Astronomy & Astrophysics manuscript no. data_center
June 29, 2021

Data processing pipelines and data center for the X-ray Imaging Spectrometer/Telescope STIX

H. Xiao[1], Shane Maloney[2], Ewan Dickson[1,4], Säm Krucker[1], László Etesi[1], Andrea Francesco Battaglia[1,3], Daniel Ryan[1], and Other STIX team members

[1] University of Applied Sciences and Arts Northwestern Switzerland (FHNW), 5200 Windisch, Switzerland
e-mail: hualin.xiao@fhnw.ch
[2] Astrophysics Research Group, School of Physics, Trinity College Dublin, Dublin 2, Ireland
[3] ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
[4] University of Graz, Universitätspl. 3, 8010 Graz, Austria

ABSTRACT

Aims. The Spectrometer/Telescope for Imaging X-rays (STIX) instrument onboard the Solar Orbiter mission launched on February 10, 2020 promises advances in the study of solar flares of various sizes. It is capable of measuring X-ray spectra from 4 to 150 keV with 1 keV resolution binned into 32 energy bins before downlinking.
Methods. Methods
Results. Results.

Key words. Solar flares –Data Center – STIX data products – Data processing pipeline

1. Introduction

Solar Orbiter is a Sun-observing mission of the European Space Agency that addresses the interaction between the Sun and the heliosphere. It was lunched on Feb. 10, 2020 for a nominal mission duration of seven years and a planned extension of three years. It carries ten sets of instruments for comprehensive remote-sensing and in-situ measurements. Solar Orbiter will perform detailed measurements of the Sun as close as 0.28 AU and for the first time look at its uncharted polar regions (Forveille & Shore (2020)). Its goal is to address the center question of heliophysics "How does the Sun create and control the heliosphere?". It is designed to identify the origins and causes of the solar wind, the heliospheric magnetic field, the solar energetic particles, the transient interplanetary disturbances, and the Sun's magnetic field. This consists of the study of energetic solar phenomena like flares, solar transients, the solar wind accelerating mechanisms, and the solar dynamo principle.

The Spectrometer Telescope for Imaging X-rays (STIX) is one of the ten instruments onboard Solar Orbiter. It measures X-rays from 4 to 150 keV and takes X-ray images with a few arcsec angular resolution by using an indirect imaging technique, based on the Moiré effect. The angular resolution is 7 arcsec and the spectral resolution 1.1 keV full-width-half-maximum at 30 keV. Its instrument consists of 32 collimators with grids and 32 pixelated Cadmium telluride detector units called Caliste-SO. STIX's main purpose is to study the extremely hot solar plasma and the high-energy electrons accelerated during solar flares. It will address the key science goals of the Solar Orbiter mission by providing information on intensity, spectrum, timing, and location of accelerated electrons near the Sun.

During its nominal operation, STIX continuously acquired data. They are processed and archived at STIX data center, established at FHNW. STIX data center processes STIX raw data and produce higher level data products, monitors telemetry data quality and instrument health, provides data products and browsing tools for the solar physics community, and support STIX operations and archives STIX data products. This paper presents the databases, data processing algorithms and the pipeline, data products and data analysis tools developed at STIX data center.

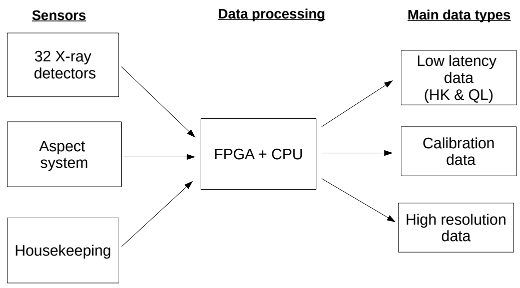

Fig. 1. STIX onboard data flow and data types.

