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 [5] (each node is equipped with two 64-core AMD EPYC 7763 CPUs and 512GB DDR4 memory).

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	1.2GB	Weather
RTM	11	449×449×235	2.0GB	Seismic Wave
NYX	6	512×512×512	3.1GB	Cosmology
SEGSalt	3	1008×1008×352	4.0GB	Geology
SCALE-LetKF	12	98×1200×1200	6.3GB	Climate

1.1.1 Baselines. Besides the QPET-integrated compressors, we included several existing solutions for QoI-preserving scientific lossy compression in the evaluations, which are in 2 categories: 1) **Parameter-search-based solutions**: general-purpose compressors (SZ3/HPEZ/SPERR) cannot directly bound specified QoI errors, so we apply parameter-search methods on top of them to figure out the best-fit (yielding highest compression ratio) data error bound for preserving the QoI with given accuracy. The parameter-search methods, including binary-search-based and FraZ [9], are from OptZConfig [8], which is the state-of-the-art parameter-search toolkit for scientific lossy compression. When applying, we slightly revised those methods to get them better adapted and accelerated in the QoI-preserving tasks. Those baselines are named SZ3/HPEZ/SPERR-OptZ-R (R is short for revised), and they represent the best (fastest) results from both parameter-search methods on every base compressor. In the parameter-search process, multiple iterations of data compression and QoI validation are performed to fit the QoI error bound, so they are typically quite slower than the original general-purpose compressors. (2) **QoI-preserving compressors**: [2] provided the SZ3-based QoI-preserving compressor QoI-SZ3, and we further ported its QoI-preserving features to HPEZ, creating QoI-HPEZ. Moreover, we evaluated MGARD-QoI [1]. Those QoI-preserving compressors have limited support for diverse QoI formats (e.g. QoI-SZ3/HPEZ only supports square and logarithm (also their block-averages), and MGARD-QoI only supports linear QoIs), so we only tested them on the QoIs supported. Other existing QoI-integrated compressors are designed for different tasks (e.g., cpSZ [3] only work for critical points in vector field data), so they are not included in our evaluation baselines.

1.1.2 QoI functions, experimental configurations, and evaluation metrics. Table 2 shows the QoI functions 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 $\tanh x$, $\frac{1}{n}\sum x^3$, 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 [2, 6, 10] of QoIs in practical scientific data analysis tasks, such as physical transform (kinetic energy as velocity's square), and clustering ($\sqrt{x^2+y^2+z^2}$ is the distance from origin when x, y, and z are coordinates).

Table 2. QoI functions in the evaluation

QoI type	QoI function	QoI type	QoI function
Pointwise	x^2	**Regional**	x (average)
	x^3		x^2 (average)
	$\log_2 x$		x^3 (average)
	$\sin 10x$	**Vector**	$x^2+y^2+z^2$
	$\tanh x$		$\sqrt{x^2+y^2+z^2}$

For the OptZ-R parameter search, we set an early termination that triggers when a maximum QoI error between 90% and 100% of the required threshold is found, because it presents near-optimal compression ratio with reasonable search time. Regarding the compression configurations, we apply the default optimization level and compression-ratio-preferred mode for HPEZ. Regarding QPET parameters, we set $c = 3$, $\beta = 0.999$ for SPERR, $c = 2$, $\beta = 0.999$ for HPEZ, and $c = 2$, $\beta = 0.99999$ for SZ3. On SZ3 and HPEZ, the autocorrelation of decompression errors are relatively high when the error bound is large [4], so we linearly dynamically decrease c when τ increases over 10^{-3}, eventually to 1.0 when τ becomes 10^{-2}.

