OGIP Calibration Memo CAL/GEN/92-002

The Calibration Requirements for Spectral Analysis

(Definition of RMF and ARF file formats)

Ian M. George[1], Keith A. Arnaud, Bill Pence, Laddawan Ruamsuwan & Michael F. Corcoran
HEASARC,
NASA/GSFC,
Greenbelt, MD20771

Version: 2020 Dec 08

SUMMARY

The approach and calibration file formats adopted by the HEASARC for spectral analysis of X-ray spectra are outlined and discussed. Intended audience: HEASARC programmers, hardware teams & authors of spectral analysis software.

Note: Users should also refer to the addendum to this memo, CAL/GEN/92-002a.

[1]Deceased

LOG OF SIGNIFICANT CHANGES

Release Date	Sections Changed	Brief Notes
1992 Oct 07		Original Version
1995 Jan 11	All	Made compatible with `LaTeX2HTML` s/w
1998 Dec 19	All	HDUVERS=1.3.0 of the RMF matrix extension
		Compatibility with OGIP/92-002a
2005 Feb 15		From KAA: fixed some ambiguities
2005 Jul 05	2	MFC: describe more fully how RMF and ARF are related
2007 Nov 1	§§3.2.1, 3.2.2	MFC: allowed F_{chan}, N_{chan} and $Chan$ to be 4 byte integers (From KAA)
2013 Jul 2	4	added description of type II ARF
2016 July 11		minor formatting. grammar corrections by MFC
2020 Dec 8		added possibility of multiple MATRIX extensions in RMF (KAA)
2022 Oct 11	All	MFC: replaced references to http://legacy with https://heasarc

Contents

1 INTRODUCTION

Calibration files within the HEASARC Calibration Database (CALDB) have been classified into 2 types:

- **BASIC CALIBRATION FILES (BCFs)** containing all the basic calibration information for a given instrument. The BCFs will contain calibration information which is both independent of time (in most cases data originating from ground calibration measurements), and information which is expected to vary throughout the mission (mainly from in-orbit measurements). In many BCFs the data will be stored in the form of large n-dimensional arrays. A more detailed discussion of the general formats and organization of the BCFs can be found in the Office of General Investigator Programs (OGIP) Calibration Memo CAL/GEN/92-003.
- **CALIBRATION PRODUCT FILES (CPFs)** are essentially rearrangements of a subset of the information within the BCFs suitable for a specific task within a given Data Analysis Package. The extraction of the necessary information from the BCFs and construction of CPFs is performed by 'Stage 2 Calibration s/w', with reference to HK data (*eg* observation date, instrument mode *etc*) as necessary.

All BCF and CPF calibration files within the High Energy Astrophysics Science Archive Research Center (HEASARC) adhere to the Flexible Image Transport System (FITS) standard. Note that the HEASARC Calibration Database can also be used to store data in non-standard, mission-specific format. Such files are called Primary Calibration Files (PCF) and are stored to provide an historical record of mission calibration and data processing, but are not meant for general data analysis.

An overview of the relationship between the BCFs, CPFs and other elements within the generic calibration dataflow is given in CAL/GEN/91-001 (George 1992).

This document describes in detail the FITS format adopted by the HEASARC for the calibration files required for the spectral analysis of X-ray spectral data.

2 OVERVIEW

2.1 The Redistribution Matrix

X-ray spectral analysis consists of convolving a model spectrum with the response of the detection system, and comparison of this convolved model with the observed data in order to constrain model parameters and thus derive physical quantities (like absorption columns, fluxes, emission measures, etc.) The "detector response" $R(I, E)$ is proportional to the probability that an incoming photon of energy E will be detected in the output detector channel I. As such, the response is a continuous function of E, while the detector output consists of a discrete number of channels. The continuous response function is converted to a discrete function by creating a "response matrix" $R_D(I, J)$ at discrete energies E_J such that

$$R_D(I, J) = \frac{\int_{E_{J-1}}^{E_J} R(I, E) dE}{(E_J - E_{J-1})} \tag{1}$$

R_D is often referred to as the "Redistribution Matrix", since it describes how a photon of energy $E_{J-1} < E < E_J$ is "redistributed" into output detector channels. The file which contains the "Redistribution Matrix" has been called the "Redistribution Matrix File", or RMF for short.

It is sometime useful to split the redistribution matrix into parts. For instance, X-ray calorimeters are high resolution but have long tails in their response down to low energies. This means that $R(I, E)$ is triangular and can be very large due to the small energy bins required because of the detector's high resolution. The size of the redistribution matrix can be reduced by dividing it into two parts. A high resolution part with small energy bins, which is mostly zeroes, and low resolution part with larger energy bins, which is triangular. For computational simplicity the larger energy bins should be made by combining the small energy bins.

2.2 Detector Effective Area

In general the response of a detector to a source of photons depends not only on the redistribution of photons but also on the sensitivity of the detector to photons of known energy. For example, for many X-ray detectors, the sensitivity is a function of off-axis angle: a source observed on-axis usually appears brighter than the same source observed away from the detector optical axis. The sensitivity of a detector to a photon of a given energy $E_{J-1} < E < E_J$ can be described by an array of values $A(J)$, and the file which contains this information is usually called the Ancillary (sometimes, Auxiliary) Response File, or ARF for short. The Ancillary Response File gives the "effective area" of the detector system, and usually includes such components as mirror vignetting, filters, etc.

