CXC-DM-004

CXC Data Model

Vol. 4

Virtual File Design

Jonathan McDowell
Chandra X-ray Center
October 22, 2001

Contents

1 Introduction

This document describes the CIAO DM ('Data Model') virtual file design. Virtual Files define a CXC DM block (see the Abstract Design document) in terms of another CXC DM block. The virtual file syntax used to specify this definition is defined in detail in Vol 12, User Reference.

2 Virtual File Concept

The Virtual File concept allows users to apply programs to an altered view of their data without the need to actually alter the data on disk. A virtual file is a DM block, and as such can be instantiated on disk as part of a dataset -in fact this is a common operation performed by

```
dmcopy virtual-file output-dataset.
```

The virtual file syntax consists of a dataset name followed by a set of 'directives', each of which is in square parentheses.

```
dataset[block directive][filter directive][select directive][bin directive][opt directiv
```

The detailed syntax of these directives is given in the User Reference.

3 The virtual and raw blocks

Every virtual file is a transformation of a single real ('raw') block of an input dataset. Once the dataset is specified, the raw block may be identified by:

- a block directive identifying the block by name.
- a block directive identifying the block by number within the dataset
- the implicit 'most interesting block' mechanism.

When only a dataset name is provided, the DM selects the 'most interesting block', which is usually the first non-null image or table. The selection use the following rules, in order of precedence:

- If there is only one block, use that one.
- If there is only one block that is not a null primary header, use that one.
- (PHA special case) If the primary header is called 'WMAP' and the second block is a table, use the table (since WMAP is auxiliary information).
- (Image case) If the primary header is not null, use it.
- (Table case) If the primary header is null, use the first block which is not excluded by the following special case:
- (GTI special case) Do not use a block whose name starts with 'GTI', 'ALLGTI', or 'STDGTI'. (since GTIs are usually auxiliary info to blocks like EVENTS).

The rules for selecting the virtual block name are:

- Use the output name specified explicitly by the block directive, if there is one.
- If a binning directive is in use, use the name of the raw block with '_IMAGE' appended.
- Use the raw block's name.

4 Column selection

The column descriptors of the virtual file are functions of the descriptors in the raw block.

- By default, the columns in the virtual file are the same as those in the raw block.
- The column select directive can be used to specify inclusion or exclusion of specific columns.
- If the select directive refers to one component of a vector column without mentioning the other, that component is promoted to a full scalar column.
- If the directive refers to a coordinate name, a new column is created with that name, whose values are generated by applying the coordinate transformation to its parent column.
- Future upgrades may include the ability to define new columns as functions of old ones in the spirit of dmtcalc, but that ability is not currently supported.

5 Table Filtering

Filtering refers to the ability to select only certain rows of the raw block table for inclusion in the virtual file. Any combination of columns or components in the raw block may be filtered on. In future releases, filtering on coordinates may also be supported. The DM filter specifies a subset of the N-dimensional space defined by the table columns; rows within this subset are visibile in the virtual file. 'Exclude filtering' is also supported, where only rows which do NOT fall in the subset appear in the virtual file.

The virtual file filter is parsed into a Subspace structure which lists each of the descriptors being filtered on. For each descriptor, an ordered set of valid ranges is defined. For two-dimensional descriptors, a region filter is also defined. With exclude filtering, the filter is explicitly inverted and the complement set is recorded in the data subspace.

Rows from the raw block are loaded into an input buffer; for each row in the buffer, the filter is applied and the row marked as good or bad. Good data is then copied to the output as requested by DM API calls.

Integer values are tested using closed intervals; real values use semi-open intervals. Bit values use a character string pattern.

It is also possible to filter using a range of the explicit input row number.

6 Image Filtering

Image filtering selects a single rectangular subset of data from an image. Coordinate systems are updated to reflect the new origin selected.

7 Table Binning

The bin directive allows users to create an image from a table. The entire image is created in memory, and the operation will fail if insufficient memory is available or if the required memory exceeds internal defaults. These defaults can be modified using an opt directive.

Coordinate systems are propagated from the table to the image. The image logical coordinates are the actual pixels in the binned image. The image physical coordinate system gives the original table values; the image world coordinate system gives the coordinate values for the table. Vector columns in the image are mapped to vector axis groups in the table.

The image range is by default the full range of the binning columns' coordinates, possibly modified by any filter on those columns. The virtual file specification can include an explicit range in each binned coordinate, and a blocking factor for each coordinate. The default data type for the binned histogram is a 2 byte integer, but this can also be explicitly selected.

7.1 Default Binning

Routines which require an image (and thus open a block using dmImageOpen) can generate one from an input virtual file even if that virtual file doesn't have an explicit binning specification. The DM guesses which columns of the table should be used and uses their full range and a blocking factor of 1. The guess is made by:

- Using the defined preferred columns, if specified in the file
- Using the sky or pos column, if present
- Using the first two columns in the file.

In the latter two cases, the resulting file is always a 2-D image.

8 Image Rebinning

A bin directive applied to an image can be used to generate a more coarsely binned image; blocking factors must be integer. The new image is created in memory.