In evaluating the compression performance, the following widely-adopted metrics [2, 7] are used: (1) Compression and decompression speeds (throughputs). (2) Compression ratio CR = $\frac{|X|}{|C|}$, which is the input data size $|X|$ divided by the data size $|C|$; (3) Bit rate BR = $\frac{|C|*8*\mathrm{sizeof}(x)}{|X|}$, 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;

2 Evaluation Results

2.1 Point-wise QoI

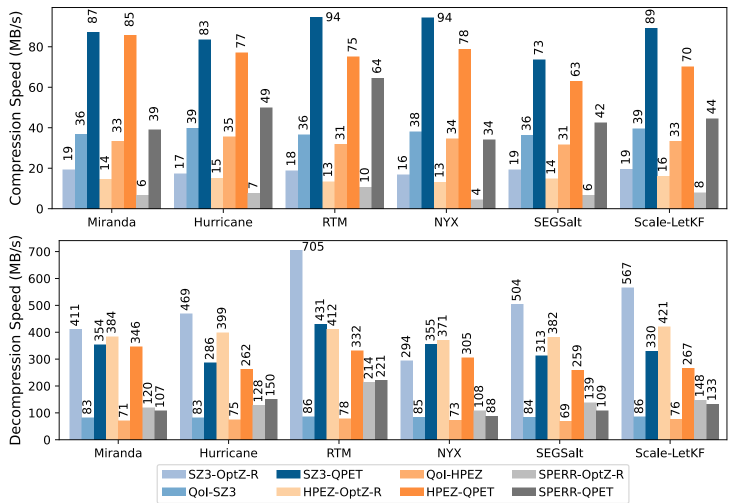

Fig. 1. Compression and decompression speed for $Q(x) = x^2$ and $\tau = 1\text{e}{-3}$.

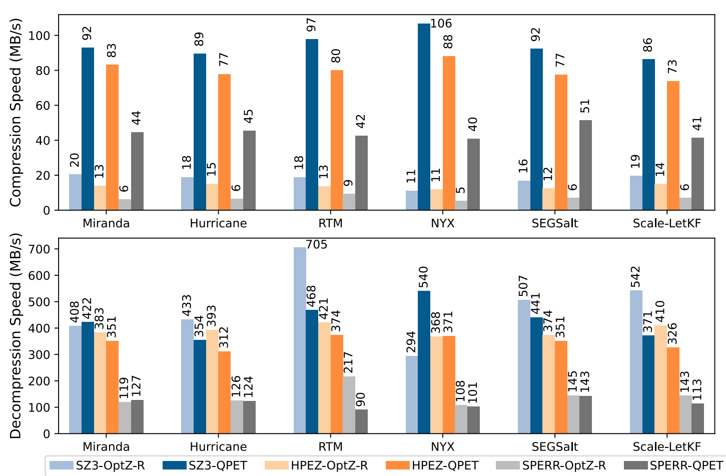

Fig. 2. Compression and decompression speed for $Q(x) = x^3$ and $\tau = 1\text{e}{-3}$.

2.2 Block-wise QoI

Table 3. Showcase of QoI-preserving error-bounded lossy compression. non-QoI compressors are manually tuned with multi-runs. ϵ: data error bound. τ: QoI error threshold. CR: Compression ratio. max e_d: maximum data error. Max e_q: maximum QoI error. All error values are relative (absolute value divided by value range).