2.3 Implementation

Spectral analysis of a file containing the X-ray spectrum of a source (often called, for historical reasons, a Pulse Height Analyzer or PHA file) using XSPEC or similar spectral analysis software requires the following Calibration Product Files:

1. **A DETECTOR REDISTRIBUTION MATRIX FILE (RMF)**

   - The RMF consists of one or more compressed 2-d (energy vs PHA channel) FITS extensions (Section 3), and another extension explicitly listing the nominal energy range of each PHA channel.
   - The RMF is created by folding together individual components due to the:
     - Detector Gain
     - Detector Energy Resolution (including the response to a monoenergetic source *eg.* escape peaks, partial charge tail *etc.*)

2. **AN ANCILLARY RESPONSE FILE (ARF)**

   - The ARF consists of a 1-D array vs. energy stored as a table extension in a FITS file (see Section 4), and includes the summed contribution of efficiency components, *ie* those not involved in the redistribution of photons such as:
     - the Effective Area of the Telescope/Collimator (including vignetting),
     - the Filter Transmission (if any)
     - the Detector Window Transmission
     - the Detector Efficiency
     - any additional energy dependent effects (*eg* correction factors for the *p.s.f.*).

Sometimes the calibration information contained in the redistribution matrix file and the ancillary response file is incorporated into a single file (see Section 6), which is usually called the "response" file (and usually denoted by a `.rsp` file extension). Combining the redistribution matrix and detector effective area information in a single response matrix was common for early missions (*EINSTEIN* and *ROSAT*, for example, or the output of the `pcarsp` tool for analysis of RXTE Proportional Counter Array Data) but is generally not done for more modern missions.

Note that, in the spectrum (PHA) file and the associated redistribution matrix (RMF) and effective area (ARF) files, the detector channels used to specify the observed spectrum and the detector calibration information refer to **unbinned detector channels**. The number of unbinned channels for a given detector is given explicitly by the `DETCHANS` keyword within the MATRIX extension (see Section 3). Rebinning of the PHA data can be specified by the GROUPING flag within the PHA file (see OGIP/92-007). It is recommended that rebinning of the data and the calibration data supplied by the RMF & ARF be done by spectral analysis software, so that the binning can be easily adapted to emphasize important regions in the spectrum.

2.4 Rationale

For many high-energy instruments, the detector gain (the mapping between detector channel and approximate X-ray energy) and energy resolution do not vary significantly with detector

coordinates or time (in particular not within an "observation"). Therefore, usually a single RMF can be constructed for a given observation and used to analyze all sources in the field-of-view. On the other hand, analysis of individual sources from a single observation will usually require customized ARFs, since the ARF depends on details of the location of the source on the detector. Spectral analysis software like `XSPEC` can use a common RMF along with source-specific ARF and PHA files. For complicated detectors there are situations in which the use of a common RMF is not possible; such instances are discussed in Section 5.

The benefit of isolating the effective area components (which are a simple function of energy for a given observation, but which may depend on time, detector position, observing mode *etc*) into the Ancillary Response File allows these components to be listed individually within the ARF if desired (see Section 4). While including the individual contributions to the ARF is optional (but recommended), the HEASARC requires that the combined product of these individual components must always be provided in the ARF.

2.5 Additional Notes

- Use of "Pulse-Invariant" channels (usually "PI" channels) are sometimes defined from some conversion of PHA channels onto a uniform scale appropriate for some standard detector gain setting. The HEASARC recommends that these "PI" channels **NOT** be used for data storage or analysis[2]. The detector gain specified within the RMF should provide, in almost all cases, the necessary information needed to convert PHA channels to a standard, uniform energy scale.

- If significant gain changes occur during a given observation (as was often the case for the Imaging Proportional Counter on the *EINSTEIN* Observatory), separate PHA files & RMFs should be constructed for times when the gain was relatively constant, prior to spectral analysis. The PHA files should be then analyzed with the appropriate (time-dependent) ARFs & RMFs.

- An RMF contains only information applicable to the redistribution of photons into detector channels. The effects of the detector effective area, detector filter (if appropriate), and Window Transmission function are included in the ARF. Additional effects, such as obscuration by a Window Support Structure, absorption due to contaminants on the detector (*eg* water ice or other substances), etc., can also be included in the ARF. This minimizes the number of RMFs required for a analysis of given observation (generally to one in most cases), minimizing disk-space requirements, and facilitating investigations of the effects of these individual components on the spectral analysis.

[2]A primary goal of the HEASARC is to standardize the format of similar files from different instruments to allow multi-mission and/or multi-wavelength analysis. Besides strictly being the incorrect method in the general case, the implicit mapping of X-ray spectral data from observed to some "standard" channel grid cannot be performed correctly for low resolution detectors. Some past and present missions have used PI channels without preserving the raw PHA values. In such cases, the data describing the entire instrument response should be stored in the same format as the RMFs described below, (see also Section 6). "PHA" datasets in which use a PI channel grid are denoted by the `CHANTYPE = 'PI'` FITS keyword within the file header (see OGIP/92-007).

3 THE HEASARC STANDARD RMF FORMAT

The standard RMF format consists of a FITS file with a null primary array and at least two extensions:

- The Redistribution (MATRIX) extension(s)
- an (EBOUNDS) extension containing the nominal energy bounds of each channel

both employing the BINTABLE FITS format.

3.1 The RMF Redistribution Extension

In general, redistribution matrices are sparse, so storing the full matrix in an RMF is not very efficient in terms of file storage and file access. In order to minimize disk-space requirements, the RMF Redistribution Matrix should be stored in a compressed format in which all matrix elements below a given threshold (specified by the `LO_THRES` FITS keyword in the RMF file) are ignored. This format is very similar to that used originally by the XSPEC `.rsp` standard-format response files.

3.1.1 Extension Header

The header must include the following (mandatory) keywords/values:

- `EXTNAME = 'MATRIX' or 'SPECRESP MATRIX'` - the name (*ie* type) of the extension
- `TELESCOP` - the"telescope" (*ie* mission/satellite name).
- `INSTRUME` - the instrument/detector.
- `FILTER` - the instrument filter in use (if any)
- `CHANTYPE` - whether the detector channels given in the matrix are uncorrected (*ie* as assigned by the detector electronics, `CHANTYPE = 'PHA'`), or have been corrected (*eg* are "pulse invariant", `CHANTYPE = 'PI'`).
- `DETCHANS` - the total number of raw detector PHA channels in the full (uncompressed) matrix.
- `HDUCLASS = 'OGIP'` - file format is OGIP standard.
- `HDUCLAS1 = 'RESPONSE'` - extension contains response data.

