GPU compressor exploration

Stefan Rua

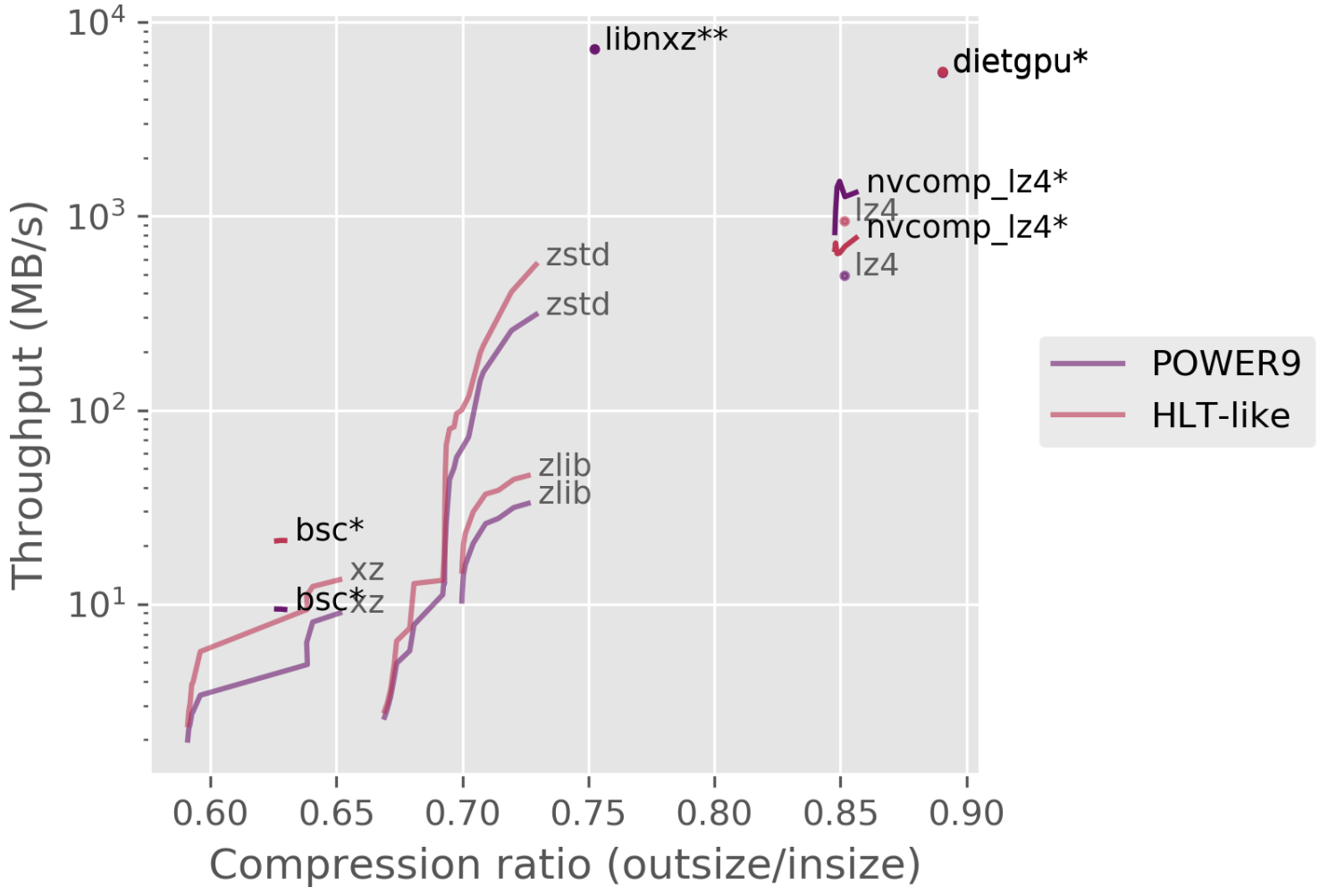
Why?

