OGIP Calibration Memo CAL/GEN/92-002

The Calibration Requirements for Spectral Analysis

(Definition of RMF and ARF file formats)

Ian M. George, Keith A. Arnaud, Bill Pence Laddawan Ruamsuwan & Michael F. Corcoran
Code 662,
NASA/GSFC,
Greenbelt, MD20771

Version: 1998 Dec 19

SUMMARY

The approach and calibration file formats adopted by the HEASARC for spectral analysis of X-ray PHA datasets are outlined and discussed.
Intended audience: primarily HEASARC programmers, hardware teams & authors of spectral analysis s/w.
Note: There is an addendum to this memo, CAL/GEN/92-002a which should also be referred to.

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

Contents

1 INTRODUCTION

All calibration files within the High Energy Astrophysics Science Archive Research Center (HEASARC) will make use of the Flexible Image Transport System (FITS; *eg* see Wells *et al.* 1981, Griesen & Harten 1981), including all the recent enhancements of the original FITS formats. Specifically, wide use will be made of 'extensions' (Grosbol *et al* 1988), 'ASCII tables' (Harten *et al* 1988), and 'binary tables' (Cotton, Tody & Pence 1995).

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.

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 format adopted by the HEASARC for the calibration files required during the spectral analysis of (PHA) 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 the 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 only a discrete number of channels. The continuous 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.

2.2 Detector Sensitivity, or Ancillary Response

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 $A(J)$. The $A(J)$ matrix is often called the "Ancillary Response Matrix", while the file which contains this matrix is usually called the Ancillary (sometimes, Auxiliary) Response File, or ARF for short. The Ancillary Response Matrix gives the "effective area" of the detector system, and usually includes such components as mirror vignetting, filters, etc.

2.3 Implementation

In the general case, spectral analysis of a PHA file (*eg* using XSPEC or equivalent Data Analysis Package) requires access to the following **Calibration Product Files**:

1. **A DETECTOR REDISTRIBUTION MATRIX FILE (RMF)**
   - 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*)
- consisting of a compressed 2-d (energy vs PHA channel) FITS extension (Section 3).
- and a second extension explicitly listing the nominal energy range of each PHA channel.

2. **AN ANCILLARY RESPONSE FILE (ARF)**
   - containing 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),
     - Filter Transmission (if any)
     - Detector Window Transmission
     - Detector Efficiency
     - any additional energy dependent effects (*eg* correction factors for the *p.s.f.*)
   - in a single 1-d (energy) FITS extension (Section 4).

However, though strongly discouraged, under certain circumstances, all the above calibration information may instead be incorporated into a single file (see Section 6).

It should be noted that the PHA channels used in all the i/p files to the spectral analysis package — *ie* within the RMF, ARF & PHA files — in all cases refer to **raw detector channels**. In the case of RMFs, the number of raw channels is given explicitly by the DETCHANS keyword within the MATRIX extension (see below). Any desired rebinning of the raw PHA data can be specified by the GROUPING flag within the PHA file (OGIP/92-007, Arnaud, George & Tennant 1992) The rebinning of the data is then performed within the spectral analysis package, along with appropriate rebinning of the calibration data supplied by the RMF & ARF.

2.4 Rationale

For many instruments, the Detector Gain & Energy Resolution do not vary significantly with detector coordinates or time (in particular not within an 'observation'). In such cases, a single RMF can be constructed for a given observation and used in conjunction with the spectral analysis of **ALL** the sources in the field-of-view. Each of the individual PHA files associated with each source will thus have its own customized ARF. The spectral analysis package will then read in the PHA/ARF pairs along with the common RMF. (Situations in which the use of a common RMF is not possible are discussed in Section 5.)

Furthermore, the isolation of those components which are not concerned with the redistribution process, and hence are a simple function of energy for a given PHA file (*ie* given time, Detector position, mode *etc*), into the ARF allows these components to be listed individually within the

ARF if desired (see Section 4). It should be noted however that the product (*Prod* array below) of the components to the ARF will always be listed.

2.5 Reminders

'PI' channels (*ie* from a redistribution of the raw PHA channel data within a PHA file onto a scale appropriate for some 'standard' Detector gain setting *etc*) should wherever possible **NOT**[1] be used: the inclusion of the Detector gain within the RMF provides the necessary PHA channel $\rightarrow$ energy conversion.

In the case of (significant) gain changes during a given 'observation' (*eg* as often the case for the *Einstein Observatory* IPC) separate PHA & RMFs should be constructed prior to spectral analysis (and the PHAs analysed in conjunction with their respective ARFs & RMFs either alone or simultaneously).