- `HDUCLAS2 = 'RSP_MATRIX'` - extension contains a response matrix.
- `HDUVERS = '1.3.0'` - version of the file format.
- `TLMIN#` - the first channel in the response. # is the column number for the F_CHAN column (see below).

If there are multiple MATRIX extensions then each one should have a different value of their EXTVER keyword (1 for the first extension, 2 for the second and so on).

The following optional keywords may be useful for programs reading the file in that they specify the amount of memory various arrays will require.

- `NUMGRP` - the total number of channel subsets. The sum of the N_GRP column.
- `NUMELT` - the total number of response elements. The sum of the N_CHAN column.

The following optional keywords supply further information:

- `PHAFILE` - name of PHA file for which this file was produced
- `LO_THRES` - minimum probability threshold used to construct the matrix (matrix elements below this value are considered to zero and are not stored)
- `HDUCLAS3` - giving further details of the stored matrix
  Allowed values are:
  - `'REDIST'` for a matrix whose elements represent probabilities associated with the photon redistribution process only
  - `'DETECTOR'` for a matrix whose elements have been multiplied by all energy-dependent effects associated with detector (eg detector efficiency, window transmission *etc*).
  - `'FULL'` for a matrix whose elements have been multiplied by all energy-dependent effects associated with detector, optics, collimator, filters *etc*.

The following keywords are now obsolete but may be included for the benefit of old software. They should be commented as obsolete.

- `RMFVERSN = '1992a'`
- `HDUVERS1 = '1.1.0'`
- `HDUVERS2 = '1.2.0'`

Finally, the following keywords are mandatory if these calibration data are ever to form an entry in a Calibration Index File (CIF; see CAL/GEN/92-008, George, Pence & Zellar 1992)

These keywords and their acceptable values are listed in more detail in CAL/GEN/92-011 (George, Zellar & Pence 1992)

However, it should be noted that there is often no such requirement for RMF or ARF files as they are usually specific to a given PHA file (but see Section 6).

- `CCLS`0001 (= `'CPF'`) - the OGIP-class of this calibration file.
- `CCNM`0001 (= `'MATRIX'`) - the (CIF) codename for this type of calibration dataset.
- `CDTP`0001 (= `'DATA'`) - the OGIP code for the form of the contents of the file ('real' data, a taskname and associated parameter inputs etc)
- `CVSD`0001 - the UTC date (in yyyy-mm-dd format) when this calibration data should first be used
- `CVST`0001 - the UTC time (in hh:mm:ss format) on the day `CVSD`0001 when this calibration data should first be used.
- `CDES`0001 - a string giving a brief descriptive summary of this dataset

3.1.2 Data Format

In the general case, the organization of the data within this extension will be as follows (with the matrix x-axis = raw PHA channel, y-axis = Energy) with each row of the BINTABLE referring to a single energy range (thus the number of rows = number of energy bins) and consist of the following columns:

1. E_{low}, a 4-byte REAL scalar for each row containing the lower energy bound of the energy bin.
   The FITS column name is `ENERG_LO`.
   The recommended units are keV.
2. E_{high}, a 4-byte REAL scalar for each row. containing the upper energy bound of the energy bin.
   The FITS column name is `ENERG_HI`.
   The recommended units are keV.
3. N_{grp}, a 2-byte INTEGER scalar for each row containing the number of 'channel subsets' for the energy bin (see below).
   The FITS column name is `N_GRP` (unitless).

4. F_{chan}, a fixed- or variable-length 2-byte or 4-byte INTEGER array for each row. Contains the channel number of the start of each "channel subset" for the energy bin.
   The FITS column name is `F_CHAN` (unitless).

5. N_{chan}, a fixed- or variable-length 2-byte or 4-byte INTEGER vector for each row. Contains the number of channels within each "channel subset" for the energy bin.
   The FITS column name is `N_CHAN` (unitless).

6. *Mat*, a (fixed- or variable-length) REAL array. Each element *Mat* is 4-byte REAL number containing all the response probability values for each 'channel subset' corresponding to the energy bin for a given row.
   The FITS column name is `MATRIX` (unitless).

These are summarized in Table 1.

A final column may be added for responses of grating instruments.

1. *Order*, a (fixed- or variable-length) INTEGER vector (array, each element within which is 2-byte) for each row containing the dispersion order of each 'channel subset' in the energy bin.
   The FITS column name is **ORDER**.
   (unitless).

This column matches the F_CHAN and N_CHAN columns and requires that every 'channel subset' be for a single order.

3.1.3 Points to Note & Conventions

- The ordering of the columns used here is recommended.
- Values of both `ENERG_LO` & `ENERG_HI` are given in each row (j) for clarity and for efficiency of access. The order should be monotonically increasing with increasing row number, starting from the minimum `ENERG_LO` value. In no case should there be any overlap between consecutive energy bins, so that for row j, `ENERG_LO`(j) $\geq$ `ENERG_HI`(j-1). In most RMFs, `ENERG_LO`(j) = `ENERG_HI`(j-1)).
- The concept of "channel subsets" is included to minimize the RMF storage requirements for instruments for which the 2-d matrix is sparse, and consists of non-zero values in two or more (unconnected) regions of channel–energy space. A channel subset therefore consists of a number (N_{chan}) of contiguous channels for which the matrix elements are above the `LO_THRES` threshold. Thus, using the above notation, a given row of the `MATRIX` array contains the elements appropriate to

channels	$F_{chan}(1)$	through	$(F_{chan}(1) + N_{chans}(1) - 1)$
followed by	$F_{chan}(2)$	through	$(F_{chan}(2) + N_{chans}(2) - 1)$
followed by	$F_{chan}(3)$	through	$(F_{chan}(3) + N_{chans}(3) - 1)$
...	...	...	...
followed by	$F_{chan}(N_{grp})$	though	$(F_{chan}(N_{grp}) + N_{chans}(N_{grp}) - 1)$

