The Consultative Committee for Space Data Systems

CCSDS Historical Document

This document's Historical status indicates that it is no longer current. It has either been replaced by a newer issue or withdrawn because it was deemed obsolete. Current CCSDS publications are maintained at the following location:

http://public.ccsds.org/publications/

The Consultative Committee for Space Data Systems

Recommendation for Space Data System Standards

LOSSLESS DATA COMPRESSION

RECOMMENDED STANDARD

CCSDS 121.0-B-2

BLUE BOOK
May 2012

CCSDS HISTORICAL DOCUMENT

The Consultative Committee for Space Data Systems

Recommendation for Space Data System Standards

LOSSLESS DATA COMPRESSION

RECOMMENDED STANDARD

CCSDS 121.0-B-2

BLUE BOOK
May 2012

CCSDS HISTORICAL DOCUMENT

DEDICATION

This document is dedicated to the memory of Mr. Warner H. Miller of NASA. Warner had been with the CCSDS since its beginning, and throughout the years he was a major contributor to numerous standards for error control coding, radio frequency modulation, and data architecture. He initiated this data compression standard and saw its publication and use by many space missions. Warner was a superb technologist, a gentleman, and a friend always ready to help, especially young colleagues. Warner and his approach to work and life in general will be deeply missed by his many friends and colleagues in the CCSDS.

AUTHORITY

Issue:	Recommended Standard, Issue 2
Date:	May 2012
Location:	Washington, DC, USA

This document has been approved for publication by the Management Council of the Consultative Committee for Space Data Systems (CCSDS) and represents the consensus technical agreement of the participating CCSDS Member Agencies. The procedure for review and authorization of CCSDS documents is detailed in *Organization and Processes for the Consultative Committee for Space Data Systems*, and the record of Agency participation in the authorization of this document can be obtained from the CCSDS Secretariat at the address below.

This document is published and maintained by:

CCSDS Secretariat
Space Communications and Navigation Office, 7L70
Space Operations Mission Directorate
NASA Headquarters
Washington, DC 20546-0001, USA

STATEMENT OF INTENT

The Consultative Committee for Space Data Systems (CCSDS) is an organization officially established by the management of its members. The Committee meets periodically to address data systems problems that are common to all participants, and to formulate sound technical solutions to these problems. Inasmuch as participation in the CCSDS is completely voluntary, the results of Committee actions are termed **Recommended Standards** and are not considered binding on any Agency.

This **Recommended Standard** is issued by, and represents the consensus of, the CCSDS members. Endorsement of this **Recommendation** is entirely voluntary. Endorsement, however, indicates the following understandings:

- o Whenever a member establishes a CCSDS-related **standard**, this **standard** will be in accord with the relevant **Recommended Standard**. Establishing such a **standard** does not preclude other provisions which a member may develop.
- o Whenever a member establishes a CCSDS-related **standard**, that member will provide other CCSDS members with the following information:
  - -- The **standard** itself.
  - -- The anticipated date of initial operational capability.
  - -- The anticipated duration of operational service.
- o Specific service arrangements shall be made via memoranda of agreement. Neither this **Recommended Standard** nor any ensuing **standard** is a substitute for a memorandum of agreement.

No later than three years from its date of issuance, this **Recommended Standard** will be reviewed by the CCSDS to determine whether it should: (1) remain in effect without change; (2) be changed to reflect the impact of new technologies, new requirements, or new directions; or (3) be retired or canceled.

In those instances when a new version of a **Recommended Standard** is issued, existing CCSDS-related member standards and implementations are not negated or deemed to be non-CCSDS compatible. It is the responsibility of each member to determine when such standards or implementations are to be modified. Each member is, however, strongly encouraged to direct planning for its new standards and implementations towards the later version of the Recommended Standard.

FOREWORD

This Recommendation establishes a common framework and provides a common basis for a Lossless data compression algorithm applicable to several different types of data.

Through the process of normal evolution, it is expected that expansion, deletion, or modification of this document may occur. This Recommended Standard is therefore subject to CCSDS document management and change control procedures, which are defined in the *Procedures Manual for the Consultative Committee for Space Data Systems*. Current versions of CCSDS documents are maintained at the CCSDS Web site:

http://www.ccsds.org/

Questions relating to the contents or status of this document should be addressed to the CCSDS Secretariat at the address indicated on page i.

At time of publication, the active Member and Observer Agencies of the CCSDS were:

Member Agencies

– Agenzia Spaziale Italiana (ASI)/Italy.
– Canadian Space Agency (CSA)/Canada.
– Centre National d'Etudes Spatiales (CNES)/France.
– China National Space Administration (CNSA)/People's Republic of China.
– Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)/Germany.
– European Space Agency (ESA)/Europe.
– Federal Space Agency (FSA)/Russian Federation.
– Instituto Nacional de Pesquisas Espaciais (INPE)/Brazil.
– Japan Aerospace Exploration Agency (JAXA)/Japan.
– National Aeronautics and Space Administration (NASA)/USA.
– UK Space Agency/United Kingdom.

Observer Agencies

– Austrian Space Agency (ASA)/Austria.
– Belgian Federal Science Policy Office (BFSPO)/Belgium.
– Central Research Institute of Machine Building (TsNIIMash)/Russian Federation.
– China Satellite Launch and Tracking Control General, Beijing Institute of Tracking and Telecommunications Technology (CLTC/BITTT)/China.
– Chinese Academy of Sciences (CAS)/China.
– Chinese Academy of Space Technology (CAST)/China.
– Commonwealth Scientific and Industrial Research Organization (CSIRO)/Australia.
– CSIR Satellite Applications Centre (CSIR)/Republic of South Africa.
– Danish National Space Center (DNSC)/Denmark.
– Departamento de Ciência e Tecnologia Aeroespacial (DCTA)/Brazil.
– European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/Europe.
– European Telecommunications Satellite Organization (EUTELSAT)/Europe.
– Geo-Informatics and Space Technology Development Agency (GISTDA)/Thailand.
– Hellenic National Space Committee (HNSC)/Greece.
– Indian Space Research Organization (ISRO)/India.
– Institute of Space Research (IKI)/Russian Federation.
– KFKI Research Institute for Particle & Nuclear Physics (KFKI)/Hungary.
– Korea Aerospace Research Institute (KARI)/Korea.
– Ministry of Communications (MOC)/Israel.
– National Institute of Information and Communications Technology (NICT)/Japan.
– National Oceanic and Atmospheric Administration (NOAA)/USA.
– National Space Agency of the Republic of Kazakhstan (NSARK)/Kazakhstan.
– National Space Organization (NSPO)/Chinese Taipei.
– Naval Center for Space Technology (NCST)/USA.
– Scientific and Technological Research Council of Turkey (TUBITAK)/Turkey.
– Space and Upper Atmosphere Research Commission (SUPARCO)/Pakistan.
– Swedish Space Corporation (SSC)/Sweden.
– United States Geological Survey (USGS)/USA.

DOCUMENT CONTROL

Document	Title	Date	Status
CCSDS 121.0-B-1	Lossless Data Compression, Blue Book, Issue 1	May 1997	Original issue, superseded
CCSDS 121.0-B-2	Lossless Data Compression, Recommended Standard, Issue 2	May 2012	Current issue: (changes affecting backward compatibility) – increases the allowed values of block length J from {8, 16} to {8, 16, 32, 64}; – increases the maximum allowed value of the reference sample interval r from 256 to 4096; – allows the use of a restricted set of code options; – defines a new subfield of the Source Configuration Field of the optional Compression Identification packet; (changes not affecting backward compatibility) – adds guidance for selection of code option when two or more code options gave the same performance for a block; – clarifies that the ID bits are to be included when selecting the coding option for a block; – clarifies some text; – adds a dedication to Warner Miller.
EC 1	Editorial Change 1	November 2017	Adds clarifying text to 3.4.3.2.

NOTE – Substantive changes from the original issue are identified by change bars in the inside margin.

CONTENTS

CONTENTS (continued)

1 INTRODUCTION

1.1 PURPOSE

The purpose of this document is to establish a Recommended Standard for a source-coding data-compression algorithm applied to digital data and to specify how these compressed data shall be inserted into source packets for retrieval and decoding.

