MEGA-PCC: A Mamba-based Efficient Approach for Joint Geometry and Attribute Point Cloud Compression

Supplementary Material

A. Attribute Distortion

Since the reconstructed geometry may be distorted, we first establish correspondences between the original and reconstructed point clouds before computing attribute distortion. To improve robustness, we compute distortion in both forward and backward directions. In the forward direction, each point p_i with attribute A_i in the original point cloud is matched to its nearest neighbor $\hat{p}_{j_i}$ with $\hat{A}_{j_i}$ in the reconstructed point cloud, and the mean squared error is computed as:

$$D_A^F = \frac{1}{N_F} \sum_{i=1}^{N_F} \left\| A_i - \hat{A}_{j_i} \right\|^2 \tag{1}$$

where N_F is the number of points in the original point cloud. Similarly, in the reverse direction, each reconstructed point $\hat{p}_j$ with attribute $\hat{A}_j$ in the reconstructed point cloud is matched to its nearest neighbor p_{j_i} with A_{j_i} in the original point cloud. The loss is computed as:

$$D_A^B = \frac{1}{N_B} \sum_{j=1}^{N_B} \left\| \hat{A}_j - A_{j_i} \right\|^2 \tag{2}$$

where N_B is the number of points in the reconstructed point cloud. The final distortion D_A is defined as the maximum of D_A^F and D_A^B. This bidirectional metric mitigates bias from asymmetric point matching and provides a more reliable measure of attribute distortion.

B. G-PCC / V-PCC parameter setting

The detailed parameter settings of G-PCCv23[2] and V-PCCv22 [1] are summarized in Table 1.

Table 1. Coding parameters in V-PCC and G-PCC

V-PCC					
	R1	R2	R3	R4	R5
Geometry QP	32	28	24	20	16
Texture QP	42	37	32	27	22
Occupancy Precision	4×4	4×4	4×4	4×4	2×2
G-PCC					
	R1	R2	R3	R4	R5
positionQuantizationScale (precision = 10)	1/8	1/4	1/2	3/4	7/8
positionQuantizationScale (precision = 11)	1/16	1/8	1/4	1/2	3/4
Texture QP	51	46	40	34	28
QPChromaOffset = -2					

Table 2. Complexity analysis in terms of model size and runtime

Method	**Model size** (MB)	**Runtime**	
		Enc (s)	**Dec (s)**
G-PCC	×	1.91	1.62
V-PCC	×	77.9	2.35
YOGA	4.4 (G) + 165.1 (A)	0.51 (G) / 7.64 (A)	0.56 (G) / 2.67 (A)
DeepPCC	5.3 (G) / 119 (A)	26.50	233.67
Unicorn	11.54 (G) / 65 (A)	0.9 (G) / 2.9 (A)	0.9 (G) / 1.6 (A)
MEGA-PCC	44.0	0.62	1.02

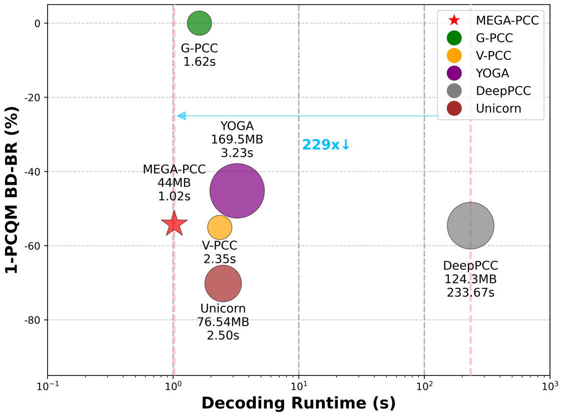

Figure 1. Comparison of compression performance and decoding time among G-PCC, V-PCC, YOGA, DeepPCC, and MEGA-PCC on the 8iVFB dataset, evaluated using the 1-PCQM metric, with G-PCC TMC13-v23 (Octree-RAHT) used as the anchor.