QoI			$Q(x) = x^2$				$Q(x) = x^3$				$Q(x) = \log_2 x$				$Q(x) = \sin 10x$				$Q(x) = \tanh x$			
Data field			SegSalt-Pressure2000				RTM-3500				NYX-Baryon Density				Miranda-Pressure				Scale-LetKF-RH			
ϵ	τ	Compressor	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d
10^{-1}	10^{-2}	SZ3-OptZ-R	9	283	21	693	9	167	19	695	23	14.1	3	226	12	101	7	379	12	9.88	8	182
		HPEZ-OptZ-R	9	448	14	413	6	231	20	428	3	14.7	21	182	12	151	6	369	12	8.78	8	214
		SPERR-OptZ-R	6	413	10	144	9	237	10	220	23	19.4	2	71	12	132	3	118	12	11.4	4	60
		QoI-SZ3		458	36	88		N/A	N/A	N/A		21.9	19	80		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		687	30	70		N/A	N/A	N/A		24	17	71		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		689	**102**	513		619	**119**	630		28.1	**46**	376		110	27	453		74.4	**77**	641
		HPEZ-QPET		**954**	75	398		**1466**	92	413		27.9	44	336		**154**	**26**	364		**196**	71	503
		SPERR-QPET		849	49	135		728	63	260		**35.6**	22	67		134	21	120		11.9	22	53
10^{-2}	10^{-3}	SZ3-OptZ-R	9	76.5	19	505	6	28	25	444	3	5.9	22	152	12	32.9	7	320	12	5.4	7	144
		HPEZ-OptZ-R	9	98.8	14	381	6	30.6	18	322	3	6.2	21	183	12	43.9	6	320	12	4.7	7	154
		SPERR-OptZ-R	10	108	5	126	10	39.6	7	136	15	6.1	2	38	12	52	3	99	12	5.8	3	39
		QoI-SZ3		169	35	86		N/A	N/A	N/A		7.4	24	112		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		206	30	70		N/A	N/A	N/A		8	20	123		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		181	**76**	332		244	**88**	505		8.8	33	126		34.4	**25**	288		18.1	**57**	350
		HPEZ-QPET		196	64	276		**320**	78	405		**9.2**	**33**	128		44.5	24	253		**23.4**	52	297
		SPERR-QPET		**224**	47	128		210	59	219		9.1	18	40		**52.5**	20	100		5.9	17	35
QoI			$Q(x) = x^2$				$Q(x) = x^3$				$Q(x) = \log_2 x$				$Q(x) = \sin 10x$				$Q(x) = \tanh x$			
Data field			Miranda-Density				Hurricane-Cloud				Scale-LetKF-T				Miranda-Viscocity				NYX-Velocity_x			
ϵ	τ	Compressor	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d
10^{-3}	10^{-3}	SZ3-OptZ-R	4	87.3	37	379	8	49.3	23	596	3	73.9	27	727	8	47.4	10	343	17	1.2	5	223
		HPEZ-OptZ-R	4	112.7	33	407	4	53	40	502	3	84.3	21	460	8	58.9	9	411	8	1.2	11	235
		SPERR-OptZ-R	4	115	12	116	5	26.4	10	98	3	142	15	154	8	63.4	5	104	7	1.3	3	15
		QoI-SZ3		80.2	39	84		N/A	N/A	N/A		67.3	30	85		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		90.9	36	73		N/A	N/A	N/A		95.8	28	77		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		96.4	**120**	450		**69.3**	**97**	353		88.6	**106**	391		51.2	**31**	400		86.5	**72**	420
		HPEZ-QPET		**121**	108	390		62.1	82	329		105	46	329		63.4	29	327		**89.2**	56	291
		SPERR-QPET		118	44	116		36.3	44	114		**158**	33	131		**65.2**	22	106		72	32	81
10^{-4}	10^{-4}	SZ3-OptZ-R	4	31.1	35	317	5	29.5	35	502	3	16	21	202	8	21.5	10	288	3	1.2	30	274
		HPEZ-OptZ-R	4	36.1	31	343	4	28.2	34	433	3	15.6	20	279	8	25.5	9	294	3	1.2	25	195
		SPERR-OptZ-R	4	49.2	11	97	2	14.3	21	73	3	28.1	12	96	8	31.5	4	86	3	1.2	6	14
		QoI-SZ3		29.5	39	80		N/A	N/A	N/A		13.2	29	75		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		33.4	35	71		N/A	N/A	N/A		15.1	28	70		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		32.4	**116**	369		**36.6**	**98**	329		16.6	**45**	192		22.2	**28**	267		15.3	**58**	202
		HPEZ-QPET		37	106	327		28.2	70	258		16.7	44	216		26.4	27	229		**16.1**	54	183
		SPERR-QPET		**50.2**	39	98		17.3	34	79		**30**	26	85		**32.1**	20	87		14.1	23	51
QoI			$Q(X) = \frac{1}{n}\sum x_i$				$Q(X) = \frac{1}{n}\sum x_i^2$				$Q(X) = \frac{1}{n}\sum x_i^3$				$Q(x,y,z) = x^2+y^2+z^2$				$Q(x,y,z) = \sqrt{x^2+y^2+z^2}$			
Data field			RTM-3200				SegSalt-Pressure3000				Scale-LetKF-V				Hurricane-UVW				Miranda-VXYZ			
ϵ	τ	Compressor	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d	#It	CR	T_c	T_d
10^{-2}	10^{-3}	SZ3-OptZ-R	7	76.1	20	617	6	66.7	24	576	14	37	11	605	10	13	11	270	8	122	14	393
		HPEZ-OptZ-R	7	99.2	14	397	7	115	15	388	9	66.8	13	434	10	14	8	260	8	179	13	411
		SPERR-OptZ-R	5	155.6	16	205	6	126	9	129	7	94	9	142	8	20.4	7	48	9	167	5	121
		MGARD-QoI		15.7	3	4		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-SZ3		53.1	32	84		104	31	85		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		55.2	25	68		109	27	70		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		93	**97**	430		156	**80**	455		329	**71**	339		26.6	**46**	**172**		163	**29**	497
		HPEZ-QPET		140	75	361		234	70	369		300	59	288		26.8	43	152		**223**	28	389
		SPERR-QPET		**238**	65	213		**254**	44	131		**351**	41	148		39	41	57		202	22	62
10^{-3}	10^{-4}	SZ3-OptZ-R	7	15.2	18	341	7	19.4	19	353	10	14.3	10	190	4	6.5	17	132	8	35.8	13	325
		HPEZ-OptZ-R	6	15.2	16	256	8	28.1	13	314	8	15.5	13	286	4	5.9	17	172	8	45.5	12	363
		SPERR-OptZ-R	6	24.4	10	106	9	30.3	5	91	13	22.7	4	87	8	8.3	5	42	9	58.6	5	102
		MGARD-QoI		5.9	3	4		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-SZ3		13.6	31	75		34.5	31	82		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		QoI-HPEZ		15.1	28	62		50	27	65		N/A	N/A	N/A		N/A	N/A	N/A		N/A	N/A	N/A
		SZ3-QPET		19	**87**	301		49.4	**59**	265		56.5	**64**	304		8.7	**46**	**130**		43.6	**29**	419
		HPEZ-QPET		20.8	71	251		67.2	54	236		82.7	57	269		9.1	43	124		54.3	28	343
		SPERR-QPET		**43.4**	50	128		**73.6**	41	116		**83.3**	38	121		13	30	32		**67.4**	21	53

