View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

Preparation of the Electronic Paper for the Proceedings of the

4th International Workshop
on
On-Board Payload Data Compression

23 & 24 October 2014
Venice, Italy

Scientific Data Compression for the Solar Wind Analyser onboard Solar Orbiter

L.Amoruso[a], C.Abbattista[a], V.Fortunato[a], C.J. Owen[b], R.Bruno[c], M.Salatti[d]

[a] Affiliation
Planetek Italia s.r.l, Via Massaua 12, Bari, Italy, I-70132
Email: amoruso@planetek.it

[b] Affiliation
University College London, Mullard Space Science Laboratory, Department of Space & Climate Physics, Holmbury St. Mary, Dorking, United Kingdom, RH5 6NT
Email: c.owen@ucl.ac.uk

[c] Affiliation
National Institute for Astrophysics, Institute for Space Astrophysics and Planetology, Via del Fosso del Cavaliere 100, Rome, Italy, 00133
Email: roberto.bruno@iaps.inaf.it

[d] Affiliation
Italian Space Agency, Via del Politecnico snc, Rome, Italy, 00133
Email: mario.salatti@asi.it

Abstract. The Solar Orbiter mission overall science goal is to discover how the Sun creates and controls the Heliosphere. Solar Wind Analyser (SWA) instrument plays a fundamental role in achieving it, providing high time resolution measurements of the three-dimensional distribution of major components of solar wind's plasma (electrons, protons, alpha particles). This results in large amounts of observation data to be sent on Ground, but has to face constraints descending from the unprecedented operational environment: available telemetry is very limited. This rises highly challenging compression issues to deal with. Classical information theory strategies have been applied in order to characterize, and then predict, the nature of the data SWA is going to measure, starting from a practical heritage on both real data acquired from similar missions and on algorithmic solutions specifically designed for the space domain. Therefore, we have defined and tailored strategies able to enhance the redundancy removal from data, exploiting their structure. The proposed approach represents a trade-off between computational resources, platform (telemetry) and science (data integrity) needs, and demonstrates to be able to meet requirements, achieving an increase of about +17.5% in the compression ratio if compared to the reference solution.

Address all correspondence to: L.Amoruso, Planetek Italia s.r.l, Via Massaua 12, Bari, Italy, I-70132; Tel: +39 080-964-4200; Fax: +39 080-964-4299; E-mail: amoruso@planetek.it

INTRODUCTION

Solar Orbiter is ESA's first M-Class mission under the Cosmic Vision 2015-2025 Program. It is funded by ESA, NASA and many European National Agencies; currently the schedule is dictated by a baseline launch date in July 2017. After the cruise phase and multiple gravity-assist manoeuvres (Venus, Earth) the spacecraft will operate on an elliptical orbit bringing it to a minimum perihelion distance of ~60 solar radii (0.28 AU), where it will make unprecedented measurements. The final goal of the mission is to establish the fundamental physical links between the highly dynamic magnetized atmosphere of the Sun and the solar wind in all its quiet and disturbed states.

The scientific payload of the Solar Orbiter mission is composed of ten experiments. Among the instruments, the Solar Wind Analyser suite (SWA), with its four sensors, will provide at high time resolution the velocity distributions for protons, alphas particles and electrons, together with measurement of minor heavy ions.

The Data Processing Unit (DPU) performs all the suite management tasks, together with scientific data processing (related to protons and electrons fluxes) in order to compress the science data stream by adapting the collected data rate to the limited telemetry bandwidth allocated to the suite. The required compression rates range from 2 to 7.6, according to the different kinds of data and associated data product' volumes.

Sample data, produced by projecting real data acquired by the CLUSTER mission to the different conditions in which SWA is going to operate, have been evaluated considering their information content, using both the entropy figure and actual compression ratios (CR). Results demonstrate the Electron Analyser Sensor (EAS) as the most critical case, considering its measured three-dimensional (acquisition angles in azimuth and elevation, and energy levels) electron distributions have to be compressed by a factor at least equal to 4.3.