C. Complexity Analysis

To evaluate the computational complexity of different compression methods, we present the model sizes and the average encoding/decoding times on the 8iVFB [3] dataset for each method in Table 2. Figure 1 and Figure 2 provides a more comprehensive comparison, presenting model size, decoding time, encoding time and objective quality in terms of 1-PCQM [10]. G-PCC [2] and V-PCC [1] are CPU-based, while YOGA, DeepPCC, Unicorn, and MEGA-PCC are GPU-accelerated. All learned methods, including MEGA-PCC, are realized on an NVIDIA RTX 3090 GPU. For YOGA [8], DeepPCC [9], and Unicorn [5, 6], the runtimes are not obtained through our own execution but are directly taken from their original publications, as official implementations are not publicly available. As shown in Table 2, V-PCC shows significantly longer encoding times than G-PCC due to patch packing and projection steps required for image-based compression. YOGA, though im-

Table 3. BD-Rate (%) comparison for geometry and attribute distortion relative to G-PCCv23

Sequence	V-PCCv22 [1]				YOGA [8]				DeepPCC [9]				JPEG Pleno VM [4]				Unicorn [5]				Proposed			
	D2	YUV	GraphSim	PQI	D2	YUV	GraphSim	PQI	D2	YUV	GraphSim	PQI	D2	YUV	GraphSim	PQI	D2	YUV	GraphSim	PQI	D2	YUV	GraphSim	PQI
longdress	-13.7	-58.2	-40.2	-55.8	-74.9	-37.6	-	-	-62.7	-41.2	-	-	-	-	-	-	-	-	-	-	-62.3	-35.0	-40.4	-54.7
loot	-45.4	-68.0	-66.2	-70.6	-79.0	-51.3	-	-	-73.6	-35.9	-	-	-	-	-	-	-	-	-	-	-70.4	-45.0	-49.1	-64.4
redandblack	-24.5	-61.1	-53.8	-61.1	-76.9	-49.0	-	-	-54.6	-39.8	-	-	-	-	-	-	-	-	-	-	-59.9	-47.4	-50.1	-53.0
soldier	-20.5	-54.2	-57.3	-60.3	-74.8	-48.9	-	-	-67.3	-48.6	-	-	-32.7	-47.4	58.8	-45.0	-	-	-	-	-66.6	-42.3	-50.2	-60.9
basketball_player	-73.6	-60.7	-58.3	-79.4	-34.9	50.3	-	-	-86.7	-22.3	-	-	-	-	-	-	-	-	-	-	-82.0	-6.5	-26.3	-64.9
dancer	-70.0	-60.8	-51.6	-78.4	-24.3	61.1	-	-	-82.2	-18.1	-	-	-	-	-	-	-	-	-	-	-79.2	-4.4	-15.9	-65.1
Average	-41.3	-60.5	-54.6	-67.6	-60.8	-12.6	-	-	-71.2	-34.3	-	-	-32.7	-47.4	-58.8	-45.0	-	-	-	-	-70.1	-30.1	-38.7	-60.5

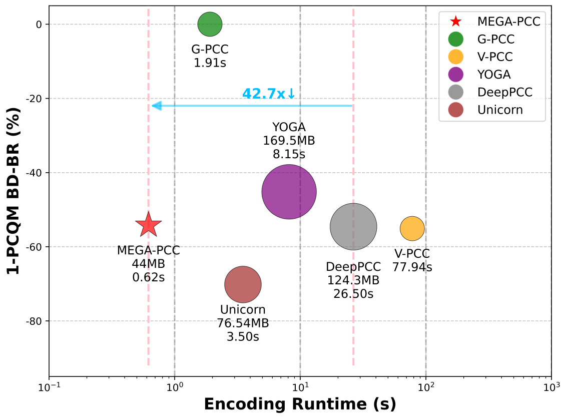

Figure 2. Comparison of compression performance and encoding time among G-PCC, V-PCC, YOGA, DeepPCC, and MEGA-PCC on the 8iVFB dataset, evaluated using the 1-PCQM metric, with G-PCC TMC13-v23 (Octree-RAHT) used as the anchor.

proving quality over G-PCC, introduces additional runtime overhead due to its two-stage pipeline, which uses G-PCC as a base layer. DeepPCC suffers from a decoding bottleneck caused by its autoregressive model, with decoding times exceeding 233 seconds. Unicorn achieves the best compression performance but relies on a complex multi-scale entropy model and a large network, leading to high computational costs and long runtimes. In contrast, MEGA-PCC offers an excellent trade-off between model size, runtime, and performance. It achieves the smallest model size (44.0 MB) and the lowest encoding and decoding times among GPU-based methods, while outperforming YOGA and DeepPCC in terms of 1-PCQM performance. Note that the reported encoding/decoding times for these learned PCC methods exclude recoloring time. MEGA-PCC avoids this processing entirely—requiring no recoloring or bit allocation—thanks to its unified design. The significant gap in runtime, faster decoding compared to existing methods clearly highlights MEGA-PCC's practicality for real-time applications.

