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

XRISM Telescope

Pre-pipeline processing at ISAS/JAXA

DTS/DAS data transfer to NASA/GSFC

Pipeline processing at NASA/GSFC

Grab data

Check data

Copy/archive data

Process data

Encrypt/compress data

Archive data at GSFC and transfer ISAS (via DTS)

Clean up data

Data available to scientists

The PL operates in a (mostly) automated manner, with17 TVac and >50 Commissioning Phase OBSIDs processed to date – typically w/ <1 day turnaround.

- 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

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

PL==pipeline; PPL==pre-pipeline; FFF == First FITS Files

XRISM Data Conventions

- 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)

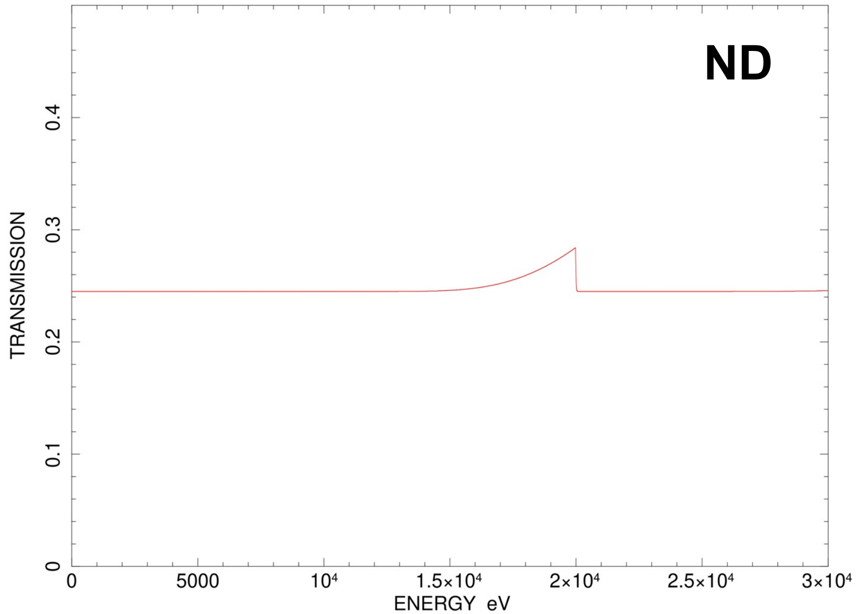
instrument	gen (General)	rsl (Resolve)	xtd (Xtend)
pointing	p=pointing	s=slew	a-all

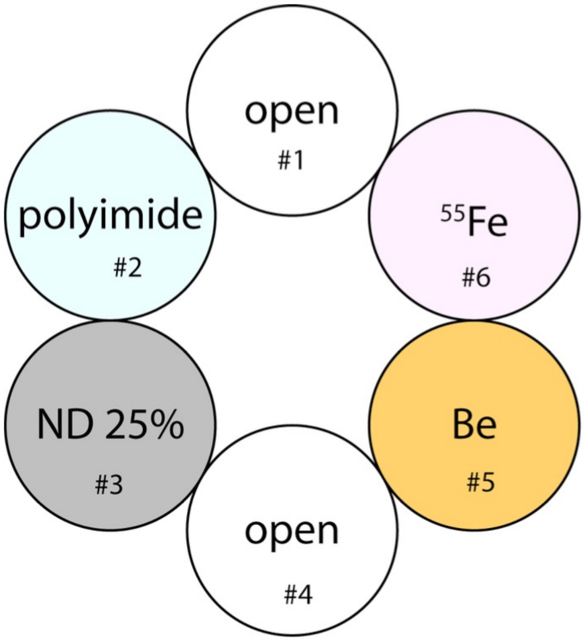
Additional specification for events files:

- 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

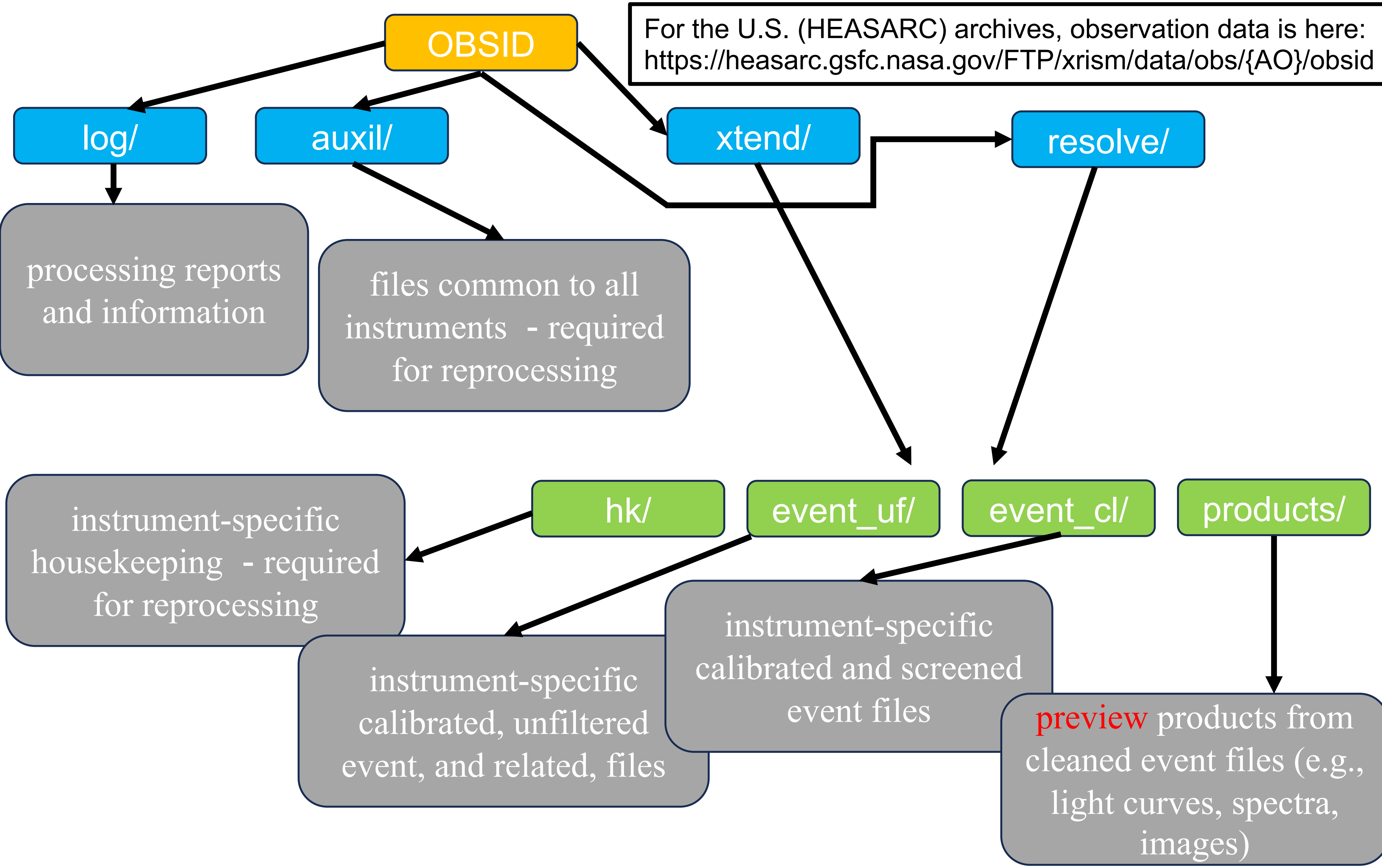
infix	FILTER	comment
px0000	UNDEF	Undefined FW position – not used for science
px1000	OPEN	Science events
px2000	Polyimide filter	Not used for science observations
px3000	ND (neutral density)	Science events with ~gray 25% transmission reduction
px4000	Be (beryllium)	Science events with soft flux reduction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