An RMF contains only information applicable to the redistribution process. The Effective area, Filter (if appropriate) & Window Transmissions are folded into the ARF. Any additional effects such as obscuration by the Window Support Structure, absorption due to deposits (*eg* water ice) upon the the Window surface *etc* will also be folded into the ARF. Again this is in order to minimise the number of RMFs required for a given observation dataset (hopefully to one in many instances). Hence this provides a reduction in the requirements for disk-space, and facilitates any investigation Users may wish to perform into the effect on their spectral analysis of varying the contribution any such components.

3 THE HEASARC STANDARD RMF FORMAT

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

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

both employing the BINTABLE FITS format.

[1]A primary goal of the HEASARC is to standardize the format of similar files from different instruments. Besides strictly being the incorrect method in the general case, the mapping of the PHA data from the observed to some 'standard' channel grid cannot be performed correctly for low resolution detectors.
However, in the case of certain past and present missions, unfortunately the use of PI channels may be unavoidable since the necessary information (*ie* to construct & use BCFs) has been lost or is unavailable from the h/w team.
However, 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'` keyword within the PHA file (see OGIP/92-007, Arnaud, George & Tennant 1992)

3.1 The RMF Redistribution Extension

In order to minimize disk-space requirements, the RMF Redistribution Matrix will be in a compressed format in which all matrix elements below a given threshold (specified by the `LO_THRES` keyword below) are not stored. In fact the format is very similar to that used currently by the XSPEC `*.rsp` SF 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).

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` - lower 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 multipled 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).

- `CCLS0001` (= `'CPF'`) - the OGIP-class of this calibration file.
- `CCNM0001` (= `'MATRIX'`) - the (CIF) codename for this type of calibration dataset.
- `CDTP0001` (= `'DATA'`) - the OGIP code for the form of the contents of the file ('real' data, a taskname and associated parameter inputs etc)
- `CVSD0001` - the UTC date (in yyyy-mm-dd format) when this calibration data should first be used
- `CVST0001` - the UTC time (in hh:mm:ss format) on the day `CVSD0001` when this calibration data should first be used.
- `CDES0001` - 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 for the energy bin (see below).
   The FITS column name is **N_GRP**.
   (unitless).

4. F_{chan}, a (fixed- or variable-length) INTEGER vector (array, each element within which is 2-byte) for each row containing 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) INTEGER vector (array, each element within which is 2-byte) for each row containing 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 vector (array, each element within which is 4-byte) containing all the response values for each 'channel subset' for the energy bin.
   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

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 integer array	2-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

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 is of course arbitrary, however that used here is strongly recommended.
- Values of both E_{low} & E_{high} are given in each row (j) for clarity, and to ease access & use. The order should be sequential, and it is suggested starting from the lowest value. In no case should there be any overlap between consecutive energy bins, *ie* $E_{low}(j) \geq E_{high}(j-1)$, (although note that in all currently known cases, $E_{low}(j) = E_{high}(j-1)$).
- The concept of 'channel subsets' is included to minimize the RMF storage requirements for instruments for which the 2-d matrix 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 raw channels for which the matrix elements are above the threshold specified by the value of the `LO_THRES` keyword. Thus, using the above notation, then a given row the Mat array contains the elements appropriate to

channels	$F_{chan}(1)$	$\rightarrow$	$(F_{chan}(1) + N_{chans}(1) - 1)$
followed by	$F_{chan}(2)$	$\rightarrow$	$(F_{chan}(2) + N_{chans}(2) - 1)$
followed by	$F_{chan}(3)$	$\rightarrow$	$(F_{chan}(3) + N_{chans}(3) - 1)$
...	...		...
etc	*etc*		*etc*
followed by	$F_{chan}(N_{grp})$	$\rightarrow$	$(F_{chan}(N_{grp}) + N_{chans}(N_{grp}) - 1)$

- In the general case, columns 4, 5 & 6 (containing F_{chan}, N_{chan} and Mat arrays) will be FITS variable-length arrays (Cotton *et al* 1995). Thus the number of elements within each vector varies between rows (clearly the number of elements within the N_{chan} array on a given row, for example, is equal to the value of N_{grp} (column 3) for that row).
  However, if the maximum number of elements within such an array (over all rows) is less than or equal to three (*i.e.* if $N_{grp} \leq 3$ for all rows), it is more efficient in terms of both disk storage requirements[2] and speed of access[3] 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. Thus, with specific regard to RMFs, the following guidelines should be adhered to:

[2]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.

[3]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.

- 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 for most instruments from which there is data currently 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 Mat 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.

