# Appendix: QPET: A versatile and portable Quantity-of-Interest-preservation framework for Error-Bounded Lossy Compression

ANONYMOUS AUTHOR(S)

This appendix reports all evaluation results we have collected. We will keep updating it as more verified results become available.

## 1 EXPERIMENTAL SETUP

### 1.1 Experimental environment and datasets

we perform the evaluations on 6 real-world scientific datasets from diverse domains (details in Table 1). Experiments are operated on Purdue Anvil computing cluster [2] (each node is equipped with two 64-core AMD EPYC 7763 CPUs and 512GB DDR4 memory).

### 1.2 Baselines

we evaluate QPET with 3 integrations (SZ3+QPET, HPEZ+QPET, SPERR+QPET). The baselines are the corresponding base compressors and QoI-SZ3 [1], the state-of-the-art error-bounded lossy compressor supporting QoI-preservation.

Table 1. Information of the datasets in experiments

| App. | # fields | Dimensions | Total Size | Domain |
|---|---|---|---|---|
| Miranda | 7 | 256×384×384 | 1GB | Turbulence |
| Hurricane | 13 | 100×500×500 | 1.3GB | Weather |
| RTM | 11 | 449×449×235 | 2.0GB | Seismic Wave |
| NYX | 6 | 512×512×512 | 3.1GB | Cosmology |
| SEGSalt | 3 | 1008×1008×352 | 4.2GB | Geology |
| SCALE-LetKF | 13 | 98×1200×1200 | 6.4GB | Climate |

### 1.3 QoI functions, experimental configurations, and evaluation metrics

Table 2 shows the QoI functions to be preserved in the evaluation tasks. Among them, there are three different categories: point-wise, regional, and vector. They have diverse mathematical formats, and for many among them (such as $e^x$, $\frac{1}{x+c}$, and vector QoIs), QPET is the first framework that supports compression with preservation of those QoIs. The selection of QoI functions in our evaluation is based on existing investigations and analysis [1, 3, 5] of QoIs in practical scientific data analysis tasks.

Table 2. QoI functions in the evaluation

| QoI type | QoI function | QoI type | QoI function |
|---|---|---|---|
| **Pointwise** | $x^2$ | **Regional** | $x^2$ (average) |
| | $\log_2 x$ | | $x^3$ (average) |
| | $e^x$ | **Vector** | $x^2 + y^2 + z^2$ |
| | $\frac{1}{x+c}$ | | $\sqrt{x^2 + y^2 + z^2}$ |
| | $x^3$ | | $xyz$ |

In the tasks of QoI-preserving error-bounded lossy compression, Both a data error bound and a QoI error threshold are required. For the QPET-integrated compressors and QoI-SZ3, those threshold values are just input parameters. For base compressors, we use binary search to acquire

the corresponding error-bounded to achieve the target QoI error threshold. Regarding the non-QPET compression configurations, we apply the optimization level of 3 (having close compression ratios to max level 4 and exhibiting better speed) and compression-ratio-preferred mode for HPEZ (*-T CR*) and the default setting for other configurations. Regarding QPET parameters in Algorithm 3 of the paper, we set $c = 3$, $\beta = 0.999$ for SPERR, $c = 2$, $\beta = 0.999$ for HPEZ, and $c = 2$, $\beta = 0.99999$ for SZ3.

In evaluating the compression performance, the following widely-adopted metrics [1, 4] are used: (1) Compression and decompression speeds. (2) Compression ratio CR = $\frac{|X|}{|C|}$, which is the input data size $|X|$ divided by the data size $|C|$; (3) Bit rate BR = $\frac{|C|*8*\text{sizeof}(x)}{|X|}$, which is the number of bits in compressed data to store each value in the input. (3) Maximum data error and QoI error between the input and output; (4) PSNR [4] of the decompressed data QoI. The PSNR of QoI is define as $PSNR_Q = 20\log_{10}\text{VRange}\left(Q\left(X\right)\right) - 10\log_{10}\text{MSE}\left(Q\left(X\right), Q\left(X'\right)\right)$, in which $X$ and $X'$ are the input/decompressed data, $Q$ is the QoI function, VRange is value range, and MSE is mean squared error. The lower mean squared error between $Q\left(X\right)$ and $Q\left(X'\right)$ lead to higher (better) PSNR. (5) Visualization of the QoI error between the original and decompressed data.

