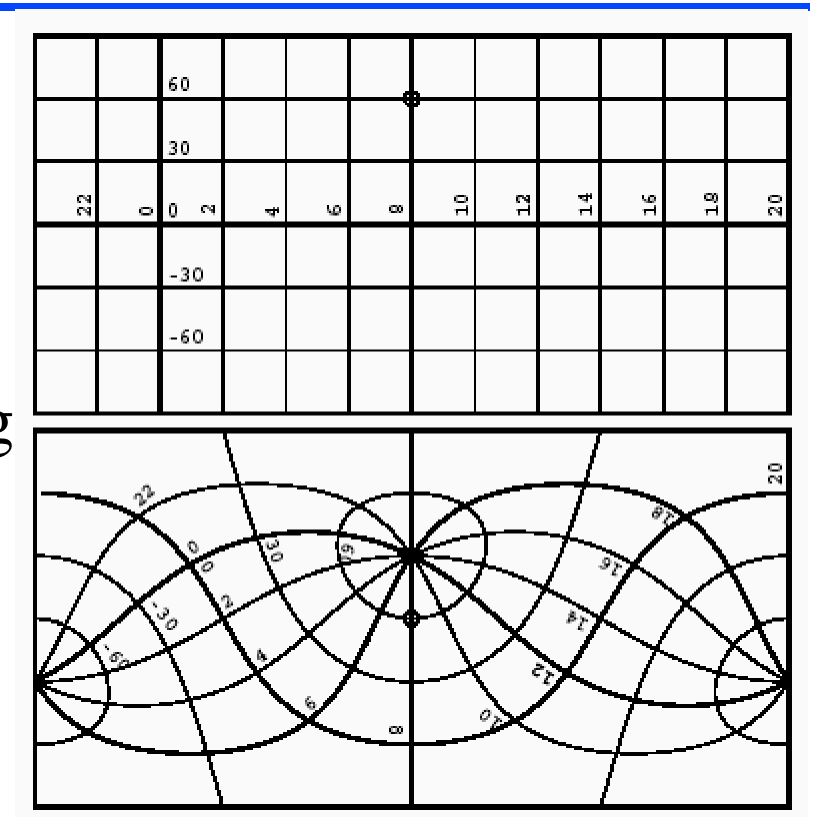
```
XTENSION= 'BINTABLE'           / Binary table extension
...
PCOUNT   =             1200348 / Size of heap
...
THEAP    =                 203 / Offset of heap
TFORM3   = '47244F '           / Float array
TDIM3    = '(124,127,3)'       / 3D array
TFORM7   = 'PB(614)'           / Pointer to byte array in heap
```

Layout of BINTABLE



World Coordinate System

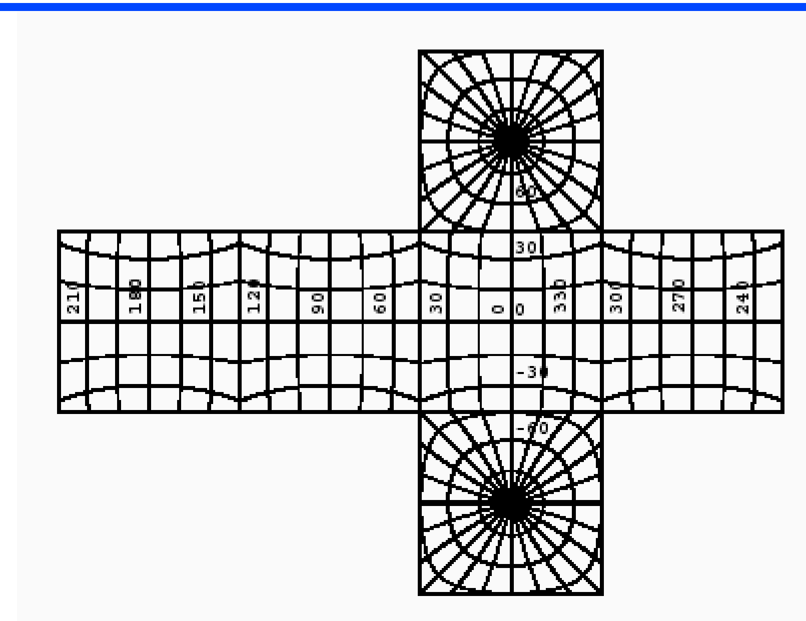
- Linear transformations
  - PC transformation matrix
  - Multiple transformations possible
- Allowed units
  - SI plus astronomical units e.g. pc, mag
- Non-linear coordinate systems
  - Type defined by '**`CTYPEn`**'
- Astronomical coordinates
  - Type and equinox



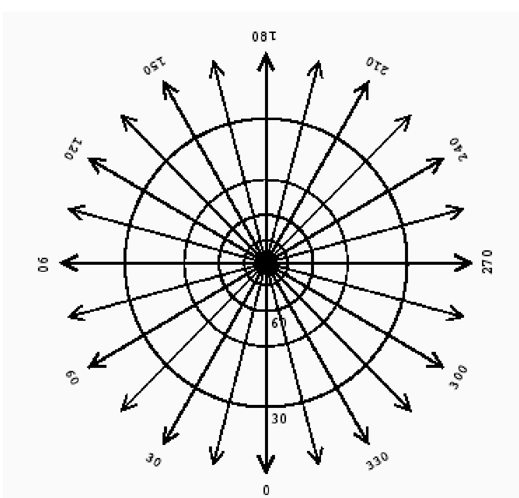

```
CRVAL1  =                  120.0 / Ref. Value for axis 1
CRVAL2  =                    0.0 / Ref. Value for axis 2
CTYPE1  = 'RA      '             / Axis 1 is RA
CTYPE2  = 'DEC     '             / Axis 2 is DEC
...
CTYPE1  = 'RA---COD'             / Conic equidistant projection
CTYPE2  = 'DEC--COD'             / Conic equidistant projection
RADESYS = 'FK5     '             / 5th fundamental catalog
EQUINOX =                 2000.0 / Default equinox
```