Table 1: OGIP format (1992a) for storing photon redistribution matrices within an RMF

Extension to *(filename)*.**RMF**
Name: RMF
Description: Photon Redistribution Matrix
Format: BINTABLE

column					
1	2	3	4	5	6
contents					
Low energy bound for row	**High energy bound for row**	**Number of channel subsets for row**	**First channel in each subset for row**	**Number chans in each subset for row**	**(non-zero) Matrix elements for row**
E_{low}	E_{high}	N_{grp}	$F_{chan}(i)$ $i = 1, N_{grp}$	$N_{chan}(j)$ $j = 1, N_{grp}$	$Mat(k)$ $k = 1, \sum_{j=1}^{N_{grp}} N_{chan}(j)$
format of each column					
4-byte real	4-byte real	2-byte integer	2-byte or 4-byte integer array	2-byte or 4-byte integer array	4-byte real array
total number of elements per row					
1	1	1	variable, or MAX(N_{grp})	variable, or MAX(N_{grp})	variable, or MAX($\sum_{j=1}^{N_{grp}} N_{chan}(j)$)
column name					
ENERG_LO	ENERG_HI	N_GRP	F_CHAN	N_CHAN	MATRIX

- Generally, the `F_CHAN`, `N_CHAN` and `MATRIX` columns will be FITS variable-length arrays (Cotton, Tody & Pence, 1995). For variable-length arrays, the number of elements within each array varies from row to row. For a given row, the number of elements within the `F_CHAN` and `N_CHAN` columnd for that row equals the value of `N_GRP` for that row. It should be noted that, if $N_{grp} \leq 3$ for all rows, it is usually more efficient in terms of both disk storage requirements[3] and speed of access[4] to designate that array as fixed-length format. This criterion has been adopted as the general policy of all files containing arrays of variable length within the HEASARC calibration database. The HEASARC recommends the following guidelines:

  - All RMFs for a given instrument should employ the same format.
  - in most cases the `F_CHAN` & `N_CHAN` columns will be fixed-length integer arrays, since commonly, for instruments for which there is data in the HEASARC archives, `N_GRP` $\leq$ 3. Note, in all cases the `F_CHAN` & `N_CHAN` columns contain the same number of elements, thus should both either be in fixed- or variable-length array format.
  - Due to the greater read-efficiency, the `MATRIX` column is also in fixed-length format **unless** this leads to a significant increase (say > 1.5) in disk-space requirements.
  - Unused elements within an array should be padded with 'null data' values.

  Spectral analysis software (like XSPEC) software should be able to handle both fixed- and variable-length arrays. Use of the FITSIO interface by spectral analysis software (Pence 1992) is recommended to read FITS files, since FITSIO can transparently interpret either format.

- If a column contains a constant value in every row, then the column can be deleted from the table and transformed into a keyword value. Analysis software should first look for a keyword value having the name of one of the required columns; if the keyword is not found, then the software should look for a column with that name.

- Each row within the RMF matrix will be normalized to 1 **detected** photon, *ie* each element of *Mat* will contain the probability of a detected photon within the appropriate energy range giving rise to a signal in that PHA channel, i.e., for a given row J,

$$\sum MATRIX(J) = 1$$

  Effects due to detector efficiencies $< 100\%$, absorption by mirrors, filters & the detector window *etc* should be included within the ARF and not the RMF. Note that, in practice, the sum of the probabilities in a given row of the `MATRIX` array may be less than unity. due to the finite probability of a photon being registered as either below or above the PHA discriminator thresholds, and/or if `LO_THRES` > 0.0.

[3] A variable-length FITS array requires at least 12 bytes of disk storage space: 4 bytes for the length of the vector, 4 bytes for the offset address, and 4 bytes for the value of each element. Thus using variable-length format for an array will lead to a saving of disk-space only if the maximum number of elements in that array is greater than 3.

[4] Due to always having to read the offset value first, before reading the value of a given element within the array, there is a significant performance inefficiency in using variable-length vectors.

3.2 The RMF EBOUNDS Extension

The RMF EBOUNDS extension lists the (nominal) energy boundaries of each of the (raw) detector channels within the redistribution matrix given above. It should be stressed that these energies are **not** necessarily the same as those given in the `ENERG_LO` & `ENERG_HI` columns of the `MATRIX` extension discussed above. The (nominal) energy boundaries are required by spectral analysis packages when (say) the user would like the results of spectral analysis (PHA data and best-fitting model) to be displayed as a function of photon energy, rather than detector channel. Providing this information in a separate FITS extension provides for more efficient access when plotting, especially if the response matrix is large. The format described here is a simple 1-dimensional list (as a function of raw detector PHA channel) which gives the nominal energy boundaries for each detector channel.

3.2.1 Extension Header

For clarity, the header must include the same mandatory keywords as the RMF extension, namely:

- `EXTNAME` (= `'EBOUNDS'`) - the name (*ie* type) of the extension
- `TELESCOP` - the"telescope" (*ie* mission/satellite name).
- `INSTRUME` - the instrument/detector.
- `FILTER` - the instrument filter in use (if any)
- `CHANTYPE` - whether the detector channels given in the matrix are PHA or PI channels (see above).
- `DETCHANS` - the total number of raw detector PHA channels in the full (uncompressed) matrix.
- `HDUCLASS = 'OGIP'` - file format is OGIP standard.
- `HDUCLAS1 = 'RESPONSE'` - extension contains response data.
- `HDUCLAS2 = 'EBOUNDS'` - extension contains a response matrix.
- `HDUVERS = '1.2.0'` - version of the file format.

along with the optional keywords

- `PHAFILE` - name of PHA file for which this file was produced

The following keywords are now obsolete but may be included for the benefit of old software. They should be commented as obsolete.