2. Onboard data processing and raw telemetry data

STIX could process a maximum of 800,000 events per second (GOES X10 flares), which corresponds to a data stream of 20 MBits/sec. The event data needs to be reconciled to an average telemetry rate of 750 bits/sec. This is done by an FPGA and a CPU located in the instrument Data Processing Unit (IDPU). The FPGA is used for highly demanding tasks such as summing and sorting the event data stream into accumulators in the SDRAM. Then application software running on the CPU processes the data in SDRAM in three parallel data paths by the flight software (FSW) running on the CPU: primary event processing, quick look processing and calibration data processing.

For the primary event processing, the high-resolution signal amplitude values are first adjusted slightly by the FPGA to compensate for the current detector temperature, and then converted into one of 32 quasi-logarithmic detector-matched 'science energy bins' by the FPGA using a lookup table. Counts as a function of energy, detector ID and pixel number are accumulated in accumulators with a variable integration time that is a multiple of 0.1 second. The accumulation is truncated when a suitable number of counts have been reached or the accumulation time exceeds the pre-defined maximum value.

The contents of these accumulators are compressed and transferred to a rotating memory. Within a timescale of minutes, the ASW transfers the data from the rotating memory to a 16 Gigabyte flash memory, which is able to store several months of data.

Note that the processing to this point, done 'blindly' by the FPGA and FSW so that there is no data selection or statistically-significant loss of science–relevant information content. Data selections are done by the processor as background tasks, based on post facto uploaded requests via ground commands, originating from the STIX operations team. There are two stages to the processing of the selected time/energy range by the instrument processor: the first step is selective averaging into science-relevant time and energy bins; The parameters are explicitly uploaded as part of user data requests. The second stage is to compress the selected data. The output of the compression algorithms is converted to telemetry packets and stored in the instrument telemetry buffer.

The second data path is quick look processing. Individual events are summed into a set of 16-bit accumulators after being corrected for temperature effect and converted to science energy channels. The accumulation time is fixed (can be adjusted by telecommand) with a default of 4 seconds. Corresponding triggers are also accumulated into 16 24-bit accumulators. The contents of these quick-look accumulators are used by the FSW for several purposes: flare identification, the determination of flare position and as input into spatially-averaged spectra and light curves for inclusion in the quick look data. The specifics of the averaging are defined by TC parameters, but the goals are to provide an overview of solar activity, and to monitor individual detector performance. The purpose of the 3rd data path is to provide data for energy calibration. In this case individual events are accumulated with their temperature-corrected scaled A/D resolution but only during quiet times and with very long accumulation times (day). This provides detector-based background spectra that include features from the calibration source that can be used on the ground to establish and monitor the detector energy gain and resolution. To avoid contamination of background spectra by periods of enhanced (but still low level) solar activity, counts are only be accumulated if no other counts occurred in the preceding TC-specified number of seconds. The data directly read from STIX hardware can be divided into three categories. The first category is housekeeping data, which include voltage and current of power supplies, temperature sensor readouts; The second category is event data read from the 32 CdTe detectors, including the digitized photon amplitudes recorded by the 384 CdTe pixels detector, as well as the triggers. The third category is aspect system photo-diode readouts, which can be used to determine the pointing directions of STIX.

