XRISM Data Processing Pipeline and Software

Michael Loewenstein for the XRISM Science Data Center (SDC)
Matt Holland (PDL), Tahir Yaqoob, Eric Miller, Bob Hill, Efrain Perez-Solis, Morgan Waddy, Mark Mekosh, Joseph Fox
Trisha Doyle, Patty Hall, Efrem Braun (alumni)
What is the SDC?

- Handles all processing of record from observation scheduling to delivery of pipeline output to data archives.
- Provides proven, standards-based community software tools needed for data processing (calibration, screening), data product extraction, and data product analysis.
- Manages XRISM ground system software and data systems involved in maintaining the XRISM Calibration Database (CalDB) derived from deliveries from the Xtend, Resolve (and XMA) instrument teams.
- Supports the XRISM Guest Observer Facility (GOF) in user-support activities.
This talk

Will cover (to different levels of detail)
• Data (archive) organization, naming conventions
• XRISM processing pipeline
• XRISM processing ftools (including reprocessing scripts) and CalDB
• Some details on Resolve energy assignment
• XRISM post-processing and data analysis ftools
• XRISM software status and plans

Will not cover
• Coordinate system definitions
• Time assignment
• Extended source details, simulations (F. Mernier)
• Bright source details (E. Hodges-Kluck)

Will focus on Resolve
How the Data Flow

Telemetry from XRISM arrives at ISAS

PPL converts the telemetry data into FFF

Data are transferred to the PL virtual machines (~bi-weekly)

Data are verified and prepared for pipeline processing

Data are processed

Data (encrypted) are transferred to the archives

The PL operates in a (mostly) automated manner, with 17 TVac and >50 Commissioning Phase OBSIDs processed to date – typically w/ <1 day turnaround.

For XRISM:
- Improved Resilience
- Improved Parallelization
- Implementation of metadata handling and archiving

PL==pipeline; PPL==pre-pipeline; FFF == First FITS Files
• All XRISM data files start with ‘xa{OBSID}’ where OBSID encapsulates the observation category (e.g., nominal GO observations, calibration observations...), observing cycle, target number for multi-target proposals, pointing number for multiple pointing proposals.

• Example: 3/00/047/03/0 (PV phase observation of XXX)
  PV phase/pre-AO/counter for 300/pointing-3/...

• Event and housekeeping files have the following structure:
  ‘xa{OBSID}{instrument}_{pointing}{#}’ where # is reserved in case files were split (not expected)

<table>
<thead>
<tr>
<th>instrument</th>
<th>gen (General)</th>
<th>rsl (Resolve)</th>
<th>xtd (Xtend)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pointing</td>
<td>p=pointing</td>
<td>s=slew</td>
<td>a-all</td>
</tr>
</tbody>
</table>

Additional specification for events files:
• Xtend: Xtend: filenames encode information on CCD operating mode
• Resolve: filenames encode whether the file contains antico (ac) or pixel array (px) data, as well as Filter information.
## XRISM Data Conventions

<table>
<thead>
<tr>
<th>infix</th>
<th>FILTER</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>px0000</td>
<td>UNDEF</td>
<td>Undefined FW position – not used for science</td>
</tr>
<tr>
<td>px1000</td>
<td>OPEN</td>
<td>Science events</td>
</tr>
<tr>
<td>px2000</td>
<td>Polyimide filter</td>
<td>Not used for science observations</td>
</tr>
<tr>
<td>px3000</td>
<td>ND (neutral density)</td>
<td>Science events with ~gray 25% transmission reduction</td>
</tr>
<tr>
<td>px4000</td>
<td>Be (beryllium)</td>
<td>Science events with soft flux reduction (GVO only)</td>
</tr>
<tr>
<td>px5000</td>
<td>Fe55 (FW cal source)</td>
<td>Calibration events</td>
</tr>
</tbody>
</table>

### Graphical Representation

![Graph showing transmission vs. energy for different filters](image)
Archive Directory structure

For the U.S. (HEASARC) archives, observation data is here: https://heasarc.gsfc.nasa.gov/FTP/xrism/data/obs/{AO}/obsid
XRISM CalDB

(Calibration Database)

Pre-fight CalDB
• A high-fidelity, theoretically-informed, self-consistent realization of all XRISM subsystems (XMA, Resolve, Xtend) based on ground measurements.
• Used in combination with the XRISM ftools to calibrate the XRISM event files and construct response files.
**What’s in the CalDB?**