2.3 Vector QoI

References

[1] Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. 2019. Multilevel techniques for compression and reduction of scientific data-quantitative control of accuracy in derived quantities. *SIAM Journal on Scientific Computing* 41, 4 (2019), A2146–A2171.

[2] 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.

[3] Xin Liang, Sheng Di, Dingwen Tao, Zizhong Chen, and Franck Cappello. 2018. An efficient transformation scheme for lossy data compression with point-wise relative error bound. In *2018 IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, 179–189.

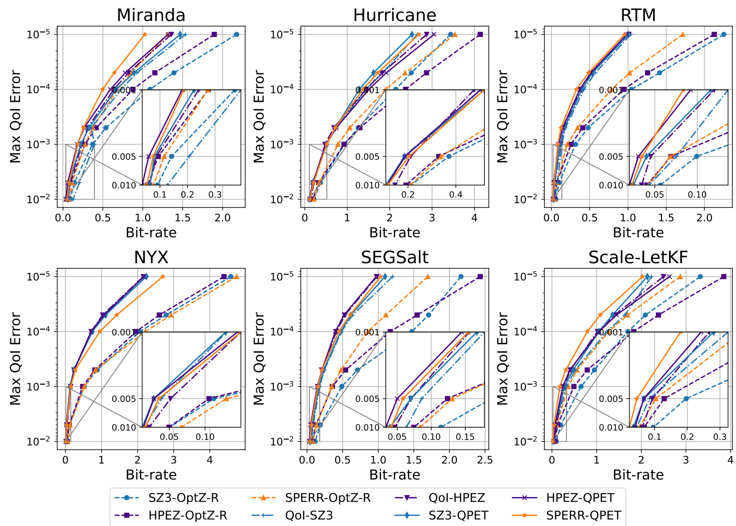

Fig. 3. Bit rate and Max QoI error plots for $Q(x) = x^2$.