Aspect data is handled by a combination of the FPGA and FSW starting from the 1 kHz 12-bit digitization of the output of the 4 aspect photodiodes. Low time resolution averaged aspect data is included in HK telemetry. In addition, user TC requests can include selected intervals with higher time resolution as part of the bulk science T/M. Details of the on-board aspect data handling can be found in RD18. No on-board aspect interpretation is required for instrument operation. In addition to these STIX specific instrument functions, the IDPU provides the normal functions of instrument level command interpretation and execution, commanding bias voltage levels in response to ground commands or internally generated criteria, commanding attenuator state changes, digitizing, monitoring and multiplexing instrument temperature sensor output, etc. Both quick-look and science data are buffered internally so that although the rate of STIX- generated data will be strongly dependent on solar activity, the data can be metered to the spacecraft at a rate and cadence determined by the spacecraft requests. The IDPU includes memories (ROM, RAM, and Flash) for data accumulation and software. It interfaces to the Payload Data Management Unit, a housekeeping system, and other STIX electronics (detector ASICs, aspect diodes readout, attenuator control, and power supplies). By use of an autonomous state machine, it manages a rotating memory. The interfaces between IDPU and the spacecraft are realized through a SpaceWire link (protocol implemented in FPGA with additional external drivers/receivers). STIX uses a robust algorithm to set a flare flag that includes information on the intensity, spectrum and location of the event. This information will be used internally as input to the data compression algorithms and sent to the s/c for optional transmission to other instruments for use in choosing to observe modes or data selection. Implementing the data flow in hardware, initial accumulation in the 64 MByte rotating memory is done autonomously using an FPGA-based hardware state machine. The same FPGA provides all necessary logic to interface the rotating memory to the main STIX processing unit. The interface is organized in such a way that the data can be accumulated by the state machine, while at the same time is available for access by the processing unit. The architecture of STIX IDPU is presented in Figure 2-23. Communication with ASICs and conversion of amplitude into 32 energy bins are performed in State Machines in Main (or Redundant) IDPUs. Initial accumulation in the rotating memory (64MByte rotating memory) is done autonomously using an FPGA-based hardware State Machine inside each of FPGAs. FPGA provides all necessary logic to interface the rotating memory to the main STIX processing unit implemented (as IP core) inside the same FPGA. The interface is organized in such a way that the data can be accumulated by the rotating memory, while at the same time is available for access by the processing unit. The total volume of the memory is 128 Mbytes, less than half of which is allocated for S/W operations. The rest of the memory can be allocated to rotating memory, implying that the 64 MByte available for rotating memory is a conservative estimate. The LEON 3FT synthesized on Actel RTAX2000S/SL core is the baseline for the digital data processing unit. It provides 20 MIPS of processing power. Since the State Machines internal the CPU FPGA handle the time-critical tasks, the LEON3FT is primarily concerned with the processing tasks mentioned in previous paragraphs.

More details descriptions of on-board data processing can be found in STIX telemetry can be classified in four categories: Housekeeping, Quick-look, Diagnostic , and user requested Bulk Science (user-requested). HK and QL will be directed to the low Latency data store in Spacecraft Solid State Mass Memory (SSMM). For the most part, STIX data are organized by data content, with each major data content type (e.g. x-ray imaging data, spatially-averaged x-ray data, HK data, etc.) associated with its own packet types. While the format of each packet type is fixed, the relative frequency of each packet type is dependent on solar activity and/or instrument mode. Although the mix of packet types will vary, the 1-day average of HK and QL will respect the allocated rate of 4.3 Mbits/day (50 bps) each. The sum of all data types will respect the orbital average of 1.8 Gbits/orbit (700 bits/second for 30 days). STIX can make use of an onboard memory to load the SSMM at a rate compatible with spacecraft requirement.

3. Data processing pipelines at STIX data center

3.1. Data link and data reception

Telemetry generated by STIX are first be stored in the spacecraft solid state disk. They are transmitted to ground data station when they are passes. After being received by the ground station of SOLAR Orbiter, it will be processed by ESA and stored in the EEDS database. ESA provides a web interface. The payload data must be created through the EEDS web page to create a data download request. After receiving the user's data request, EDDS displays the original payload data in the form of hex code and sends it to the stix data center server through the RSYNC protocol. Users can define the time interval of data request. Generally speaking, STIX server requests data from EDDS once a day. The data STIX obtains from EDDS is a HEX code with a time stamp, which is the same as the data sent by STIX to the spacecraft platform after being converted into binary data.

3.2. Data processing pipeline

3.2.1. Raw data parsing

include a decompression error map here

3.2.2. Background level monitoring

3.2.3. Flare detection

The flare detection threshold is given by

$$Threshold = median + 2\sigma \tag{1}$$

Median and sigma are calculated using the latest LCs of quiet period.

