Improving lossy compression for climate datasets with SZ3

Franck Cappello, Sheng Di,
Robert Underwood

Argonne National Laboratory

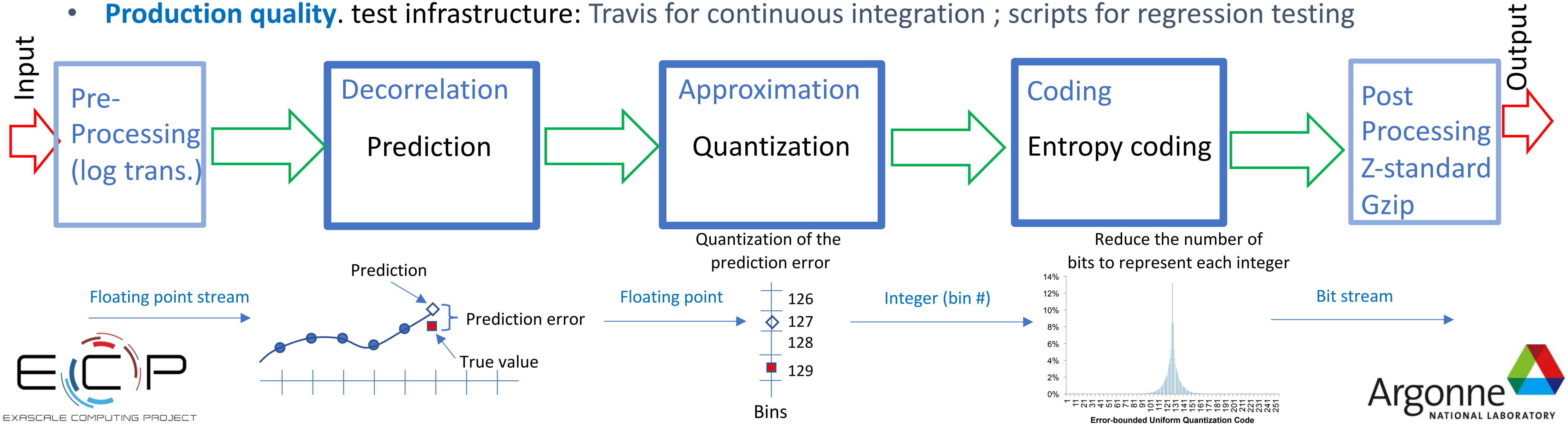
https://szcompressor.org

EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY

SZ lossy compression framework

- **Error bounded: Provides different error controls (Absolute, relative, value range error bounds, PSNR)**
- **Multi-stages, Multi-strategies** (temporal, spatial), **Multi-algorithms**
- **Prediction based** lossy compressor (differs from transform based lossy compressors)
- **Compress/decompress by blocks** for nearly **random-access decompression**
- Can compress **1D, 2D, 3D datasets.** and unstructured datasets as 1D
- Multiple // implementations: CPU Core (Vector Instructions), Multi-core (OpenMP), GPU (Cuda, Kokkos, *HIP*, DPC++*),* FPGA (proto)
- Integrated into the main I/O libraries: **HDF5, ADIOS** and **PnetCDF**
- **Production quality**. test infrastructure: Travis for continuous integration ; scripts for regression testing

