Hurford 16.04.2015
Corrected and expanded, 11.05.2015

Integer Compression in STIX T/M

1. Scope

This document recommends the parameterized algorithm to be used by flight software to compress integers in the STIX t/m. The content of this document is adapted from a .ppt description discussed internally in December 2014.

2. Introduction

The expected values of integers in the STIX t/m (e.g. event counts, trigger counts, visibilities) are expected to range from just a few to more than 10^9. To save t/m and to avoid having to dynamically adapt t/m formats to the values themselves, compression of these values to a smaller, fixed number of bits is required. Such compression should not, however, be a limiting factor in the interpretation of such numbers.

In general, compression is a common problem for which various compression algorithms have been identified. The choice of algorithm depends on what characteristic(s) of the data one is willing to sacrifice. After considering several options (square root, STIX-specific approaches, etc) , it was decided to use a single family of quasi-floating point algorithms. These have the characteristics that they can cover a large range of values and that their maximum error can be expressed as a fraction of the value of the datum. The latter feature is appropriate for STIX since systematic (calibration) errors (as opposed to statistics) will limit the interpretation in many cases. Provided the compression error is smaller than the expected systematic error or accuracy requirements in the datum, such a choice is well-suited to STIX requirements.

This write-up first outlines the range of numerical values to be accommodated for different t/m data items, describes the quasi-floating point concept and algorithm, and then suggests parameterization of the compression algorithms for each data type.

3. Expected range of STIX integer t/m values

The estimates are based on the following assumptions, which refer to t/m values, not to accumulator values.

Table 1. Assumed maximum values

Detector count rate	25000/s/det
Count rate per pixel	10000/s/pixel
Trigger count rate	50000/s/det
Max integration time for event t/m	32 s
Max calibration event rate	500/s
Max calibration bins to sum	128
Calibration oscillator	1000 Hz
Max calibration integration time	2 days
Quick look light curve integration time	4 s
Quick look spectral integration time	1024 s

The assumed maximum values are realistic in that they can be expected to occur under well-defined and plausible conditions. Therefore these values are adopted to avoid having to change the t/m format dynamically. The rationale for the assumed maximum values can be summarized as follows:

- Current baselined value of the median detector rate for changing rate control regimes is 22000/s/det. 10% margin is added to allow for detector-to-detector variations
- When one row of large detectors only is used (as in the 3rd rate control regime), 40% of the counts can occur in one pixel for a course subcollimators when the source is unresolved.
- The maximum trigger count rate corresponds to the maximum detector count rate.
- The maximum integration time is consistent with an integration time that can comfortably characterize the flare decay.
- The maximum calibration rate represents ~2x the (somewhat uncertain) current estimated background plus calibration source rate.
- Maximum calibration bins to average permits a low resolution version of the background spectrum.
- The maximum calibration integration time is based on the current plan by which the calibration data will be downloaded every 16 to 32 hours.
- The maximum spectral integration time supports the option of transmitting one single detector spectrum every 32 seconds.

Table 2. Max values of Telemetry items subject to compression

t/m DATUM	Calculation of max value	Max value	Uncompressed bits	Signed?
Summed calibration counts in 1 spectral datum	500 x 0.25(E) x 2days x 86400 / 8 large pixels / 32 det	8x10^4	17	No
Calibration counter	1000Hz x 2 days x 86400	1.7x10^8	28	No
QL light curve datum	25000 x 30det x 4s * 0.5(E)	1.5x10^6	21	No
QL spectral datum	25000 x 1024s x 0.1(E)	2.5x10^7	25	No
Individual trigger accumulators	50000 x 2det x 100s	1 x 10^7	24	No
Summed Trigger accumulators	50000 x 32det x 100s	1.6 x 10^8	28	No
Pixel counts	10000 x 32s	3 x 10^5	19	No
Visibility count differences	9000 x 32s x 0.8(E)	2.3 x 10^5	18	Yes
Variance *	½ x40bins x (30det x25000 x.1s /4)^2	7 x 10^9	33	No

* Assumes input values are scaled by ¼ prior to variance calculation.
(E) is the assumed max fraction of counts in one energy bin

Aspect data, temperatures, voltages and currents are not to be compressed.
Compressed values should be used only for t/m and not for any internal processing.
Input values that exceed the maximum allowable compressed value are represented by all 1's in the compressed output.