- Data from the HLT is compressed
- This is done on CPUs
- GPUs can be very fast

What I've been doing

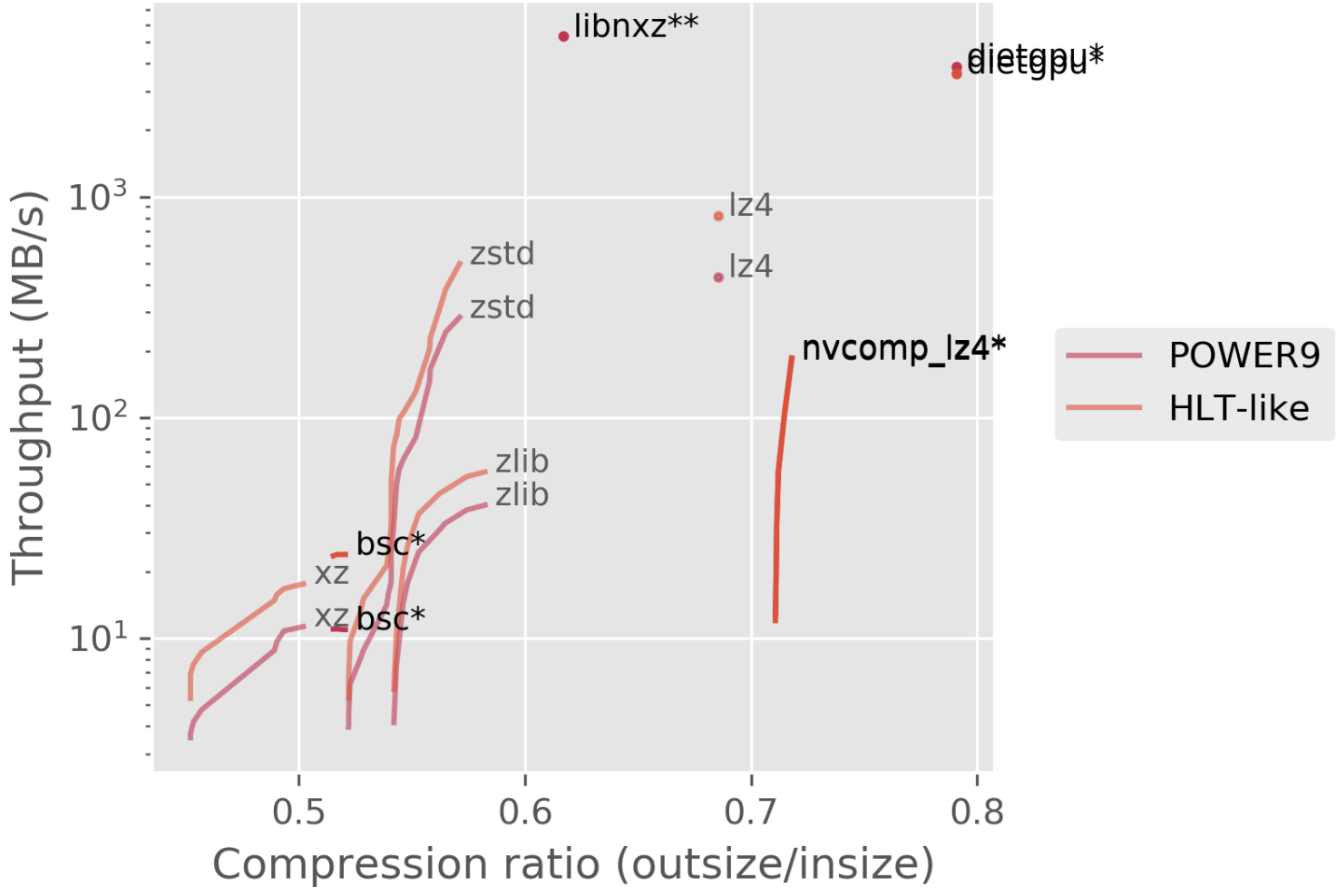
- Looking for GPU compressor implementations
- Adding them to a benchmarking program
- Comparing them to CPU compressors

Results - PP events



* = GPU
** = NX

Results - HI events



* = GPU
** = NX

Results - details

Data:

- 100 PP events
- 100 HI events

Timing:

- Wall time
- Disk $\rightarrow$ RAM **excluded**
- RAM $\rightarrow$ GPU memory **included**
- Fastest from 5 repeats

Results - details

POWER9 machine:

- IBM POWER9
- Nvidia Tesla V100

HLT-like machine:

- AMD EPYC 75F3
- Nvidia Tesla T4

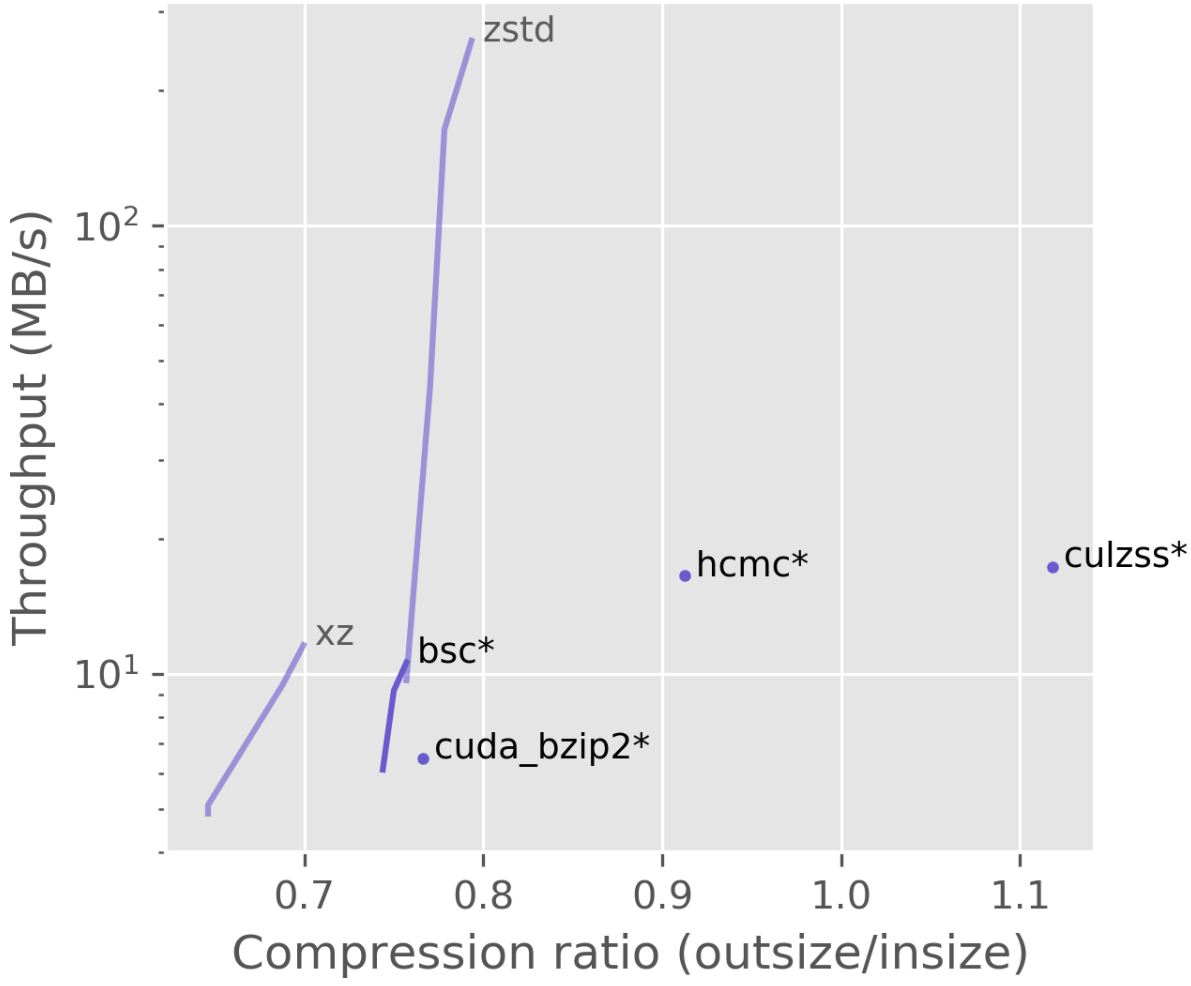
That's it

Contact info

Stefan Rua

stefan.elias.rua@cern.ch
stefan.rua@iki.fi

Extra - less promising ones



Extra - problems

Nvidia

From version 2.3 onwards, the compression / decompression source code will not be released. We'll continue to maintain this Github for documentation and code sample purposes.

Researchers

Hi Stefan

It has been a while that we have worked on that. The best I can find is this code for CULZSS not for bit one, but it might be working or not, i am not sure.

hope that helps.