EAGE

EAGE DIGITAL [2020]

FIRST EAGE DIGITALIZATION CONFERENCE AND EXHIBITION

30 NOV - 3 DEC 2020 I VIENNA I AUSTRIA

SPONSORED BY

OMV

WesternGeco

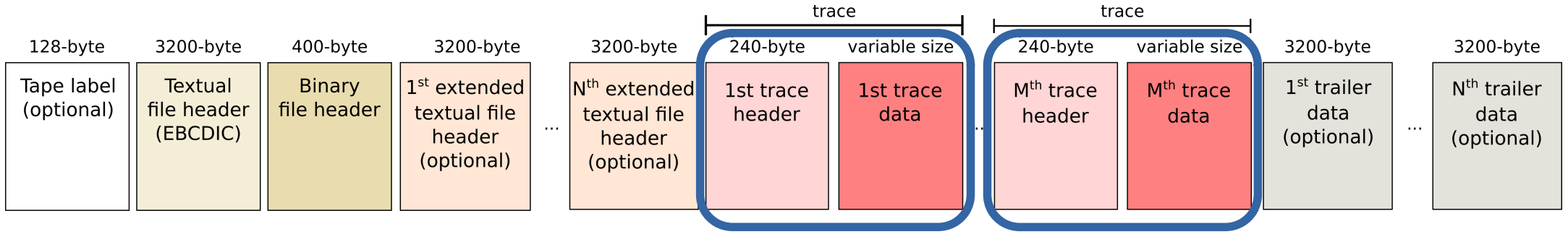
EAGE

Seamless Translation of Modern File Formats to SEG-Y through the File System Interface

Lucas C. Villa Real
Maximilien de Bayser

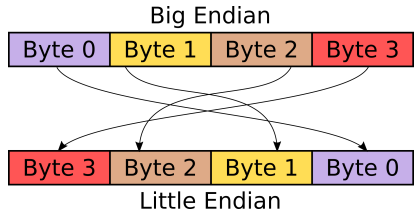
IBM Research

SEG-Y is an industry standard that dates back to 1975



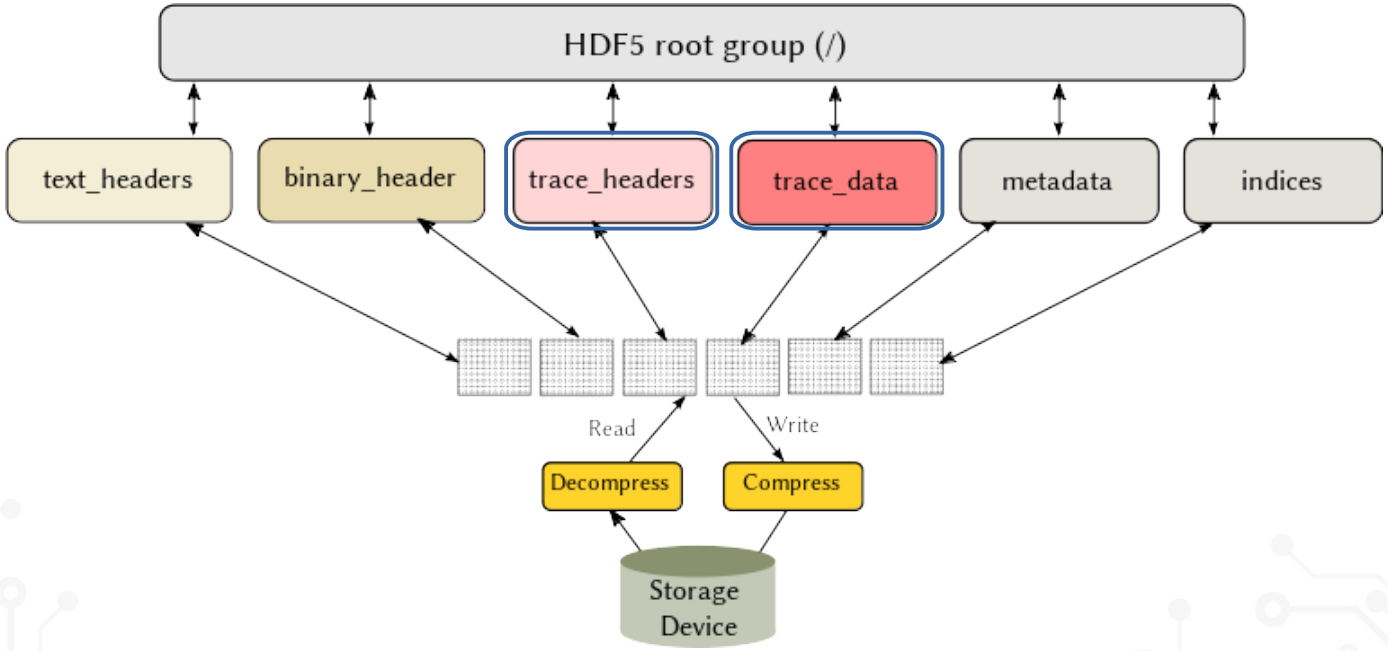
Even though it is widely adopted, it is inneficient in many ways:

