4 CDS Data formats - data structure

4.1 Rasters

A *raster* is a series of GIS or NIS exposures grouped together as one observation "unit". Rasters may be

- Single exposures, yielding a single spectrum (for the GIS) or a single slit spectrogram (for the NIS).
- Repeated exposures at the same position, yielding a time series of spectra/slit spectrograms.
- A spatial scan covering an extended area, with one spectrum at each pixel position.

A schematic illustration of a raster is shown in Figure 4, where the length of the slit is 240 arcsec (120 pixels on the detector) and the width of the line windows is 10 pixels for all windows. The slit has been moved parallel to itself with 2 arcsec stepsize, and 120 individual spectra were recorded as illustrated in Figure 4. Thus the data file contains 120 individual exposures of each of the line windows and they cover an area of the solar surface of 240 $\times$ 240

4.2 Data format

The SOHO Science Operations Working Group (SOWG) has selected FITS as the standard file format for all the scientific data files generated by the PI teams. The reasons for this are two-fold: First of all, it provides compatibility with existing datasets and software used in the astronomical community, including ground-based solar data. Secondly, it facilitates interoperability between the various instrument teams at the SOHO Experimenters Operation Facility (EOF), since the file structure is independent of computer architecture and operating system. In other words, FITS files looks the same regardless of what computer the file is sitting on, and can be copied from computer to computer without modification.

An analysis of the CDS requirements by William Thompson at GSFC concluded that the best way to store the CDS data in FITS files was to use *binary tables* (Thompson, 1993). We also refer to CDS Software Note No. 3 (Thompson 1993a) for more details about the FITS binary tables. During operations, the science data, as well as associated details about the observations, are extracted from the telemetry files and written to FITS files.

The format of a CDS FITS binary table file as illustrated in Figure 5 can be described as having three parts:

1. The main FITS header. This will describe the observation as a whole. There will be no data array associated with the main header.
2. The FITS header for the binary table. Except for those records peculiar to the FITS structure, this will replicate the information in the main header. It will also describe the columns of the binary table.