- `RMFVERSN = '1992a'`
- `HDUVERS1 = '1.0.0'`
- `HDUVERS2 = '1.1.0'`

Finally, if these calibration data are ever to form an entry in a Calibration Index File, the mandatory `C***` keywords listed in Section 3.1.1 are also mandatory, but in this case with `CCNM0001 = 'EBOUNDS'`.

3.2.2 Data Format

A BINTABLE FITS format has been chosen whereby each each row refers to a single detector channel. The number of rows is thus the number of (raw) detector channels **and must correspond exactly** to the channels within the PHA file and hence also to the value of the `DETCHANS` keyword in the RMF `MATRIX` extension described above. Thus, we have

1. *Chan*, a 2-byte or 4-byte INTEGER scalar giving the raw channel number for each row.
   The FITS column name is **CHANNEL**.
   (unitless)
2. E_{min}, a 4-byte REAL scalar for each for each row containing the nominal energy corresponding to the lower boundary of the detector channel.
   The FITS column name is **E_MIN**.
   The recommended units are keV.
3. E_{max}, a 4-byte REAL scalar for each for each row containing the nominal energy corresponding to the upper boundary of the detector channel.
   The FITS column name is **E_MAX**.
   The recommended units are keV.

Table 2 summarizes the organization of this extension

3.2.3 Points to Note & Conventions

- The ordering of the columns is arbitrary, however the order presented here is strongly recommended.
- Note that `E_MIN` and `E_MAX` in the EBOUNDS extension are different from `ENERGO_LO` and `ENERGO_LO` given in the MATRIX extension. `E_MIN` and `E_MAX` are determined by the characteristics of the detector; values of `ENERG_LO` and `ENERG_HI` are generally selected by the calibration scientist, who may choose to have much finer energy resolution than the detector offers in order to oversample the detector response.

Table 2: OGIP format (1992a) for the EBOUNDS extension within RMFs

Extension to *(filename)***.RMF**
Name: EBOUNDS
Description: Nominal energy boundaries for each detector channel
Format: BINTABLE

column		
1	2	3
contents		
Detector channel number	**Low energy bound for row**	**High energy bound for row**
Chan	E_{min}	E_{max}
format of each column		
2-byte or 4-byte integer	4-byte real	4-byte real
total number of elements per row		
1	1	1
column name		
CHANNEL	E_MIN	E_MAX

- Because pulse-height analysers generally oversample the true spectral response of most X-ray detectors, there is **no** guarantee that an incident X-ray with an energy between `E_MIN` and `E_MAX` will be detected in the corresponding detector channel given in the EBOUNDS extension. Determining in which channels an incident X-ray photon may be detected requires a full spectral analysis using the redistribution matrix and an assumed intrinsic source spectrum.

4 THE HEASARC STANDARD ARF FORMAT

The ARFs are relatively straightforward, consisting of a simple 1-dimensional list (as a function of energy) of the product of the various components required for spectral analysis not involved in the photon redistribution process (see Section 2.4).

4.1 The ARF Extension

4.1.1 Extension Header

The header must include the following (mandatory) keywords:

- `EXTNAME` (= `'SPECRESP'`) - the name (*ie* type) of the extension
- `TELESCOP` - the "telescope" (*ie* mission/satellite name).
- `INSTRUME` - the instrument/detector.
- `FILTER` - the instrument filter in use (if any)
- `HDUCLASS = 'OGIP'` - file format is OGIP standard.
- `HDUCLAS1 = 'RESPONSE'` - extension contains response data.
- `HDUCLAS2 = 'SPECRESP'` - extension contains an ARF.
- `HDUVERS = '1.1.0'` - version of the file format.

The following optional keywords supply further information:

- `PHAFILE` - name of PHA file for which this file was produced

The following keywords are now obsolete but may be included for the benefit of old software. They should be commented as obsolete.

- `ARFVERSN = '1992a'`
- `HDUVERS1 = '1.0.0'`
- `HDUVERS2 = '1.1.0'`

As for the RMF file, if these calibration data are ever to form an entry in a Calibration Index File, the `CC***` keywords listed in Section 3.1.1 are also mandatory, but in this case with `CCNM0001 = 'SPECRESP'`. If (in addition to the *Prod* array) the ARF `SPECRESP` extension also lists each of the individual contributing components (see below), and these components are to be be listed in the Calibration Index File, then each component must have its own unique set of `C***` keywords (denoted by `CCNM`$XXXX$ *etc* where $XXXX$ is a number of the form `0002`, `0003` *etc* In this case, the `CCNM`$XXXX$ must conform to the appropriate standards given in CAL/GEN/92-011 (George, Breedon and Corcoran 1992).

4.1.2 Data Format: Type I - a single arf

The general HEASARC standard for ARFs also makes use of the BINTABLE FITS format, and thus the data again resides in a single extension of a FITS file (though generally as a file separate from the RMF) with a null primary array. As in the case of the RMFs, each row of the BINTABLE refers to a single energy range. The ARF SPECRESP extension must use the same energy binning as the RMF MATRIX extension to which its associated, that is, the ARF should have the same `ENERG_LO` and `ENERG_HI` values as its associated RMF file, and so the ARF SPECRESP and RMF MATRIX extensions should have the same number of rows.

In all cases the following columns are included in the SPECRESP extension of the ARF (preferably as the first 3 columns within the table):

1. E_{low}, a 4-byte REAL scalar for each row containing the lower energy bound of the energy bin.
   The FITS column name is **ENERG_LO**.
   The recommended units are keV.

2. E_{high}, a 4-byte REAL scalar for each row containing the upper energy bound of the energy bin.
   The FITS column name is **ENERG_HI**.
   The recommended units are keV.

