

Szip Compression in HDF Products

Please be aware that SZIP compression has been replaced by AEC compression as of HDF5-1.10.7. AEC is the open source version of SZIP.

The AEC library can be obtained from the [German Climate Computing Center](https://gitlab.dkrz.de/k202009/libaec) from:

<https://gitlab.dkrz.de/k202009/libaec>

Szip compression software, providing lossless compression of scientific data, has been provided with HDF software products as of HDF5 Release 1.6.0 and HDF4 Release 2.0.

Szip is an implementation of the extended-Rice lossless compression algorithm. The Consultative Committee on Space Data Systems (CCSDS) has adopted the extended-Rice algorithm for international standards for space applications[1,6,7]. Szip is reported to provide fast and effective compression, specifically for the EOS data generated by the NASA Earth Observatory System (EOS)[1]. It was originally developed at University of New Mexico (UNM) and integrated with HDF4 by UNM researchers and developers.

As the graphs to the right illustrate, the primary gain with Szip compression is in speed of processing. Szip also provides some advantage in compression ratio over other compression methods shown here. These results, including the data presented in the graphs below, are from tests conducted by Pen-Shu Yeh, et al. [1], with the HDF4 Szip integration.

Szip and HDF5

Using Szip compression in HDF5: Szip is a stand-alone library that is configured as an optional filter in HDF5. Depending on which Szip library is used (encoder enabled or decode-only), an HDF5 application can create, write, and read datasets compressed with Szip compression, or can only read datasets compressed with Szip.

Applications use Szip by setting Szip as an optional filter when a dataset is created. If the Szip encoder is enabled with the HDF5 library, data is automatically compressed and decompressed with Szip during I/O. If only the decoder is present, the HDF5 library cannot create and write Szip-compressed datasets, but it automatically decompresses Szip-compressed data when data is read.

See this [sample HDF5 program](#) for an illustration of the use of Szip compression with HDF5.

Details of the required and optional parameters are provided in the [H5Pset_szip](#) entry in the HDF5 Reference Manual.

Software distribution: Starting with Release 1.6.0, HDF5 has been distributed with Szip enabled, making it easier to use Szip compression. The software is distributed as follows:

- Pre-compiled HDF5 binaries are provided with the Szip encoder enabled.
- Szip source code will be distributed by The HDF Group.
- Users who cannot legally use Szip encoding should download and install the decode-only Szip library.

Szip and HDF4

Using Szip compression in HDF4: Szip is a stand-alone library that is configured as an optional filter in HDF4. To use Szip with HDF4, the Szip library must be downloaded and the HDF4 library must be configured and compiled with Szip support and with the Szip library.

Applications use Szip by setting Szip as an optional filter when a dataset is created. Once enabled, data is automatically compressed and decompressed with Szip during I/O.

See this [sample HDF4 program](#) for an illustration of the use of Szip compression with HDF4.

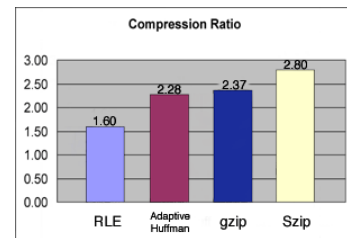
Details of the required and optional parameters are provided in the [SDsetcompress](#), [SDgetcompress](#), [SDsetchunk](#), [SDgetchunkinfo](#), and [HCget_config_info](#) entries in the HDF4 Reference Manual.

Software distribution: HDF4 is distributed with Szip enabled, making it easier to use Szip compression:

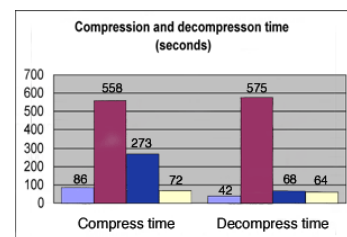
- Szip will be enabled and pre-configured in pre-compiled HDF4 binaries.
- Users who cannot legally use Szip encoding should download and build the HDF4 library

To Obtain: SZIP Source Code

Comparison of compression performance in tests with HDF4



Size of compressed output [1]



Speed of compression operation [1]

with the decode-only Szip library.

