SZ Lossy Compression for Scientific Datasets

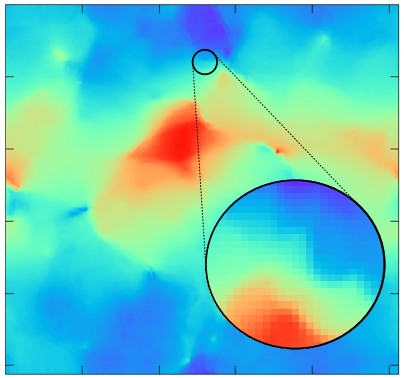
Sheng Di, Franck Cappello

sdi1@anl.gov, cappello@mcs.anl.gov

MCS, Argonne National Laboratory

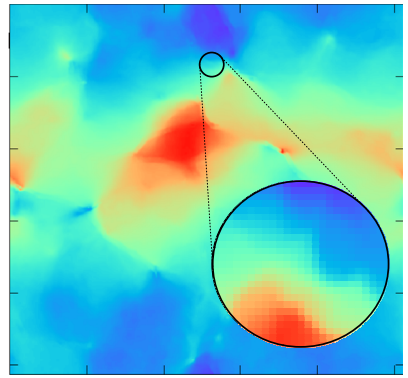
ECP EXASKY/NYX:
Cosmology simulation
X velocity field

Original



SZ Compression ratio: 143
PSNR=64dB

Signal/error ~10^3



U.S. DEPARTMENT OF ENERGY | Office of Science

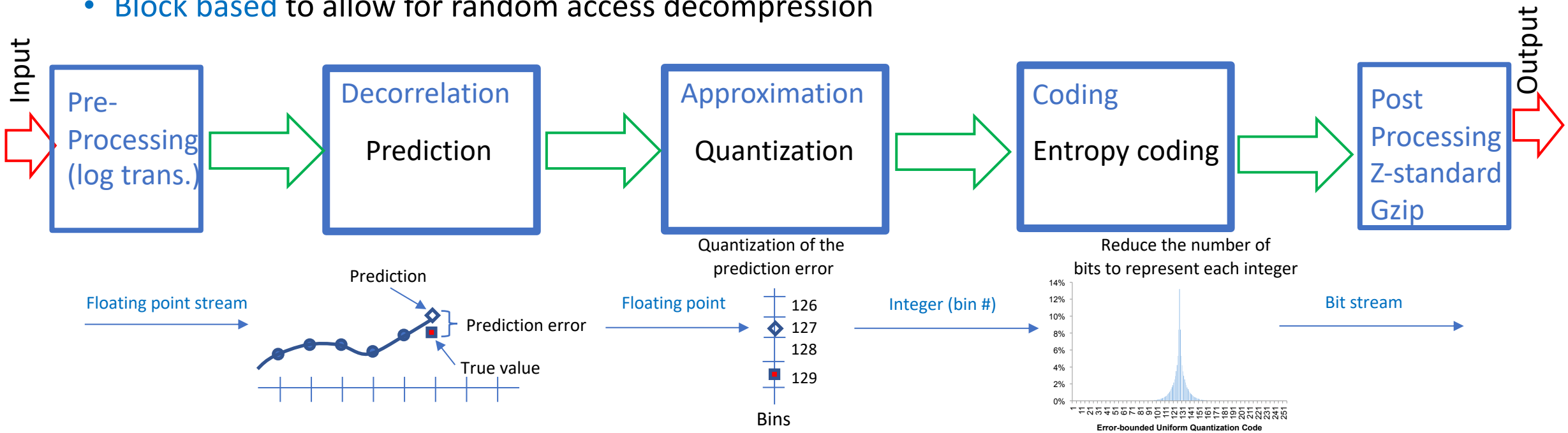
Argonne NATIONAL LABORATORY

ECP EXASCALE COMPUTING PROJECT

ANL SZ Framework Design Principles

- Error bounded (point-wise, PSNR)
- Multi-stages, Multi-algorithms, Multi-strategies (space, time)
- Prediction based : allowing customization of predictor
- Block based to allow for random access decompression

