CTA RAO investigation

Vladimír Bahyl

IT Department

February 2021

Agenda

What is RAO?

Spectra Logic TAOS

CERN RAO

Tests and Comparison

Conclusion

What is RAO?

Recommended access order (RAO):

A feature of the 3592E07 and later drives is the ability to accept a list of User Data Segments (see 4.27.1.1) and reorder those User Data Segments into a recommended access order that minimizes the time to read those User Data Segments.

A User Data Segment (UDS) is defined as a grouping of contiguous logical objects (i.e., logical blocks and filemarks) and is described by partition number, beginning logical object identifier, and ending logical object identifier.
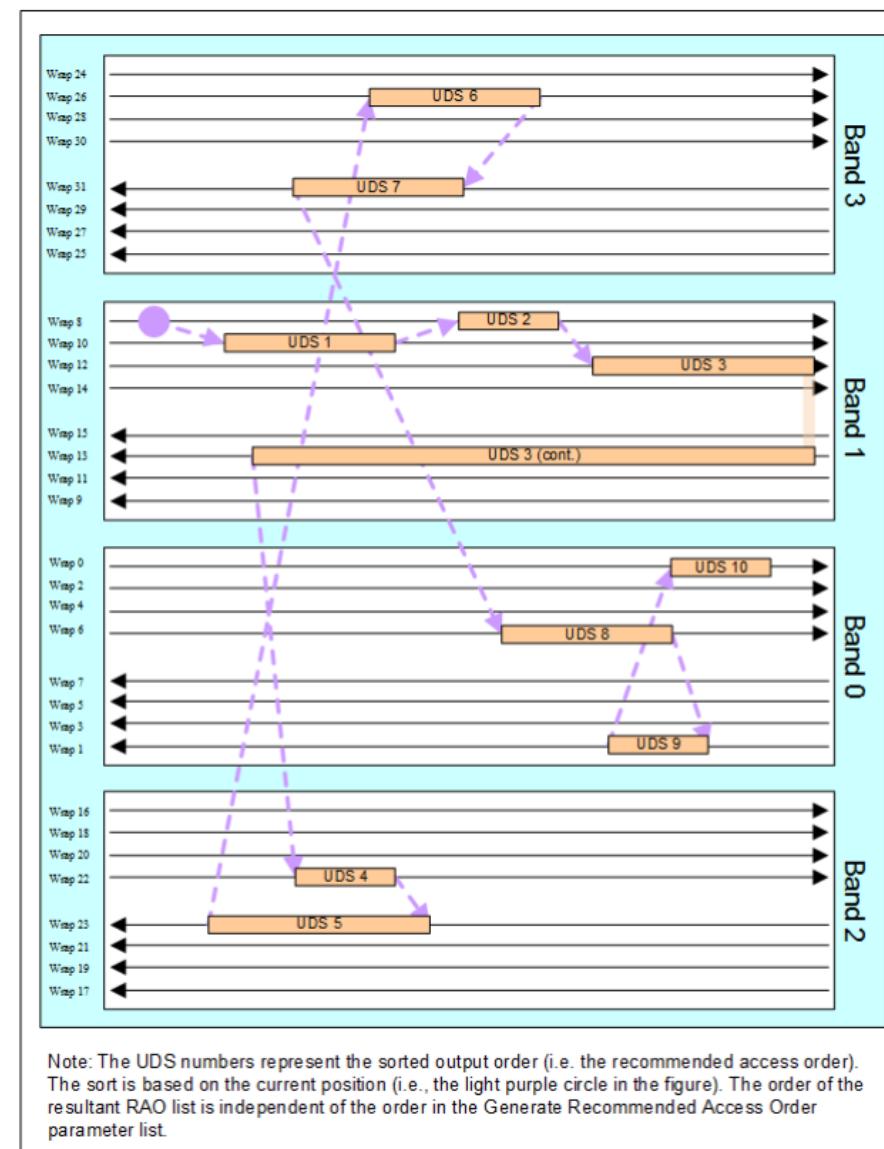

