* Related Work:

1. In this research paper[1] they describe the design, analyze the structure, and evaluate the performance of SPDP, an automatically synthesized lossless compression algorithm for single- and double-precision floating-point data. It is the best-compressing out of the 9,400,320 possible four-stage algorithms that can be built from our set of 48 algorithmic components that does not include any bit-level coders. SPDP yields the highest compression ratio on eleven of the 26 tested datasets and outperforms all of the evaluated compressors except Zstd.

More importantly, their analysis represents a first step in a new direction aimed at improving their understanding of how to build effective domain-specific compression algorithms. First, by systematically generating candidates and analyzing the structure of the best resulting algorithm, they were able to gain insight into its operation and learned how to handle mixed precision datasets. Second, they were able to demonstrate that a competitive algorithm can be created based solely on transformations that do not process data at bit granularity.

1. In this paper author propose simple, effective and very fast compression/decompression methods for FASTQ genomic data using general purpose compression components in a columnar compression model, robust to eventual changes in the FASTQ format specification.The compression speed as observed for ERR317482\_1 FASTQ data (≈ 7GB) in our tests is about 40X better than default gzip and 4.5X-7.4X faster than the latest, best proposals like DSRC v2 and slimfastq as featured in [2]. Decompression speeds are 4X-5X faster than gzip and 5X-17X faster than the specialized FASTQ compressors, DSRC and slimfastq- a more robust re-implementation of Fqzcomp/Fastqz. Compression ratios are 5.6%-11.4% better than gzip flavors and just 6.4% - 9.7% less than the best, more complex coders as reported in [2].
2. Data compression is an important technique to improve the performance and space efficiency for flash-based storage systems .In this research paper Author[3] describe the employing fixed compression algorithms, as in most current flash-based storage products that incorporate data compression, fails to recognize and exploit the significant diversity in compressibility and access patterns of data and misses the opportunity to improve system performance, space efficiency or both. EDC is proposed in this paper to exploit the compression diversity of the workload characteristics .More specifically, for compressible data blocks EDC employs algorithms with higher compression ratios in time periods with lower system utilization and algorithms with lower compression ratios in time periods with higher system utilization. For non-compressible (or very lowly compressible) data blocks, it will write them through to the flash storage directly without any compression. Their extensive trace-driven evaluations on a lightweight implementation of the EDC prototype show that EDC achieves a much better trade-off between performance and space efficiency than the state-of-the-art schemes with fixed algorithms.

* Refrences:

1. SPDP: An Automatically Synthesized Lossless Compression Algorithm for Floating-Point Data,Steven Claggett, Sahar Azimi, and Martin Burtscher ,Department of Computer Science,Texas State University ,San Marcos, TX 78666, USA.
2. Ibrahim Numanagic, James K Bonfield, Faraz Hach, Jan Voges, J¨orn Ostermann, Claudio Alberti, Marco Mattavelli, and S Cenk Sahinalp. Comparison of high-throughput sequencing data compression tools. Nature Methods, 13(12):1005–1008, October 2016.