Source coding for data compression is a method utilized in data systems to reduce the volume of digital data to achieve benefits in areas including, but not limited to,

a) reduction of transmission channel bandwidth;

b) reduction of the buffering and storage requirement;

c) reduction of data-transmission time at a given rate.

1.2 SCOPE

The characteristics of source codes are specified only to the extent necessary to ensure multi-mission support capabilities. The specification does not attempt to quantify the relative bandwidth reduction, the merits of each approach discussed, or the design requirements for coders and associated decoders. Some performance information is included in reference [C2].

This Recommended Standard addresses only Lossless source coding, which is applicable to a wide range of digital data, both imaging and non-imaging, where the requirement is for a moderate data-rate reduction constrained to allow no distortion to be added in the data compression/decompression process. The decompression process is not addressed. See reference [C2] for an outline of an implementation.

1.3 APPLICABILITY

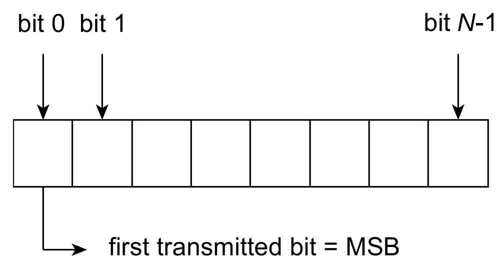
This Recommended Standard applies to data compression applications of space missions anticipating packetized telemetry cross support. In addition, it serves as a guideline for the development of compatible CCSDS Agency standards in this field, based on good engineering practice.

1.4 RATIONALE

The concept and rationale for the Lossless source coding for data compression algorithm described herein may be found in reference [C2].

1.5 BIT NUMBERING CONVENTION AND NOMENCLATURE

In this document, the following convention is used to identify each bit in an N-bit word. The first bit in the word to be transmitted (i.e., the most left justified when drawing a figure) is defined to be 'Bit 0', the following bit is defined to be 'Bit 1', and so on up to 'Bit *N*-1'. When the word is used to express an unsigned binary value (such as a counter), the Most Significant Bit (MSB) shall correspond to the highest power of two, i.e. 2^{N-1} .



In accordance with modern data communications practice, spacecraft data words are often grouped into 8-bit 'words' which conform to the above convention. Throughout this Recommended Standard, the following nomenclature is used to describe this grouping:

8-Bit Word = 'Byte'

1.6 PATENTED TECHNOLOGIES

The Consultative Committee on Space Data Systems (CCSDS) draws attention to the fact that it is claimed that compliance with this document may involve the use of a patent concerning the method for coding low entropy data given in section 3.

The CCSDS takes no position concerning the evidence, validity, and scope of these patent rights.

The holders of these patent rights have assured the CCSDS that they are willing to negotiate licenses under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with CCSDS. Information can be obtained from the CCSDS Secretariat at the address indicated on page i. Contact information for the holder of these patent rights is provided in annex A.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those identified above. The CCSDS shall not be held responsible for identifying any or all such patent rights.

1.7 REFERENCES

The following documents contain provisions which, through reference in this text, constitute provisions of this Recommended Standard. At the time of publication, the editions indicated were valid. All documents are subject to revision, and users of this Recommended Standard are encouraged to investigate the possibility of applying the most recent editions of the documents indicated below. The CCSDS Secretariat maintains a register of currently valid CCSDS Recommended Standards.

[1] *TM Synchronization and Channel Coding*. Recommendation for Space Data System Standards, CCSDS 131.0-B-2. Blue Book. Issue 2. Washington, D.C.: CCSDS, August 2011.

[2] *Space Packet Protocol*. Recommendation for Space Data System Standards, CCSDS 133.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2003.

[3] *Lossless Multispectral & Hyperspectral Image Compression*. Recommendation for Space Data System Standards, CCSDS 123.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, May 2012.

2 OVERVIEW

2.1 GENERAL

This Recommended Standard defines for standardization a particular adaptive source coding algorithm that has widespread applicability to many forms of digital data. In particular, the science data from many types of imaging or non-imaging instruments are well suited for the application of this algorithm.

There are two classes of source coding methods: Lossless and Lossy.

A Lossless source coding technique preserves source data accuracy and removes redundancy in the data source. In the decoding process, the original data can be reconstructed from the compressed data by restoring the removed redundancy; the decompression process adds no distortion. This technique is particularly useful when data integrity cannot be compromised. The price it pays is generally a lower Compression Ratio, which is defined as the ratio of the number of original uncompressed bits to the number of compressed bits including overhead bits necessary for signaling parameters.

On the other hand, a Lossy source coding method removes some of the source information content along with the redundancy. The original data cannot be fully restored and data distortion occurs. However, if some distortion can be tolerated, Lossy source coding generally achieves a higher compression ratio. By controlling the amount of acceptable distortion and compression, this technique may enable acquisition and dissemination of mission data within a critical time span.

This Recommended Standard addresses only Lossless source coding and does not attempt to explain the theory underlying the operation of the algorithm.

2.2 THE SOURCE CODER

The Lossless source coder consists of two separate functional parts: the preprocessor and the adaptive entropy coder, as shown in figure 2-1.

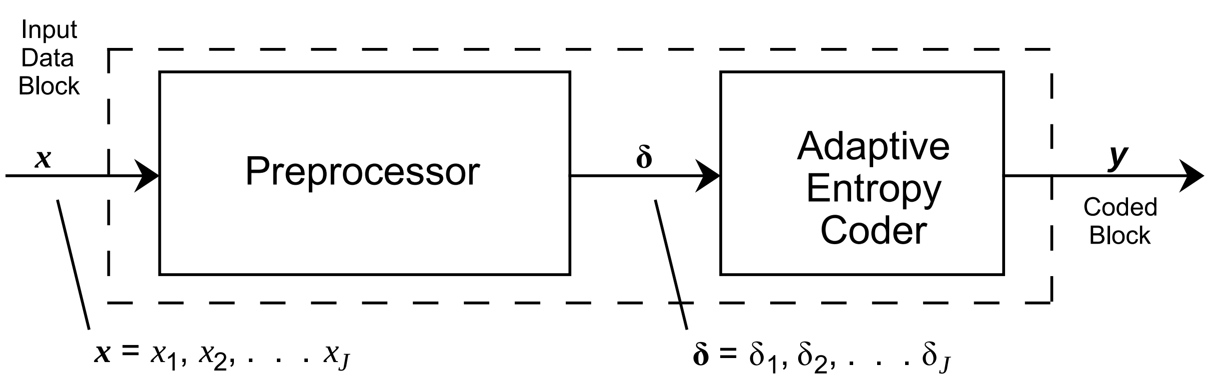


Figure 2-1: Schematic of the Source Coder

The inputs to the source coder are

$$\boldsymbol{x} = x_1, x_2, \dots x_J$$

which is a block of J n-bit samples, where n is a constant value.

Preprocessor:

The preprocessor applies a reversible function to input data samples $\boldsymbol{x}$, to produce a 'preferred source':

$$\boldsymbol{\delta} = \delta_1, \delta_2, \dots \delta_i, \dots \delta_J$$

where each δ_i is an n-bit integer, $0 \le \delta_i \le (2^n-1)$. For an ideal preprocessing stage, $\boldsymbol{\delta}$ will have the following properties:

a) the $\{\delta_i\}$ is statistically independent and identically distributed;

b) the preferred probability, p_m, that any sample δ_i will take on integer value m is a nonincreasing function of value m, for $m = 0, 1, \dots (2^n-1)$.

The preprocessor function is a reversible operation, and, in general, the best Lossless preprocessor will meet the above conditions and produce the lowest entropy, which is a measure of the smallest average number of bits that can be used to represent each sample.

This Recommended Standard does not attempt to explain methods for choosing a preprocessing stage. This Recommended Standard does provide the definition of a basic preprocessing stage that may be suitable for many applications. However, it is important that the user carefully address this issue since careful selection of an appropriate preprocessing stage is essential for efficient compression and depends on the source-data characteristics. Interested users should refer to reference [C2].