<table>
<thead>
<tr>
<th>GEN (mission-independent)</th>
<th>Atomic data (2), rigidity (1), leap seconds (1), solar geophysical data (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XRISM (all instruments)</td>
<td>Timing files (3), XMA optical constants (1), XMA vignetting (1), atomic line data (1), filter file configuration (1), event selection expressions (1), SAA bounds (1), col. defs. (1)</td>
</tr>
<tr>
<td>RESOLVE</td>
<td>Teldef (1), QE (1), blocking filter (1), contamination (1), bad pixels (1), instrument map (1), pixel map (1), rmf params (1), FW files (5), gate valve (1), gain (1), antico (1), time coefficients (1), mxsparam (1), secondary pulses (1), pulse templates (1), XMA (5), clipping (1)</td>
</tr>
<tr>
<td>XTEND</td>
<td>Teldef (1), QE (1), contamination (1), instrument map (1), bad pixels (1), mask (1), rmf params (1), video temp (1), charge trail (1), CTI (1), split thresh (1), pixel pattern (1), gain (1), XMA (5)</td>
</tr>
</tbody>
</table>
Pipeline (PL) Processing
The Pipeline processing script proceeds as follows:

- Check input files for validity
- Determine nominal pointing - aberattitude, attconvert, aspect
- Calculate Optical Axis (sky) – coodpnt
- Calculate quartz clock temperature trend (backup timing) – xatrendtemp
- Collect necessary HK, orbit, and attitude information into “make-filter” (mkf) and “extended HK” (ehk) files (for screening, coordinate assignment) – xamkehk
- Calibrate, cleans for each instrument; extracts preview products
Resolve PL Processing

Detector channels (fixed)

<table>
<thead>
<tr>
<th>30</th>
<th>32</th>
<th>34</th>
<th>26</th>
<th>24</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>31</td>
<td>33</td>
<td>25</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>35</td>
<td>18</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>17</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
<td>15</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>8</td>
<td>16</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

(look down)
## Resolve Event File Columns

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITYPE</td>
<td>Resolution Grade; ITYPE=0: HP, ITYPE=1: MP, ITYPE=2: MS, ITYPE 3: LP, ITYPE 4: LS, ITYPE 5: BL, ITYPE 6: EL, ITYPE 7: Rj</td>
<td>0-7</td>
</tr>
<tr>
<td>PIXEL</td>
<td>Pixel number (pixel 12 is the calibration pixel)</td>
<td>0-35</td>
</tr>
<tr>
<td>RISE_TIME*</td>
<td>Measured time from baseline to peak for Resolve pulse</td>
<td>0-255</td>
</tr>
<tr>
<td>DERIV_MAX*</td>
<td>Maximum value of pulse derivative</td>
<td>32768-32767</td>
</tr>
<tr>
<td>TICK_SHIFT*</td>
<td>Shift of pulse peak</td>
<td>-8-7</td>
</tr>
<tr>
<td>QUICK_DOUBLE*</td>
<td>Double-pulse flag</td>
<td>b0-b1</td>
</tr>
<tr>
<td>SLOPE_DIFFER*</td>
<td>Pulse shape validity</td>
<td>b0-b1</td>
</tr>
<tr>
<td>PI</td>
<td>Linearized Energy Channel</td>
<td>0-60000</td>
</tr>
<tr>
<td>STATUS</td>
<td>16 bit Event Flag (14 in use)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STATUS[1]: in (0) or out (1) of all-pixel GTI file</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[2]: in (0) or out (1) of individual-pixel GTI</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[3]: coincident with antico events (1)</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[4]: coincident with other event within a temporal proximity</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[5]: coincident with pixel 12 event</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[6]: coincident with pixel 12, and recoil energy test satisfied</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[7]: coincident with event in wiring proximity (electrical crosstalk) – short timescale</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[8]: largest PHA in electrical crosstalk group – short timescale</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[9]: coincident with MXS, direct mode</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[10]: coincident with MXS afterglow, direct mode</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[11]: coincidence with MXS, indirect mode</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[12]: coincident with MXS afterglow, indirect mode</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[13]: coincident with event in wiring proximity (electrical crosstalk) – long timescale</td>
<td>0-1</td>
</tr>
<tr>
<td></td>
<td>STATUS[14]: largest PHA in electrical crosstalk group – long timescale</td>
<td>0-1</td>
</tr>
</tbody>
</table>

*pulse shape diagnostics, may be used to filter problematic events
**Resolve Grade (ITYPE)**