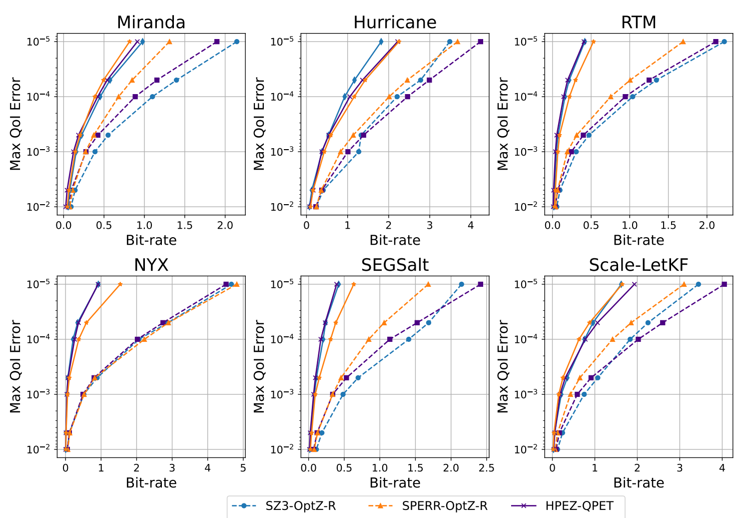

Fig. 4. Bit rate and Max QoI error plots for $Q(x) = x^3$.

[4] Jinyang Liu, Sheng Di, Kai Zhao, Xin Liang, Zizhong Chen, and Franck Cappello. 2022. Dynamic quality metric oriented error bounded lossy compression for scientific datasets. In *2022 SC22: International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*. IEEE Computer Society, 892–906.

[5] 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.

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

[7] 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),

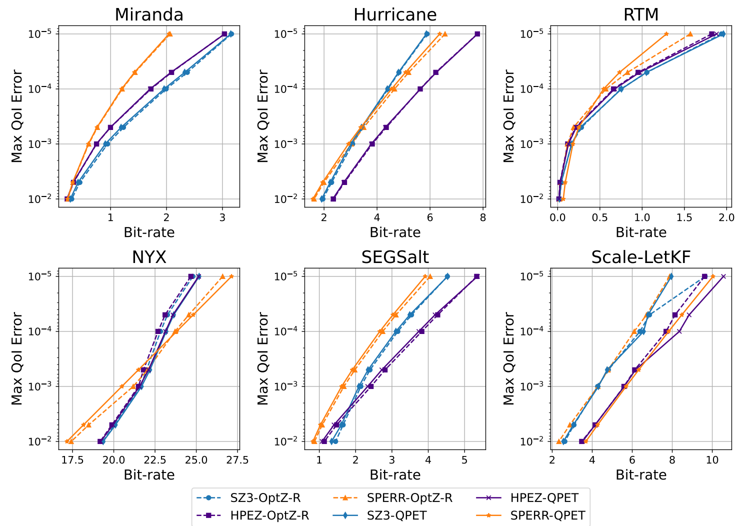

Fig. 5. Bit rate and Max QoI error plots for $Q(x) = \sin 10x$.