Analyses on the compression performances have been carried out on an extended range of algorithms, considering solutions specifically designed for the space domain (CCSDS 121, 122 and 123) and more general ones (as lzma and JPEG2000). The one demonstrating the best compression performance was CCSDS 123 (performing the encoding step under the block-adaptive entropy coding approach), able to outperform 4.3 on all projected EAS datasets (with one exception only reaching only 4.27).

The DPU however is equipped with a LEON2 processor (running at 100MHz) and in charge of the whole suite management (four sensors commanding, housekeeping, S/C communications, FDIR…), so computational resources available to compression tasks are limited. Computational load required by CCSDS 123 is not actually sustainable and hence drives the need for a compression scheme combining efficient performance, both in terms of achievable ratios and computational load.

SOLAR WIND ANALYSER INSTRUMENTS SUITE

The SWA suite composes of four sensors and the DPU. Sensors are:

- the Electron Analyser System, with its two heads (EAS-1 and EAS-2), intended to make the high resolution determination of the core, halo and strahl electron velocity distributions in the solar wind (energy ranging from 1 eV to 5 KeV) and their moments;
- the Proton & Alpha Sensor (PAS), sampling the velocity distribution of proton and alpha particles (energy ranging from 0.2 to 20 KeV/q) at high time resolution equivalent to the ambient proton cyclotron period;
- the Heavy Ions Sensor (HIS), measuring major charge states of C, O and Fe, 3D velocity distributions of prominent heavy solar wind ions, suprathermal ions and pick-up ions of various origins, such as weakly ionized species (He+, O+).

Thus each of the sensors collects (and counts) different particles carried by the solar wind, sampling a portion of the full sky with a scan of azimuth angles, elevation angles and energy levels.

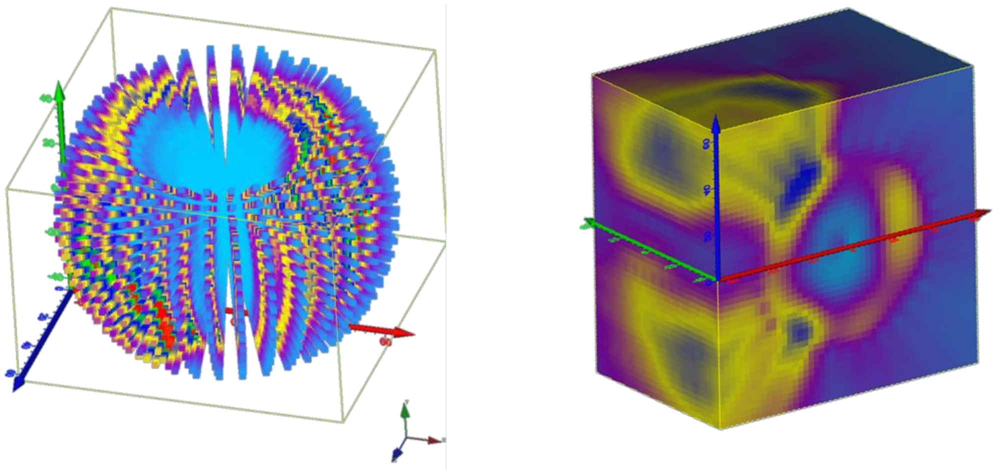
The DPU performs scientific data processing on-board, from taking the raw counts generated by the sensors to transmitting the spacecraft SSMM the data packets to be transmitted to ground for all the sensor but HIS, which is provided of its own processing, and packing logic. The data processing strategy is twofold: the bandwidth allocated to SWA is actually limited to a 14,5 Kbit/s average rate, thus a double data reduction approach has been adopted, with a regular (continuous) wind flux characterization via moments computation, and a full raw data transmission at longer intervals. Moments produce a statistical characterization of the wind (considered as a plasma flux) with synthetic parameters, while full raw data allow checking such characterization as well as possible unexpected conditions.