Adaptive Entropy Coder:

The function of the Adaptive Entropy Coder is to calculate uniquely decipherable, variable-length codewords corresponding to each block of samples input from the preprocessor. The entropy coder incorporates multiple coding options, each exhibiting efficient performance over different yet overlapping ranges of entropy. The coder selects the coding option that gives the highest compression ratio among the various options on the same block of J samples. A code-option 'identifier', requiring only a few bits, is attached before the first codeword bit in a coded block to signal the coding option to the decoder for proper decompression. Since the block size J can be small and a new code option is selected for each block, the overall coding can adapt to rapid changes in data statistics.

2.3 PACKETIZATION OF CODED DATA

The variable-length encoded bit stream representing a *J*-sample block forms a Coded Data Set (CDS). CCSDS telemetry source packet structure is recommended to transport the CDSes, which will be contained in the source data field of the packet. The information related to, for example, the sensor, mission, time, and other mission-specific details necessary for the routing and accounting of the packets, will be contained in the Packet Primary Header and (if present) in the Packet Secondary Header (see reference [2]).

2.4 ERROR CONTROL

Individual channel bit errors have greater consequences when data are compressed. Even then, the consequences need not be catastrophic. For this reason, to limit error propagation when utilizing the source coding algorithm described in this document, the following is recommended:

a) use telemetry channel coding as described in reference [1];

b) use packetized telemetry as described in reference [2].

3 ADAPTIVE ENTROPY CODER

3.1 CODE SPECIFICATION

3.1.1 Figure 3-1 represents the general-purpose Adaptive Entropy Coder with a preprocessor. Basically, such a coder chooses one of a set of code options to use to represent an incoming block of preprocessed data samples, **δ**. A unique identifier (ID) bit sequence is attached to the code block to indicate to the decoder which decoding option to use.

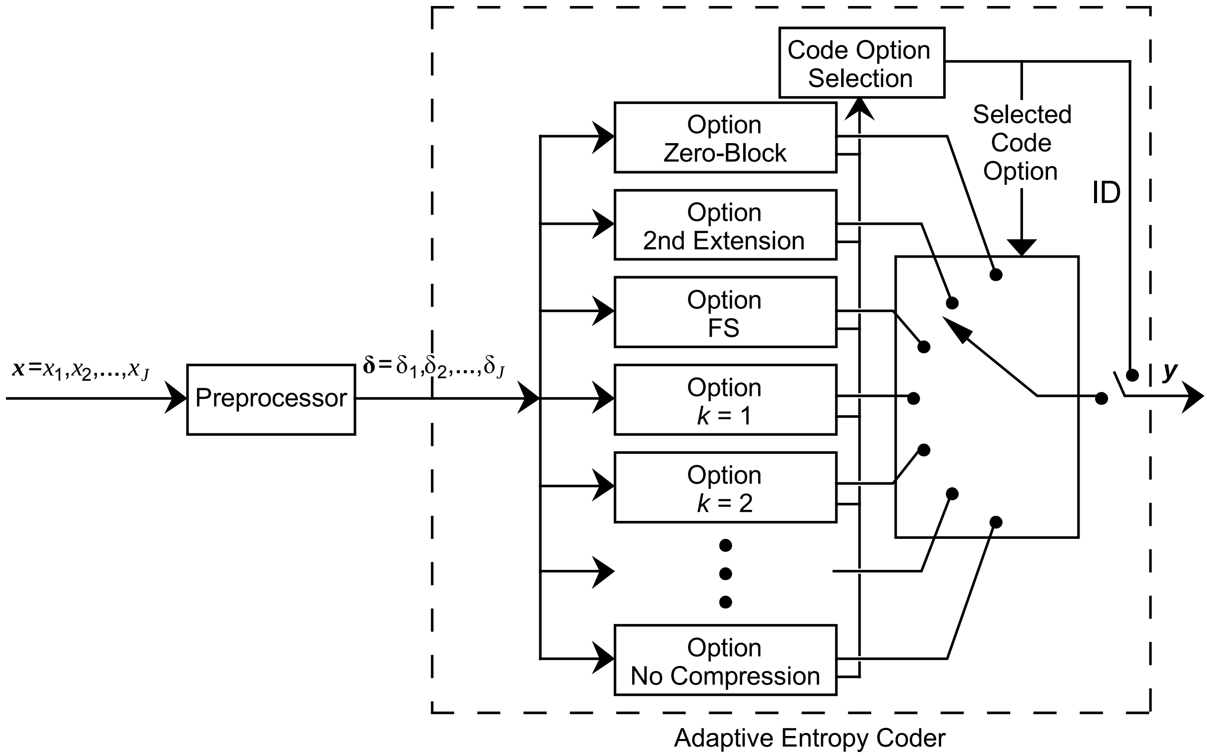


Figure 3-1: The Adaptive Entropy Coder with a Preprocessor

NOTE – Figure 3-1 illustrates the principle of the Adaptive Entropy Coder with a preprocessor; it does not illustrate an implementation.

3.1.2 The basic code selected is a variable-length code that utilizes Rice's adaptive coding technique (refer to reference [C2]). In Rice's coding technique, several algorithms are concurrently applied to a block of J consecutive preprocessed samples. The algorithm option that yields the shortest encoded length for the current block of data is selected for transmission. The zero-block option is a special case in that a single codeword sequence represents one or more consecutive blocks of J preprocessed samples (see 3.4.3). In all other options, the codeword sequence represents a single block of J consecutive preprocessed samples.

3.1.3 The following variables are required by Rice's adaptive coding technique:

- block size, J;
- resolution, n (number of input bits/sample);
- the ID bit sequence of the selected code option.

3.1.4 The following constraints shall apply to the Entropy Coder's variable-length adaptive coding scheme:

J = 8, 16, 32, or 64 samples per block;

n = resolution with a maximum of 32 bits per sample with digital signal values from 0 to 2^n-1, or from -2^{n-1} to 2^{n-1}-1.

3.2 FUNDAMENTAL SEQUENCE

The most basic option is a variable-length Fundamental Sequence (FS) codeword, which consists of m zeros followed by a one when preprocessed sample $\delta_i = m$. Table 3-1 illustrates the FS codewords. A Fundamental Sequence is the concatenation of J FS codewords.

Table 3-1: Fundamental Sequence Codewords As a Function of the Preprocessed Samples

Preprocessed Sample Values, δ_i	FS Codeword
0	1
1	01
2	001
.	.
.	.
.	.
2^n-1	0000 … 00001 (2^n-1 zeros)

3.3 SAMPLE SPLITTING

3.3.1 The kth split-sample option is obtained by removing the k least-significant bits (LSBs) from the binary representation of each preprocessed sample, δ_i, and encoding the remaining bits with an FS codeword (see figure 3-2). This produces a varying codeword length. The FS codewords for the current block of *J* preprocessed samples are transmitted along with the removed LSBs, preceded by an ID field indicating the value of *k*. This process enables the adaptation of codeword length to source-data statistics.

3.3.2 The FS option described in 3.2 is a special case of sample splitting where $k = 0$.

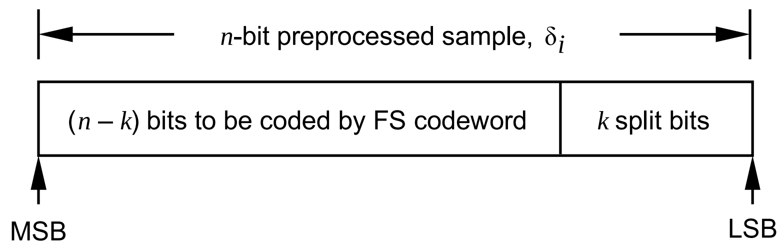


Figure 3-2: Split-Sample Format

3.4 LOW ENTROPY OPTIONS

3.4.1 GENERAL

Two code options, the Second-Extension option[1] and the Zero-Block option, provide more efficient coding than other options when the preprocessed data are highly compressible.

3.4.2 THE SECOND-EXTENSION OPTION

When the Second-Extension option is selected, each pair of preprocessed samples in a *J*-sample block is transformed and encoded using an FS codeword. A pair of consecutive samples δ_i and δ_{i+1} from a *J*-sample preprocessed data block are transformed into a single new symbol γ by the following equation.

$$\gamma = (\delta_i + \delta_{i+1}) (\delta_i + \delta_{i+1} + 1)/2 + \delta_{i+1}$$

The *J*/2 transformed symbols in a block are encoded using the FS codeword of table 3-1. The above process requires *J* to be an even integer which the recommended values in 3.1.4 obey (J = 8, 16, 32, or 64).