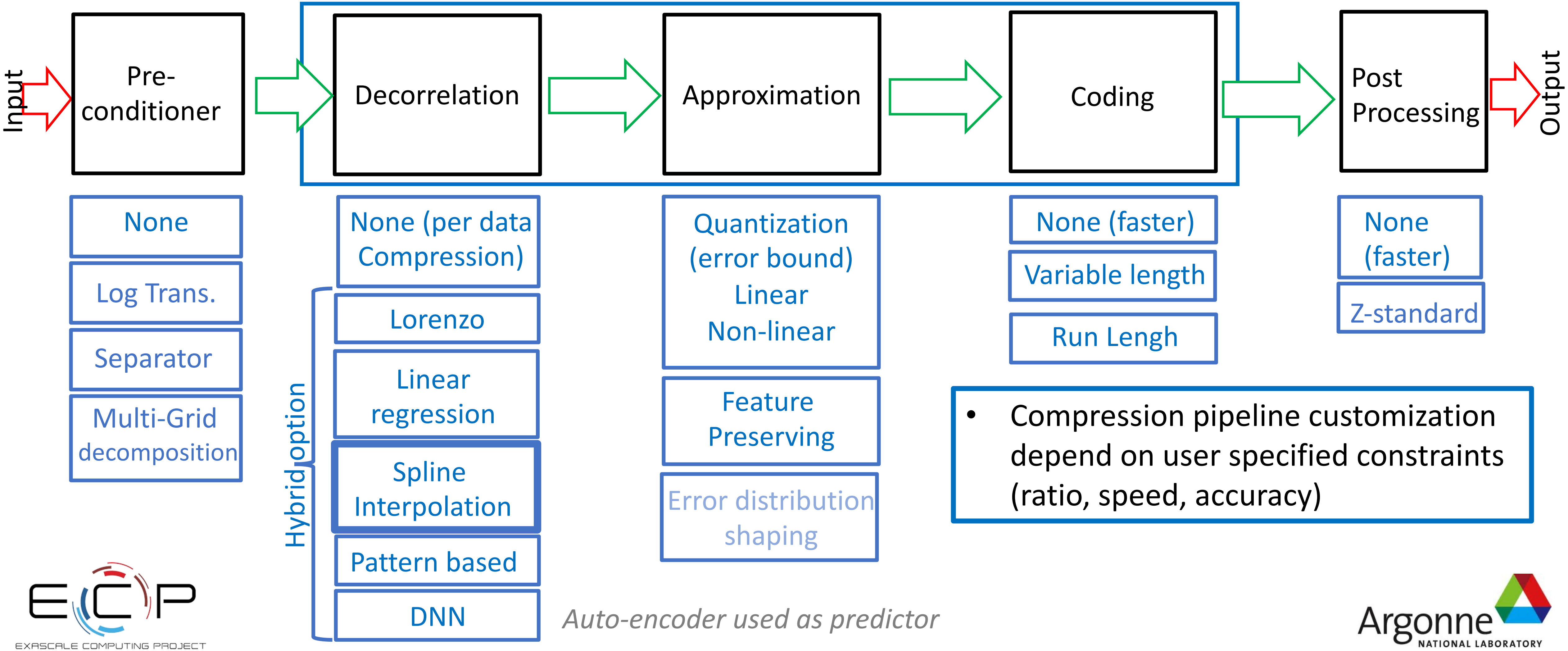


EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY

SZ3 Customization

Custom versions of SZ are built from modules



- Compression pipeline customization depend on user specified constraints (ratio, speed, accuracy)