ANALYSES

DPU operates the SWA sensors in different "modes", designed to meet the science goals and, producing a wide variety of data products. Modes selection controls the "raw" telemetry rates defining time resolutions, on-board processing and

data compression, i.e. the specific data products. The duty cycle among modes is designed to comply with SWA suite telemetry budget limit, assuming the expected compression ratio for each of the products is achieved.
Our analyses aim to assess processing performances, from both data compression ratios and its computational load points of view, considering that processing time is a limited resource, as is the telemetry volume. Analyses have been conducted over a number of steps: first of all the worst case for compression was identified considering all the data products, their volumes and their generation rates. Each of the EAS sensors produces 32768 samples (@ 16 bit/sample), resulting in 512 Kbit/s which, in one of the modes, have to be compressed in real time. This is the case demonstrated to be the most critical.

Secondly, the test data set was identified. Considering that real data available at 0,28 AU are very limited and that Solar Wind models have yet to be assessed (this is one of the aims of the mission), defining relevant data sets was not a straightforward task. Acquisition conditions are very variable and moreover CR figures are demonstrated to be very sensitive to solar wind parameters, mainly to particles density and to solar wind velocity.



a) **b)**

Figure 1 Example of a simulated 3D EAS data distribution in sensor's elevation-azimuth geometry (a) and as an acquired data cube (b).

Data acquired by the PEACE instrument (Plasma Electron And Current Experiment 0) onboard the Cluster mission, made similar measurements @1 AU, have been re-conditioned in order to adapt them to different acquisition conditions of EAS in the inner heliosphere. Thus the main change is related to the Sun distance (from 1,0 to 0,28 AU). The list of the available datasets which produced results presented within this analysis is reported in the following Table 1. They have been selected in order to cover a wide range among the possible acquisition conditions the EAS sensor is expected to face in operations; a new assessment on test datasets is now on-going on this set according also to data collected from the Helios mission @0.5 AU.

Table 1 EAS Simulated Datasets

Dataset #id	*Particles Density [ppcm^3]*	*Wind Velocity [Km/s]*
DS#20080308	~3,0	~450
DS#20040302	***~2,5***	***~700-900***
DS#20070324	~2,5	~400
DS#20080406	~1,0	~700

DS#20080508	~1,0	~550
DS#20080313	~0,8	~600

Analyses on the compression performances, in a preliminary assessment step, was carried out using an extended set of algorithms, considering solutions specifically designed for the space domain (CCSDS 121, 122 and 123) and general ones (as lzma and JPEG2000), and based on a wide set of possible compression approaches: wavelet based, dictionary based, sorting and prediction. Tests have been performed considering «off-the-shelf» SW implementations (tools and libraries); a summary of the results is reported in the following Table 2.

Table 2 Compression algorithms performances trade-offs

Algorithm	***Compression Ratio (CR)***
JPEG-2000 (Mathworks Matlab)	2.26
LZMA (7-zip.org)	2.85
SZIP (© M.Schindler [8]-[9])	3.13
Rice (Basic Compression Library [2])	2.71
CCSDS-121 (HDF-group [7])	2.82
CCSDS-121 (ESA WhiteDwarf [4])	3.21
CCSDS-121 (Custom implementation)	3.23
CCSDS-122 (ESA WhiteDwarf [4])	2.22
CCSDS-123,SA-Modeoption (ESA [3])	4.09
CCSDS-123,BA-Modeoption (ESA [3])	4.27

According to these results, detailed analyses focused mainly on CCSDS-121 and CCSDS-123 algorithms: full tests on the identified worst case data have been considered and results shown in the following Table 3. They demonstrated the latter (with the block-adaptive entropy coding option for encoding) being able to achieve the requested rate of 4.3 on all EAS datasets (the only exception reached a 4.27 figure which is still acceptable). CCSDS-121 reached instead figures ranging from 3.23 to 4.17. It has to be remarked how the single absolute worst distribution has been considered adopting a conservative approach, while (hopefully) more significant (and conservative) estimates have been used when considering best values.

