The LC Framework for Generating Efficient Data-Compression Algorithms

Overview
LC is a framework for automatically generating customized lossless and guaranteed-error-bounded lossy data-compression algorithms for individual files or sets of files. The resulting compressors and decompressors are parallelized and produce bit-for-bit the same result on CPUs and GPUs.

General Features
- Automatic search (exhaustive and GA)
- Pareto-optimal speed vs. compression ratio
- Lossy and lossless algorithms
- Full CPU/GPU compatibility
- OpenMP & CUDA parallelization (some HIP)
- Optimized code generators
- Extensive library of components and preprocessors (data transformations)
- Regular expression support for subsetting
- User customizable and extensible
- Four word sizes (1, 2, 4, and 8 bytes)

Lossy-Mode Features
- 32-bit single and 64-bit double precision
- Guaranteed point-wise error bounds
- Absolute, relative, relative to range, and combined absolute & relative error bounds
- Lost bits replaced by zeros or by random bits to minimize autocorrelation

Publications
HPDC’22, ICS’23, SC’23, IPDPS’24, DCC’24

Lossless FP32 Compression
- “FPratio” and “FPspeed” created by LC
- Evaluated on SDRBench inputs
- FPratio yields highest compression ratio
- FPspeed yields over 500 GB/s throughput

Lossy FP32 Compression
- “PFPL” algorithm created by LC
- Evaluated on SDRBench inputs
- PFPL compresses more than cuSZp
- PFPL is much faster than other compressors

Collaboration Opportunities
- Looking for users who need compression
- Looking for interesting/challenging data sets
- Looking for feedback and suggestions for additional features

Team
- PI: Martin Burtscher (Texas State University)
- Co-PI: Sheng Di (Argonne National Lab)
- Senior advisor: Franck Cappello (ANL)
- Students: Noushin Azami (TXST), Alex Fallin (TXST), Brandon Burtchell (TXST), Yiqian Liu (TXST), Benila Jerald (TXST), Kayla Wesley (TXST), Yafan Huang (UIowa), Jiajun Huang (UC Riverside), and Milan Shah (NCSU)

Contact Info
burtscher@txstate.edu