Figure 12 — Example Reordering of UDS's
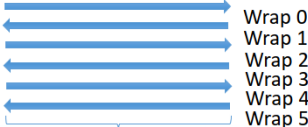
Spectra Logic TAOS

Time-Based Access Order System (TAOS™)

- https://spectralogic.com/wp-content/uploads/white_paper_increasing_file_read_throughput_on_tape_with_taos.pdf
- https://www.fujifilmsummit.com/wp-content/uploads/2018/06/Spectra-Logic-Matt-Starr.pdf
- https://hpc.csiro.au/users/dmfug/Meeting_Jun2018/Presentations/Spectra_2018.pdf
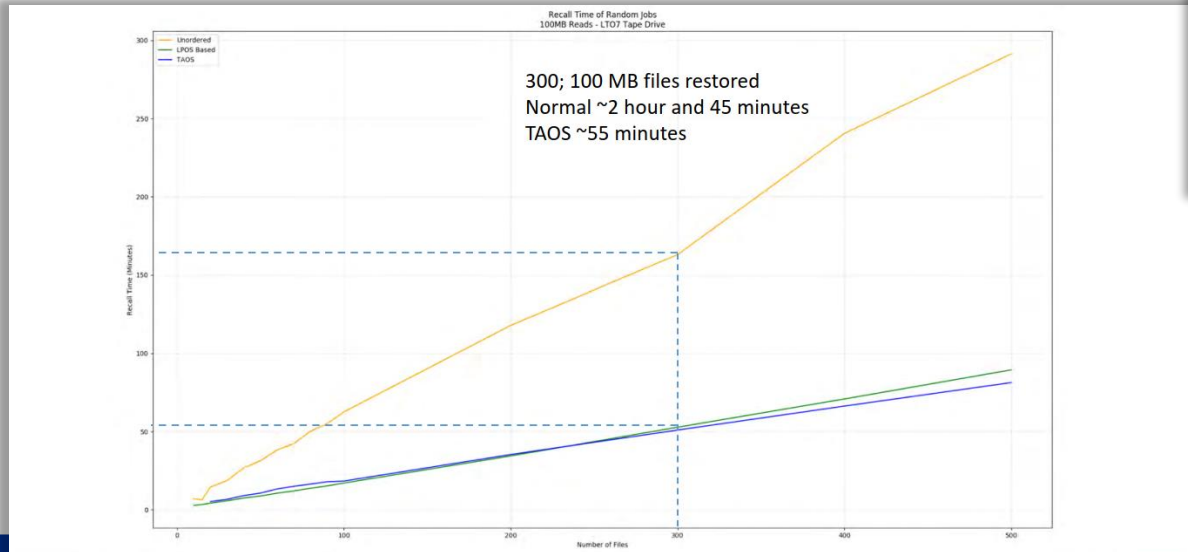
Media and Drive Wear Reduction Testing with TAOS