- Current version: SZ 2.1.11
  (Previous versions: SZ 1.1, SZ 1.4)



- **Lorenzo (1-layer non-linear prediction)**
- **extended Lorenzo (2-layer prediction)**
- **Linear-regression based prediction**
- **Tri-cubic interpolation**
- **pattern-aware prediction**
- **……**

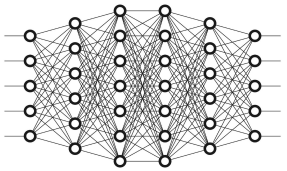
Argonne NATIONAL LABORATORY

SZ Applications

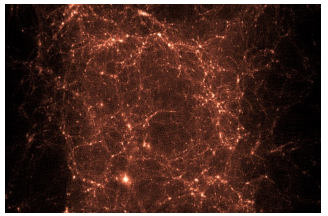
- Integrated into multiple scientific applications
- Evaluated/used by 20+ institutes/universities.
- 2000+ downloads/year
- Integrated in Spack.
- 40+ papers documenting each progress, optimization, application

SZ compressor:
szcompressor.org

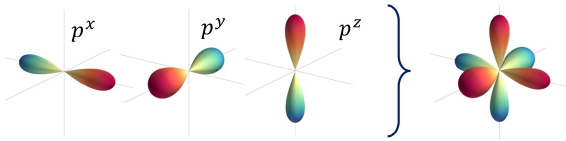
Deep Learning



Cosmology

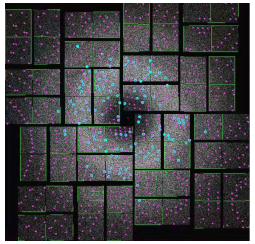


Quantum Chemisty



p-type basis functions p-shell

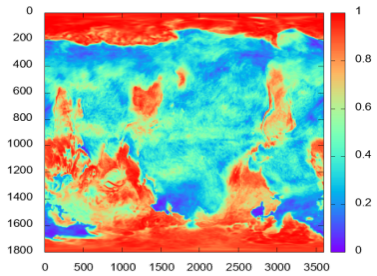
Light Source X-ray



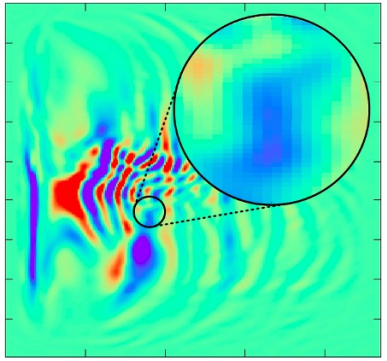
SZ

Molecular Dynamics

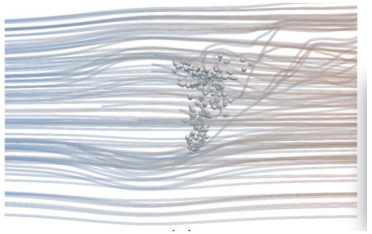
Climate/Weather



Seismology Wave Fields



Vector Fields

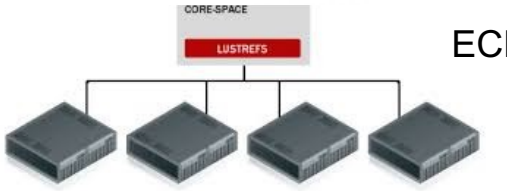


SZ Use-cases

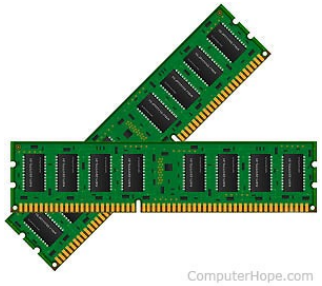
ECP apps: **GAMESS**