Table 3 CCSDS-based compression performances

Dataset #id	***CCSDS 121***			***CCSDS 123***		
	Worst	Worst-50	Best-50	Worst	Worst-50	Best-50
DS#20080308	3,57	3,80	3,99	4,45	4,69	4,99
DS#20040302	3,23	3,48	3,85	4,27	4,53	4,98

DS#20070324	3,67	3,73	3,74	4,65	4,70	4,72
DS#20080406	4,06	4,14	4,61	5,33	5,51	6,01
DS#20080508	4,34	4,52	4,77	5,56	5,79	6,11
DS#20080313	4,17	4,36	5,02	5,55	5,79	6,55

However, as already mentioned, the CR was only one of the constraints: computational load on the LEON 2 core had to be taken into proper consideration as well. Due to complexity of the processing tasks scheme (not all compressions have to be performed at the same rate), the actual figure considered was the ratio between the computing time, evaluated on the ASIC reference platform, for each task and its periodic cadence. This figure demonstrated how only CCSDS-121 produces loads compatible to the constraints in the worst processing conditions (equal to 17.6% vs 68.7% produced by CCSDS-123).

METHOD

Final aim of the analysis performed is to obtain higher compression ratios while still maintaining a limited computational load (also considering LEON 2 limited performances in floating point operations). Methods taken into consideration were focused on possible improvements of the pre-processing scheme, trying to identify one specifically customized to SWA data, which might provide a solution which is simpler than CCSDS-123 but still more effective than CCSDS-121.

First of all a full analysis over the data and their structure has been performed: their 3D organization was investigated to evaluate compression performance on data sequences re-arranged wrt the sensor's acquisition order. Results show how a "simple" re-ordering scheme is able to improve the compression ratios by approx. 10%. The performances is strictly related to the prediction scheme: Unit Delay predictor actually provides differences to the encoder and thus largely benefits of similarity (a slower variation rate) between next adjacent samples. According to the figures reported in the following Table 4 (showing percentage CR improvements compared to the acquisition order, nominally equal to 100%), elevation angle has been identified as preferential dimension, i.e. the one demonstrating the highest (spatial) correlation between samples. Second ordering dimension comes out less evident (actually changing with the specific datasets), and different criteria, related to sensors' acquisition process, can be considered instead.

Table 4 Simple data reordering impact: discovering preferential direction in particles counts acquisition scheme

Re-Arranged Sequence (slow to fast variations)	**CCSDS 121**	**CCSDS 123**
El-En-Az	100%	100%
El-Az-En	83%	98%
Az-En-El	**109%**	**106%**
Az-El-En	83%	97%
En-Az-El	**111%**	**115%**
En-El-Az	101%	94%

Simple reordering schemes however still entail periodical jumps in samples' order each time sensor steers back from the last to the first elevation angle or energy level. Data can be re-ordered instead in such a way that jumps are avoided completely, always considering a sample that in the 3D space is next to the previous one, varying only one of the three indices per time. This, let's say "complex", re-ordering scheme provides a total improvement equal to up to 17.5% if compared to the custom predictor stand-alone performances, and so is able to bring the CRs to the required figure with most datasets (and the two remaining exceptions can be compensated in an overall average reasoning). The improvements on the compression performance are shown, and moreover evident, in the following Table 5.