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?

XRISM X-Ray Imaging and Spectroscopy Mission

GEN (misson-independent) **5 files**
Atomic data (2), rigidity (1), leap seconds (1), solar geophysical data (1)

XRISM (54 files)

GEN (all instruments) **10 files**
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)

RESOLVE **26 files**
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)

XTEND **18 files**
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)

Pipeline (PL) Processing

Data Processing - General

The Pipeline processing script proceeds as follows:

- Check input files for validity
- Determine nominal pointing - aberattitude, attconvert, aspect
- Calculate Optical Axis (sky) – coordpnt
- 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)

30	32	34	26	24	23
29	31	33	25	22	21
27	28	35	18	20	19
1	2	0	17	10	9
3	4	7	15	13	11
5	6	8	16	14	

12

(look down)

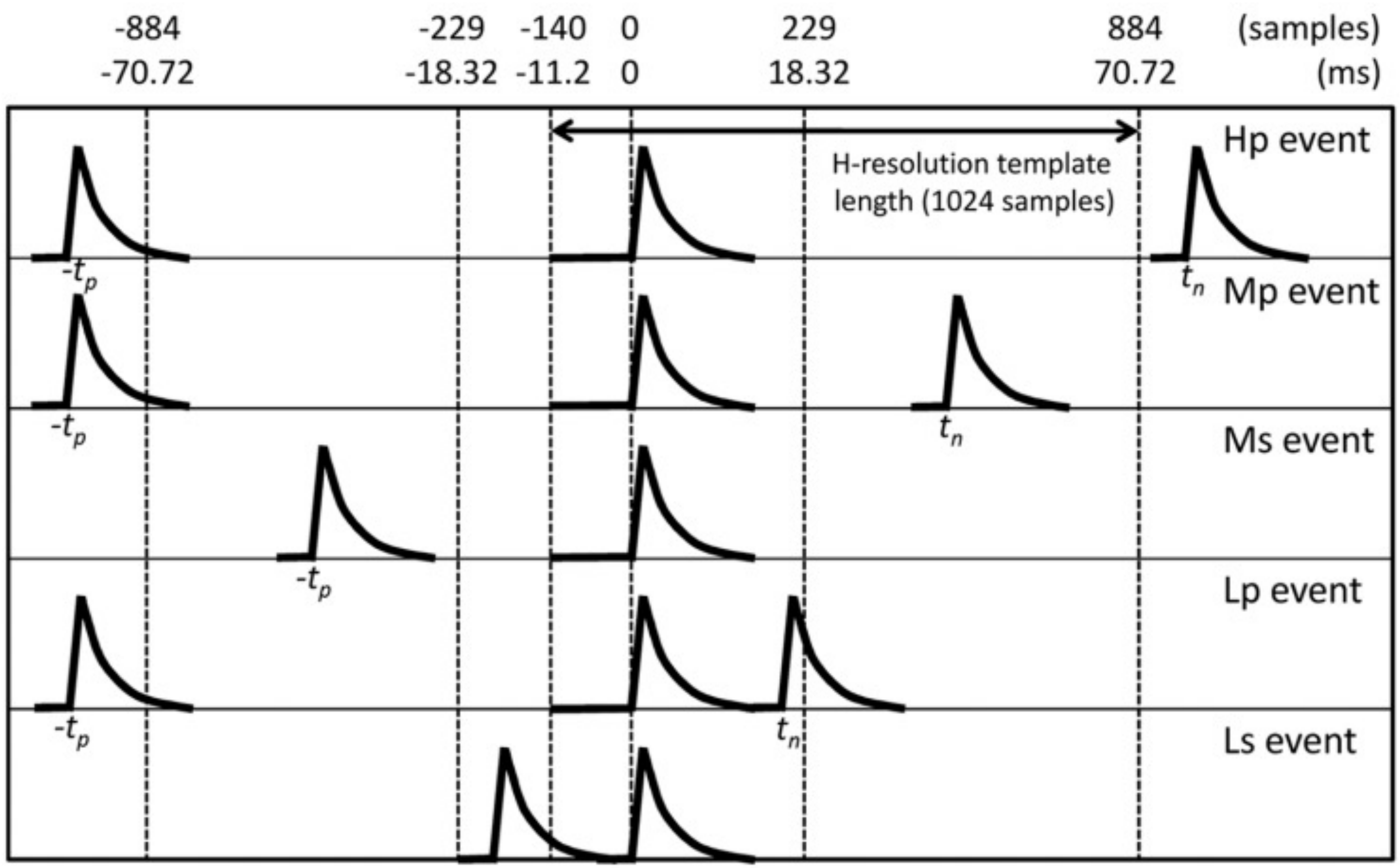
Resolve Event File Columns

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

***pulse shape diagnostics, may be used to filter problematic events**

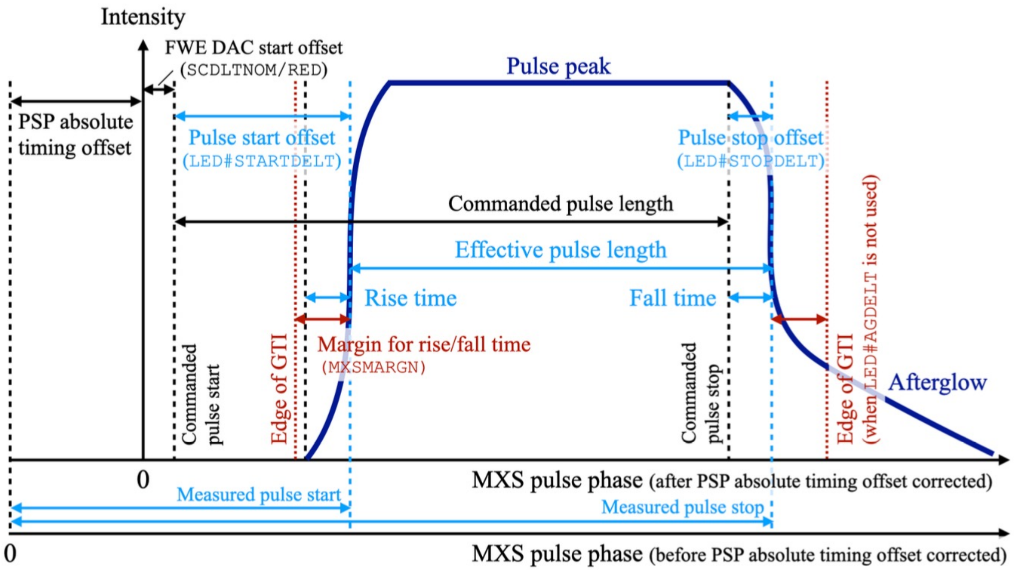
Resolve Grade (ITYPE)

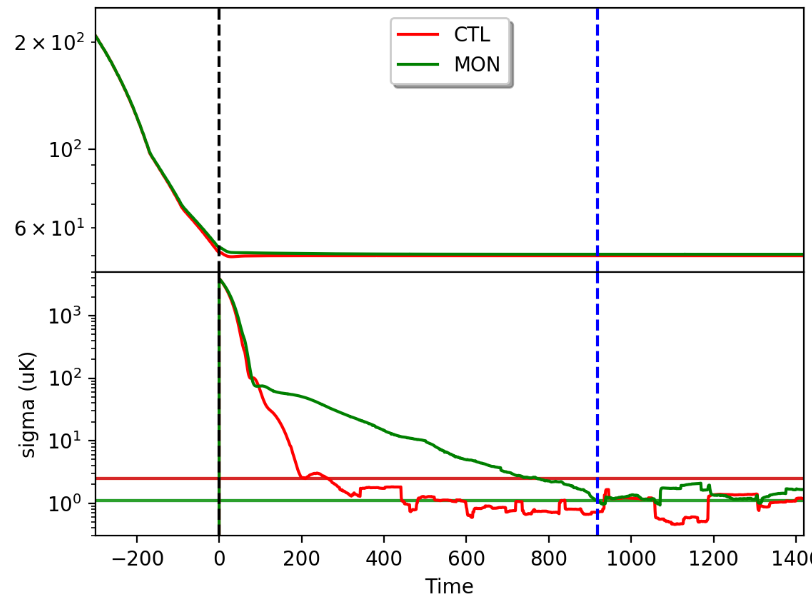
	$t_p \le \delta t_2$	$\delta t_2 < t_p \le \delta t_1$	$\delta t_1 < t_p$
$t_n \le \delta t_2$	Ls	Ls	Lp
$\delta t_2 < t_n \le \delta t_1$	Ls	Ms	**Mp**
$\delta t_1 < t_n$	Ls	Ms	**Hp**



Resolve PL Prelims

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

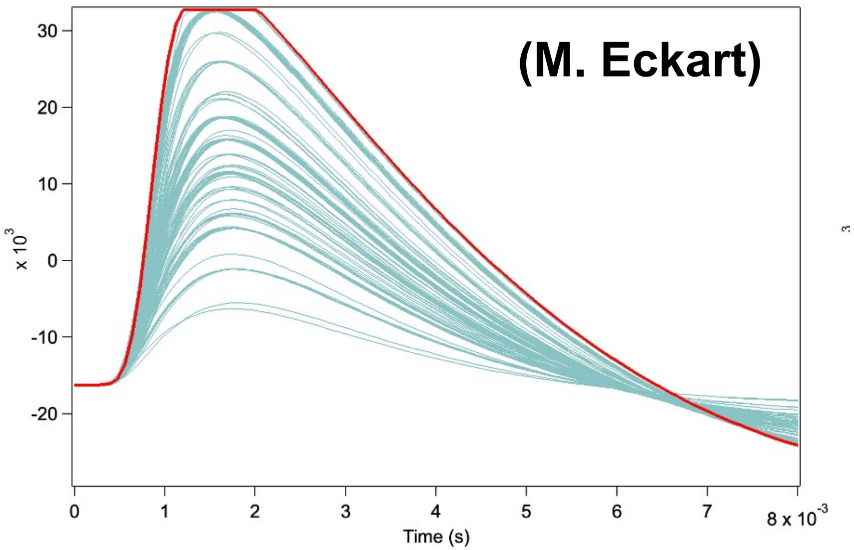




- Invert LOST GTI file (expo map) – gtiinvert
- Assign antico energy (PI) – rslanticolpi (also light curve; rslanticolc)

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!]