## 2 EVALUATION RESULTS

### 2.1 Point-wise QoI

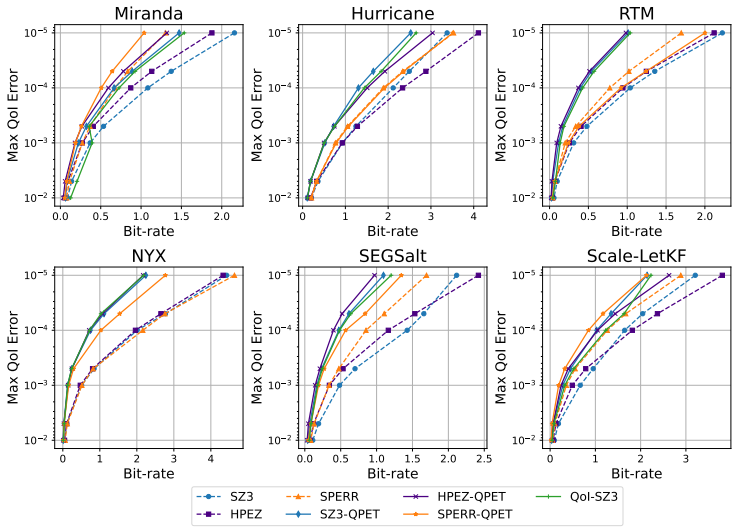


Fig. 1. Bit rate and Max QoI error plots for $Q(x) = x^2$.

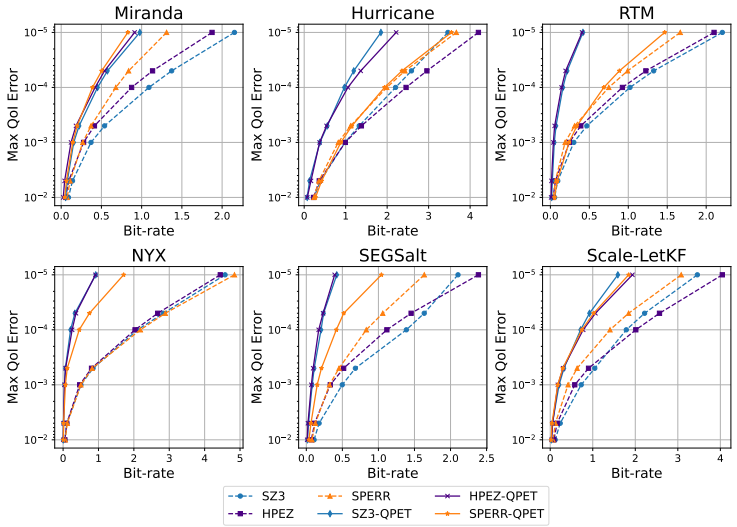


Fig. 2. Bit rate and Max QoI error plots for $Q(x) = x^3$.

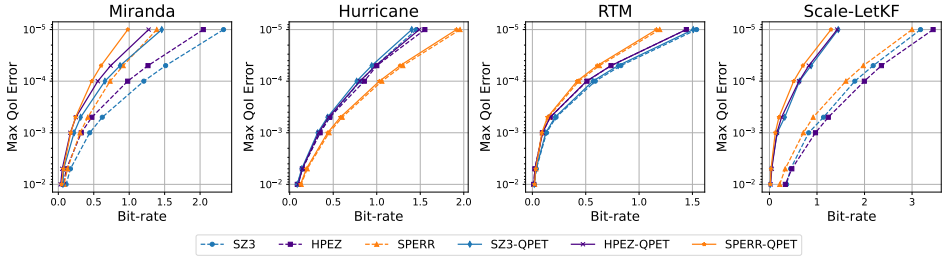


Fig. 3. Bit rate and Max QoI error plots for $Q(x) = e^x$.