3.4.3 ZERO-BLOCK OPTION

3.4.3.1 The Zero-Block option is selected when one or more blocks of preprocessed samples are all zeros. In this case, a single codeword may represent several blocks of preprocessed samples, unlike other options where an FS codeword represents only one or two preprocessed samples.

[1] The first extension of a preprocessed sample is the preprocessed sample itself.

3.4.3.2 The set of *r* blocks between consecutive reference samples (or possibly fewer than *r* blocks at the end of the input sequence), as described in 4.3, is partitioned into one or more segments. Each segment, except possibly the last, contains *s* blocks. The recommended value of *s* is 64.

3.4.3.3 Within each segment, each group of adjacent all-zeros blocks is encoded by the FS codewords, specified in table 3-2, which identify the length of each group. The Remainder-Of-Segment (ROS) codeword in table 3-2 is used to denote that the remainder of a segment consists of five or more all-zeros blocks.

Table 3-2: Zero-Block Fundamental Sequence Codewords As a Function of the Number of Consecutive All-Zeros Blocks

Number of All-Zeros Blocks	FS Codeword
1	1
2	01
3	001
4	0001
ROS	00001
5	000001
6	0000001
7	00000001
8	000000001
.	.
.	.
.	.
63	0000 … 0000000001 (63 0s and a 1)

3.5 NO COMPRESSION

The last option is to not apply any data compression. If it is the selected option, the preprocessed block of samples receives an attached identification field but is otherwise unaltered.

3.6 CODE SELECTION

3.6.1 The Adaptive Entropy Coder includes a code selection function, which selects a coding option that minimizes the number of bits (including ID bits) used to encode the current block of samples. The ID bit sequence specifies which option was used to encode the accompanying block of samples. The ID bit sequences are shown in table 5-1.

3.6.2 When two or more coding options minimize the length of an encoded block, the option selected for the block should be chosen as follows:

a) the ‘no compression’ option should be chosen when it minimizes the encoded length for the block; otherwise

b) the Second-Extension option should be chosen when it minimizes the encoded length for the block; otherwise

c) the coding option having the smallest code parameter value k (where the FS option is treated as k=0) should be chosen.

4 PREPROCESSOR

4.1 PREPROCESSOR FUNCTION

4.1.1 Two of the factors contributing to the coded bit rate performance (in bits/sample) of this Lossless data compression technique are the amount of correlation removed among data samples in the preprocessing stage, and the coding efficiency of the entropy coder. The function of the preprocessor is to decorrelate data and reformat them into non-negative integers with the preferred probability distribution. There are situations when a preprocessor is not necessary (see reference [C2]), and may be bypassed to provide better compression performance.

4.1.2 A preprocessor contains two functions, prediction and mapping, as shown in figure 4-1. The preprocessor subtracts the predicted value, $\hat{x}_i$, from the current data value, x_i. The resultant (n+1)-bit prediction error, Δ_i, is then mapped to an n-bit integer value, δ_i, based on the predicted value, $\hat{x}_i$. When a predictor is properly chosen, the prediction error tends to be small, and for some sources, has a probability distribution approaching Laplacian, for which the Adaptive Entropy Coder is optimal. There are several preprocessing techniques, of which only one, the Unit-Delay Predictor as described in 4.2, is presented in this Recommended Standard (see reference [C2] for predictor examples).

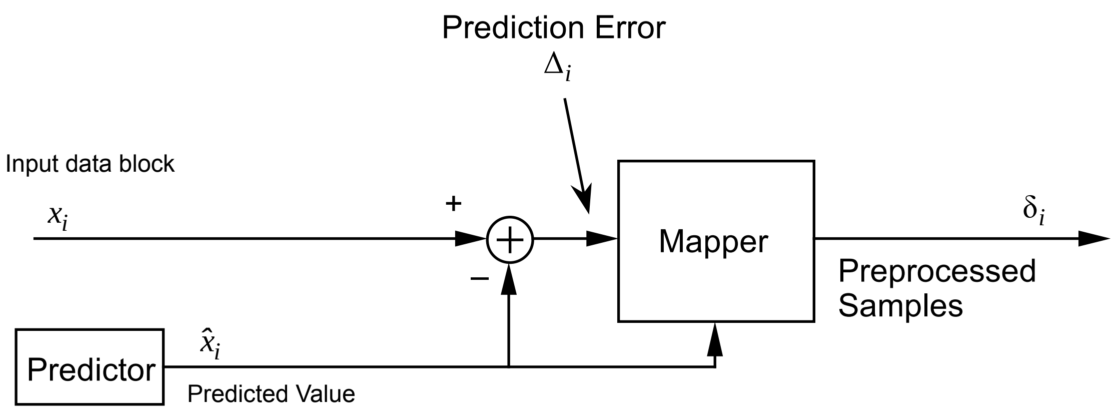


Figure 4-1: A Preprocessor

4.2 PREDICTORS

4.2.1 PREDICTION TECHNIQUES

Several preprocessing techniques can be used with the Adaptive Entropy Coder. One technique, using the Unit-Delay Predictor, is specified in 4.2.2 below. An application-specific predictor may be used instead of the unit-delay predictor, but such a predictor is unique and not specified in this Recommended Standard.

4.2.2 UNIT-DELAY PREDICTOR

The unit-delay prediction technique illustrated in figure 4-2 uses the one-sample delayed input data signal as the predictor for the current data signal, and the prediction error is passed to the following mapper along with the predicted value for mapping to a nonnegative integer.

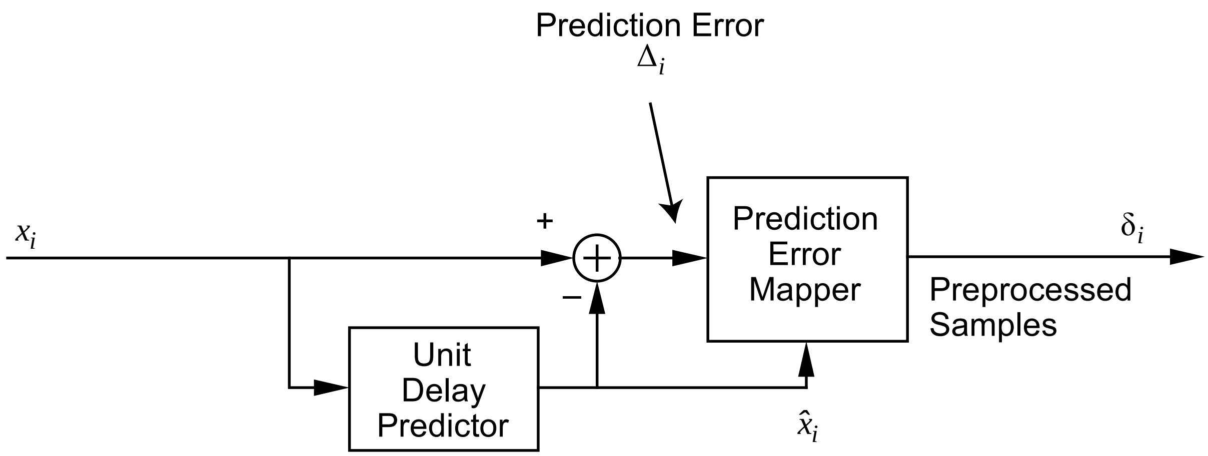


Figure 4-2: Preprocessor Using a Unit-Delay Predictor

4.3 REFERENCE SAMPLE

A reference sample is an unaltered input data sample upon which succeeding sample prediction is based. When a unit-delay predictor or other higher-order predictors that use the previous data signal in prediction are used, reference samples are required by the decoder to recover the sample values from decoded predictor errors. When the reference sample is inserted, there are $J – 1$ preprocessed samples in the CDS. The user must determine how often to insert references. The reference sample interval, r, is limited to a maximum value of 4096 CDSes (e.g., 262144 samples when $J = 64$). When a reference sample is not required by the preprocessor, the parameter r serves to define an interval of input data sample blocks that will be further segmented under the zero-block option described in 3.4.3.2.