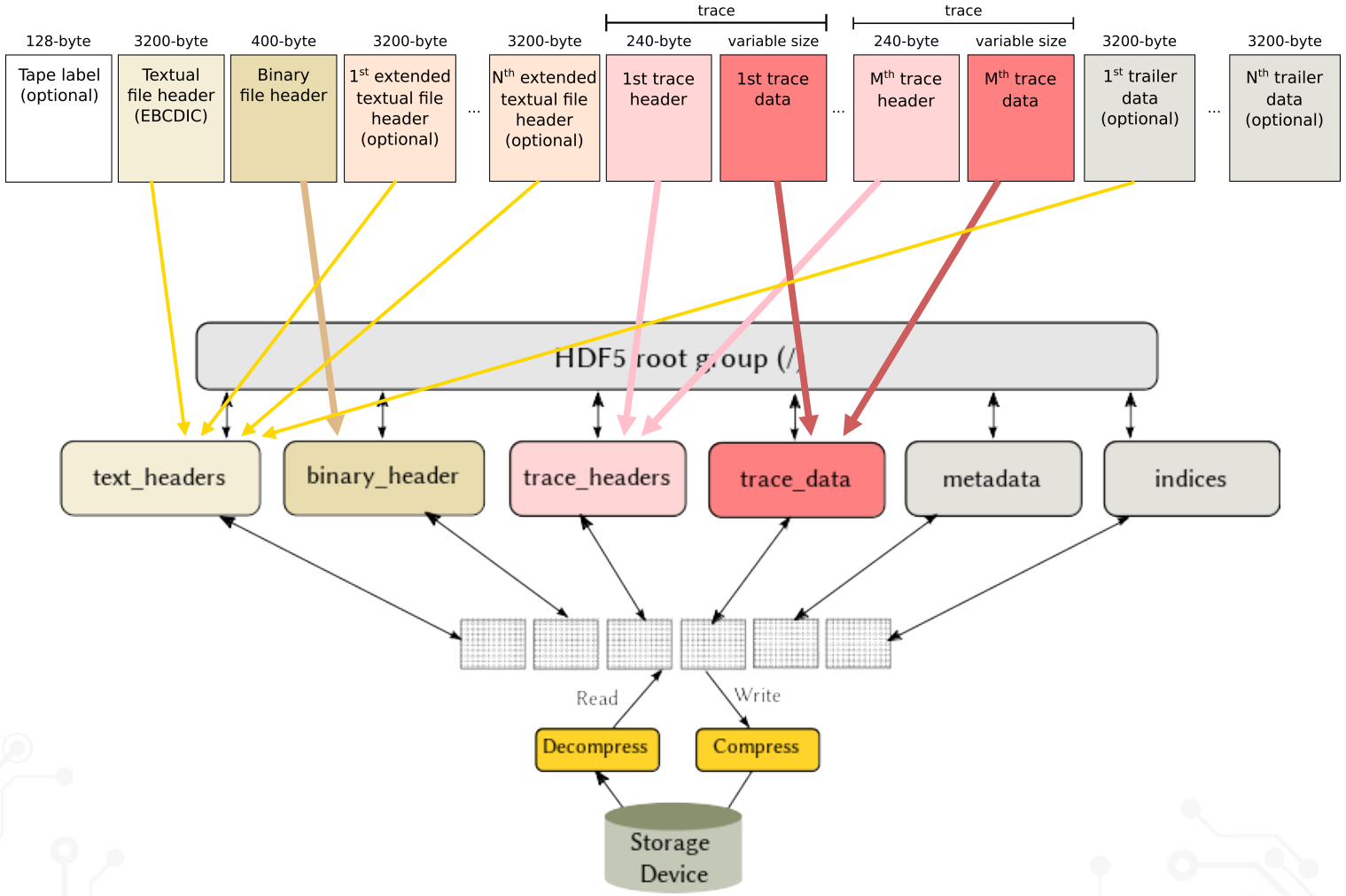
- Trace headers and data are interleaved on disk
- Big Endian byte ordering (prior to rev 2.0)
- No support for data compression
- No spatial indexes
- Important metadata are optional, thus are often filled up with zeros