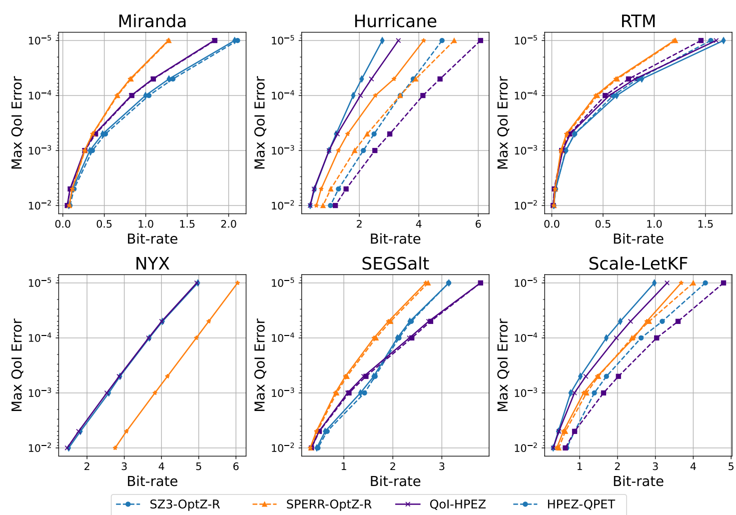

Fig. 6. Bit rate and Max QoI error plots for $Q(x) = \tanh x$.

285–303. https://doi.org/10.1177/1094342017737147

[8] Robert Underwood, Jon C Calhoun, Sheng Di, Amy Apon, and Franck Cappello. 2022. OptZConfig: Efficient Parallel Optimization of Lossy Compression Configuration. *IEEE Transactions on Parallel and Distributed Systems* (2022).

[9] Robert Underwood, Sheng Di, Jon C Calhoun, and Franck Cappello. 2020. Fraz: A generic high-fidelity fixed-ratio lossy compression framework for scientific floating-point data. In *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 567–577.

[10] 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.

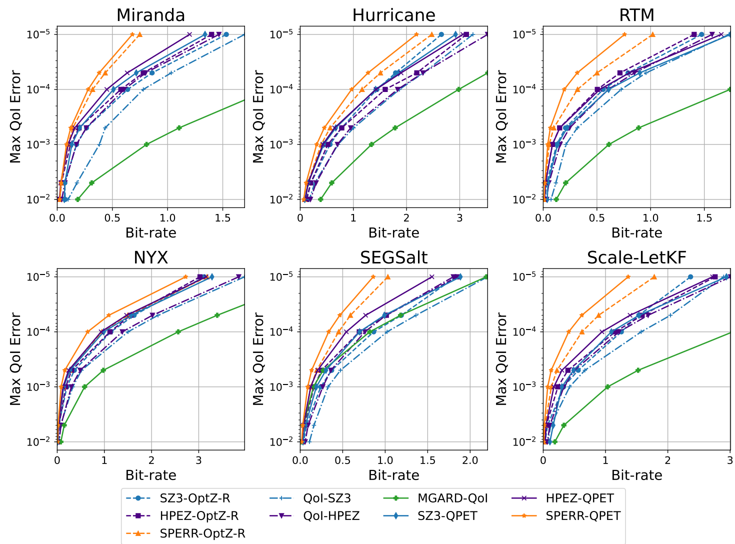

Fig. 7. Bit rate and Max QoI error plots for $Q(X) = \frac{1}{n_b}\sum x$, $n_b = 4^3$, i.e. average x on 4x4x4 blocks.

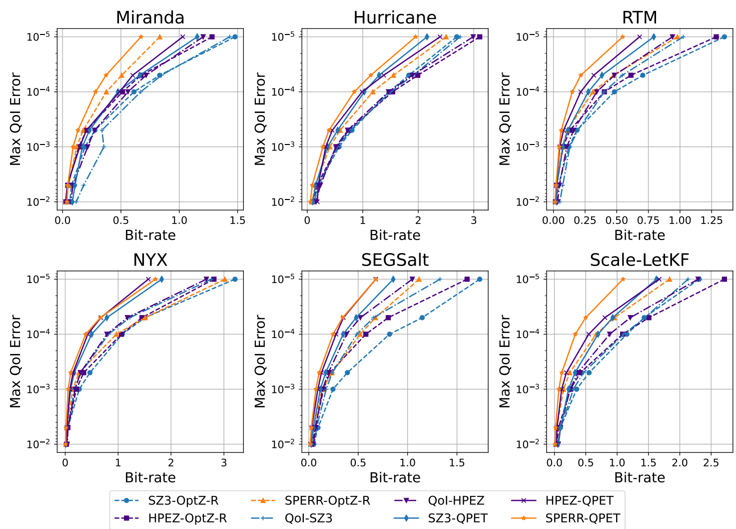

Fig. 8. Bit rate and Max QoI error plots for $Q(X) = \frac{1}{n_b}\sum x^2$, $n_b = 4^3$, i.e. average x^2 on 4x4x4 blocks.