- A test was performed with 100 files with a file size between 1-100MB:
- TAOS Meters of Tape Across the Drive Head: 2,470
- Unordered Meters of Tape Across the Drive Head: 31,878
- That's a ~13x reduction in meters of tape. Or 8.4% of the original meters of tape.
- Spectra currently estimates that TAOS will reduce tape and drive wear by ~10X on media read operations when used in conjunction with HPSS 7.5 or later
- Releasing with major archive packages
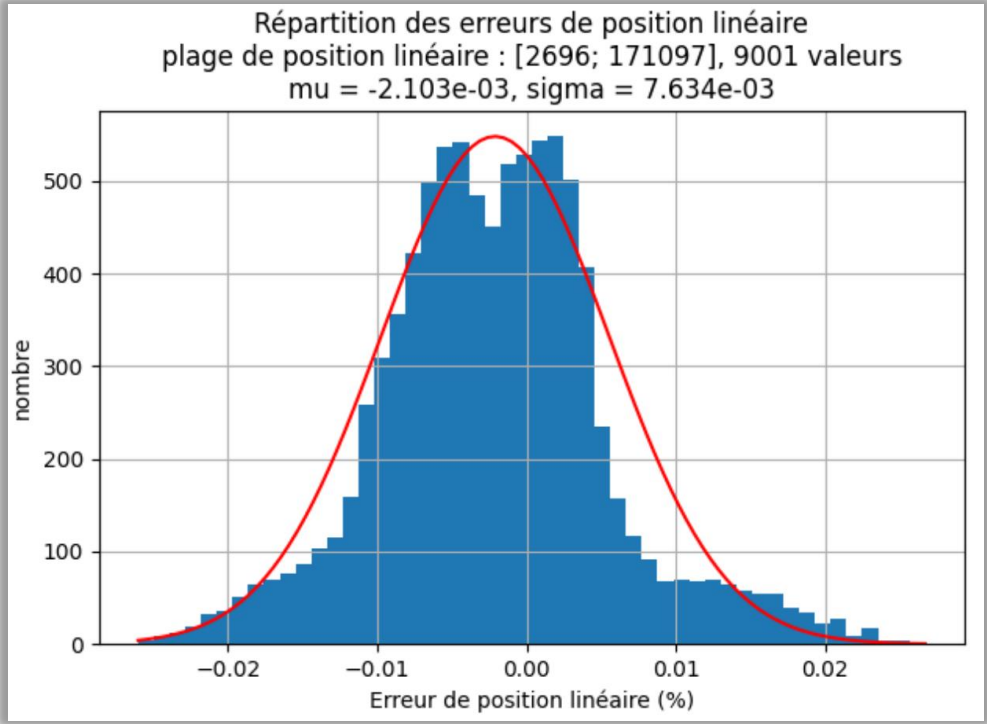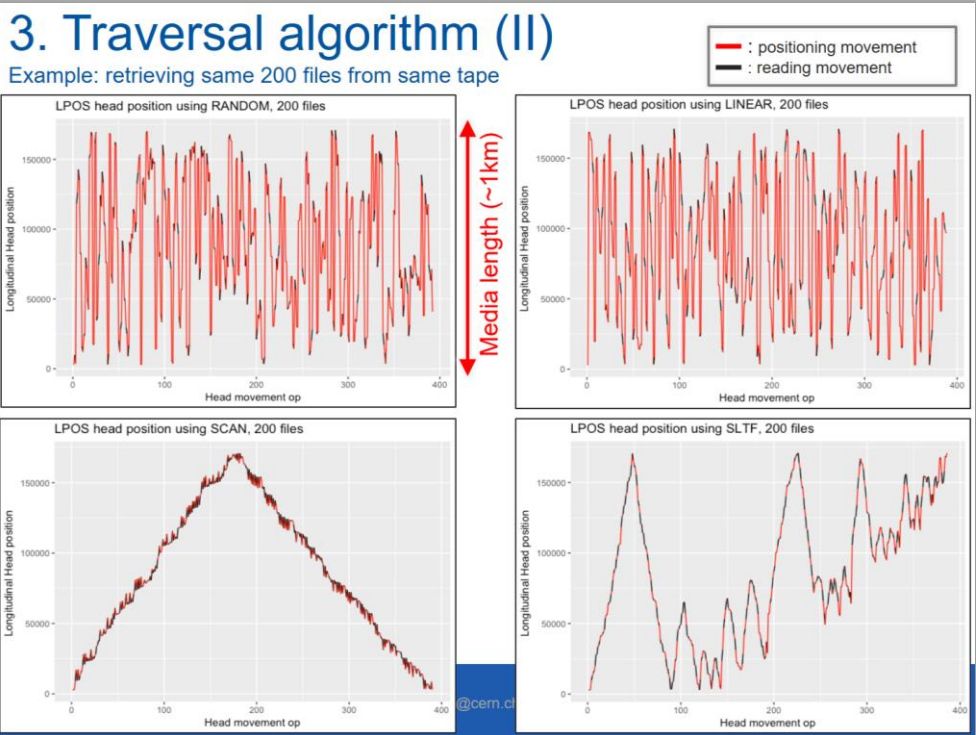