4.4 PREDICTION ERROR MAPPER

The Prediction Error mapper takes the prediction error values and maps them into non-negative integers suitable for the Adaptive Entropy Coder. The prediction error Δ_i resulting from taking the difference between a signal value, x_i, and a predictor value, $\hat{x}_i$, both n-bit integers, will have an (n+1)-bit dynamic range of $[-2^n+1, 2^n-1]$. However, for every predictor value, there are only 2^n possible prediction error values. With a properly chosen predictor, the most probable prediction error value is zero, followed by +l and –1, +2 and –2, …, etc. The smallest prediction error value is the difference between the minimum signal value, x_{min}, and the predictor value, $\hat{x}_i$: $x_{min} - \hat{x}_i$. The largest prediction error value is the difference between the maximum signal value, x_{max}, and the predictor value, $\hat{x}_i$: $x_{max} - \hat{x}_i$. To map the possible 2^n prediction error values into non-negative integers, the following equation is used:

$$\delta_i = \begin{cases} 2\Delta_i & 0 \le \Delta_i \le \theta_i \\ 2\,|\Delta_i| - 1 & -\theta_i \le \Delta_i < 0 \\ \theta_i + |\Delta_i| & \text{otherwise} \end{cases}$$

where

$$\theta_i = minimum(\hat{x}_i - x_{min}, x_{max} - \hat{x}_i);$$

and for signed *n*-bit signal value,

$$x_{min} = -2^{n-1}, \; x_{max} = 2^{n-1} - 1;$$

for non-negative *n*-bit signal value,

$$x_{min} = 0, \; x_{max} = 2^n - 1.$$

5 DATA FORMAT

5.1 LOSSLESS DATA STRUCTURES

5.1.1 GENERAL

Several parameters are required in order to transfer the adaptive variable-length losslessly coded data between the coder and the telemetry channel packet formatter.

5.1.2 OPTION IDENTIFICATION

5.1.2.1 Users shall choose to use either the Basic or Restricted set of code options. When sample resolution $n \leq 4$, the use of the Restricted set of code options reduces the number of available coding options, thus allowing the use of shorter ID bit sequences.

5.1.2.2 The ID Field specifies which of the options was used for the accompanying set of samples. The ID-code keys for each of the options are shown in table 5-1.

Table 5-1: Selected Code Option Identification Key

Code Option	**Resolution**					
	Basic:	**–**	**–**	$n \leq 8$	$8 < n \leq 16$	$16 < n \leq 32$
	Restricted:	$n = 1, 2$	$n = 3, 4$	$4 < n \leq 8$	$8 < n \leq 16$	$16 < n \leq 32$
Zero Block		00	000	0000	00000	000000
Second Extension		01	001	0001	00001	000001
FS		—	01	001	0001	00001
k=1		—	10	010	0010	00010
k=2		—	—	011	0011	00011
k=3		—	—	100	0100	00100
k=4		—	—	101	0101	00101
k=5		—	—	110	0110	00110
k=6		—	—	—	0111	00111
k=7		—	—	—	1000	01000
k=8		—	—	—	1001	01001
k=9		—	—	—	1010	01010
k=10		—	—	—	1011	01011
k=11		—	—	—	1100	01100
k=12		—	—	—	1101	01101
k=13		—	—	—	1110	01110
k=14		—	—	—	—	01111
k=15		—	—	—	—	10000
⋮		⋮	⋮	⋮	⋮	⋮
k=29		—	—	—	—	11110
no compression		1	11	111	1111	11111
NOTE – ‘ – ’ indicates no applicable value						

5.1.2.3 For applications not requiring the full entropy range of performance provided by the specified options, a subset of the options at the source may be implemented. The ID is always required, even if a subset of the options is used.

5.1.3 REFERENCE SAMPLE

When the preprocessor is present and the reference sample is required, the first CDS of the Source Packet Data Field shall contain a reference sample. References shall then be inserted in the Source Packet Data Field at least every 4096 CDSes as specified in 4.3. When the preprocessor is absent, or it does not require a reference sample, the reference sample shall not be inserted in the CDS.

5.1.4 CODED DATA SET FORMAT

5.1.4.1 The CDS format when a sample-splitting option is selected is shown in figure 5-1. Figure 5-1a shows the case where there is a reference sample; figure 5-1b shows the format when no reference sample is present. The CDS has the following structure when a sample-splitting option is selected: 1) ID bit sequence optionally followed by an *n*-bit reference sample, 2) compressed data, and 3) concatenated *k* least-significant bits from each sample.

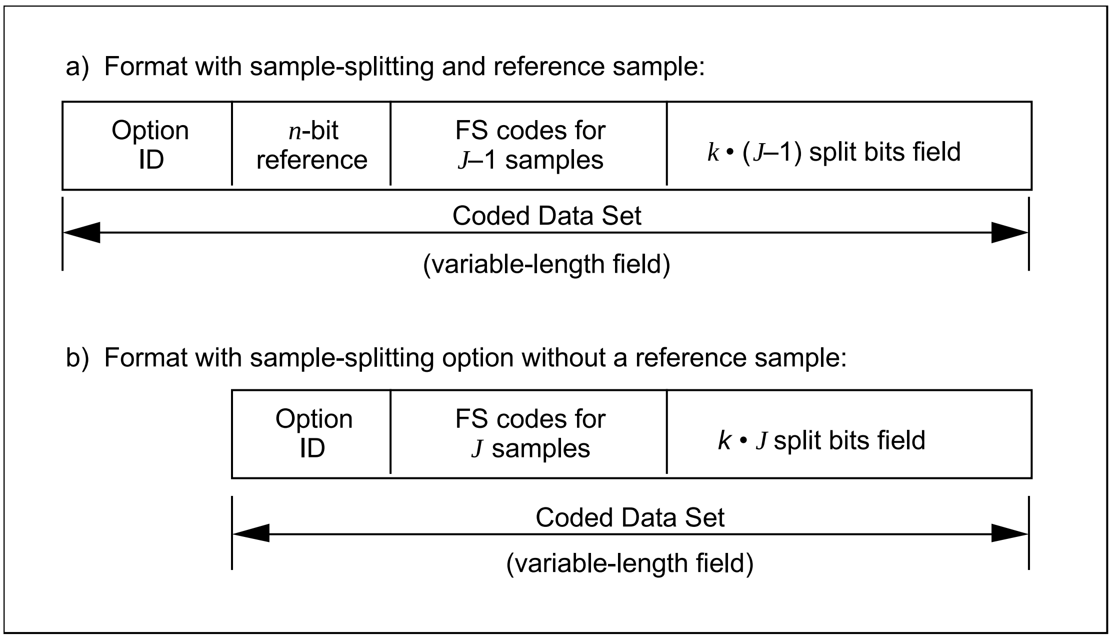


Figure 5-1: CDS Format When Sample-Splitting Option Is Selected

5.1.4.2 When the no-compression option is selected, the CDS is fixed length containing the option ID field, optionally followed by an *n*-bit reference sample, and *J* preprocessed samples. The case where a reference is present is shown in figure 5-2a; the non-reference case is shown in figure 5-2b.