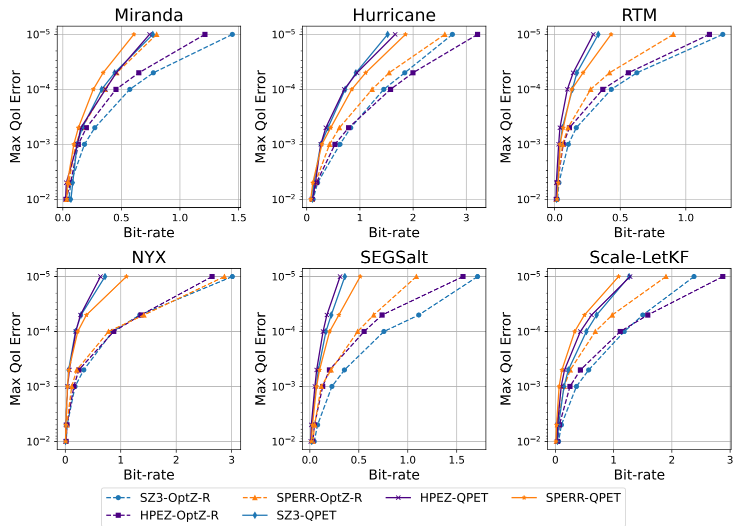

Fig. 9. Bit rate and Max QoI error plots for $Q(X) = \frac{1}{n_b}\sum x^3$, $n_b = 4^3$, i.e. average x^3 on 4x4x4 blocks.

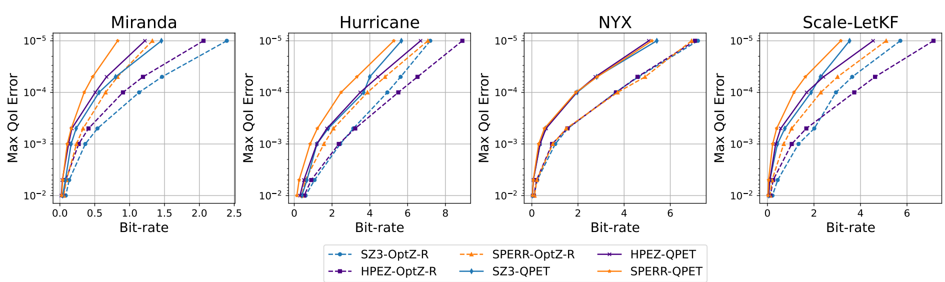

Fig. 10. Bit rate and Max QoI error plots for $Q(x, y, z) = x^2 + y^2 + z^2$.

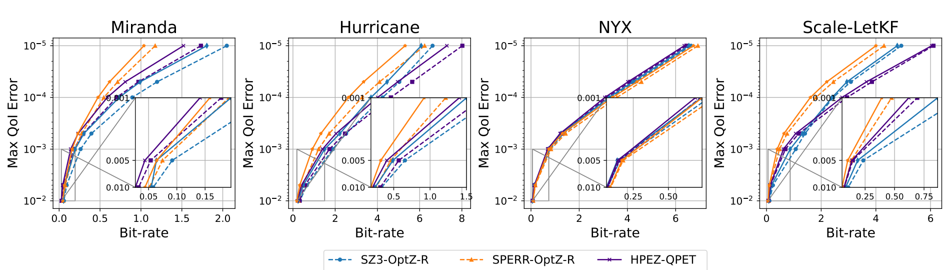

Fig. 11. Bit rate and Max QoI error plots for $Q(x, y, z) = \sqrt{x^2 + y^2 + z^2}$.