Reducing I/O time for HPC simulations

ECP apps: **HACC**, **NWCHEMeX**



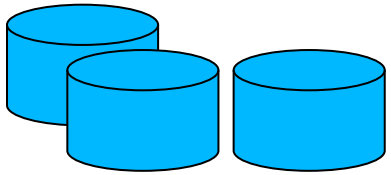
Reducing memory footprint



SZ Lossy Compression

Reducing storage footprint

ECP apps: **HACC, NYX**



Reducing Communication cost

Reducing data stream indensity

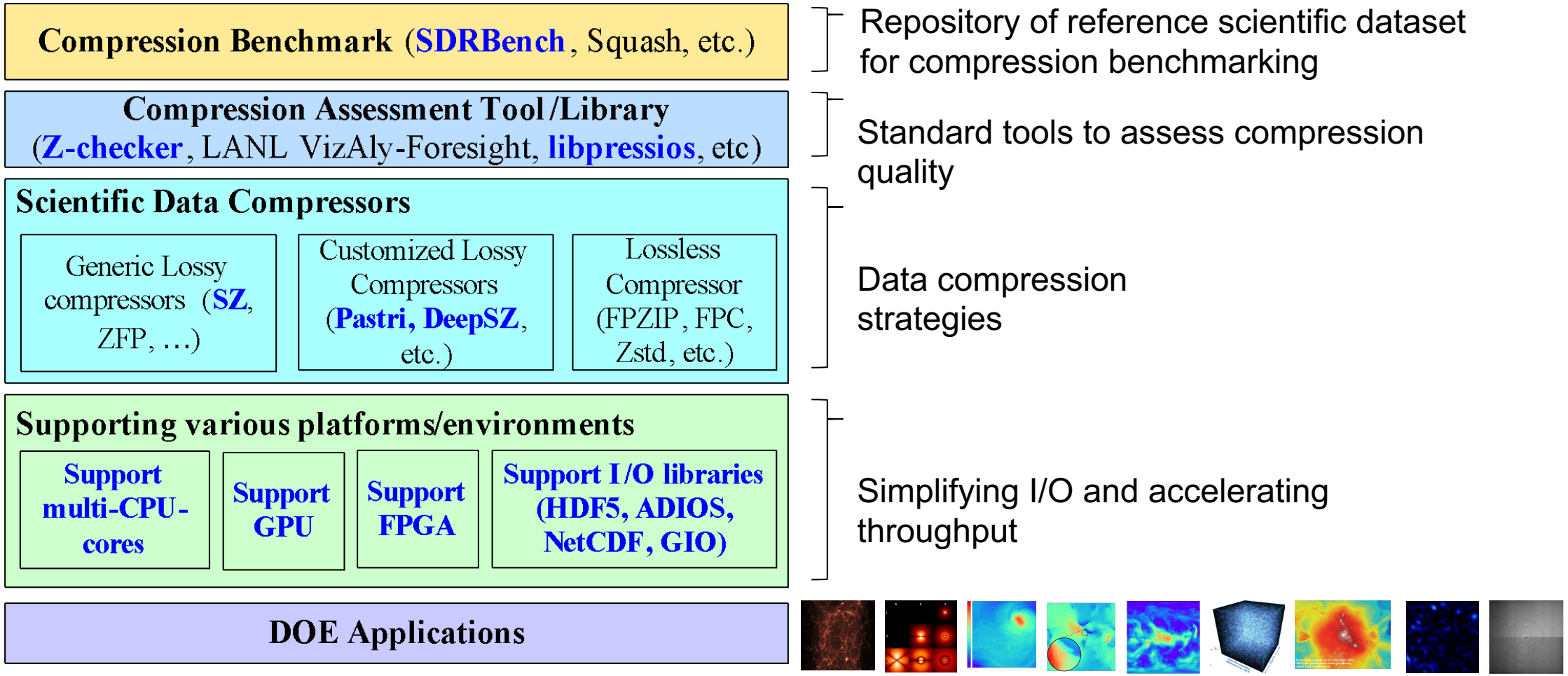
ECP apps: **EXAFEL**

Eliminating recomputation cost

ECP apps: **GAMESS**

Publication:
IJHPCA2017

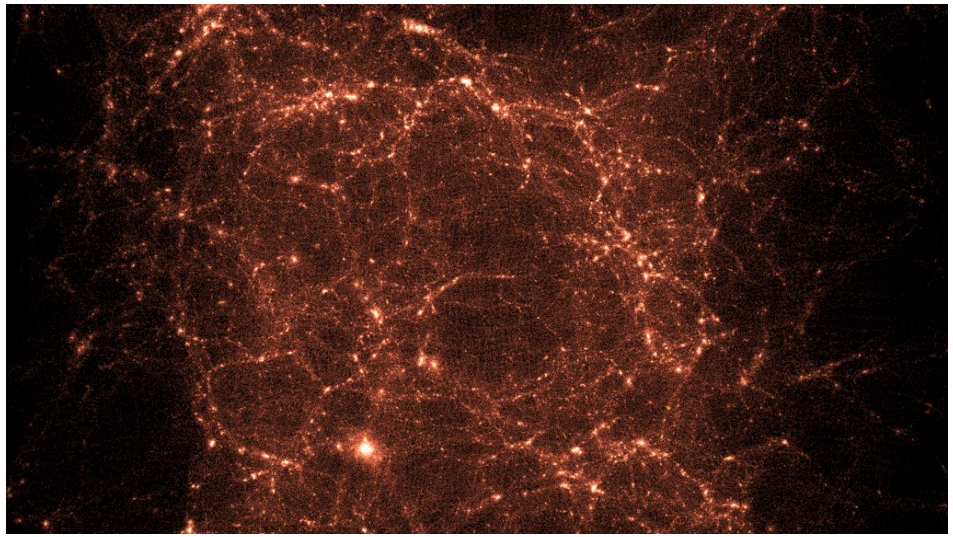
SZ Software Ecosystem



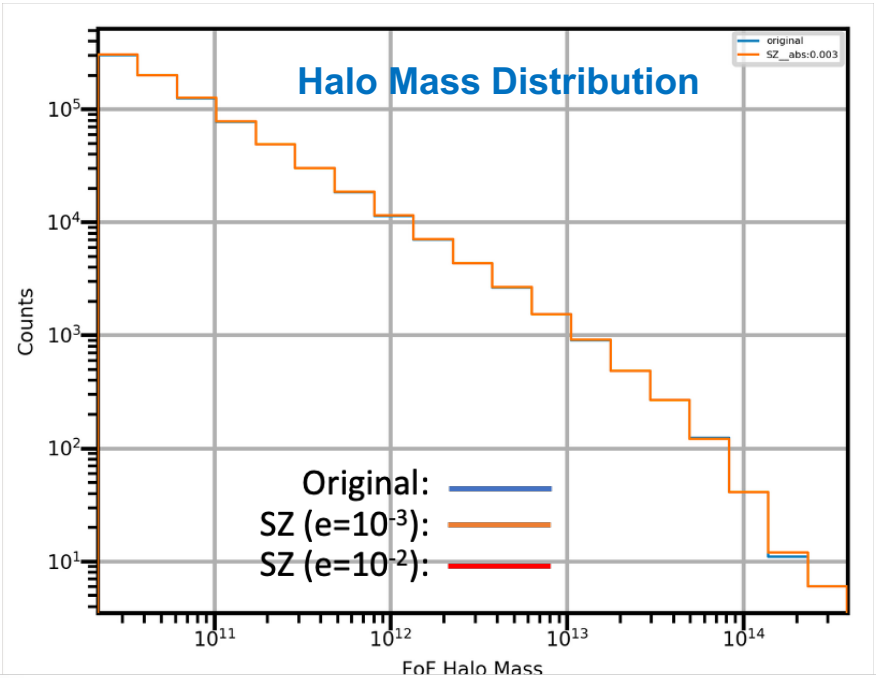
The items in blue are developed/focused in our team.