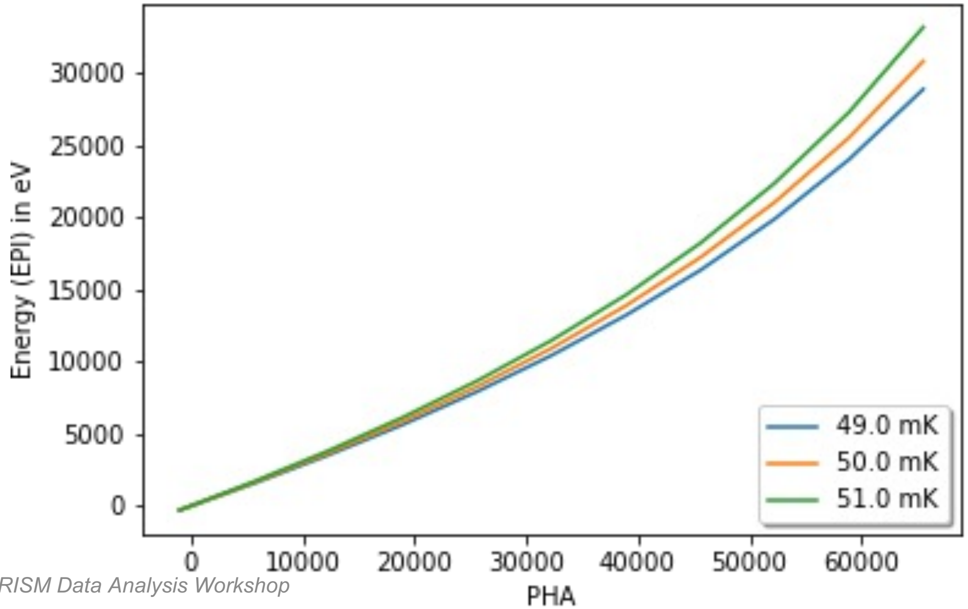


- Associate groups of secondary with primary events (energy assignment) – rslsecid

Resolve PL Energy assignment

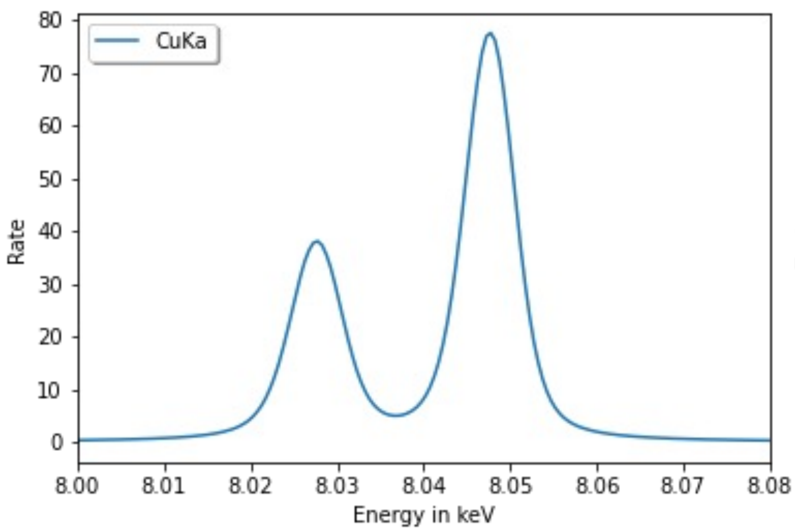
- Merge event files and screen for drift correction
  - Improved pre-screening **planned**
- Compute the energy scale drift correction (energy assignment) – rslgain
  - Parameter optimization in **progress**
  - MXS rate correction **planned**
- Assign initial energy (PI) – rslpha2pi

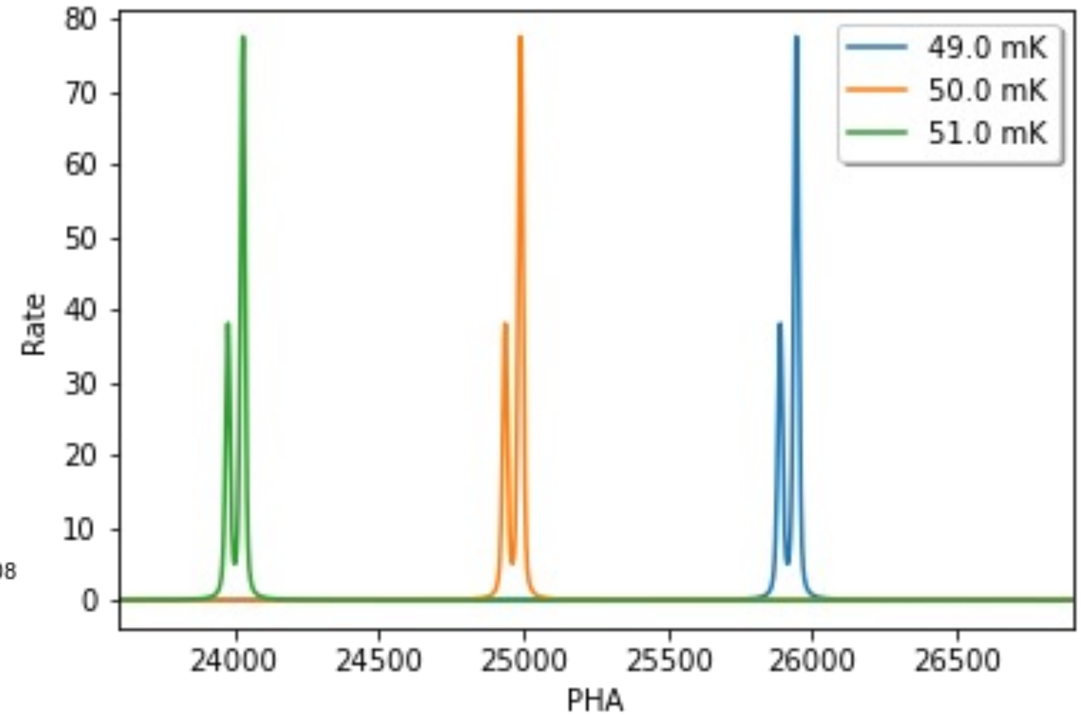
- Gain curve: Energy as a function of PHA (pulse height amplitude) and ‘temperature’
- Separately calculated for High, Mid, and Low resolution events for each pixel