300; 100 MB files restored
Normal ~2 hour and 45 minutes
TAOS ~55 minutes

CERN SCSI REOWP RAO

Germán Cancio Meliá

*   https://indico.cern.ch/event/730908/contributions/3153156/

Paul Vriet

*   http://cds.cern.ch/record/2728741/

CERN SCSI REOWP RAO

Based on estimating the block position within the wrap
Works reasonably well with incompressible files
Needs further enhancements if not at BOT

```
Example of sg_raw -r 10k /dev/sg2 a3 1f 45 2 00 00 ff ff ff ff 00 00 output:

SCSI Status: Good

Received 1996 bytes of data:
 00     07 ca 00 00 00 00 00 00  00 00 00 00 00 03 2a 9c    ..............*.
 10     00 01 00 00 00 00 00 00  00 06 57 22 00 02 00 00    ..........W"....
 20     00 00 00 00 00 09 85 e6  00 03 00 00 00 00 00 00    ................
 30     00 0c b4 3e 00 04 00 00  00 00 00 00 00 0f e0 e2    ...>............
 40     00 05 00 00 00 00 00 00  00 13 0f 47 00 06 00 00    ...........G....
 50     00 00 00 00 00 16 3e 02  00 07 00 00 00 00 00 00    ......>.........
 60     00 19 6c 43 00 08 00 00  00 00 00 00 00 1c 9b 39    ..lC...........9
 70     00 09 00 00 00 00 00 00  00 1f c9 9a 00 0a 00 00    ................
 80     00 00 00 00 00 22 f8 92  00 0b 00 00 00 00 00 00    ....."..........
 90     00 26 26 fa 00 0c 00 00  00 00 00 00 00 29 55 b8    .&&..........)U.
 a0     00 0d 00 00 00 00 00 00  00 2c 84 10 00 0e 00 00    .........,......
 b0     00 00 00 00 00 2f b2 f8  00 0f 00 00 00 00 00 00    ...../..........
 c0     00 32 e1 2b 00 10 00 00  00 00 00 00 00 36 10 19    .2.+.........6..
 d0     00 11 00 00 00 00 00 00  00 39 3e 80 00 12 00 00    .........9>.....
 e0     00 00 00 00 00 3c 6d 59  00 13 00 00 00 00 00 00    .....<mY........
 f0     00 3f 9b 92 00 14 00 00  00 00 00 00 00 42 ca 5f    .?...........B._
 100    00 15 00 00 00 00 00 00  00 45 f8 cc 00 16 00 00    .........E......
```

```
wrapId,endOfWrapBlockId
0,207516
1,415522
2,624102
3,832574
4,1040610
5,1249095
6,1457666
7,1666115
8,1874745
9,2083226
10,2291858
```

Description of RAO tests

Using LTO-8 tapes
- But the base research was done using LTO-7M cartridges

Two types of data:
1. Larger incompressible files
2. Many small highly compressible files

GRAO / RRAO TAOS queries and subsequent recalls always start from BOT
- `mt eject ; mt load ; mt rewind`