3. *Prod*, a 4-byte REAL scalar for each row containing the product of all the components (effective area, filter transmission, correction factors *etc*) specific to a given PHA file (*ie* the spectral response of the instrument as a whole).
   The FITS column name is **SPECRESP**.
   The recommended units are cm^2.

Other columns can be added to show the various components out of which *Prod* was constructed but these are optional.

Table 3 summarizes the organisation of an ARF.

4.1.3 Data Format: Type II - multiple arfs

It is sometime convenient to store many ARFs in the same file so we provide a type II format analogous to that for spectral files. As with the case of a single ARF, the data are stored in a BINTABLE extension. However, vector arrays are used in place of scalars for the columns. Thus the energies and response are vectors and each row of the table contains a single ARF. The notation below is the same as that used in the description of the type I format.

1. Num, a 2-byte INTEGER giving the reference number of the spectrum stored in this row.
   The FITS column name is **ARF_NUM**.
   (unitless)

2. E_{low}, a 4-byte REAL array containing the lower energy bound of the energy bin.
   The FITS column name is **ENERG_LO**.
   The recommended units are keV.

3. E_{high}, a 4-byte REAL array containing the upper energy bound of the energy bin.
   The FITS column name is **ENERG_HI**.
   The recommended units are keV.

4. *Prod*, a 4-byte REAL array containing the product of all the components (effective area, filter transmission, correction factors *etc*) specific to a given PHA file (*ie* the spectral response of the instrument as a whole).
   The FITS column name is **SPECRESP**.
   The recommended units are cm^2.

If the **TELESCOP**, **INSTRUME**, **DETNAM**, **FILTER** differ between the ARFs then these keywords can be included as columns although this is not recommended.

4.1.4 Points to Note & Conventions

- The ordering of the columns used here is recommended.
- Values of both `ENERG_LO` & `ENERG_HI` are given in each row (j) of the SPECRESP extension for clarity and for efficiency of access. The order of the energy bins should be monotonically increasing with increasing row number, starting from the minimum `ENERG_LO` value. In no case should there be any overlap between consecutive energy bins, so that for row j, `ENERG_LO`(j) $\geq$ `ENERG_HI`(j-1). In most RMFs, `ENERG_LO`(j) = `ENERG_HI`(j-1)).

Table 3: OGIP format (1992a) for ARFs

Extension **to** *(filename)*.**ARF**
Name: ARF
Description: Ancillary Response File
Format: BINTABLE

column			
1	2	3	4, 5, 6 ... (as necessary)
contents			
Low energy bound for row	**High energy bound for row**	**Product of components for row**	**Component 1, Component 2, Component 3, ... for row** (as necessary)
E_{low}	E_{high}	$Prod$	C_1, C_2, C_3, ...
format of each column			
4-byte real	4-byte real	4-byte real	4-byte real (or 2-byte integer if more appropriate)
total number of elements per row			
1	1	1	1, 1, 1, ...
column name			
ENERG_LO	ENERG_HI	SPECRESP	(as in original BCF file)

- The dimension of the data within the *Prod* column will be length^2 (due to the inclusion of the effective area).

5 USAGE: TYPICAL SCENARIOS

For non-imaging devices (assuming a time-stable detector gain) a User extracts a PHA file, then runs software to generate a redistribution matrix file and the ancillary response file appropriate for the observation. The user than inputs these files along with the PHA file, into (*eg* XSPEC) or some other analysis software and chooses an appropriate model of the emission. The model is folded through the redistribution matrix multiplied by the effective area given in the ARF, and then compared directly to with the data. A new redistribution matrix and/or effective area curve can be generated to explore the effects of the observational parameters and/or to employ a new detector gain relation. Spectral analysis could continue using the original PHA file & RMF in conjunction with the new ARF, for example. In some cases, for a given detector, a single RMF may be applicable to several datasets within the HEASARC archive.

For imaging instruments (with a time- and position-stable Detector Gain etc.) a User performs an almost identical set of actions as above for each PHA file extracted (*ie* from each source in the image). Each PHA file therefore has an associated ARF file, which the User creates and/or customizes using a response generator. However all the PHA files will normally share a single RMF.

For imaging instruments (with a time-stable Detector Gain *etc*, but which varies with detector position) a User has to construct both a RMF & ARF for each PHA file extracted (*ie* from each source in the image). Users can customize the individual ARFs as desired.

For instruments (of any type) for which the Detector Gain *etc* is not stable with time (*ie* significantly varies over the course of a pointing), the observational dataset should be broken-down into a series of periods for which all Detector-related quantities **are** considered sufficiently constant. Separate PHA files, RMFs and associated ARFs can then be constructed for each of these periods (with each RMF obviously containing the matrix constructed using the gain setting appropriate to its time-window). Spectral analysis is then performed on these files either individually or simultaneously.

6 SPECIAL CASES

Sometimes the effective area curve has been combined with the redistribution matrix to produce a "spectral response" file, or the spectra in raw PHA channels has been "corrected" to a standard channel system (as denoted by the `CHANTYPE = 'PI'` keyword within the PHA file), requiring use of a "corrected" redistribution matrix. In such cases, the response file format should follow

the redistribution matrix file format given above.

If no ARF is specified within the PHA file (via the `ANCRFILE` keyword – see OGIP/92-007 (Arnaud, George & Tennant 1992), then the Spectral Analysis Package should assume that the instrument spectral response (*i.e.* the *Prod* array from Table 3) has been folded in with the redistribution matrix (*Mat* from Table 1), and this information is what is stored in the RMF extension. In this case the following changes to the list of mandatory keywords/values given in Section 3.1.1 are necessary to the header of the (new) RMF extension:

- `EXTNAME` (= `'SPECRESP MATRIX'`) - the name (*ie* type) of the extension
- `CCNM0001` (= `'SPECRESP MATRIX'`) - the (CIF) codename for this type of calibration dataset.

Again, it is emphasized that this is generally not recommended, especially in the case of future missions.