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.





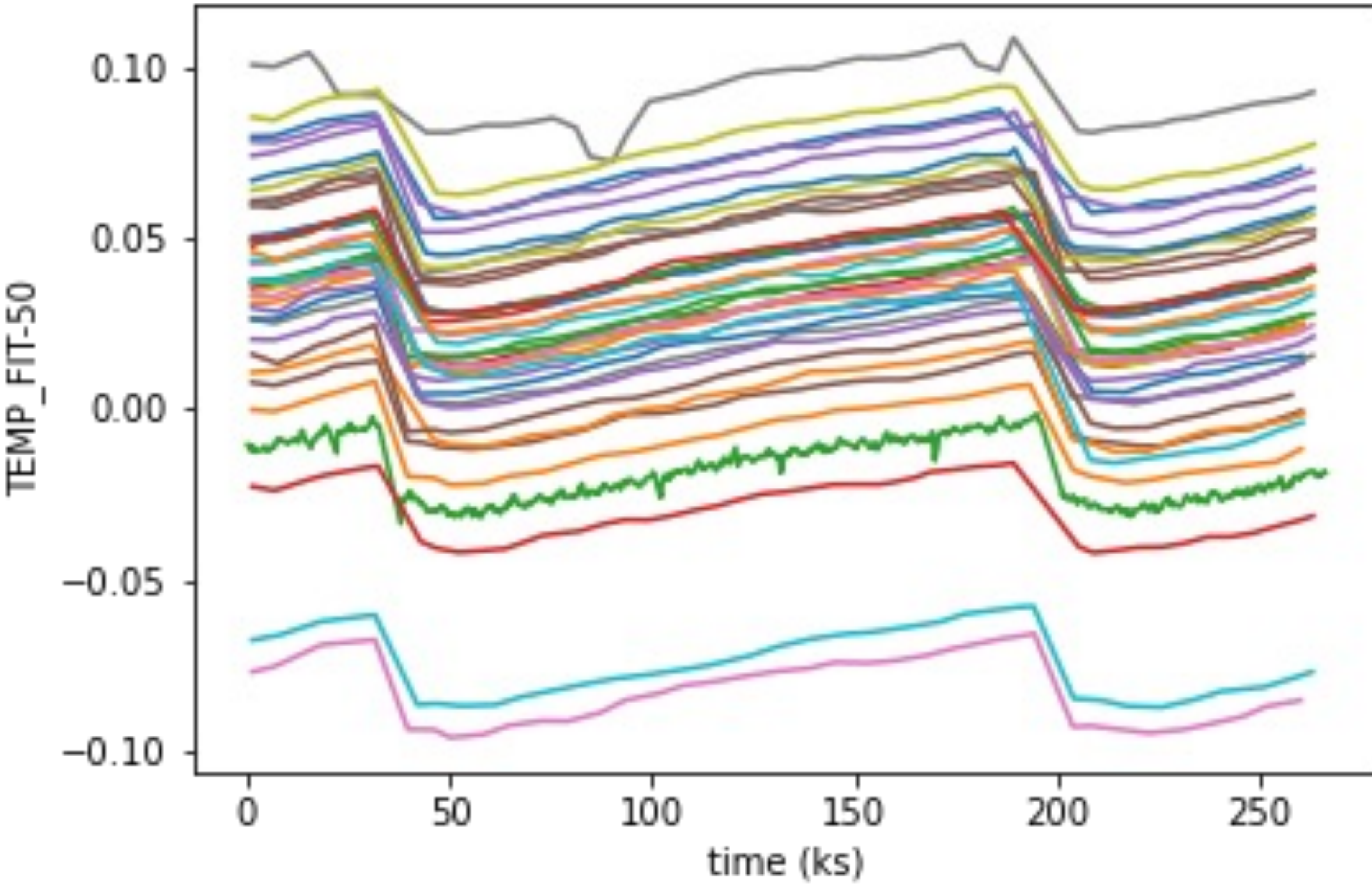
Methods and lines

XRISM X-Ray Imaging and Spectroscopy Mission

Cal method	Cal lines	Pipeline Gain History	Pipeline energy assignment
Fe55	MnKa, MnKb	If FW set to Fe55 (MnKa)*	Primary method
MXS (direct)	Cuka, Cukb, Crka, Crkb	If MXS direct on (CuKa, CrKa)*	If FW Fe55 not in optical path
MXS (indirect)	AlKa, AlKb, MgKa, MgKb	If MXS indirect on (AlKa, MgKa)*	For calibration only
Cal-pix	MnKa, MnKb	Always (MnKa)	If no MXS direct and no Fe55