<table>
<thead>
<tr>
<th>$t_n \leq \delta t_2$</th>
<th>$\delta t_2 &lt; t_p \leq \delta t_1$</th>
<th>$\delta t_1 &lt; t_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_S$</td>
<td>$L_S$</td>
<td>$L_P$</td>
</tr>
<tr>
<td>$\delta t_2 &lt; t_n \leq \delta t_1$</td>
<td>$L_S$</td>
<td>$M_P$</td>
</tr>
<tr>
<td>$\delta t_1 &lt; t_n$</td>
<td>$L_S$</td>
<td>$M_P$</td>
</tr>
</tbody>
</table>

![Graph showing resolution grades and time intervals](image)

2024.01.18  
2nd XRISM Data Analysis Workshop
Resolve PL Prelims

• Calculate ADR temperature fluctuations (for ADR GTI) – rslctsfluct (new for XRISM)
  - Update to rsladrgti to account for gain settling planned

• Calculate ADR GTI (screening) – rsladrgti (new for XRISM)

• Calculate MXS GTI (screening, energy assignment) – rslmxsgti
  - Major updates to MXS tasks implemented to account for afterglow and all offsets, and their dependencies on pulse parameters and the state of the LED

• Invert LOST GTI file (expo map) – gtiinvert
• Assign antico energy (PI) – rslanticolpi (also light curve; rslanticolc)
Resolve PL ‘Pre-run’

- Assign higher level coordinates (up to RA-DEC) – coordevt
- Set energy-independent data quality STATUS flags (screening) – rslflagpix
  - Handling of Resolve ‘contingency mode’
  - Tighter coincidence window for antico
- Assign pulse clipping flag (screening) – rslplsclip (new for XRISM) [affects >~20 keV events only!]

- Associate groups of secondary with primary events (energy assignment) – rslsecid
Resolve PL Energy assignment

- Merge event files and screen for drift correction
  - Improved pre-screening **planned**
- Compute the energy scale drift correction (energy assignment) – rslgain
  - Parameter optimization in **progress**
  - MXS rate correction **planned**
- Assign initial energy (PI) – rslpha2pi

- Gain curve: Energy as a function of PHA (pulse height amplitude) and ‘temperature’
- Separately calculated for High, Mid, and Low resolution events for each pixel
Energy assignment

• Accumulate spectra from an onboard calibration source at different times over the observation exposure.
• Use knowledge of the intrinsic line energy and shape, and the measured PHA to derive the effective temperature.
<table>
<thead>
<tr>
<th>Cal method</th>
<th>Cal lines</th>
<th>Pipeline Gain History</th>
<th>Pipeline energy assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe55</td>
<td>MnKa, MnKb</td>
<td>If FW set to Fe55 (MnKa)*</td>
<td>Primary method</td>
</tr>
<tr>
<td>MXS (direct)</td>
<td>Cuka, Cukb, Crka, Crkb</td>
<td>If MXS direct on (CuKa, CrKa)*</td>
<td>If FW Fe55 not in optical path</td>
</tr>
<tr>
<td>MXS (indirect)</td>
<td>AlKa, AlKb, MgKa, MgKb</td>
<td>If MXS indirect on (AlKa, MgKa)*</td>
<td>For calibration only</td>
</tr>
<tr>
<td>Cal-pix</td>
<td>MnKa, MnKb</td>
<td>Always (MnKa)</td>
<td>If no MXS direct and no Fe55</td>
</tr>
</tbody>
</table>

*per-pixel*
Gain Histories (per pixel)
The Resolve MXS

- The modulated X-ray source (MXS) illuminates each pixel with X-rays generated using LEDs operating with a short pulse period and low duty cycle. The pulsed emission dominated by narrow features of known energy.
- There are direct (LED1/3 on the nominal/redundant side) and indirect (LED2/4 on the nominal/redundant side). The K-shell lines emission lines directly generated by the Cr/Cu anode may be used for energy assignment.
- This enables continuous pixel-by-pixel energy gain drift monitoring while avoiding contamination of the astrophysical source; although, there is an afterglow.
- Because of complications due to gate-valve shadowing and effect of the MXS rate on the gain, the FW/Fe55 is currently the calibrator of choice.

https://doi.org/10.1117/1.JATIS.4.1.011204
Methods and lines

Resolve Main Pipeline Gain Calibration Method Decision Tree

- Make per-pixel Fe55 MnKα gain history, apply.
- Make per-pixel MXS indirect mode Mg Kα and Al Kα gain history.
- Make per-pixel MXS direct mode Cr Kα and Cu Kα gain history.