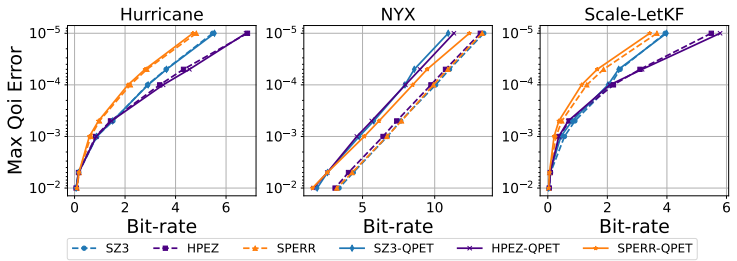


Fig. 4. Bit rate and Max QoI error plots for $Q(x) = \frac{1}{x+c}$.

## 2.2 Block-wise QoI

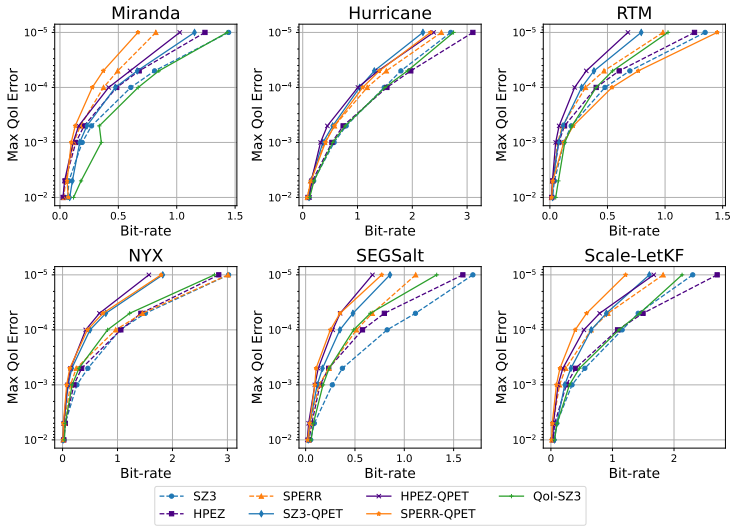


Fig. 5. Bit rate and Max QoI error plots for $Q(X) = \frac{1}{n_b} \sum x^2$, $n_b = 4^3$, i.e. average $x^2$ on 4x4x4 blocks.

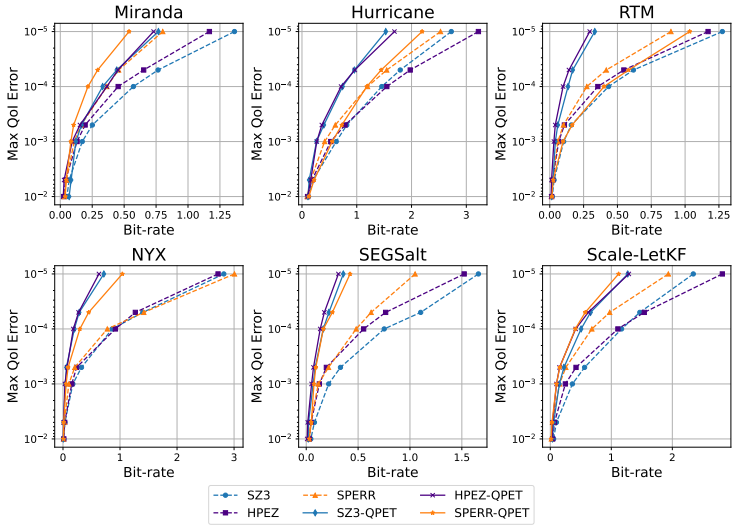


Fig. 6. Bit rate and Max QoI error plots for $Q(X) = \frac{1}{n_b} \sum x^3$, $n_b = 4^3$, i.e. average $x^3$ on 4x4x4 blocks.

## 2.3 Vector QoI

## REFERENCES


[1] Pu Jiao, Sheng Di, Hanqi Guo, Kai Zhao, Jiannan Tian, Dingwen Tao, Xin Liang, and Franck Cappello. 2022. Toward Quantity-of-Interest Preserving Lossy Compression for Scientific Data. *Proceedings of the VLDB Endowment* 16, 4 (2022), 697–710.

[2] X Carol Song, Preston Smith, Rajesh Kalyanam, Xiao Zhu, Eric Adams, Kevin Colby, Patrick Finnegan, Erik Gough, Elizabett Hillery, Rick Irvine, et al. 2022. Anvil-system architecture and experiences from deployment and early user operations. In *Practice and experience in advanced research computing*. 1–9.