4. Quasi-floating point compression

This class of algorithms can be described by 3 input parameters:

- S = 1 implies the datum may be signed; = 0 if datum is always positive
- K = the number of exponent bits to be used
- M = the number of mantissa bits to be used.

The algorithm is applied to an integer value, V, to yield a compressed value, C, as follows:

1. If V < 0, set s=1 and set V = -V.
2. If V < 2^(M+1), set C = V and skip to step 6.
3. If V >= 2^(M+1), shift V right until 1's (if any) appear only in LS M+1 bits
3. Exponent, e = number of shifts+ 1
4. Mantissa, m = LS M bits of the shifted value.
5. Set C = m + 2^M * e
6. If S=1, set msb of C = s.

The algorithm family has the following properties:

- The compressed value is exact for input absolute values up to 2^(M+1)
- RMS fractional error is in range: 1./(sqrt(12)*2^M) to 1/(sqrt(1/12)*s^(M+1))
- The maximum absolute value that can be compressed is given by:

  2^(2^K-2) x (2^(M+1) -1)
- The number of bits required to hold the compressed value is S+K+M

An important facet of the algorithm is that after shifting, the compressed value does not need to include the leading 1, which can be assumed.
Decompression on the ground is done by table lookup with the compressed value as the index.

5. Examples of compression:

K=4, M=4 compression of integer, 25.

Value = 0000 0001 1001;   Compressed value 0001 1001

K=4, M=4 compression of unsigned value, 374.

Value number of shifts

- 000101110110 0
- 000010111011 1
- 000001011101 2
- 000000101110 3
- 000000010111 4
- Mantissa = 0111
- Exponent = 0101
- Compressed value = 0101 0111

With these compression parameters:

- Max exact value: 31
- Max compressible value (1111 1111) corresponds to about 2^14 x 31 = 2^19 or ~500,000
- <RMS error> = < 1/(sqrt(12)*16), 1/(sqrt(12)*31)> =~1.5%

K=4, M=3 compression of a signed integer, -374

- Absolute Value number of shifts
- 000101110110 0
- 000010111011 1
- 000001011101 2
- 000000101110 3
- 000000010111 4
- 000000001011 5
- Mantissa = 011
- Exponent = 0110
- Sign bit = 1
- Compressed value = 1 0110 011

- With these compression parameters:
    - Max exact value: +-15
    - Max compressible value (s1111 111) corresponds to about +- 2^14 x 15 = 2^18 or +-250,000
    - <RMS error= < 1/(sqrt(12)*8) , 1/(sqrt(12)*15)> =~3%

6. Application to STIX Telemetry

The table below summarizes the suggested default parameters for the STIX data compression. Compression to 1 octet can be used in all cases. There will be TC commands to change the compression parameters for each case if necessary, but the size of the output (8-bits) can be considered frozen so that the t/m format will not be affected. Note that the parameters themselves can be condensed to a 1 byte parameter descriptor for inclusion in the relevant t/m packets.

Table 3. Suggested default parameters for STIX t/m Integer Compression

t/m DATUM	S	K	M	Max allowable input value	Max exact input value	Rms compression error	Parameter Descriptor*
Summed calibration counts in 1 spectral datum	0	4	4	2^19	31	~1.5%	0 100 100
Calibration counter	0	5	3	2^34	15	~3%	0 101 011
QL light curve datum	0	5	3	2^34	15	~3%	0 101 011
QL spectral datum	0	5	3	2^34	15	~3%	0 101 011
Individual trigger accumulators	0	5	3	2^34	15	~3%	0 101 011
Summed Trigger accumulators	0	5	3	2^34	15	~3%	0 101 011
Pixel counts	0	4	4	2^19	31	~1.5%	0 100 100
Visibility counts	1	4	3	2^18	15	~3%	1 100 011
Variance	0	5	3	2^34	15	~3%	0 101 011

* Format = S kkk mmm