In-house seismic file format based on HDF5

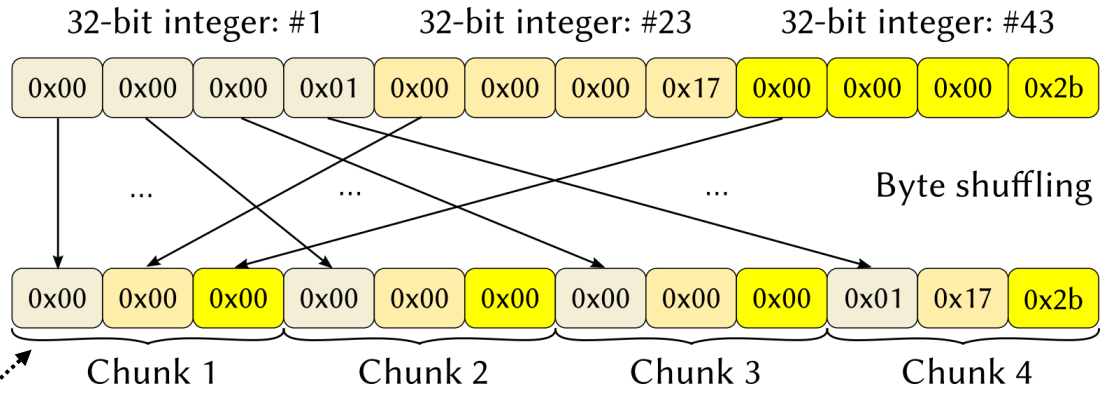
- 1:1 mapping to SEG-Y (both datasets and metadata)
- Traces are stored contiguously on disk on fixed-sized blocks
- Supports optional indexing of metadata fields and compression

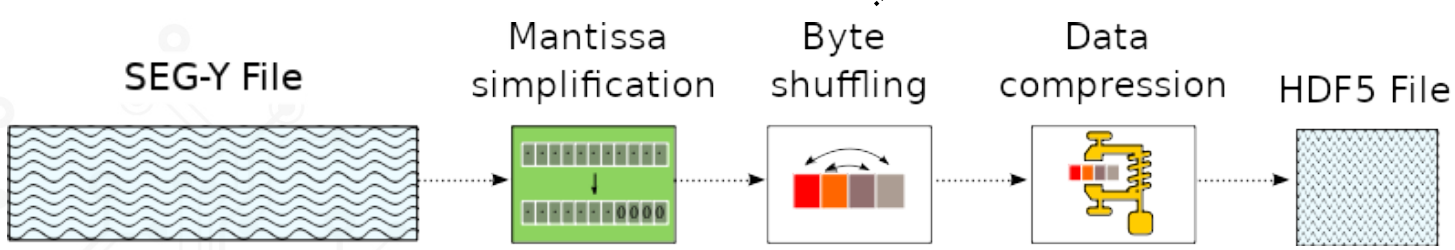



trace
trace
128-byte
3200-byte
400-byte
3200-byte
3200-byte
240-byte
variable size
240-byte
variable size
3200-byte
3200-byte
Tape label (optional)
Textual file header (EBCDIC)
Binary file header
1st extended textual file header (optional)
...
Nth extended textual file header (optional)
1st trace header
1st trace data
...
Mth trace header
Mth trace data
1st trailer data (optional)
...
Nth trailer data (optional)
HDF5 root group (/)
text_headers
binary_header
trace_headers
trace_data
metadata
indices
Read
Write
Decompress
Compress
Storage Device