Therefore the threshold is a fixed value. In Nov. 2020, the threshold was about 350

and now 281.

3.2.4. Calibration data processing

Ba134 radioactive sources with a total activity of about 4000 Bq are placed at the front of each detector. The total activity of the radioactive source is approximately 4000 Ba. When the radioactive source decays, gamma rays are generated. These gamma rays can form peaks in the energy spectrum of the detector. The corresponding energy is known, and the corresponding relationship between ADC and real energy can be calibrated through the position of the peak in the energy spectrum, that is, the calibration coefficient. The figure below is a typical Ba133 gamma-ray energy spectrum measured by STIX CdTe detector. There are three obvious peaks in the energy spectrum, and they correspond to three energies of 30 keV, 35 keV and 81 keV. There are many ways to determine the position of each peak, you can use the ECC method, or use the Gaussian function to fit the left part of the peak.

3.2.5. Flare location using coarse flare data

3.2.6. Flare classification using GOES x-ray flux

3.2.7. Fit file creation

4. Data levels

- Raw data.
- L1 data.
- L2 data

The latest level 1 FITS IO from Shane has been integrated into the data processing pipeline on pub023 server.

I have recreated fits products for all old telemetry data with the upgraded SW.

The L1 fits files created by this pipeline have a different data level: L1A ('A' here means prerelease/alpha version).

The idea behind L1A data sets is to allow for quicker access to STIX data in the fits format instead of grabbing data from plots or using JSON requests,

for operations, debugging etc.

The L1A data sets can be generated within a few minutes after the arrival of a new raw telemetry file.

The differences between L1A and L1 available in Shane's ftp include:

1. Two different L1A data files may have duplicated data

2. L1A data sets are still created for incomplete packets (L1 checks for data completeness)

3. SPICE kernel data for telemetry files always arrives after one or two days later.

So there may be a sub-second difference between the UTC time in fits files (same to times on web pages)

and the real time.

Shane's formal L1 release can avoid this issue if they are produced on a later date.

4. L1A contains housekeeping data

The Level0 archive contains TM which has been parsed or decommuated into readable structures but no additional external information is include:

times are not converted to UTC no calibration or conversions applied for STIX we need to decide if we decompress / combine X-ray L0 the count/trigger data at this stage or in the next level L1

copy manual, tree like, json formats name, raw value, eng value, children look-up table, to know description

estimat mongodb benchmark Mongodb benchmark, key value, index, performance

5. Data request procedure

created of data requests manuall checked **data request naming convention yyddmm00 to yyddmm 11**

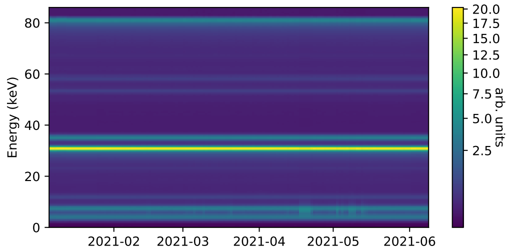

Fig. 2. Caption

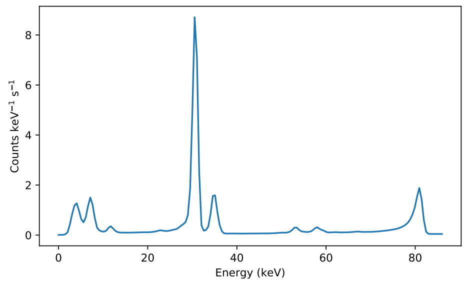

Fig. 3. Spectrogram of STIX calibration spectra

6. Database

6.1. Raw data packet database

6.2. Configuration database

7. Web data browsers

7.1. Quick-look light curve

7.2. Science data quick analysis

7.2.1. Calibration data

7.2.2. Solar Orbiter orbit viewer

8. Data access and APIs

9. Future work

10. Conclusions

References

Forveille, T. & Shore, S. 2020, A&A, 642, E1