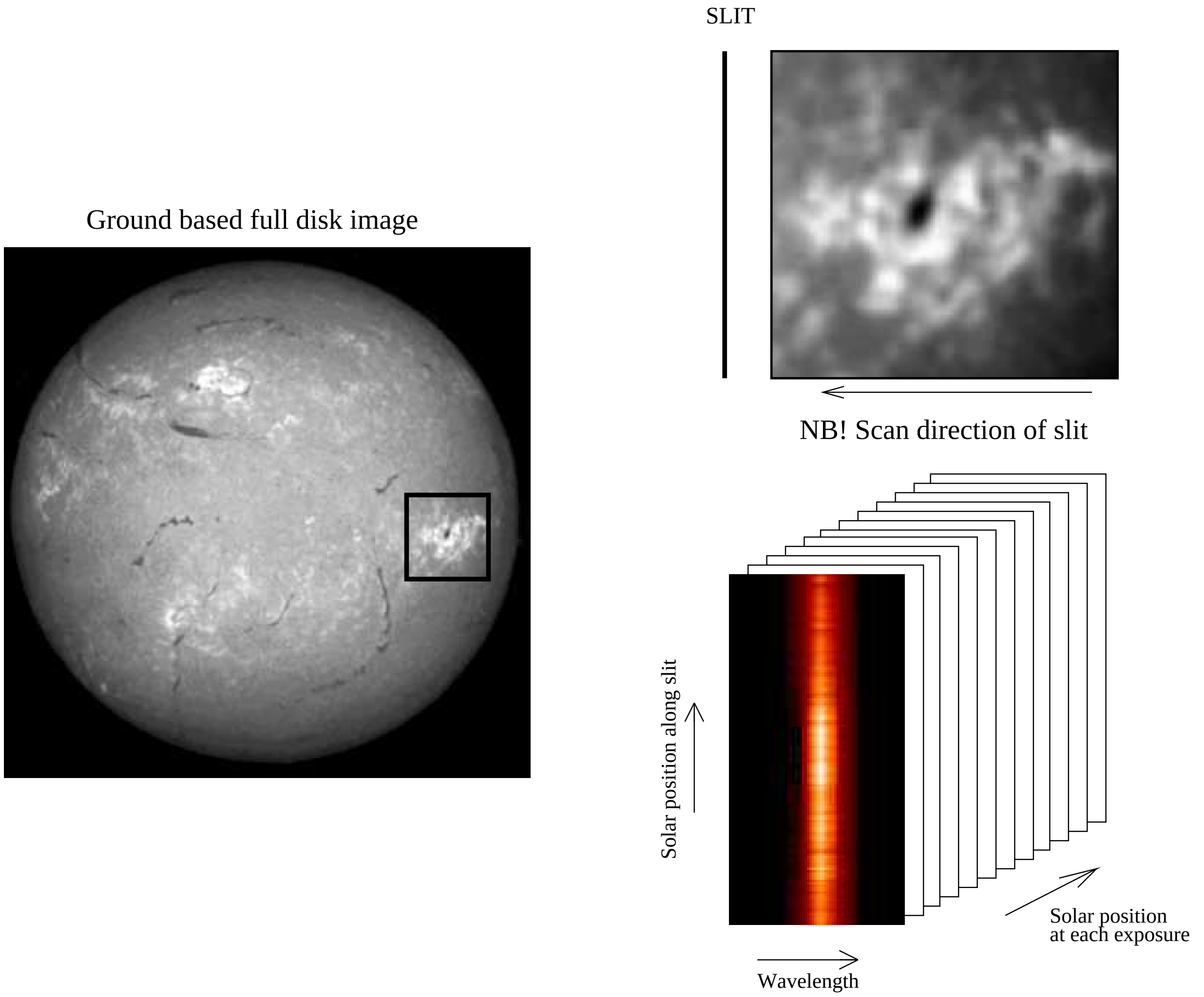

Figure 4: Schematic illustration of an observing sequence with CDS. The 240 arcsec slit was moved in steps of 2 arcsec to cover the marked area of the solar disk. Note that the slit is moved $west \rightarrow east$. The exposures are rearranged before the FITS file is created to appear as if they have been taken the other way around. A total of 120 exposures was obtained resulting in a $10 \times 120 \times 120$ data cube for each line window.

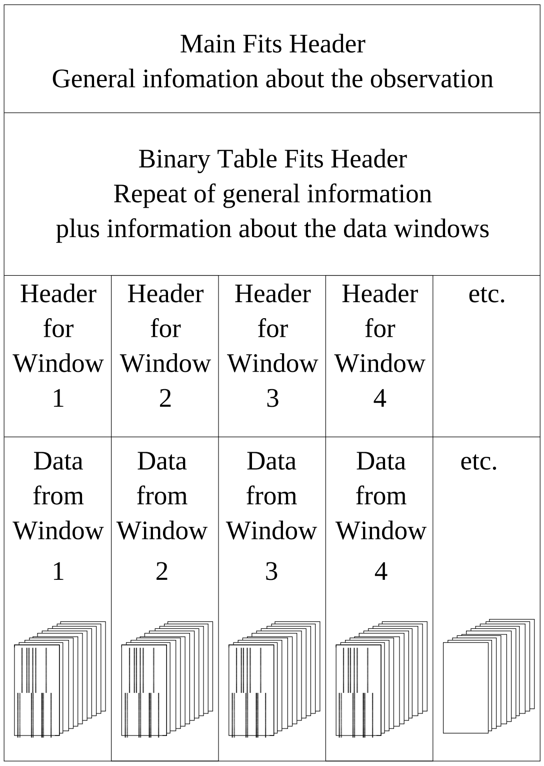

Figure 5: Schematic illustration of the FITS binary table file format. Each data window is stored in a separate column.

3. The body of the binary table. This will contain one complete raster sequence (a series of exposures). Every data window will be stored in a separate column. Additional columns will be used to store data about the individual exposures (e.g. time, pointing, etc.).

If the data on the spacecraft were compressed before transmission, the de-compression will have been applied before the creation of the FITS files. An example of a CDS FITS header with all the keywords is given in Appendix B.

4.3 Data structure in IDL

The Quicklook software for the CDS project is partially menu driven, partially command line driven. To simplify command line operations and construction of specialized analysis software, we felt the need to construct a simple and coherent internal data structure for the data in the FITS files.