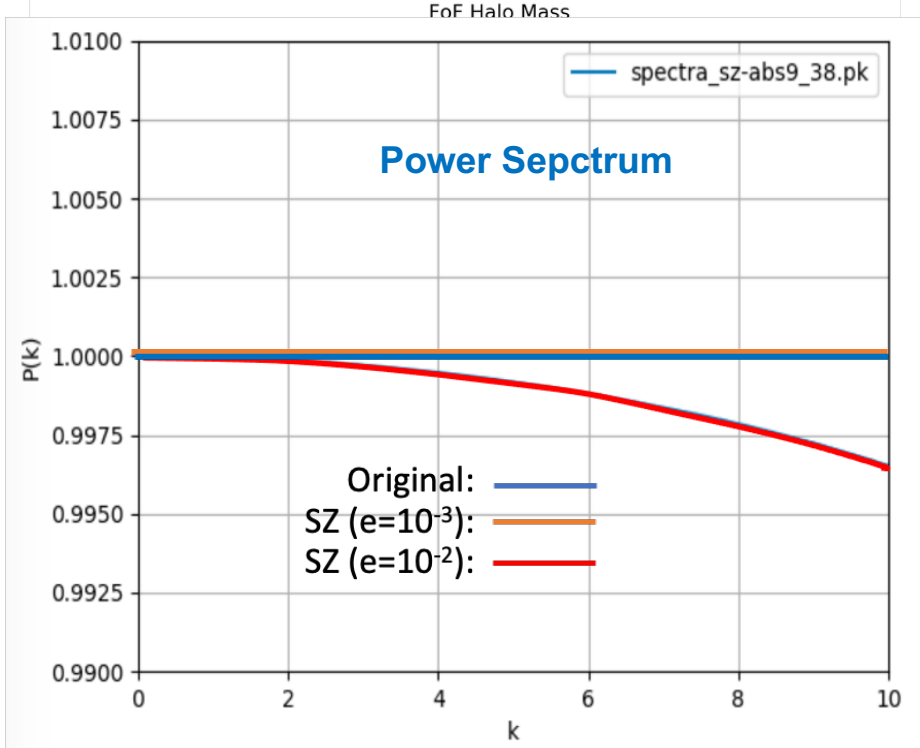
Success story: Cosmology simulation
ECP HACC

N-body problem with domain decomposition, medium/long-range force solver (particle-mesh method), short-range force solver (particle-particle/particle-mesh algorithm).



SZ 2.0: **CR ~5** (~6bits/value) at 10^{-3} error bound





Particle dataset: 6 x 1D array (x, y, z, vx, vy ,vz)