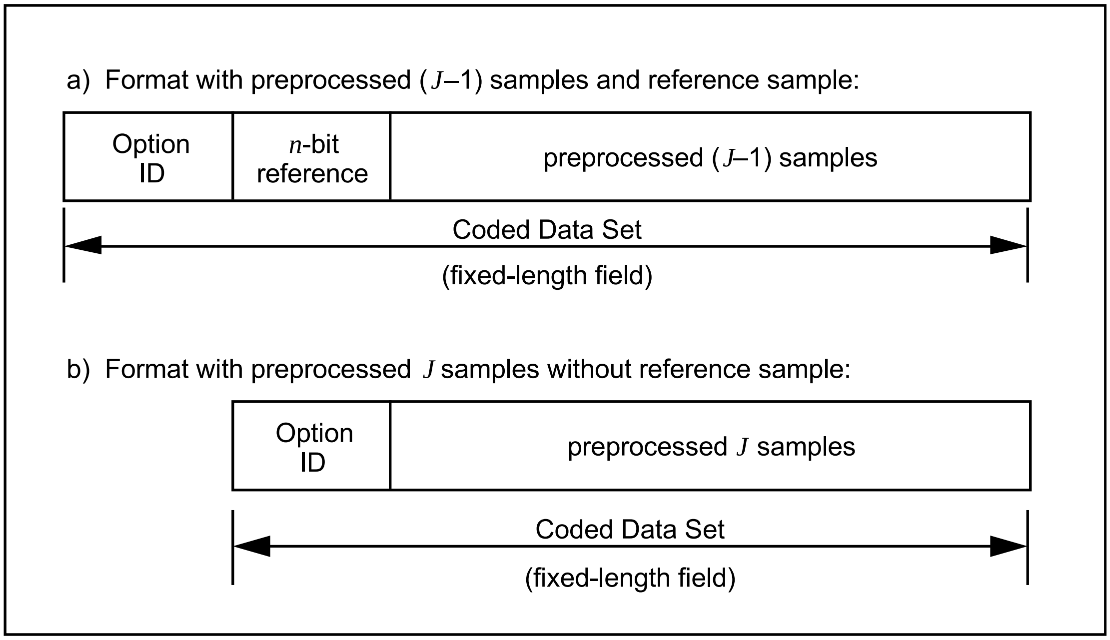


Figure 5-2: CDS Format When No-Compression Option Is Selected

5.1.4.3 When the Zero-Block option is selected, the CDS contains the option ID field, optionally followed by an *n*-bit reference sample, and a required FS codeword specifying the number of concatenated zero valued blocks or the ROS condition as described in 3.4.3. The case where a reference is present is shown in figure 5-3a; the non-reference case is shown in figure 5-3b.

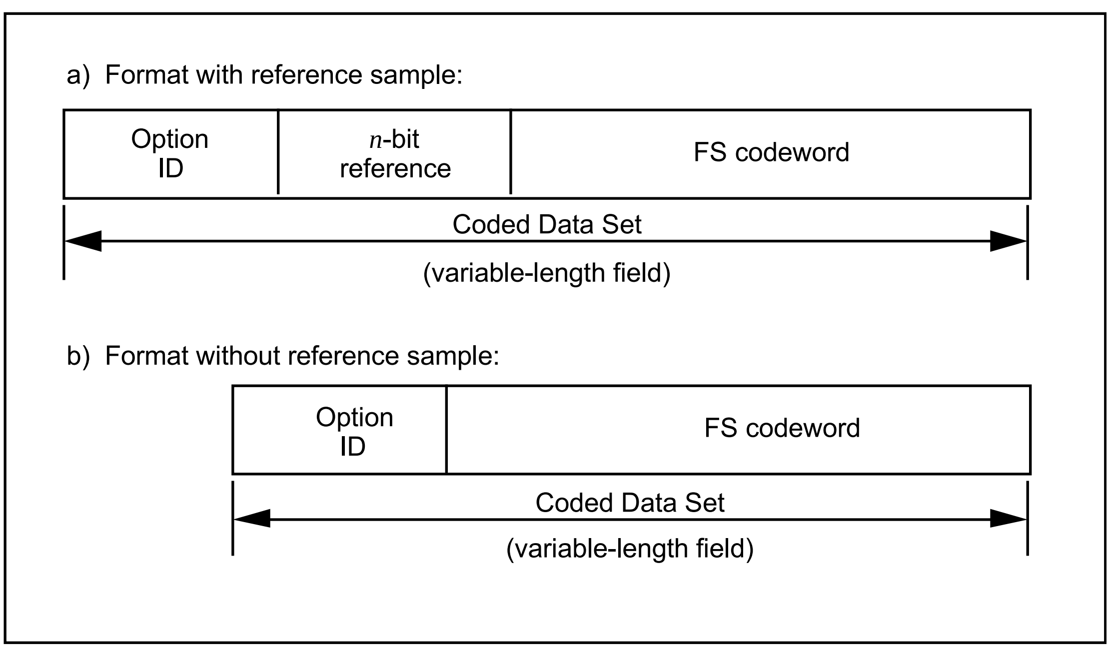


Figure 5-3: CDS Format When Zero-Block Option Is Selected

5.1.4.4 When the Second Extension option is selected, the CDS contains the option ID field, optionally followed by an n-bit reference sample, and required FS codewords for $\frac{J}{2}$ transformed samples. The case where a reference is present is shown in figure 5-4a; the non-reference case is shown in figure 5-4b. In the case when a reference is inserted, a '0' sample is added in front of the J–1 preprocessed samples, so $\frac{J}{2}$ samples are produced after the transformation.

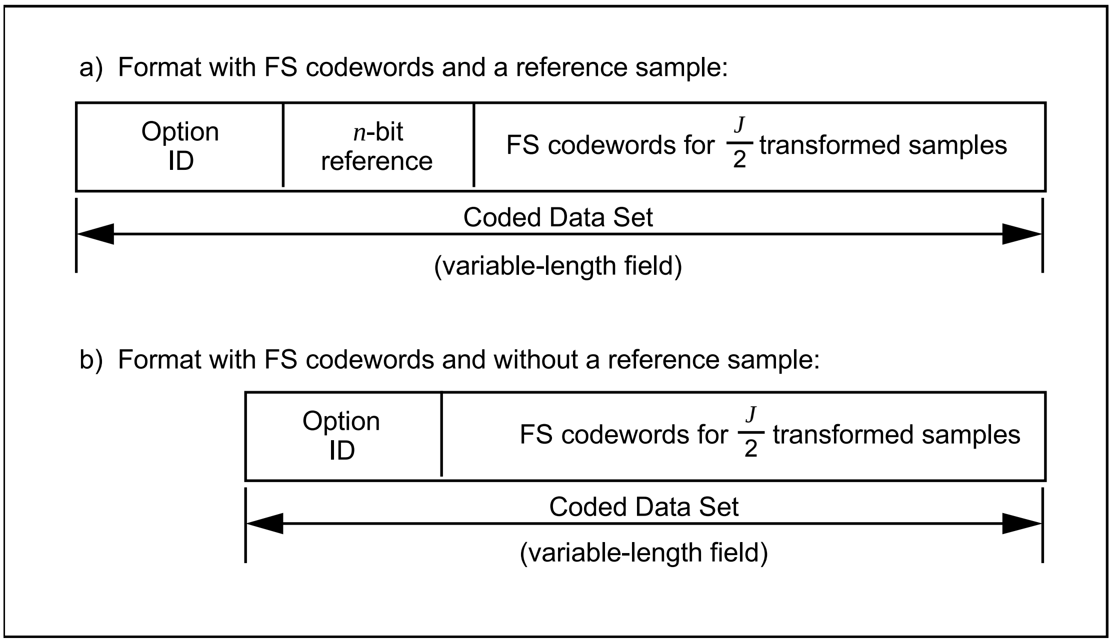


Figure 5-4: CDS Format When the Second-Extension Option Is Selected

5.2 PACKET FORMAT

5.2.1 LOSSLESS PACKET FORMAT

Lossless data compression packets shall be formatted as shown in figure 5-5 (see reference [2]). The packet formatter uses the parameter provided by the source data coder to form one or more CDSes to determine the packet size in bytes. Fill bits of zero value may be needed to force the packet to end on a byte boundary.

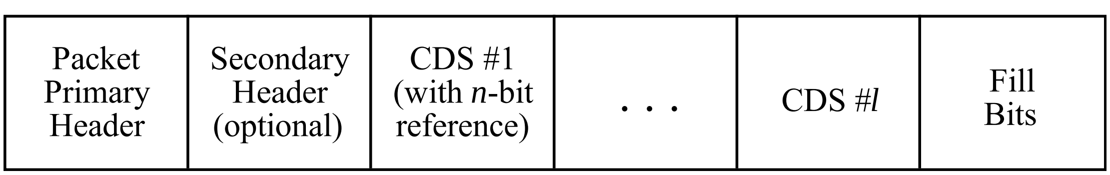


Figure 5-5: Packet Format for *l* CDSes

5.2.2 PACKET REQUIREMENTS

5.2.2.1 A Source Packet Data Field must meet the following requirements:

- a CDS within a packet must meet the format requirements defined in 5.1;
- when the reference sample is used, the Source Packet Data Field shall begin with a CDS that contains this reference, followed by one or more CDSes; when the reference sample is not required in the preprocessor, or the preprocessor is absent, a reference sample shall not be inserted in the first CDS in the Source Packet Data Field;
- several CDSes can be put in sequence within a source packet;
- fill bits are allowed only at the end of the Source Packet Data field, not within the body of compressed data;
- each packet must end on a byte boundary.