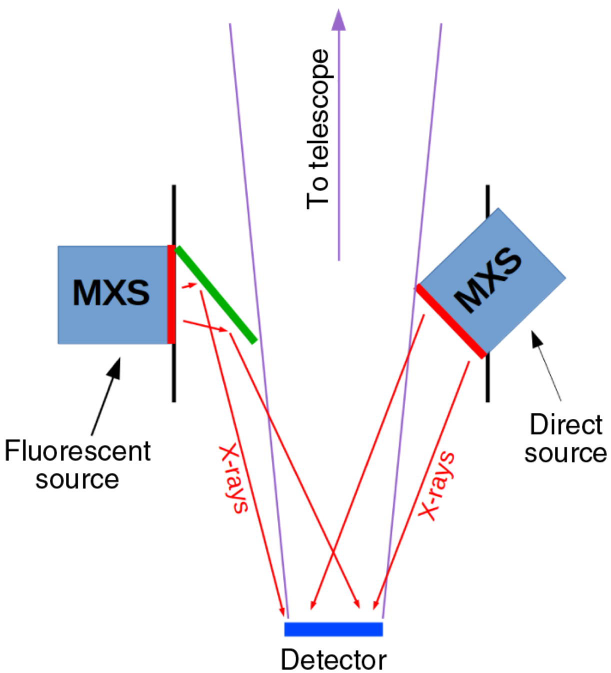
***per-pixel**

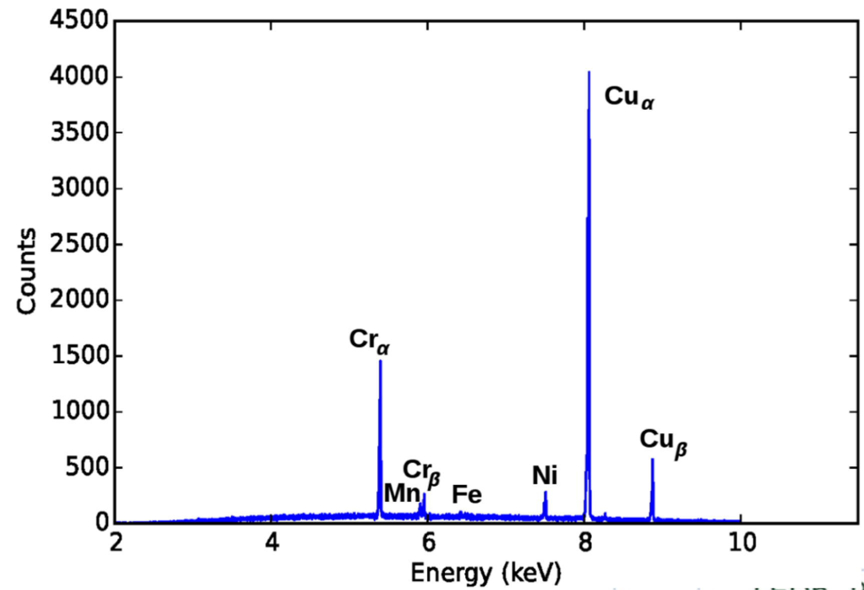
Gain Histories (per pixel)

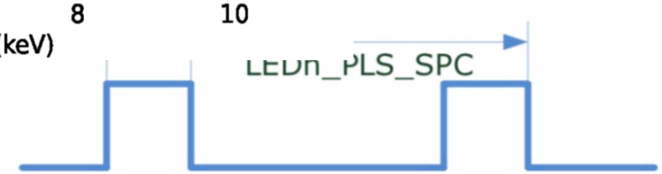


The Resolve MXS

- 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.