Is Fe 55 FW Cal source in optical path?
- YES
  - Make per-pixel Fe55 MnKα gain history, apply.
- NO
  - Is MXS indirect mode on?
    - YES
      - Apply per-pixel MXS direct mode Cu Kα gain history.
      - Do nothing.
    - NO
      - Make gain history files for calpixel for: Mn Kα
      - Apply cal-pixel Mn Kα gain history.

Is MXS direct mode on?
- YES
  - Apply per-pixel MXS direct mode Cu Kα gain history.
- NO
  - Do nothing.

Make all possible gain history files.

- Use FW/Fe55 if in optical path
- Use MXS if direct mode is enabled, FW/Fe55 not in optical path
- Otherwise use cal-pixel.
Final Calibration, Screening

- Set energy-dependent data quality STATUS flags – rslflagpix
  - Improved handling of recoil cross-talk
  - Improved precision for electrical crosstalk (asymmetric window) and frame event flagging planned

- Re-assign groups of secondary events with primary – rslsecid
  - Improved labeling of ‘orphans’ – enables filtering of uncorrected secondaries

- Correct secondary event raw pulse heights - rslseccor
  - Algorithm (and CalDB) improved for XRISM

- Assign final energy (PI) – rslpha2pi

- Filter times for
  - HK (nominal instrument status)
  - Telemetry (not saturated)
  - Pointing accuracy, stability
  - Safe angle above Earth, sunlit Earth limb
  - Away from SAA in orbit
  - ADR (not recycling, settling down)
  - MXS (not pulsing, not in afterglow)

- Screen events for
  - Grade (ITYPE<5)
  - Data quality (based on STATUS)
  - Pulse shape (SLOPE_DIFFER, QUICK_DOUBLE)
  - Screening optimization for XRISM in progress

- Additional post-PL screening is necessary (rise-time, frame events)

![Graph of MnKa Hp fwhm=4.4 eV]
Xtend PL Processing
(in brief)
**Xtend Calibration, Screening**

- Calculate Xtend CCD mode GTI (for expo map) – xtdmodegti
- Assign higher level coordinates - coordevt
- Merge inner 3x3 and outer 5x5 pulse heights – xtdphas
- Set initial data quality STATUS flags (for screening) – xtdflagpix
- Assign grade and initial energy (PI) – xtdpi
- Update grade and PI w/ grade-dependent CTI correction – xtdpi
- ID and output flickering pixels among clean events – searchflickpix/coordevt
- Update STATUS flags w/ flickering pixel information – xtdflagpix
- Filter times
  - HK (nominal instrument status)
  - Telemetry (not saturated)
  - Pointing accuracy, stability
  - Safe angle above Earth, sunlit Earth limb
  - Away from SAA in time and orbit
- Screen for good events
  - Grade
  - Data quality (STATUS; not in a “bad” detector area, pixel, row, etc)

*ahtigen + ahscreen*
XRISM Users should reprocess data when there has been an update to pipeline software and/or CalDB since the data was archived; or, if a user wants to apply a non-standard calibration or screening criteria.

Resolve and/or Xtend data may be reprocessed using the xapipeline script that is part of the standard software library (set the instrument parameter to ALL, Xtend, or Resolve).

Resolve/Xtend data may also be reprocessed using the rslpipeline/xtdpipeline scripts.

Users may start/finish the reprocessing at the calibration, screening, or preview product generation stage. The ‘pre-calibration’ steps of mkf/ehk and ADR GTI and MXS GTI file generation may be skipped (in which case the pre-existing files may be used).

Users essentially have access to all of the parameters for all of the tools that comprise the pipeline – so the full sequence of individual tools rarely, if ever, need to be run (except by the SDC and instrument teams for verification).

The pipeline tools include a randomization seed – output from multiple, identical runs may not be precisely identical.
Post-pipeline – RMF/ARF

• Resolve RMF improvements
  - Additional pixel and grade dependencies of LSF components
  - Si fluorescence
  - Electron loss continuum response function
  - Split matrix (sparse/hi-resolution + dense/lo-resolution)
  - RMF library option planned

• Effective area and PSF tools – updates and plans
  - New task (dustyarfmod) makes an effective area correction for X-ray point sources with dust halos.
  - A new tool, xmasim, is being developed that adds new functionalities to xrtraytrace, including detector effects, ability to include objects in the optical path (such as Resolve gate valve and filters), custom image outputs, automation of calculation of statistical errors on raytracing results, and pseudo-parallelization/distributed computing capability.
  - Updates are planned for the ARF generator that will include the features described for xmasim, enabling studies and analysis for more complex scenarios.
  - PSF library planned for Resolve, enabling spectroscopy that accounts for cross-contribution of spatial regions due to the PSF spatial redistribution.
  - Prototype work has started in order to run xrtraytrace on AWS Batch.
  - Improvements have been made to the core raytracing program xrtraytrace to accommodate geometrical changes to the inner foil structure, and to implement more accurate thermal shield modeling.
• The non-X-ray background (NXB) is an irreducible background that is not due to X-ray photons (mostly high-energy charged particles) collected by the optics and focused on the detector.