7 EXAMPLE FITS HEADERS

As an example, below we list the relevant keywords from an *ASCA* SIS0 RMF and ARF.

7.1 *ASCA* RMF

7.1.1 RSP_MATRIX Extension

```
XTENSION= 'BINTABLE'           / binary table extension
BITPIX  =                    8 / 8-bit bytes
NAXIS   =                    2 / 2-dimensional binary table
NAXIS1  =                   34 / width of table in bytes
NAXIS2  =                 1180 / number of rows in table
PCOUNT  =              1031160 / Number of bytes acumulated in heap
GCOUNT  =                    1 / one data group (required keyword)
TFIELDS =                    6 / number of fields in each row
TTYPE1  = 'ENERG_LO'           / label for field   1
TFORM1  = 'E       '           / data format of field: 4-byte REAL
TUNIT1  = 'keV     '           / physical unit of field
TTYPE2  = 'ENERG_HI'           / label for field   2
TFORM2  = 'E       '           / data format of field: 4-byte REAL
TUNIT2  = 'keV     '           / physical unit of field
TTYPE3  = 'N_GRP   '           / label for field   3
TFORM3  = 'I       '           / data format of field: 2-byte INTEGER
```

```
TTYPE4  = 'F_CHAN  '           / label for field   4
TFORM4  = 'PI(2)   '           / data format of field: variable length array
TTYPE5  = 'N_CHAN  '           / label for field   5
TFORM5  = 'PI(2)   '           / data format of field: variable length array
TTYPE6  = 'MATRIX  '           / label for field   6
TFORM6  = 'PE(418) '           / data format of field: variable length array
EXTNAME = 'MATRIX  '           / name of this binary table extension
TLMIN4  =                    0 / First legal channel number
TLMAX4  =                  511 / Highest legal channel number
TELESCOP= 'ASCA    '           / mission/satellite name
INSTRUME= 'SIS0    '           / instrument/detector
FILTER  = 'NONE    '           / filter information
CHANTYPE= 'PI      '           / Type of channels (PHA, PI etc)
DETCHANS=                  512 / Total number of detector PHA channels
LO_THRES=             1.00E-07 / Lower probability density threshold for matrix
HDUCLASS= 'OGIP              ' / Keyword information for Caltools Software.
HDUCLAS1= 'RESPONSE          ' / Keyword information for Caltools Software.
HDUCLAS2= 'RSP_MATRIX        ' / Keyword information for Caltools Software.
HDUVERS = '1.3.0             ' / Keyword information for Caltools Software.
HDUCLAS3= 'DETECTOR          ' / Keyword information for Caltools Software.
CCNM0001= 'MATRIX            ' / Keyword information for Caltools Software.
CCLS0001= 'CPF               ' / Keyword information for Caltools Software.
CDTP0001= 'DATA              ' / Keyword information for Caltools Software.
CVSD0001= '1993-02-20        ' / Keyword information for Caltools Software.
CVST0001= '11/11/11          ' / Keyword information for Caltools Software.
CDES0001= 'SISRMGv1.10:1180x512 S0C0 G"0234"  V100 P40 E1.6                   '
CBD10001= 'CHAN(0- 511)      ' / Keyword information for Caltools Software.
CBD20001= 'ENER(0.2-12.0)keV'  / Keyword information for Caltools Software.
CBD30001= 'GRADE("0234" )         ' / Keyword information for Caltools Software.
CBD40001= 'RAWX(-95- 325)     ' / Keyword information for Caltools Software.
CBD50001= 'RAWY( 22- 444)     ' / Keyword information for Caltools Software.
RMFVERSN= '1992a   '           / Obsolete
HDUVERS1= '1.0.0             ' / Obsolete
HDUVERS2= '1.2.0             ' / Obsolete
END
```

7.1.2 EBOUNDS Extension

```
XTENSION= 'BINTABLE'           / binary table extension
BITPIX  =                    8 / 8-bit bytes
NAXIS   =                    2 / 2-dimensional binary table
NAXIS1  =                   10 / width of table in bytes
NAXIS2  =                  512 / number of rows in table
PCOUNT  =                    0 / size of special data area
```

```
GCOUNT  =                    1 / one data group (required keyword)
TFIELDS =                    3 / number of fields in each row
TTYPE1  = 'CHANNEL '           / label for field   1
TFORM1  = 'I       '           / data format of field: 2-byte INTEGER
TUNIT1  = 'channel '           / physical unit of field
TTYPE2  = 'E_MIN   '           / label for field   2
TFORM2  = 'E       '           / data format of field: 4-byte REAL
TUNIT2  = 'keV     '           / physical unit of field
TTYPE3  = 'E_MAX   '           / label for field   3
TFORM3  = 'E       '           / data format of field: 4-byte REAL
TUNIT3  = 'keV     '           / physical unit of field
EXTNAME = 'EBOUNDS '           / name of this binary table extension
TLMIN1  =                    0 / First legal channel number
TLMAX1  =                  511 / Highest legal channel number
TELESCOP= 'ASCA    '           / mission/satellite name
INSTRUME= 'SIS0    '           / instrument/detector
FILTER  = 'NONE    '           / filter information
CHANTYPE= 'PI      '           / Type of channels (PHA, PI etc)
DETCHANS=                  512 / Total number of detector PHA channels
SMOOTHED=                    0 / 0 = raw, 1-12 = smooth, -1 = ep-lin, -2 = mean-
HDUCLASS= 'OGIP              ' / Keyword information for Caltools Software.
HDUCLAS1= 'RESPONSE          ' / Keyword information for Caltools Software.
HDUCLAS2= 'EBOUNDS           ' / Keyword information for Caltools Software.
HDUVERS = '1.2.0             ' / Keyword information for Caltools Software.
CCNM0001= 'EBOUNDS           ' / Keyword information for Caltools Software.
CCLS0001= 'CPF               ' / Keyword information for Caltools Software.
CDTP0001= 'DATA              ' / Keyword information for Caltools Software.
CVSD0001= '1993-02-20        ' / Keyword information for Caltools Software.
CVST0001= '11/11/11          ' / Keyword information for Caltools Software.
CDES0001= 'SISRMGv1.10:1180x512 S0C0 G"0234"  V100 P40 E1.6                 '
CBD10001= 'CHAN(0- 511)      ' / Keyword information for Caltools Software.
CBD20001= 'ENER(0.2-12.0)keV'  / Keyword information for Caltools Software.
CBD30001= 'GRADE("0234" )        ' / Keyword information for Caltools Software.
CBD40001= 'RAWX(-95- 325)     ' / Keyword information for Caltools Software.
CBD50001= 'RAWY( 22- 444)     ' / Keyword information for Caltools Software.
RMFVERSN= '1992a   '             / Obsolete
HDUVERS1= '1.0.0             ' / Obsolete
HDUVERS2= '1.2.0             ' / Obsolete
END
```