EXASCALE COMPUTING PROJECT

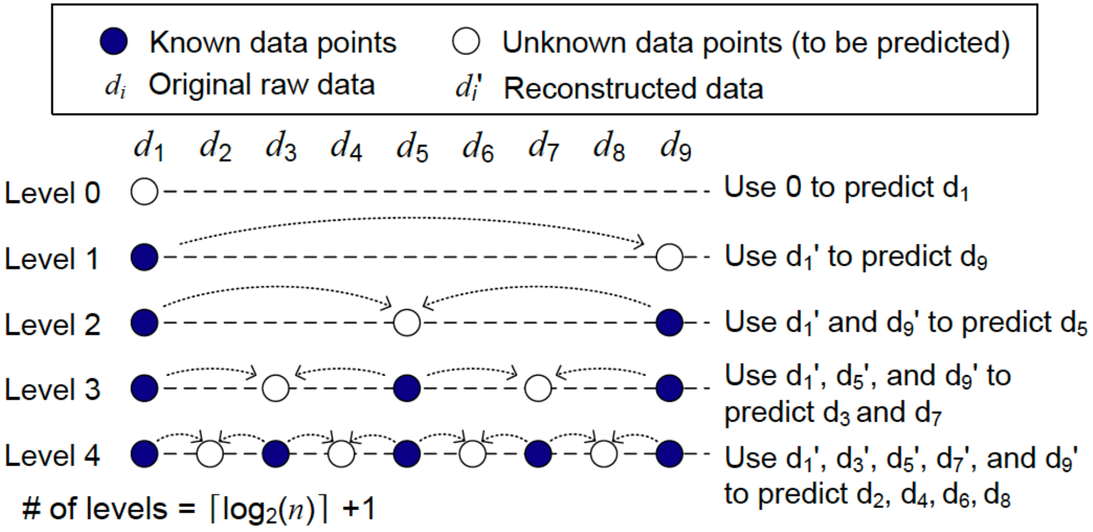
Argonne NATIONAL LABORATORY

SZ3 Spline interpolation Predict.

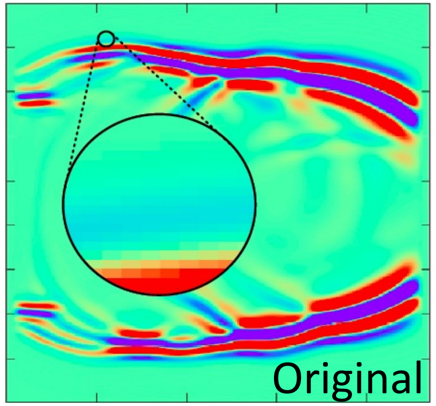
Predictor based on multilevel, multidimensional tri-cubic spline interpolation

Spline method	*Prediction Value* p_i
Linear spline	$p_i = \frac{1}{2}d_{i-1} + \frac{1}{2}d_{i+1}$
Cubic spline	$p_i = -\frac{1}{16}d_{i-3} + \frac{9}{16}d_{i-1} + \frac{9}{16}d_{i+1} - \frac{1}{16}d_{i+3}$

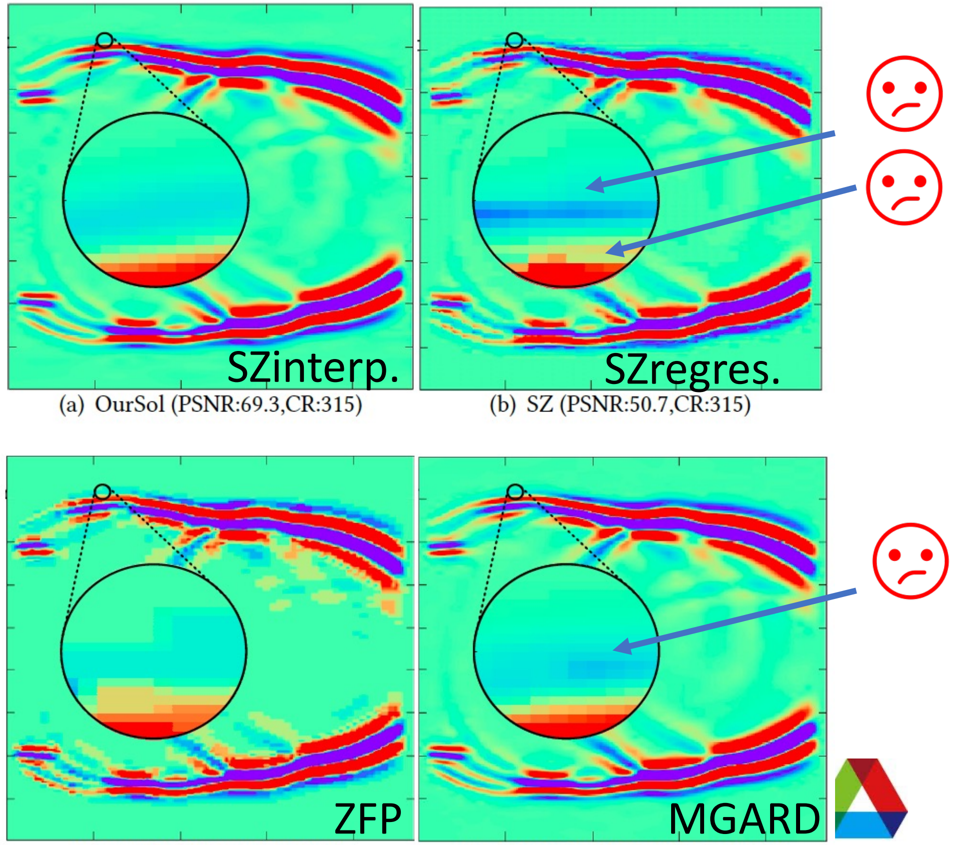
1D case (linear spline):