Licensing terms

The version of Szip distributed with HDF products is free for non-commercial use, which may occur in two sets of circumstances:

- Non-commercial users may use the Szip software integrated with HDF products to both encode (compress) and decode (uncompress) data. This applies to educational and research applications.
- Commercial users may use the software to decode any data. Further, they may use the software in internal activities that do not involve or result in the development of an Szip-based software product.

Commercial licenses are available for commercial users who wish to distribute an Szip-based software product or engage in commercial uses that are not allowed above. For further licensing information or to view a copy of the Szip copyright statement, see [Commercial use terms and the copyright and license notice](#) pertaining to Szip in HDF products.

Further information

See the following materials for further information:

- To obtain: [SZIP Source Code](#)
- HDF5: The [H5Pset_szip](#) entry in the HDF5 Reference Manual for HDF5 Szip usage information
- The [SDsetcompress](#), [SDgetcompress](#), [SDsetchunk](#), [SDgetchunkinfo](#), and [HCget_config_info](#) entries in the HDF4 Reference Manual for HDF4 Szip usage information
- Examples of Szip usage in [HDF5](#) and in [HDF4](#)
- An "Szip Performance Study" (PDF)
- [Commercial use terms and the copyright and license notice](#) pertaining to Szip in HDF products
- The [HDF/Szip collaborative agreement](#) (PDF)
- References to publications discussing Szip or its use. (below)

References

1. Pen-Shu Yeh, Wei Xia-Serafino, Lowell Miles, Ben Kobler, Daniel Menasce, "Implementation of CCSDS Lossless Data Compression in HDF," *Earth Science Technology Conference-2002*, 11-13 June 2002, Pasadena, California.
Paper: [http://esto.nasa.gov/conferences/estc-2002/Papers/A3P2\(Yeh\).pdf](http://esto.nasa.gov/conferences/estc-2002/Papers/A3P2(Yeh).pdf)
Conference: <http://esto.nasa.gov/conferences/estc-2002/>
2. J. Venbrux, P.S. Yeh, G. Zweigle, J. Vesel, "A VLSI Chip Solution for Lossless Medical Imagery Compression," *SPIE Conference on Medical Imaging 1994*. Vol. 2164, pp. 561-572, February 13-14, 1994, Newport Beach, California.
3. J. Venbrux, J. Gambles, D. Wiseman, G. Zweigle, W.H. Miller, P.S. Yeh, "A VLSI Chip Set Development for Lossless Data Compression," *AIAA Computing in Aerospace 9 Conference*. October 1993, San Diego, California.
4. J. Venbrux, G. Zweigle, J. Gambles, D. Wiseman, W. Miller, P. Yeh, "An Adaptive, Lossless Data Compression Algorithm and VLSI Implementations," *NASA Symposium on VLSI Design*. Pp 1.2.1-1.2.16, November 1993.
5. J. Venbrux, P.S. Yeh, and M. N. Liu, "A VLSI Chip Set for High Speed Lossless Data Compression," *IEEE Transactions on Circuits and Systems for Video Technology*. Pp. 381-391, December 1992.
6. CCSDS 120.0-G-1. *Lossless Data Compression*. Green Book. Issue 1, May 1997. This Report presents a summary of the key operational concepts and rationale underlying the requirements for the CCSDS Recommendation, Lossless Data Compression. Supporting performance information along with illustrations are also included. This Report also provides a broad tutorial overview of the CCSDS Lossless Data Compression algorithm and is aimed at helping first-time readers to understand the Recommendation.
Appears In: CCSDS Publications TGannett 02/04/2003 387K Services Check Out View
7. CCSDS 121.0-B-1. *Lossless Data Compression*. Blue Book. Issue 1, May 1997. This Recommendation defines a source-coding data-compression algorithm and specifies how data compressed using the algorithm are inserted into source packets for retrieval and decoding.
Appears In: CCSDS Publications