Transformation pipeline (optional steps):

- Reduction of floating point precision
- Byte/Bit shuffling
- Compression of traces and headers
- Parallel I/O

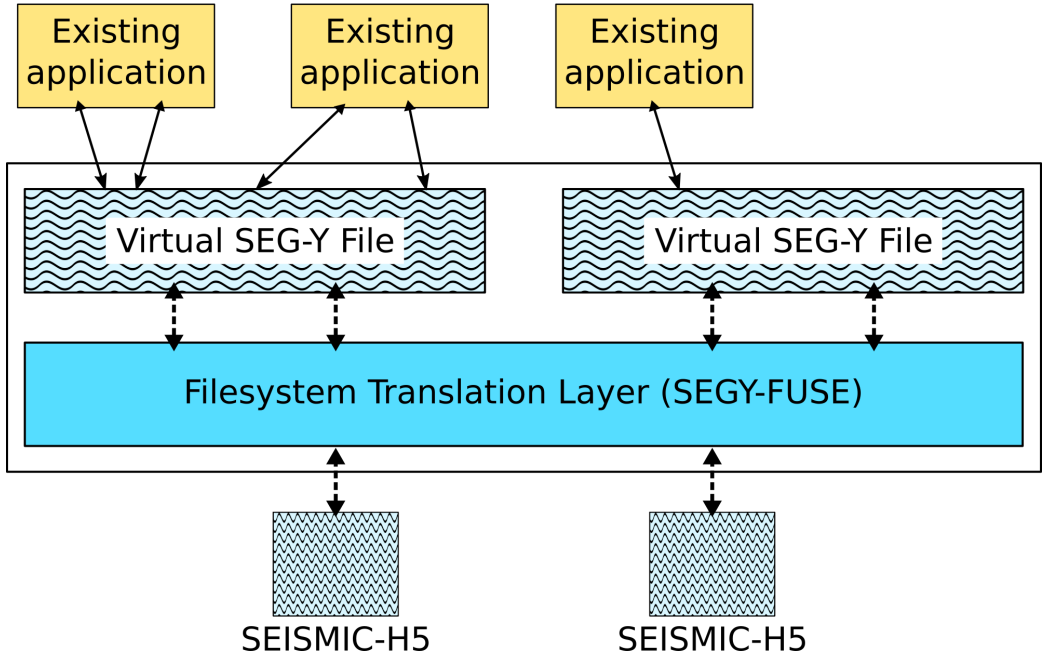




EAGE

SEGY-FUSE

- Disguises SEISMIC-H5 files as SEG-Y
- We know the original SEG-Y file size and the size of each structure, so mapping SEG-Y ↔ SEISMIC-H5 is done in constant time
- Legacy programs are made consumers of a modern file format with no change to their code



EAGE

PERFORMANCE CONSIDERATIONS

- *SEISMIC-H5 is very efficient*

Reading mode	File size	Time (seconds)	Observations
SEG-Y	5.8 GB	7.6 ± 0.9	–
Compressed SEG-Y (LZ4)	5.7 GB	23.0 ± 1.45	3x slower
SeismicH5 (LZ4)	5.1 GB	0.27 ± 0.06	3% of original time
SEGY-FUSE (LZ4)	5.1 GB	6.75 ± 0.66	Faster than original

I/O over virtualized SEG-Y is faster than over the original file

EAGE

Thank you

Lucas C. Villa Real (lucasvr@br.ibm.com)
Maximilien de Bayser (mbayser@br.ibm.com)

IBM Research