To meet these needs, an IDL data structure has been designed to allow all the data in a FITS file to be accessed through a single variable, instead of having, for instance, one variable for each of the detector windows.

The IDL function `READCDSFITS` is used to read a FITS file (e.g., `s3300r00.fits`) into such an IDL structure and is invoked by the statement:

```
IDL> a = readcdsfits('s3300r00')
```

It returns a single anonymous IDL structure variable encompassing all the science data and the auxiliary information that goes with the spectral data in the FITS file. This variable is sent as a parameter to the other Quicklook routines. We have provided routines to "interface" with the data structure. These will return, for instance, the data array corresponding to one particular detector window, or a single spectrum from a given wavelength band. For further information we refer to CDS Software Note No. 9, *CDS IDL Data Structure*, and CDS Software Note No. 41, *Analyzing CDS Data in IDL: An Observers Guide*. In the following the CDS structure returned from `READCDSFITS` is called the Quicklook Data Structure – QLDS. For a discussion of the data structure we refer to CDS Software Note No. 9.

4.4 Specifying Individual Parts of a Raster

Despite the differences between the GIS and the NIS detectors, the data in *any* CDS raster may be viewed as a set of 4-dimensional arrays (one array for each line window). For simplicity, the term "window" is always used, even when "windowing" is not used (full readout).

By convention, the first array dimension is the "dispersion" dimension. The second dimension is the "solar X" dimension, the third is the "solar Y" dimension, and the fourth (when present) is the "time" dimension.

Note that although the nominal solar X/Y dimensions are *usually* aligned with the true solar X/Y directions, this is *not* always the case. Data taken during roll manouvers must be interpreted with care, as the Quicklook software often does not handle these cases properly (resulting in erroneous values for reported solar X/Y coordinates etc.).

To specify parts of the data in a raster, the Quicklook software uses the following terms/keywords:

`WINDOW`, or window index, refers to a specific window in a given raster/QLDS. Windows are referred by numbers (0..NWINDOWS-1).

`OFFSET` refers to the *offset* dispersion pixel within a given line window. The pair `(WINDOW,OFFSET)` uniquely defines a wavelength band and a specific dispersion pixel.

`BAND` is used to refer to a given *wavelength band*, numbered 1 and 2 for NIS, 1 through 4 for GIS.

`DETX` refers to the detector X position. The pair `(BAND,DETX)` has the same function as the pair `(WINDOW,OFFSET)`

`XIX` refers to the X index, i.e., the index of the second dimension of the data array associated with a line window.

`YIX` refers to the Y index, i.e., the index of the third dimension of the data array associated with a line window.

`TIX` refers to the time index, i.e., the index of the fourth dimension of the data array associated with a line window.

5 CDS Quicklook Software

5.1 How to get started

After entering IDL we need to read a FITS file containing science data. If you know the name of the FITS file(s), you may use `READCDSFITS` directly from the IDL prompt (as in Section 4.3 to read in the data.

If you don't know which FITS files you'd like to read in, there are two routines that may be used to browse through lists of observations, `XCAT`, and `PICKFITS`:

Using `XCAT` requires that your system is set up so that `XCAT` can locate the as-run database (see CDS Software Note No. 17).

Using `PICKFITS` requires that you (or someone at your site) use the routine `CFITSLIST` to produce/maintain a list of available FITS files (see the on-line documentation).

Let's assume your're using `XCAT`. Then you'd start by using a command like

```
IDL> xcat,q
```

where `q` is the output parameter, to receive the QLDS.

Select a line with a the desired study, and then select from the rasters in that study. You then select `Read and Exit with QL data structure` from the pulldown menu called `Read & View Current FITS File`. This will read in the FITS file and exit from XCAT, and the variable `q` will contain the Quicklook Data Structure (QLDS), with the data from the selected fits file.

You may also start the Quicklook "Display Menu" directly from `XCAT` by selecting `Quicklook` from the `Read & View Current FITS File` pulldown menu. You may even read several rasters after each other, and have one Quicklook Display Menu for each of them. You may exit `XCAT` without killing the Quicklook Display Menus.

NOTE: To enable `XCAT` and the Quicklook Display Menu to be functioning side by side, allowing *several* files to be read and displayed on screen simultaneously, it is impossible[1] to make the data

[1] well, *virtually* impossible