Table 5 Overall custom data pre-processing performances

Dataset #id	***acquisition order***			***«simple» re-ordering***			***«complex» re-ordering***		
	Worst	Worst-50	Best-50	Worst	Worst-50	Best-50	Worst	Worst-50	Best-50
DS#20080308	3.67	3.85	4.01	3.90	4.04	4.24	4.18	4.32	4.52
DS#20040302	3.30	3.47	3.90	3.71	3.93	4.33	3.88	4.12	4.57
DS#20070324	3.75	3.80	3.81	4.08	4.16	4.19	4.30	4.38	4.41
DS#20080406	4.16	4.25	4.71	4.37	4.62	5.02	4.71	4.90	5.36
DS#20080508	4.45	4.68	4.90	4.66	4.82	5.11	4.97	5.17	5.50
DS#20080313	4.24	4.43	5.16	4.62	4.85	5.38	4.90	5.16	4.81

The improvement becomes evident when comparing the distributions of residuals (differences between samples and their prediction); pre-processing and specifically mapping (i.e. the second step in pre-processing) are designed to fit on Laplacian distributions and they perform better as the actual residual distribution comes closer to the ideal one. The following Figure 2 compares cumulative distributions on predictions with the different re-ordering schemes considered: it shows how the data distribution after unit delay prediction (i.e. the prediction errors) progressively concentrates around zero.

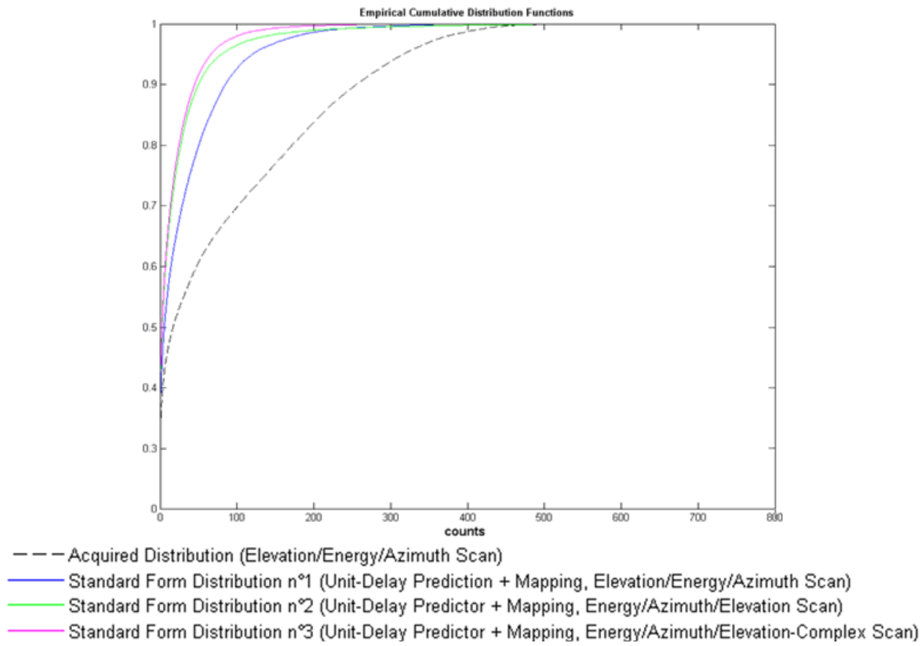


Figure 2 Empirical cumulative distribution functions of the original and pre-processed data.

The method does not affect computational load, in terms of mathematical operations, while the possible increased amount of memory accesses has to be compensated for, including data re-ordering in data acquisition low-level logic (which can use a dedicated FPGA).

1.1. EVOLUTIONS

Further possible improvements were considered, introducing an increased complexity level in prediction (using time and/or more space correlations). A prediction scheme involving two samples was able to produce interesting results: instead of the difference between current and previous sample, the prediction was based on a triangular filter (as in a linear interpolation algorithm). In this way the prediction was considered as an interpolation, current sample's expected value is estimated supposing the knowledge of previous and next ones. This approach demonstrated a further improvement equal to 15% on test data (and exceeds +30% if compared to the original scheme). Also from a computational point of view the increased complexity of the prediction scheme demonstrated a further advantage: the additional load needed in prediction are totally compensated for in the mapping and encoding steps, so that the overall load promises to be even better than the standard' one.