Increasing the number of selected random files from each tape:
- 50, 100, 250, 500, 1000, 1500, 2000 (TAOS results are only up to 500 files due to problems with my setup)
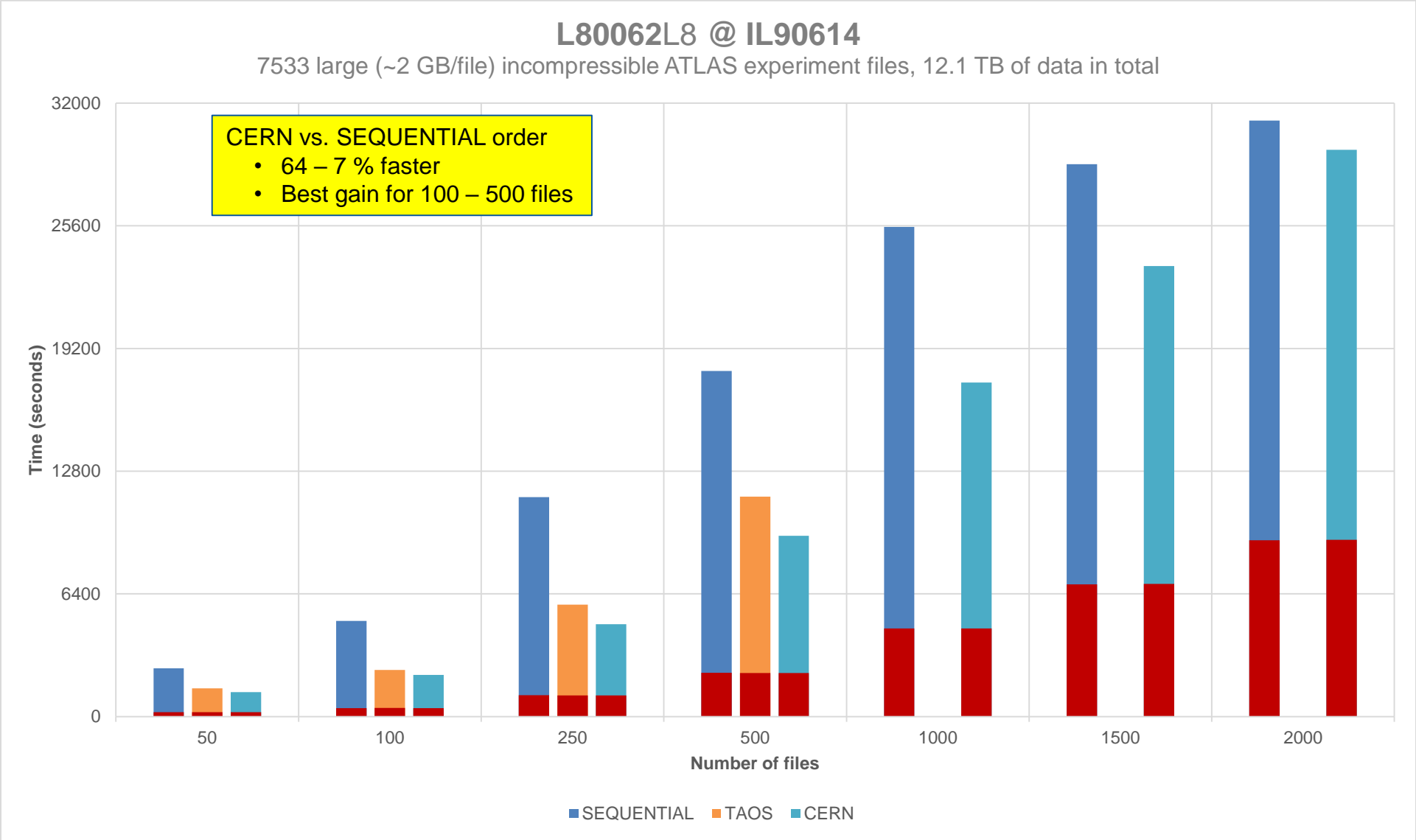- Using same list of files with each solution

Position to the 1st block of each file according to the given order:
- Sequential ascending, TAOS, CERN

Read the file until EOF mark

Measure total seek and read time (in seconds)

CTA RAO investigation   1/2

L80062L8 @ **IL90614**

7533 large (~2 GB/file) incompressible ATLAS experiment files, 12.1 TB of data in total

CERN vs. SEQUENTIAL order
- 64 – 7 % faster
- Best gain for 100 – 500 files

Time (seconds)

Number of files

■ SEQUENTIAL ■ TAOS ■ CERN

CTA RAO investigation

L80088L8 **@ IL90613**

167944 small (~200 MB/file) highly compressible nTOF experiment files, 34.5 TB of data in total

CERN vs. SEQUENTIAL order
• 62 – 18 % faster
• Best gain for 100 – 500 files

Time (seconds) vs. Number of files

Legend: SEQUENTIAL, TAOS, CERN

Conclusion

Several solutions now exist to increase the reading performance of LTO.

CTA has successfully implemented RAO for drives that do not support it natively and read performance improvements have been demonstrated.

When recalling less than 500 files from tape, CTA RAO solution can decrease the positioning time by 50% or more.