At level 0, 0 to predict $d1$, → Store quantized error (e0)
At level 1, $d1$+e0 to predict $d9$, → Store quantized error (e9)
At level 2, $d1$+e0 and $d9$+e9 to predict $d5$, → Store error (e5)
…



CR:~315 **Figure 25: Visualization of RTM, original data**



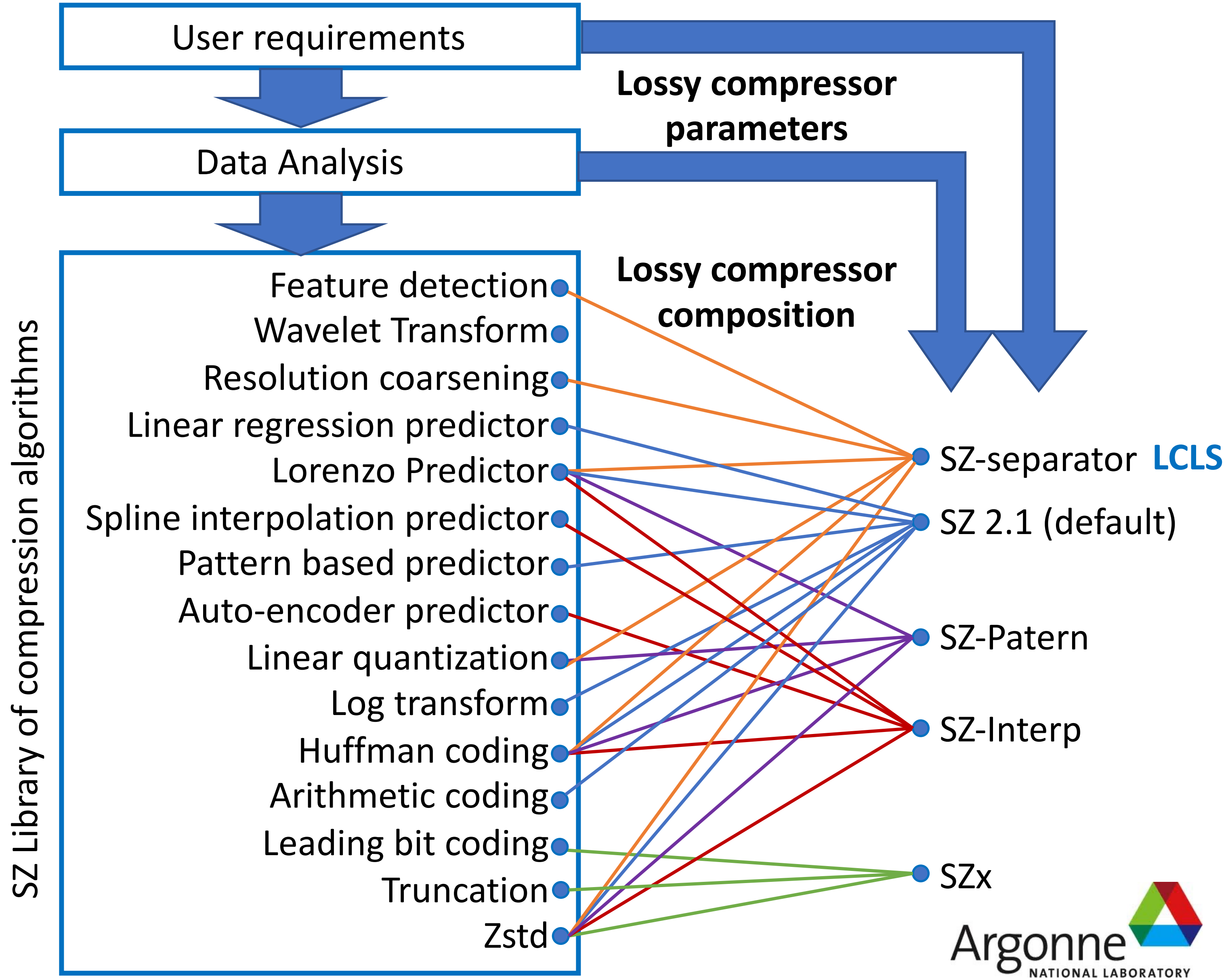
(a) OurSol (PSNR:69.3,CR:315) (b) SZ (PSNR:50.7,CR:315)