7.2 *ASCA* ARF

7.2.1 SPECRESP Extension

```
XTENSION= 'BINTABLE'           / binary table extension
BITPIX  =                    8 / 8-bit bytes
NAXIS   =                    2 / 2-dimensional binary table
NAXIS1  =                   12 / width of table in bytes
NAXIS2  =                 1180 / number of rows in table
PCOUNT  =                    0 / size of special data area
GCOUNT  =                    1 / one data group (required keyword)
TFIELDS =                    3 / number of fields in each row
TTYPE1  = 'ENERG_LO'           / label for field   1
TFORM1  = '1E      '           / data format of field: 4-byte REAL
TUNIT1  = 'keV     '           / physical unit of field
TTYPE2  = 'ENERG_HI'           / label for field   2
TFORM2  = '1E      '           / data format of field: 4-byte REAL
TUNIT2  = 'keV     '           / physical unit of field
TTYPE3  = 'SPECRESP'           / label for field   3
TFORM3  = '1E      '           / data format of field: 4-byte REAL
TUNIT3  = 'cm**2   '           / physical unit of field
EXTNAME = 'SPECRESP'           / name of this binary table extension
TELESCOP= 'ASCA    '           / Telescope (mission) name
INSTRUME= 'SIS0    '           / Instrument name
FILTER  = 'NONE    '           / Instrument filter
HDUCLASS= 'OGIP    '           / Organisation devising file format
HDUCLAS1= 'RESPONSE'           / File relates to response of instrument
HDUCLAS2= 'SPECRESP'           / effective area data is stored
HDUVERS = '1.1.0   '           / Version of file format
RESPFILE= 'test.rmf'           / RMF file used to get the energies
WAOAA   =          7.68984E+00 / WMAP-wgtd avg off-axis ang
HISTORY   ARF created by ascaarf v3.00
HISTORY   from test.sp
HISTORY   using test.rmf
HISTORY   with extended source algorithm
HISTORY   XRT effec area from /FTP/caldb/data/asca/xrt/bcf/xrt_ea_v2_0.fits
HISTORY       PSF from        /FTP/caldb/data/asca/xrt/bcf/xrt_psf_v2_0.fits
HISTORY    Input WMAP array has size   28 by   28 bins
HISTORY                  expanded to   28 by   28 bins
HISTORY    First WMAP bin is at detector pixel  336  648
HISTORY      8 detector pixels per WMAP bin
HISTORY    WMAP bin size is   0.21600 mm
HISTORY                       0.21216 arcmin
HISTORY    Selected region size is   1.8701     arcmin^2
HISTORY    Optical axis is detector pixel 662.72 559.02
```

```
HISTORY    1180 energies from RMF file
HISTORY    Effective area fudge applied
HISTORY    Arf filter applied
ARFVERSN= '1992a   '           / Obsolete
HDUVERS1= '1.0.0   '           / Obsolete
HDUVERS2= '1.1.0   '           / Obsolete
END
```

ACKNOWLEDGMENTS

We thank the numerous people, both inside and outside the OGIP, who have contributed ideas and suggestions. In particular we thank Alan Smale.

REFERENCES

Information regarding on-line versions of any of the following references with an OGIP Memo number (*ie* documents starting `OGIP/..` or `CAL/..`) can most easily be found via the WorldWide Web by following the links from the URL:
https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/caldb_doc.html

Most OGIP Calibration Memos of general community interest will eventually appear as articles in *Legacy*, but are also available on request from The Office of Guest Investigator Programs, Code 660.2, NASA/GSFC, Greenbelt, MD 20771, USA.

Arnaud, K.A., George, I.M. & Tennant, A., 1992.
(OGIP/92-007)
Cotton, W.D., Tody, D. & Pence, W.D. 1995. *Astron. Astrophys. Suppl.*, **113**,159.
George, I.M., 1992. *Legacy*, **1**, 56.
(CAL/GEN/91-001)
George, I.M., Pence, W. & Zellar, R. 1992.
(CAL/GEN/92-008)
George, I.M., Zellar, R. & Pence, W., 1992.
(CAL/GEN/92-011)
George, I.M., Arnaud, K.A. & Ruamsuwan, L., 1992.
(CAL/SW/92-004)
Griesen, E.W. & Harten, R.H., 1981. *Astron. Astrophys. Suppl.*, **44**, 371.
Grosbol, P., Harten, R.H., Greisen, E.W. & Wells, D.C., 1988. *Astron. Astrophys. Suppl.*, **73**,365.
Pence, W., 1992. *Legacy*, **1**, 14..
Wells, D.C., Griesen, E.W. & Harten, R.H., 1981, *Astron. Astrophys. Suppl.*, **44**, 363.

USEFUL LINKS

- CAL/GEN/92-002a, the addendum to this memo
- The OGIP Calibration Database Home Page
- The HEASARC Home Page
- The OGIP FITS Working Group Home Page