Very hard to compress

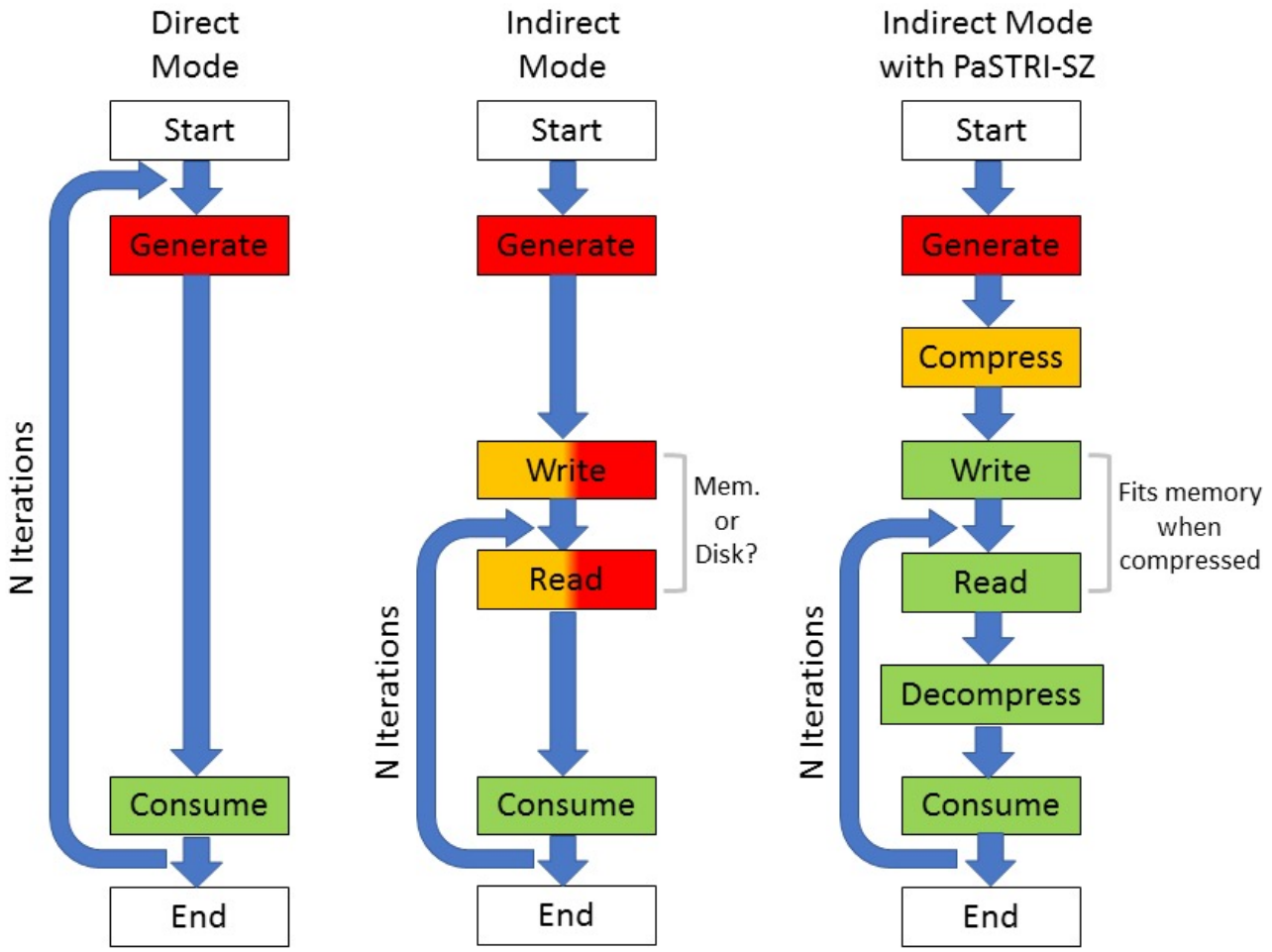
Preferred error controls:

- Point wise max error (Relative) bound
- Absolute (position), Relative (Velocity)

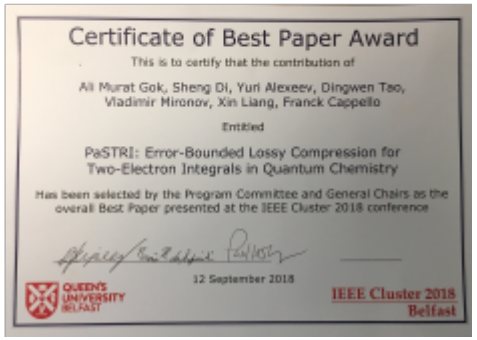
ECP EXASCALE COMPUTING PROJECT

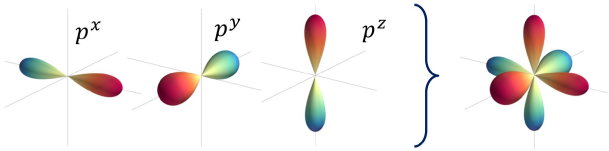
Success story: Quantum Chemistry
ECP GAMESS

- Two-Electron Integrals in Quantum Chemistry



Best paper award
at IEEE Cluster 2018





p-type basis functions p-shell

The goal is to obtain the **wavefunction** of a chemical system by solving the **Schrödinger equation**.