(c) ZFP (PSNR:51.7,CR:258) (d) MGARDx (PSNR:62.5,CR:310)

SZ3: Customization Process

SZ 3 library of algorithms for lossy compression and examples of SZ compressors built from the library of algorithms.

To compose and tune a compression pipeline we analyze the data to compress and user requirements in compression speed, ratio and accuracy.



EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY

SZ3 use-cases

We are seeing an increasing diversity/number of use-cases

"Classic" use-cases:

1) Visualization
2) Reducing storage footprint (on-line, in situ compression)
3) Reducing I/O time (on-line, in-situ compression)

Recently identified use-cases:

4) Reducing streaming intensity (recent for generic floating-point compressors)
5) Lossy checkpoint/restart from lossy state
   - reduce checkpoints footprint on storage – adjoint, accelerate checkpointing
6) Re-computation Avoiding by reducing the memory footprint → GAMESS

Cappello, F., Di, S., Li, S., Liang, X., Gok, A. M., Tao, et Al., Use cases of lossy compression for floating-point data in scientific data sets. *The International Journal of High Performance Computing Applications*, *33*(6), 1201–1220, 2019

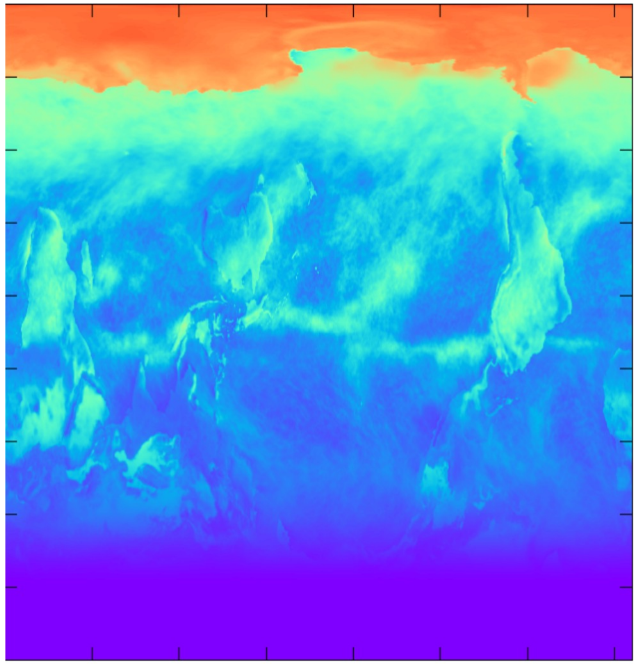
SZ3 on Climate Data

SZ can reach high compression ratios on the Atmosphere and Ocean datasets while maintaining quality metrics documented by **Pinard et al (2020):**