Coordinate Systems

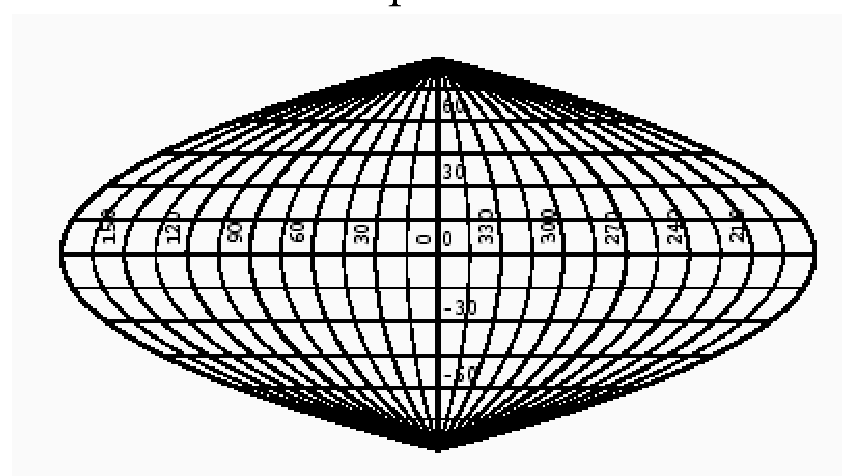
- 26 coordinate systems defined
  - Mathematical transformation
- Work in progress on
  - Representation of spectra
    - Non-linear spectral coordinates
    - Dispersive elements e.g. prisms
    - Distortions from analytical form



CSC: COBE quadrilateralized spherical cube



TAN: Gnomoic



PAR: Parabolic

Usage in Astronomy

- Export format for new data
  - All major observatories provide observational data in FITS
  - Instrument setup and description of observations
    - Encoded in FITS headers
    - Observatories define data dictionaries
- Archival data format
  - All archives offer data in FITS format
  - Many archive use FITS as internal format
- Input for pipeline data reduction
  - Observatories make standard data reductions
  - Reduced data provided in FITS

FITS and Data Processing

- Astronomical data processing system
  - AIPS, AIPS++, IRAF, MIDAS, GIPSY, IDL, ...
  - All have interface to FITS image data
    - Some use FITS as internal format
    - Some use internal formats for efficiency
  - Many provide interface to FITS tables
    - Interferometric data (radio and optical) use **`BINTABLE`** format
    - X-ray projects use table for photon lists
    - Used for catalogs e.g. target lists
  - Open source libraries for accessing FITS data
    - Available for FORTRAN, C, C++, Java, Python, ...
  - Standard display tools for FITS files
  - Compression
    - Science data compress badly if no loss of information
    - Standard compression (e.g. bzip2) on entire FITS files (factor of ~2)

The Good and the Bad

- Good features
  - Common format accepted for all astronomical data
  - Makes it easy to interchange data
  - Support vast majority of astronomical data types
  - Self-documenting
- Problem areas
  - Header has very fixed format
    - No easy cross reference between different data entities
    - No good support of hierarchical keywords
    - No support for multi-values keywords
    - Only limited set of standard keywords
      - Need to define data dictionaries
  - Format too flexible (careful design of new file structures!!)
  - Support only ASCII characters (none issues)

New Directions

- Virtual Observatories (VO)
  - Provide facilities to search for useful data in archives
    - Use archives as a virtual telescope
    - Interfaces to archive data through FITS
  - Defines XML for for small tables (**VOTables**)
    - More flexible due to XML data description
    - Less efficient for large data sets
  - XML definitions of collection of data
    - Global description of data in XML
    - Pointers to FITS data files
    - Solve some problems with FITS limitations of headers

Bibliographic References

- Wells, D. et al. 1981 "*FITS: A Flexible Image Transport System*", Astronomy & Astrophysics Suppl., 44, p.363
- Grosbøl, P., et al., 1988: "*Generalized extensions and blocking factors for FITS*", Astronomy & Astrophysics Suppl., 73, p.359
- Hanisch, R., et al., 2001: "*Definition of the Flexible Image Transport System (FITS)*", Astronomy & Astrophysics, 376, p.359

- FITS Support Office and other useful URL's:
  - http://www.heasarc.gsfc.nasa.gov/docs/heasarc/fits.html
  - http://www.cv.nrao.edu/fits/
  - http://archive.stsci.edu/fits/
  - http://fits.gsfc.nasa.gov/

Conclusions

- FITS is only accepted interchange format for astronomy
  - Major advantage of international science collaborations
- Self-documented
- Support multi-dimensional arrays and complex tables
- Controlled by International Astronomical Union
  - Backwards compatible developments
    - Once FITS, always FITS
- Default format for new data from observatories
- Used by many archives as storage format
- Supported by all astronomical data processing systems