We customized an efficient error-bounded compressor for GAMESS two-electron integrals dataset and successfully integrated it into GAMESS Fortran code

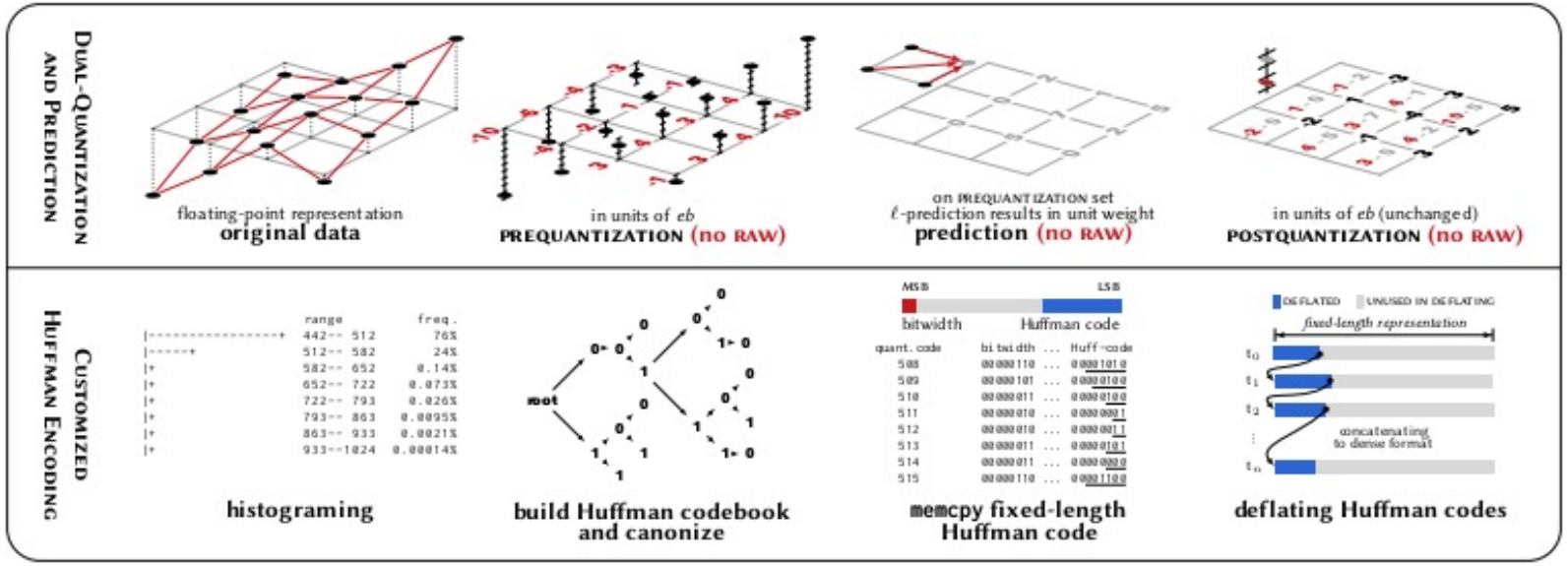
CR: 17 at 10^-11

1.5X end-to-end overall execution performance gain is observed.

GPU performance

cuSZ: cuda based SZ , **kSZ**: kokkos based SZ

Key techniques: **dual-quantization/prediction, Huffman on GPU**



Compression Performance on GPU:

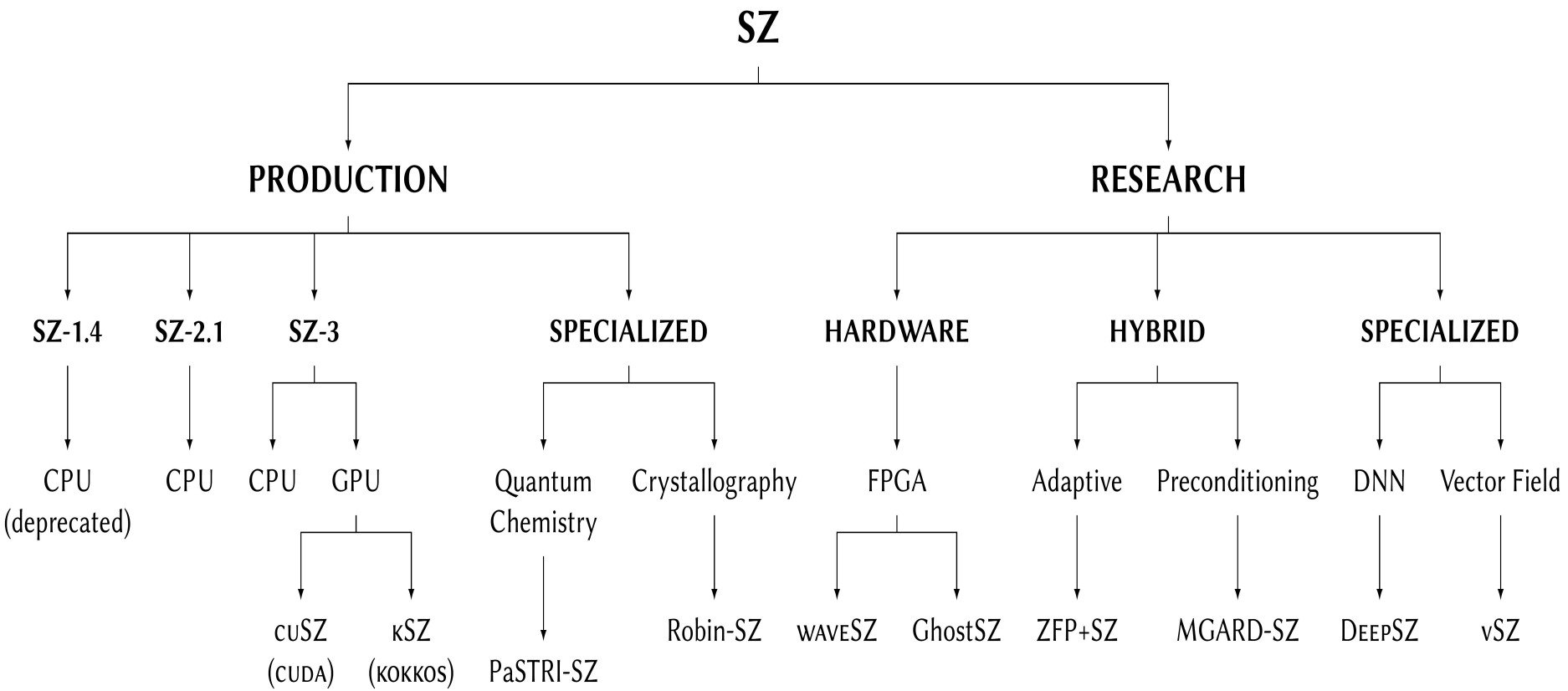
- Overall: HACC (A100): 66.7GB/s, NYX (A100): 66.9GB/s, QMCPack (A100): 62.6GB/s

- Fastest Huffman encoding implementation on GPU: 135GB/s~175GB/s on A100.

J. Tian, S. Di, S. Di, K. Zhao, C. Rivera, M. Hickman, R. Underwood, S. Jin, X. Liang, J. Calhoun, D. Tao, and F. Cappello, "cuSZ: An Efficient GPU Based Error-Bounded Lossy Compression Framework for Scientific Data", PACT 2020.

J. Tian, C. Rivera, S. Di, J. Chen, X. Liang, F. Cappello, , "Revisiting **Huffman** Coding: Toward Extreme Performance on Modern GPU Architectures.", IPDPS 2021.

SZ as a community software



Core-group:

- (Lead) Argonne National Laboratory
  Dr. Franck Cappello, Dr. Sheng Di, Dr. Ali Murat Gok
- Washington State University
  Dr. Dingwen Tao, Jiannan Tian, Cody Rivera, Sian Jin, Chengming Zhang
- University of California, Riverside
  Kai Zhao, Sihuan Li, Jinyang Liu
- Clemson Univerity
  Dr. Jon Calhoun, Robert Underwood
- Missouri University of Science and Technology
  Dr. Xin Liang

Other Contributors:
Dr. Martin Herbordt, Dr. Qingqing Xiong, Dr. Wen Xia, Xiangyu Zou

Thanks

This research was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative.

Argonne NATIONAL LABORATORY