• Xtend and Resolve will accumulate a database – that will be accessible online and downloadable - of NXB events during Earth occultations when XRISM is pointed toward the dark limb of the Earth.

• Xtend and Resolve NXB generators (xtdnxbgen, rslnxbgen) extract events from the NXB database, matching the spacecraft and instrument conditions and settings.

• The Resolve NXB is expected to be low in cleaned event lists (~0.0003 cts/sec/pixel for Hitomi SXS), but may have an additional MXS component.
Final Remarks

- SDC is coordinating incorporation of in-flight calibration analysis into CalDB.
- Hot fixes to the software and CalDB are in the process of development and incorporation in an improved processing pipeline and software-library to serve as a candidate for the first public release.
- When validated this will be integrated into HEASoft (early in the PV phase). All Hitomi tasks are maintained as part of a dual architecture where common code may be shared.
- Comprehensive Functional and Lightweight Field Test for automated regression-checks of validated unit-tests.
  - LFT delivered to end users as part of ftool release
  - CFT to be used for internal (HEASARC) validation
- Documentation
  - Improved help files
  - Doxygen - bridges the code and help files
    - Easy access and documentation of XRISM software
    - Full access to task code and associated codes/librarie
    - Easy to navigate code components
    - Structure follows the directory structure of HEASoft.
CalDB Status, Plans

- Pre-fight CalDB: provides a high-fidelity, self-consistent realization of the XRISM subsystems (XMA, Resolve, Xtend).
- SDC is coordinating incorporation of in-flight calibration analysis into CalDB.
- XMA
  - Status: A fully self-consistent model of the XMA that reproduces XMA ground calibration.
  - Upcoming: Simple update to Gate-valve closed transmission (account for mesh)
  - Future: Verification and possible updates from in-flight calibration
- Xtend
  - Status: A model adequate for processing in-flight data.
  - Upcoming: Teldef (Xtend aimpoint) update.
  - Future: Updates and verification from in-flight calibration (including bad pixel list)
- Resolve
  - Status: Accurate but incomplete realization of ground measurements
  - Upcoming: Gain (pixel 16)
  - Future:
    - Updates and verification from in-flight calibration
    - Time coefficients (windows for screening) from commissioning + in-flight calibration
    - RMF parameters (including extended LSF), blocking filter transmission and quantum efficiency refinements from additional analysis of ground data
- General
  - Status: Accurate of ground measurements, Hitomi-like screening.
  - Upcoming: SAA
  - Future: Updated screening criterion
Gain history, application

**Schematic for Resolve gain application and and energy scale assignment**

**START**

- **rslgain**
  - Find PHA(i) (shift) at cal. line energy E0; convert to kT(i)

- **Gain History File; kT(i), time(i)**

- **Non-calibration events**
  - **rslgain** Group cal events into spectra at time(i)

- **Resolve (pre-run) Event file:**
  - Cal. Events: PHA(k), time (k);
  - Non-cal. Events: PHA(j), time(j)

**STOP**

- **rslpha2pi** Convert event E(j) to PI(j) energy channel

- **rslpha2pi** Convert event PHA(j) to E(j) using T(j) and gain CaldB
  - Group cal events into spectra at time(i)

- **rslpha2pi** Interpolate to get event kT(j) vs. time(j)
Gain Histories (per pixel)
Graphical Overview of the Suite

Real pointing, attitude external source image, or source type, regions
**ARF generator**

Energy grid, off-axis & azimuthal angles, source type
**xmasim**

**xrtraytrace**

Resolve filters, gate valve

Detector effects

ARF (observation-specific)

Effective area, images, raytracing event files

Observation-specific “tiled” matrix of ARFs & spectra for sub-regions: fit simultaneously to solve spatial/spectral PSF mixing problem.

**Resolve PSF libraries**
(GV open/closed, FW open, ND, Be, or Al/Polyimide)