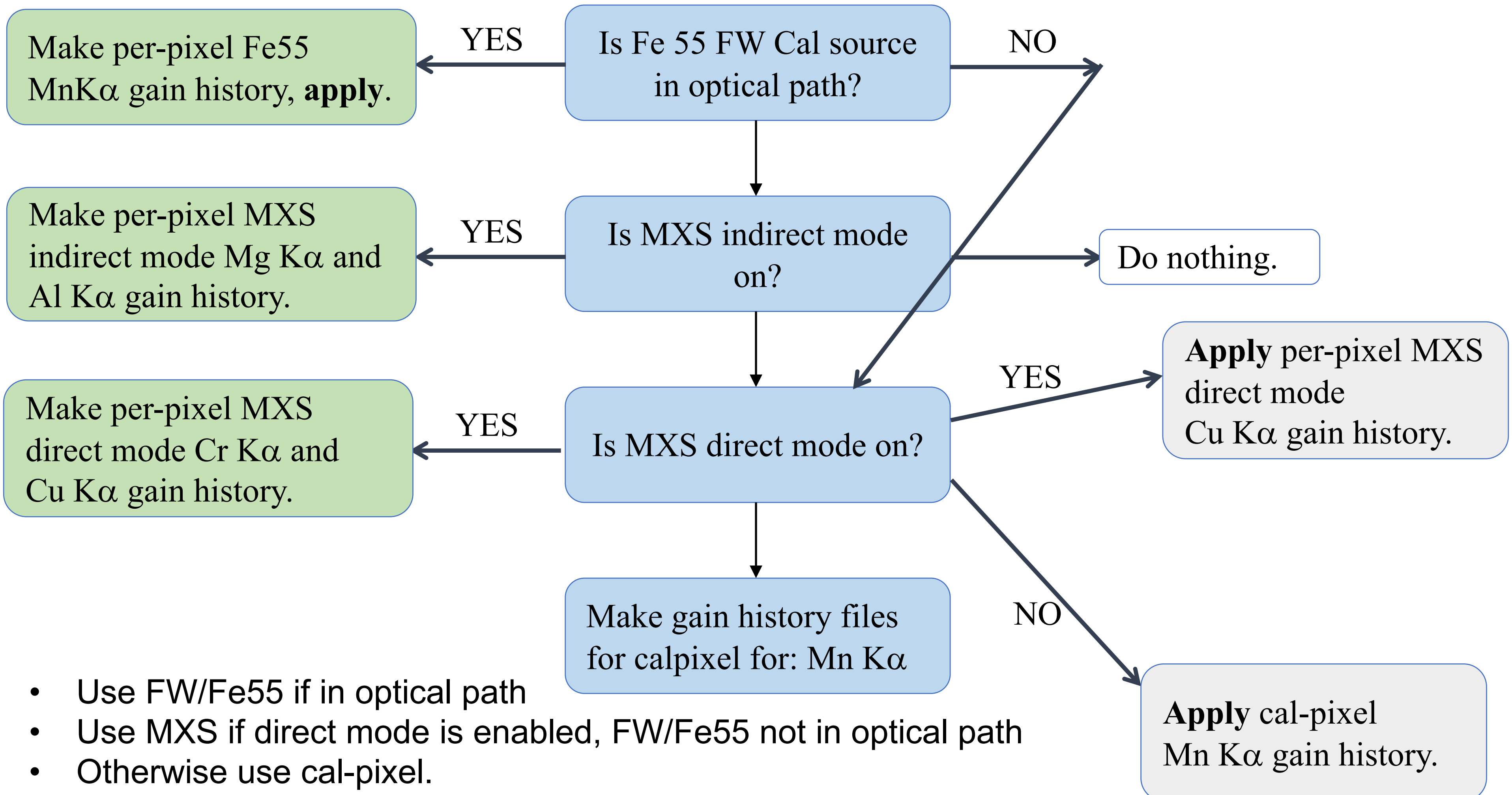




https://doi.org/10.1117/1.JATIS.4.1.011204

Methods and lines

Resolve Main Pipeline Gain Calibration Method Decision Tree



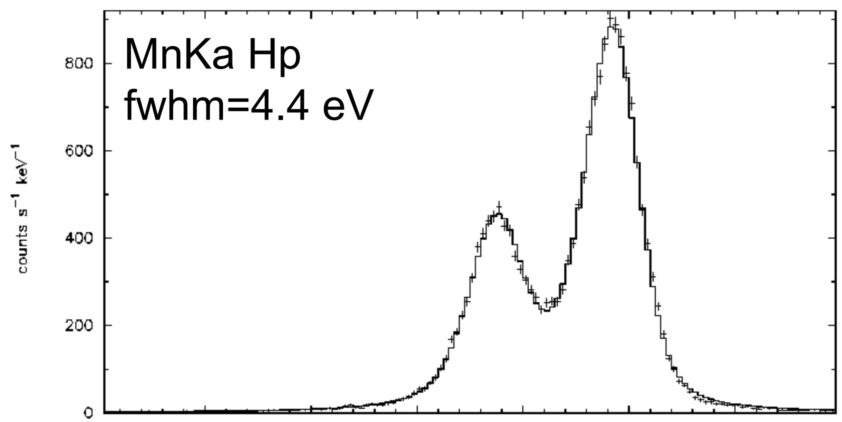
- Use FW/Fe55 if in optical path
- Use MXS if direct mode is enabled, FW/Fe55 not in optical path
- Otherwise use cal-pixel.
- Make all possible gain history files

Final Calibration, Screening

XRISM X-Ray Imaging and Spectroscopy Mission

- 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**

ahtigen + ahscreen



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

Xtend PL Processing

(in brief)

Xtend Calibration, Screening

XRISM X-Ray Imaging and Spectroscopy Mission

- 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

User Data Reprocessing

- 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

X-Ray Imaging and Spectroscopy Mission

- 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.

Post-pipeline - NXB

- 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

X-Ray Imaging and Spectroscopy Mission

- 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.