[3] Zhaoyuan Su, Sheng Di, Ali Murat Gok, Yue Cheng, and Franck Cappello. 2022. Understanding impact of lossy compression on derivative-related metrics in scientific datasets. In *2022 IEEE/ACM 8th International Workshop on Data*

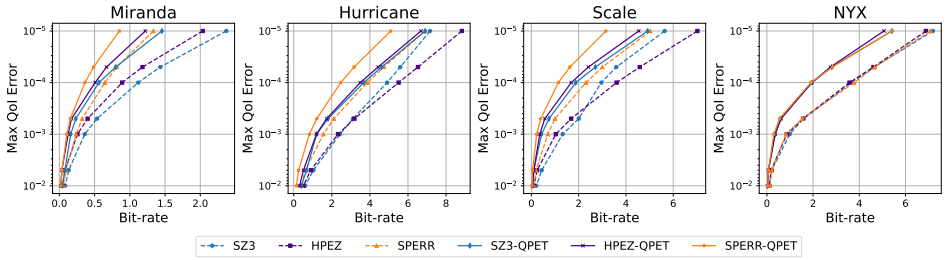


Fig. 7. Bit rate and Max QoI error plots for $Q(x, y, z) = x^2 + y^2 + z^2$.

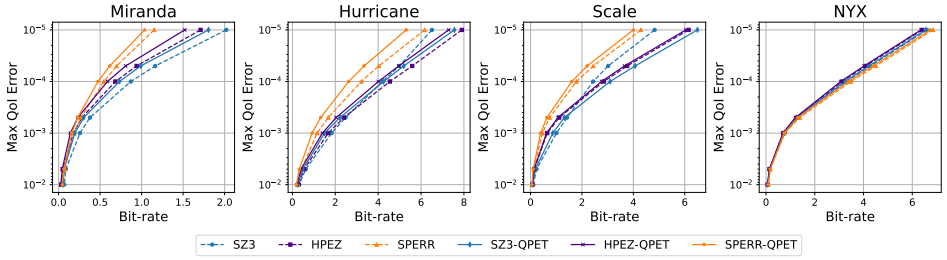


Fig. 8. Bit rate and Max QoI error plots for $Q(x, y, z) = \sqrt{x^2 + y^2 + z^2}$.

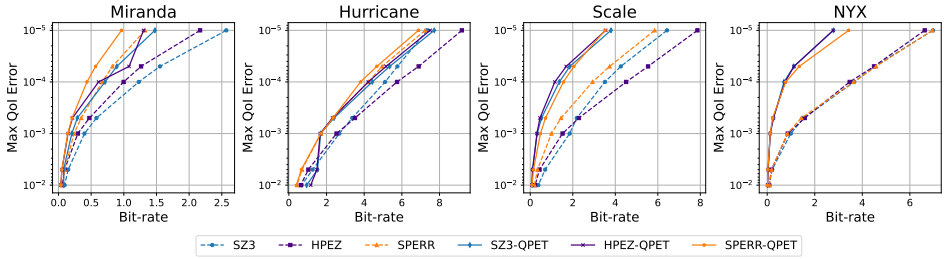


Fig. 9. Bit rate and Max QoI error plots for $Q(x, y, z) = xyz$.

*Analysis and Reduction for Big Scientific Data (DRBSD).* IEEE, 44–53.

[4] Dingwen Tao, Sheng Di, Hanqi Guo, Zizhong Chen, and Franck Cappello. 2019. Z-checker: A framework for assessing lossy compression of scientific data. *The International Journal of High Performance Computing Applications* 33, 2 (2019), 285–303. https://doi.org/10.1177/1094342017737147

[5] Xuan Wu, Qian Gong, Jieyang Chen, Qing Liu, Norbert Podhorszki, Xin Liang, and Scott Klasky. 2024. Error-controlled Progressive Retrieval of Scientific Data under Derivable Quantities of Interest. In *2024 SC24: International Conference for High Performance Computing, Networking, Storage and Analysis SC*. IEEE Computer Society, 1368–1383.