- Structural Similarity Image Metric **(DSSIM): >.995**
- Spatial Relative Error (SRE): <5% at **REL=1e-4**
- Pearson's Coefficient of Determination **(R^2): >.99999**
- P-value of the Kolmogorov-Smirnov test: **>.05**

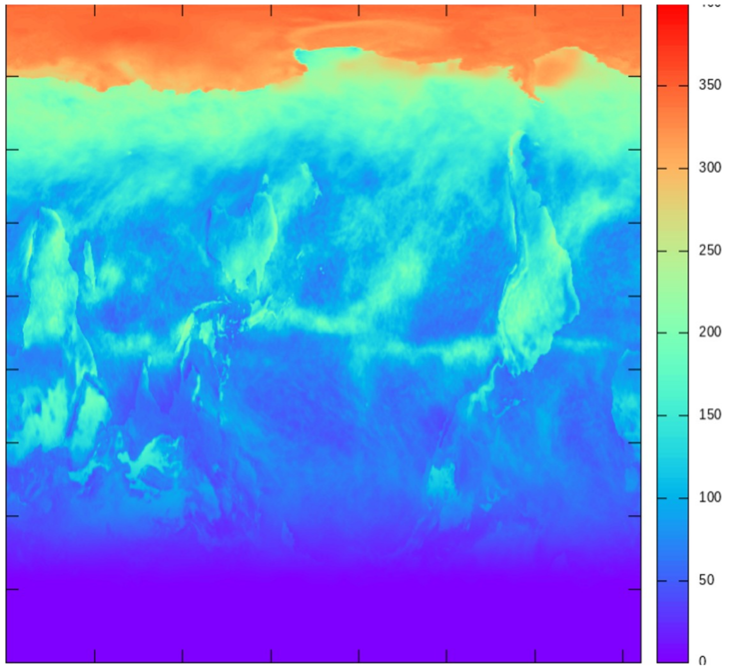
Values accurate to machine precision

Future work is still needed to improve performance on the Ice and Land datasets which have many small buffers



Original, CESM/ATM/ FSUTOA

Upwelling solar flux at top of atmosphere



SZ3 compressed (interp predictor):
93x smaller (without KS test)
21x smaller (with KS test)

Discussion: (statistician comment: KS test is too sensitive to sample size and not sensitive enough to outliers. Other tests e.g. Anderson-Darling might be better).

SZ3 Performance WRT Pinard et al. Requirements

Largest Compression Ratio For Each Compressor that Satisfies Each Pinard et al (2020) Requirements

- Structural Similarity Image Metric **(DSSIM): >.995**
- Spatial Relative Error (SRE): <5% at **REL=1e-4**
- Pearson's Coefficient of Determination **(R^2): >.99999**

Compressors	Pearson R^2	Spatial Relative Error	dSSIM
SZ	30.65	31.49	39.86
SZ_Interp	**93**	**93**	**59.81**
ZFP	13.27	13.27	18.87
MGARD	27.1	4.69	X
MGARDx	14.7	6.49	X
TThresh	16.1	16.1	2.98
BitGrooming	1.51	1.51	1.51
Digit Rounding	1.86	1.86	1.86
FPZip	1.95	1.95	1.95
NDZip	1.64	1.64	1.64
Zstd	1.35	1.35	1.35

All together: R^2, SRE, dSSIM

→ CR~59.81

EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY

SZ3 Contributors

Core-group (SZ team)
(Lead) Argonne National Laboratory
Franck Cappello, Sheng Di, Robert Underwood, Julie Bessac, Xiaodong Yu
Washington State University
Dingwen Tao, Jiannan Tian, Sian Jin, Chengming Zhang
University of California, Riverside
Kai Zhao, Jinyang Liu
Clemson University
Jon Calhoun, Griffin Dube
Missouri University of Science and Technology
Xin Liang
University of Alabama
Cody Rivera
Alumni
Ali Murat Gok, Dr. Sihuan Li

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

ECP EXASCALE COMPUTING PROJECT