EXTRA SLIDES

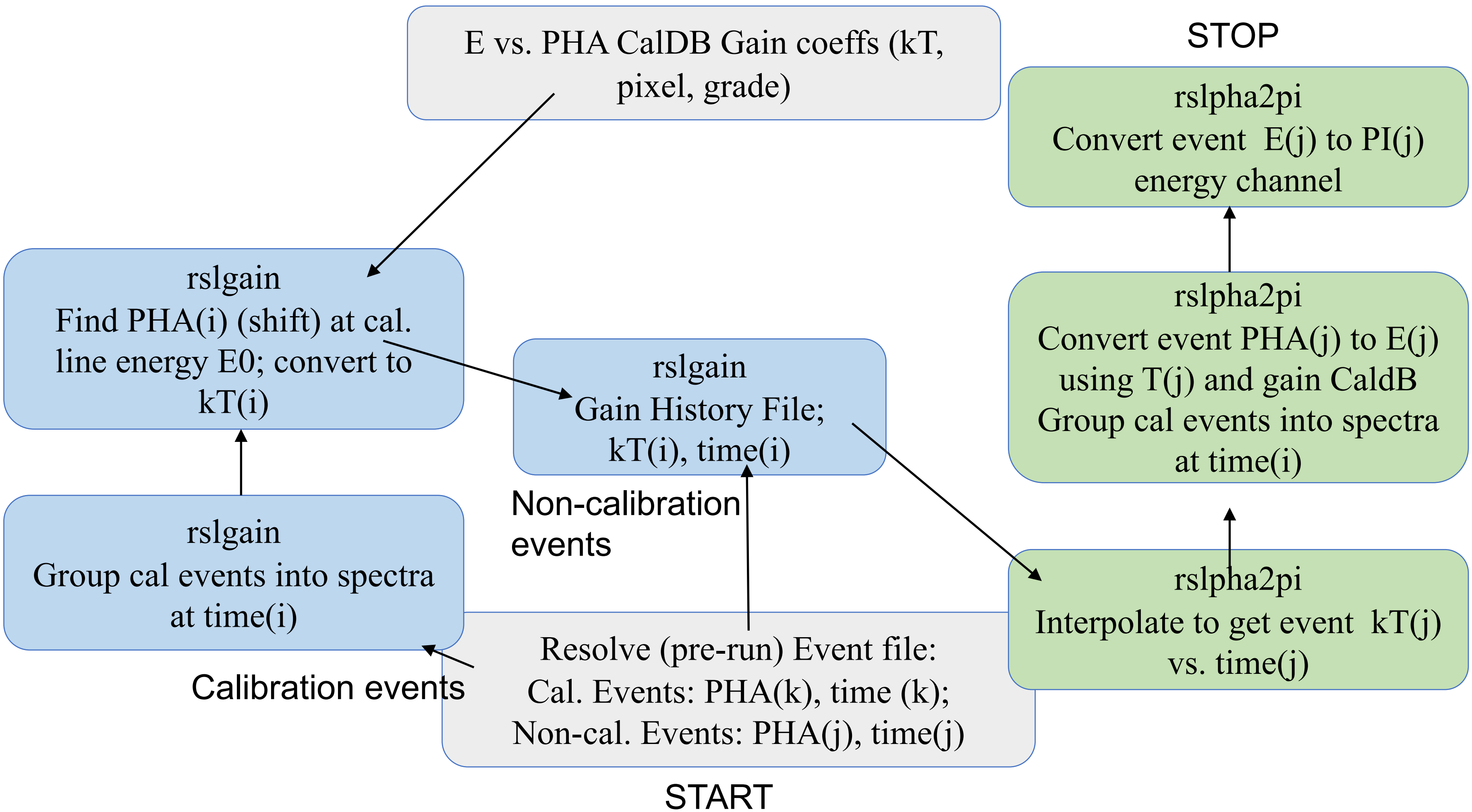
CalDB Status, Plans

XRISM X-Ray Imaging and Spectroscopy Mission

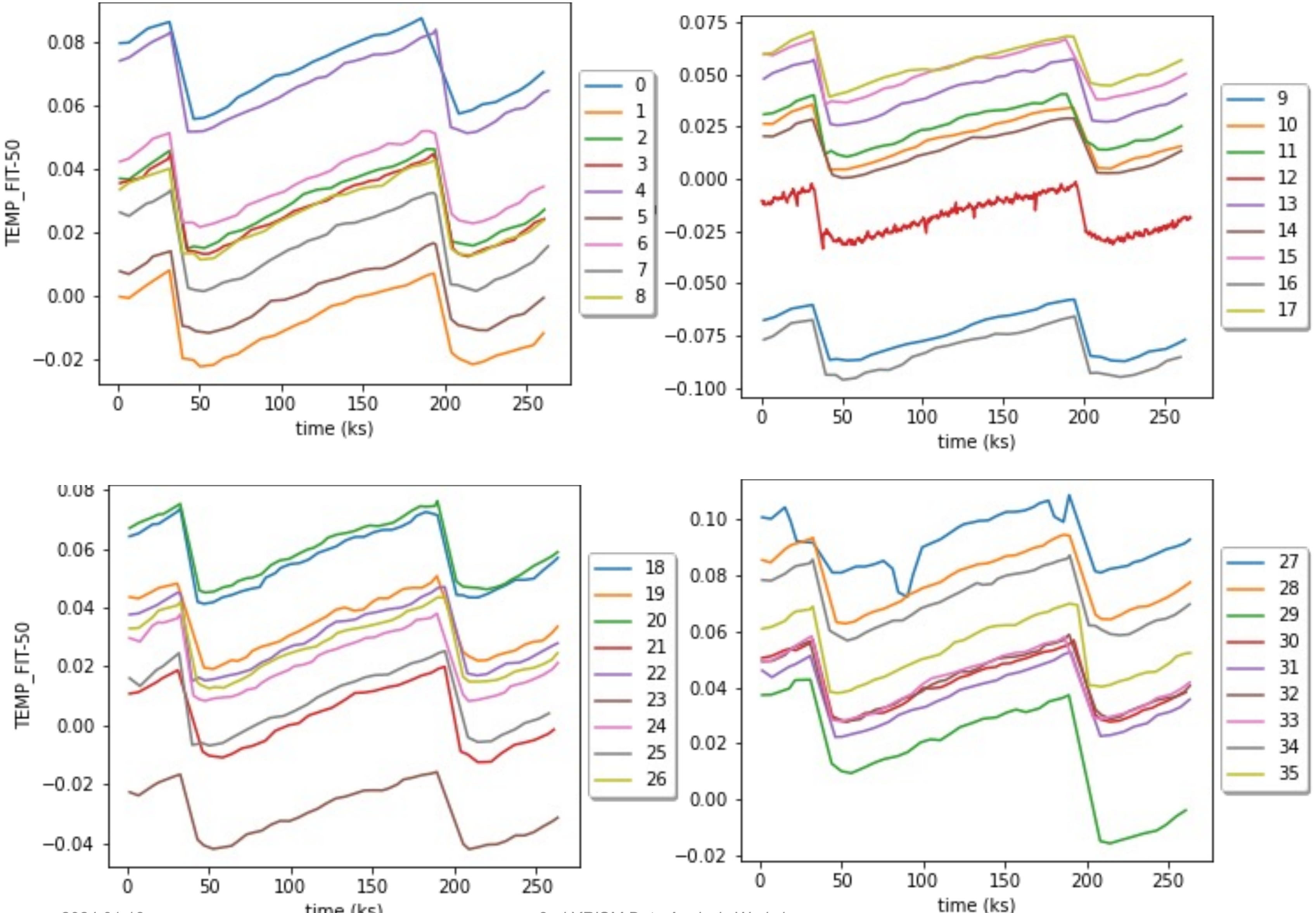
- Pre-fight CalDB: provides a high-fidelity, self-consistent realization of the XRISM subsystems (XMA, Resolve, Xtend).
- SDC is coordinating incorporation of in-flight calibration analysis into CalDB.
- XMA
  - Status: A fully self-consistent model of the XMA that reproduces XMA ground calibration.
  - 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 cal
    - 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



Gain Histories (per pixel)



Graphical Overview of the Suite

XRiSM X-Ray Imaging and Spectroscopy Mission