It should be emphasized, however, that the choice of fixed- or variable-length arrays should make no difference to downstream s/w (*eg* XSPEC)

as the two formats look identical when using FITSIO interface (Pence 1992).

- If a column contains a constant value in every row, then the column is deleted from the table and transformed into a keyword value. The downstream s/w should then first look for a keyword value; if the keyword is not found, then the s/w will 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. Effects due to detector efficiencies $< 100\%$, absorption by mirrors, filters & the detector window *etc* are included within the ARF. However, due to the finite probability of a photon being registered as either below or above the PHA discriminator thresholds, and when the value of `LO_THRES` > 0.0, the sum of the values within the Mat array for a given row may be less than unity.

3.2 The RMF EBOUNDS Extension

This 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** the same as those given in the `ENERG_LO` & `ENERG_HI` columns of the `MATRIX` extension above.) This information is required by spectral analysis packages when (say) the results of spectral analysis (PHA data and best-fitting model) are required to be displayed as a function of photon energy, rather than detector channel. Since the number of raw channels in the matrix of many detectors is large, clearly a great saving in processing time during such spectral analysis tasks is offered if this information is provided explicitly by the RMF.

The format described here is relatively straightforward, being a simple 1-dimensional list (as a function of raw detector PHA channel) listing the appropriate energy boundaries.

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 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 organisation of this extension

3.2.3 Points to Note & Conventions

- The ordering of the columns is if course arbitrary, however that used here is strongly recommended.
- E_{min} and E_{max} should never be confused with E_{low} and E_{high} within the RMF extension.
- E_{min} and E_{max} should be used with extreme care. Inexperienced users/programmers are reminded that the pulse-height analyser vastly oversamples the true spectral response in most X-ray detectors. Thus there is **no** guarantee that an incident X-ray with an energy between E_{min} and E_{max} will be detected in the corresponding channel (hence the requirement of the RMF in the first place).

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).

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

$\boxed{\textbf{Extension}}$ **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 integer	4-byte real	4-byte real
total number of elements per row		
1	1	1
column name		
CHANNEL	E_MIN	E_MAX

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 `C***` keywords listed in Section 3.1.1 are also mandatory, but in this case with `CCNM0001 = 'SPECRESP'`. Furthermore however, if (in addition to the *Prod* array) the ARF `SPECRESP` extension also lists each of individual contributing components (see below), and these components are too be be listed in the 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, Zellar & Pence 1992).

4.1.2 Data Format

The general http://heasarc.gsfc.nasa.gov/ 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 **NOT** 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 number of rows hence equals the number of energy bins **and must directly correspond to those in the corresponding RMF**. In all cases the following columns are included (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 Points to Note & Conventions

- The ordering of the columns is of course arbitrary, however that used here is strongly recommended.
- Values of both E_{low} & E_{high} are given in each row (j) for clarity, and to ease access & use. The order should be sequential, and it is suggested starting from the lowest value, and should be identical to those in the corresponding RMF. In no case should there be any overlap between consecutive energy bins, *ie* $E_{low}(j) \geq E_{high}(j-1)$, (although note that in all currently known cases, $E_{low}(j) = E_{high}(j-1)$).
- The dimension of the data within the $Prod$ column will be length^2 (due to the inclusion of the effective area).

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)

5 USAGE: TYPICAL SCENARIOS

For non-imaging devices (with a time-stable Detector Gain etc) a User extracts a PHA file, runs a response generator and constructs a RMF & ARF. These are read into their favourite Data Analysis Package (*eg* XSPEC) along with the PHA file, and spectral analysis attempted. Should the User then wish to investigate the effect of alternate calibrations, a response generator can be run again/spawned. The User then specifies the desired calibration, and a new ARF written. Spectral analysis could continue using the original PHA file & RMF in conjunction with the new ARF. In some cases a single RMF may be applicable to several contiguous observational 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 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 will obviously be able to 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 time-stable. 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

It should be stressed that the spectral response matrix of instruments for which either:

- it is necessary to assume a PHA channel → PI channel conversion has been performed on the PHA data[4]; and/or
- the components of the matrix are unavailable in a suitable format, thus requiring the *Prod* array to be folded into the RMF,

can (and should) be written adopting the RMF format described above.

[4]This is denoted by the `CHANTYPE = 'PI'` keyword within the PHA file (see OGIP/92-007, Arnaud, George & Tennant 1992).
Examples include the current versions of the *ROSAT* PSPC and *Einstein Observatory* IPC datasets within the HEASARC archive

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 and Mike Corcoran.

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:
`http://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.