Unfortunately, such prediction scheme also has a significant disadvantage: it entails the resolution of a non-linear system of equations for data de-compression. This system is demonstrated to be not invertible without a reliable initial guess and, though not yet having exploited the time correlation, a convenient way to define such a guess (without transmitting additional information) is still a matter of research.

RESULTS

The approach finally defined, according to the results of tests and analyses performed, foresees a CCSDS 121 scheme, to be applied on a custom pre-processing which exploits a "complex" data reordering scheme.

The overall results with the proposed scheme, if compared on the standard CCSDS 121 pre-processing, demonstrated an improvement in CR from 3,30 to 3.88, equal to 17.5% in the worst case. Figures have been evaluated as an average on 891 EAS simulated acquisitions at the worst case solar wind's conditions, considering the best representative compression product. It has to be remarked as well how for six, out of the eight data sets considered, the scheme has been able to provide a CR at least equal to the required 4.3. This result allows us anyway to adopt the proposed approach, because the worst solar wind's conditions will affect only a limited percentage of data acquisitions and the two exceptions are then compensated for during average acquisition cycle. Finally the customized compression scheme, designed against electrons data particles counts, and still valid for protons, is able to meet the goal.

REFERENCES

[1] Johnstone, A. D., Alsop, C., Burge, S., Carter, P. J., Coates, A. J., Coker, A. J., Fazakerley, A. N., Grande, M., Gowen, R. A., Gurgiolo, C., Hancock, B. K., Narheim, B., Preece, A., Sheather, P. H., Winningham, J. D., and Woodliffe, R. D.: "Peace: A plasma electron and current experiment", *Space Science Reviews*, 79, 351-398, doi:10.1023/a:1004938001388, 1997

[2] Basic Compression Library, A set of open source implementations of several lossless compression algorithms (Huffman, RLE, Rice, Lempel-Ziv-77 and Shannon-Fano), http://www.bcl.comli.eu

[3] ESA, Data Compression Tools, An implementation of the new CCSDS 123 Recommended Standard http://www.esa.int/TEC/OBDP/SEM069KOXDG_0.html

[4] ESA, Data Compression Tools, WhiteDwarf data compression evaluation tool, http://www.esa.int/TEC/OBDP/SEM069KOXDG_2.html

[5] Rice R., "Some Practical Universal Noiseless Coding Techniques-Part III". Module PSI14.K, *Jet Propulsion Laboratory*, JPL Publication 91-3, Pasadena, CA, 1991

[6] University of Nebraska Lincoln, Department of Electrical Engineering, An implementation of CCSDS 122.0-B-1 Recommended Standard http://hyperspectral.unl.edu/

[7] University of New Mexico, Szip compression software, an implementation of the extended-Rice lossless compression algorithm http://www.hdfgroup.org/doc_resource/SZIP/

[8] Szip, A general purpose lossless compression program http://www.compressconsult.com/szip/

[9] Schindler M., “Method and apparatus for sorting data blocks”, *U.S. Patent 6199064*, March 6, 2001

REFERENCES

Number citations consecutively in square brackets [1]. The sentence punctuation follows the brackets [2]. Refer simply to the reference number, as in [3]. Do not use “Ref.[3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first....” The title of the book or the name of the journal shall be typed in italic.

Give all authors’ names; do not use “et al” unless there are six authors or more. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

Sample References

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans.. Roy. Soc. London*, vol. A247, pp. 529-551, April 1955.
[2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, pp. 68-73, 1892.
[3] I.S. Jacobs and C.P. Bean, “Fine Particles, Thin Films and Exchange Anisotropy,” in *Magnetism*, vol. III. G.T. Rado and H. Suhl, Eds. New York: Academic, pp. 271-350, 1963.
[4] K. Elissa, “Title of paper,” unpublished.
[5] R. Nicole, “Title of Paper,” in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface*,” IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740-741, August 1987.