NOTE – Some implementations may require an adequate number of fill bits be added in order to end a packet on an even-numbered byte boundary.

5.2.2.2 Unless the option to use the CIP is chosen (see section 6), in order to decode packets that may include fill bits, several pieces of information must be communicated to the decoder a priori. This information will be mission specific and fixed for a given Application Process Identifier (APID) per mission:

- l, the number of CDSes that are in a packet;
- r, the reference sample interval, equaling the number of CDSes counted from one CDS containing a reference sample up to but not including the next consecutive CDS containing a reference sample;
- n, the resolution;
- J, the number of samples per block;
- whether the Basic or Restricted set of code options is used (when $n \leq 4$).

5.2.2.3 A Packet Secondary Header is optional and can be used, for example, to relate observation time and position information to the user (see reference [2]).

5.2.2.4 The use of the Sequence Flags in the Packet Sequence Control Field is optional and can be used, for example, to signal a group of compressed data packets. Their use is governed by reference [2].

6 COMPRESSION IDENTIFICATION PACKET (OPTIONAL)

6.1 COMPRESSION IDENTIFICATION PACKET STRUCTURE

6.1.1 When the compressed data are transmitted as grouped source packets, a Compression Identification Packet (CIP) is an optional packet that, if used, shall precede and provide configuration information for a group of compressed application data packets. The CIP will be transmitted from an application process in space to one or several sink processes on the ground.

6.1.2 The CIP shall be the first packet of the group.

6.1.3 The CIP shall consist of two major fields positioned contiguously in the following sequence: Packet Primary Header and the Packet Data Field. See figure 6-1.

6.1.4 The CIP shall contain information that would allow the decompressor to be automatically configured to acquire a group of compressed application data packets without the need for managing a-priori information. The CIP shall be utilized to configure the decompressor automatically only if there is a reliable system for file transfer.

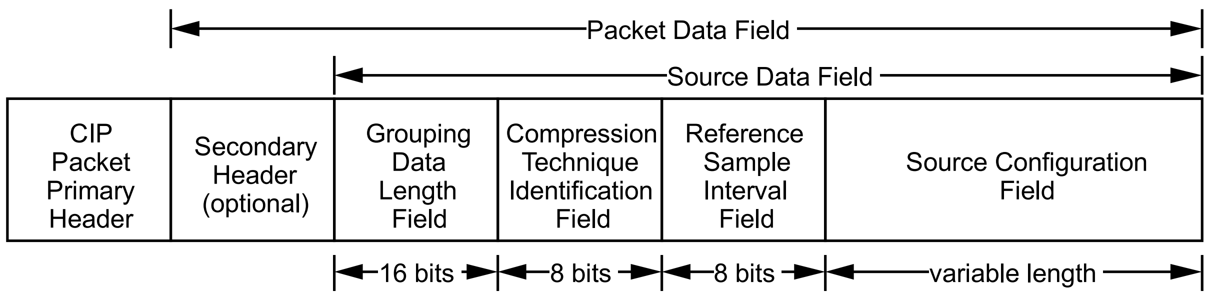


Figure 6-1: Compression Identification Packet Structure

6.2 CIP PRIMARY HEADER

6.2.1 GENERAL

The Packet Primary Header is mandatory for the CIP and its structure shall conform to the CCSDS Space Packet Protocol Blue Book, reference [2]. The CIP Packet Primary Header Field shall contain the source data APID. The use of the CIP will be mission specific and fixed for a given APID.

6.2.2 SEQUENCE FLAGS

6.2.2.1 The Sequence Flags are in the packet Sequence Control field, as specified in reference [2]. The field is located in the Packet Primary Header of packets encapsulating compressed user data. As indicated below, the field is always ‘01’ for the CIP Primary Header.

6.2.2.2 The Sequence Flags shall be set as follows:

– ‘01’ for the group’s first packet, which is the CIP;

– ‘00’ for the continuing source packets with compressed data of the group;

– ‘10’ for the last source packet with compressed data of the group.

6.2.2.3 For a source packet not belonging to a group of source packets with compressed data, the Sequence Flags shall be set to ‘11’.

6.3 PACKET DATA FIELD

6.3.1 GENERAL

The Packet Data Field of a CIP shall consist of two fields positioned contiguously in the following sequence: Packet Secondary Header (optional) and the Source Data Field.

6.3.2 SECONDARY HEADER (OPTIONAL)

The Secondary Header is a means for placing ancillary data such as time and spacecraft position/attitude information with the CIP.

6.3.3 SOURCE DATA FIELD

6.3.3.1 General

The Source Data Field for the CIP shall consist of four fields positioned contiguously in the following sequence:

	Length (bits)
The Grouping Data Length	16
Compression Technique Identification	8
Reference Sample Interval	8
Source Configuration	(Variable)

6.3.3.2 Grouping Data Length

The Grouping Data Length is a 16-bit field of which the first 4 bits are reserved. The remaining 12 bits of the field shall contain a binary number equal to the number of packets containing compressed data within the group minus one, with the number of packets containing compressed data ranging from 1 to 4096. The number of packets in the group with the CIP included shall range from 2 to 4097.

6.3.3.3 Compression Technique Identification Field

6.3.3.3.1 The Compression Technique Identification (CTI) field shall signal the compression technique in use for the group of source packets identified by the CIP.

6.3.3.3.2 When no compression technique for the current group is used, the CTI field shall be set to all zeros.

6.3.3.3.3 Only the Lossless data compression technique is currently defined, and is signaled by the value 1 in the CTI field. Other values are reserved for future use by CCSDS and are not permitted.

6.3.3.4 Reference Sample Interval Field

6.3.3.4.1 The reference sample interval, r, equals the number of CDSes counted from one CDS containing a reference sample up to but not including the next consecutive CDS containing a reference sample. When the preprocessor is absent, or it does not require a reference sample, the reference sample shall not be inserted in the CDS; nevertheless, parameter r serves to define the interval of input data sample blocks for the zero-block option as described in 4.3.

6.3.3.4.2 The 8-bit Reference Sample Interval field shall contain a binary number equal to $(r–1) \bmod 256$. That is, this field encodes the modulus of $(r–1)$ with respect to divisor 256.

6.3.3.5 Source Configuration

6.3.3.5.1 Subfield Partitions

6.3.3.5.1.1 The Source Configuration field shall be partitioned into four subfields, which should appear in the following order: Preprocessor, Entropy Coder, Extended Parameters, and Instrument Configuration (see figure 6-2). The Preprocessor and Entropy Coder subfields are required, whereas the Instrument Configuration subfield is optional. The Extended Parameters subfield is required whenever any of the following conditions hold:

a) the block length J satisfies $J>16$;

b) the reference sample interval r satisfies $r>256$;

c) the Restricted set of code options is used (see 5.1.2.1);

If none of the above conditions holds, then the Extended Parameters subfield shall not be included.

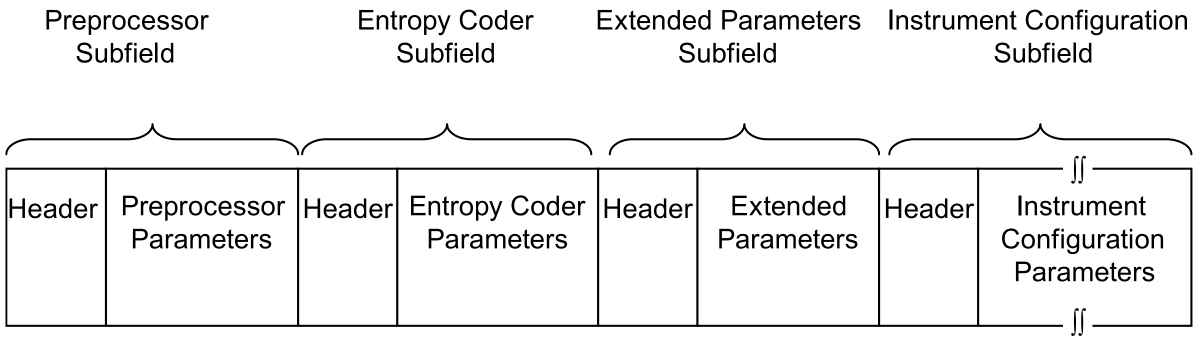


Figure 6-2: Source Configuration Field

6.3.3.5.1.2 Each subfield of the Source Configuration field shall have a header as the first two bits to identify the subfield type. These subfield header bits shall be set as follows:

00 – Preprocessor
01 – Entropy Coder
10 – Instrument Configuration
11 – Extended Parameters

6.3.3.5.2 Preprocessor

6.3.3.5.2.1 The length of the Preprocessor subfield shall be two bytes, the first two bits of which shall be the header as described in 6.3.3.5.1.2.

6.3.3.5.2.2 The Preprocessor parameters for the Lossless data compressor shall be partitioned into six areas and shall be positioned contiguously following the 2-bit Preprocessor header. See 3.1 and section 4 for preprocessor parameter definitions. The six areas are

a) Preprocessor Status (1 bit)

   0 – absent
   1 – present

   The preprocessor status shall be set to '0' (absent) when the source coder is being used as the block-adaptive encoder defined in 5.4.3.3 of reference [3].

b) Predictor type (3 bits); ignore if preprocessor status is '0':

   000 – bypass predictor
   001 – unit delay predictor
   111 – application-specific predictor
   All other codes are reserved by CCSDS for future preprocessing options.

c) Mapper type (2 bits); ignore if preprocessor status is '0':

   00 – Prediction Error mapper described in 4.4

01 – reserved
10 – reserved
11 – application-specific mapper

d) Block size (*J*) (2 bits):

	Number of Samples/Block
00	J=8
01	J=16
10	J=32 or J=64
11	application specific

NOTE – When J=32 or J=64, the value of J is encoded in the Extended Parameters subfield.

e) Data sense (l bit):

0 – two's complement
1 – positive; mandatory if preprocessor is bypassed

f) Input data sample resolution (*n*) (5 bits):

The 5-bit Input Data Sample field shall contain a binary number equal to the input data sample resolution minus one, with the data sample resolution ranging from 1 to 32.

6.3.3.5.3 Entropy Coder

6.3.3.5.3.1 The length of the Entropy Coder subfield shall be two bytes, the first two bits of which shall be the header as described in 6.3.3.5.1.2.

6.3.3.5.3.2 The Entropy Coder parameters subfield shall be partitioned into two areas and shall be positioned contiguously following the 2-bit Entropy Coder header. The two areas are

a) Data resolution range (2 bits):

00 — Spare
01 — for $n \leq 8$
10 — for $8 < n \leq 16$
11 — for $16 < n \leq 32$

b) Number of CDSes per packet, *l* (12 bits):

The 12-bit field indicating the number of CDSes per packet (l) shall contain a binary number equal to $l - 1$.

6.3.3.5.4 Instrument Configuration

The Instrument Configuration Subfield (ICS) is an instrument-unique field used to address unique instrument configuration parameters. The contents of this field are mission specific and are beyond the scope of this Recommended Standard. If used, the first two bits shall be the header as specified in 6.3.3.5.1.2.

6.3.3.5.5 Extended Parameters

6.3.3.5.5.1 The length of the Extended Parameters subfield shall be two bytes, the first two bits of which shall be the header as described in 6.3.3.5.1.2.

6.3.3.5.5.2 The Extended Parameters subfield shall be partitioned into six areas. The six areas are:

a) Reserved (2 bits):

   These two bits shall be set to '00'.

b) Block size (J) (4 bits):

	Number of Samples/Block
0000	J=8
0001	J=16
0010	J=32
0011	J=64
1111	Application Specific

   Other values are reserved.

c) Reserved (1bit):

   This bit shall be set to '0'.

d) Restricted code options flag (see 5.1.2.1) (1 bit):

   0 – Basic set of code options are used;

   1 – Restricted set of code options are used.

e) Reserved (2 bits):

   These two bits shall be set to '00'.

f) Reference sample interval extension (4 bits):

   This field shall encode the value $\lfloor (r-1)/256 \rfloor$. That is, the largest integer less than or equal to $(r–1)/256$ shall be encoded.

CCSDS HISTORICAL DOCUMENT

ANNEX A

SECURITY, SANA, AND PATENT CONSIDERATIONS

(INFORMATIVE)

A1 SECURITY CONSIDERATIONS

A1.1 SECURITY BACKGROUND

It is assumed that security is provided by encryption, authentication methods, and access control to be performed at the application and/or transport layers. Mission and service providers are expected to select from recommended security methods suitable to the specific application profile. Specification of these security methods and other security provisions is outside the scope of this Recommended Standard.

A1.2 SECURITY CONCERNS

Security concerns in the areas of data privacy, integrity, authentication, access control, availability of resources, and auditing are to be addressed in the appropriate layers and are not related to this Recommended Standard. The use of lossless data compression does not affect the proper functioning of methods used to achieve such protection.

The use of lossless data compression slightly improves data integrity because the alteration of even a single bit of compressed data is likely to cause conspicuous and easily detectible corruption of the reconstructed data, thus making it more likely that malicious data alteration will be detected.

A1.3 POTENTIAL THREATS AND ATTACK SCENARIOS

An eavesdropper will not be able to decompress compressed data if proper encryption is performed at a lower layer.

A1.4 CONSEQUENCES OF NOT APPLYING SECURITY

There are no specific security measures prescribed for compressed data. Therefore consequences of not applying security are only imputable to the lack of proper security measures in other layers.

A2 SANA CONSIDERATIONS

The recommendations of this document do not require any action from SANA.

A3 PATENT CONSIDERATIONS

Implementers should be aware that the method for coding low entropy data specified in this Recommended Standard is covered by U.S. Patent 5448642. Potential user agencies should direct their requests for licenses to the U.S. Patent 5448642 patent rights holder, whose contact information is:

Bryan A. Geurts
Department Office of Patent Counsel
NASA Goddard Space Flight Center
Code 140.1
8800 Greenbelt Road, Greenbelt, MD 20771
Phone: +1 301 286 7352
Fax: +1 301 286 9502
E-mail: bryan.a.geurts@nasa.gov

CCSDS HISTORICAL DOCUMENT

ANNEX B

GLOSSARY OF ACRONYMS AND TERMS

(INFORMATIVE)

B1 PURPOSE

This annex defines key acronyms and terms that are used throughout this Recommended Standard to describe source coding for data compression.

B2 ACRONYMS

Acronym	Definition
CDS	coded data set
CIP	compression identification packet
CTI	compression technique identification
FS	fundamental sequence
ICS	instrument configuration subfield
LSB	least significant bit
MSB	most significant bit
ROS	remainder of segment
SANA	Space Assigned Numbers Authority

B3 TERMS

ADAPTIVE ENTROPY CODER: An entropy coder codes the source samples with uniquely decodable codewords that, upon decoding, reconstruct the source samples. With an Adaptive Entropy Coder, the average codeword length also follows closely the information content of the source.

ENTROPY: Entropy is a quantitative measure of the average amount of information per source sample, expressed in bits/sample.

FUNDAMENTAL SEQUENCE: The variable-length Fundamental Sequence (FS) code represents the non-negative integer m with a binary codeword of m zeros followed by a 1. Application of the FS code to a block of J samples produces a sequence of J concatenated codewords called the Fundamental Sequence.

RICE'S ADAPTIVE CODING: The basic Rice adaptive coding algorithm chooses the best of several code options to use on a block of data. These options are targeted to be efficient over different ranges of data activity. The options are implemented using a combination of FS coding and the splitting of preprocessed samples into their most-significant and least-significant bit parts.

SAMPLE SPLITTING: Sample splitting is a procedure for separating the binary representation of a sample into two groups of adjacent bits, one for lower-order bits, the other for higher-order bits.

SPLIT BITS: Split bits are the lower-order bits separated by sample splitting from the binary representation of a sample.

ANNEX C

INFORMATIVE REFERENCES

(INFORMATIVE)

[C1] *Organization and Processes for the Consultative Committee for Space Data Systems*. Space Data System Standards, CCSDS A02.1-Y-3. Yellow Book. Issue 3. Washington, D.C.: CCSDS, July 2011.

[C2] *Lossless Data Compression*. Report Concerning Space Data System Standards, CCSDS 120.0-G-2. Green Book. Issue 2. Washington, D.C.: CCSDS, December 2006.