D. Details of Rate-Distortion Curve

Table 3 summarizes the BD-rate results of MEGA-PCC alongside several existing joint compression methods, evaluated in terms of D2-PSNR, YUV-PSNR, GraphSIM [7], and PQI [10], using G-PCC v23 as the anchor. Figure 6 further illustrates the corresponding rate-distortion curves for D1-PSNR, Y-PSNR, GraphSIM, and PQI. Due to limited public availability of implementations, not all metrics are reported for every method. The results show that YOGA and DeepPCC achieve the best geometric compression in terms of D1-PSNR, while V-PCC leads in attribute compression as measured by Y-PSNR. MEGA-PCC ranks second in Y-PSNR, and notably, its performance gap over YOGA and Deep-PCC becomes even more pronounced on high-precision (11-bit) sequences, demonstrating strong generalization and robustness across varying data precision. Notably, due to the non-convex nature of DeepPCC's R-D curve, the BD-rate values reported for Y-PSNR and YUV-PSNR on 11-bit sequences may not fully reflect its actual performance where DeepPCC visually underperforms G-PCC in the corresponding R-D curves. Moreover, MEGA-PCC achieves compression quality comparable to V-PCC under the PQI metric, which jointly considers geometric and attribute fidelity. For the GraphSIM metric, which emphasizes local structural consistency rather than individual attribute values, MEGA-PCC underperforms on sequences with relatively monotone colors (e.g., loot, basketball_player, dancer). This is partly due to the advantage of methods like V-PCC and JPEG Pleno, which project point clouds to 2D images and apply mature image compression techniques, improving local color smoothness.

In contrast, PQI adopts a multiscale strategy to identify salient features and captures both global and local distortions across geometry and attributes equally. Unlike GraphSIM's radius-based local neighborhoods, PQI provides a more holistic and perceptually aligned assessment of reconstruction quality. Under this metric, MEGA-PCC performs comparably to V-PCC, reinforcing its strength as a balanced and efficient solution for joint geometry and attribute compression

E. Qualitative Comparisons

Figure 3 and 4 shows visualizations of the reconstructed soldier sequence using several methods under both high- and low-quality settings, with their respective bit allocations illustrated. Notably, G-PCC allocates a greater portion of bits to geometry, whereas V-PCC and JPEG Pleno prioritize attribute information. These differing allocation strategies lead to varying reconstruction quality, with V-PCC and

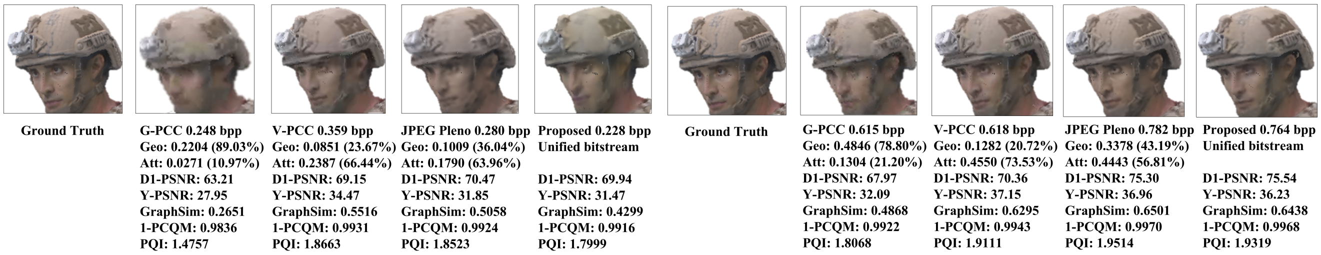

Figure 3. Visual comparison on *soldier*.

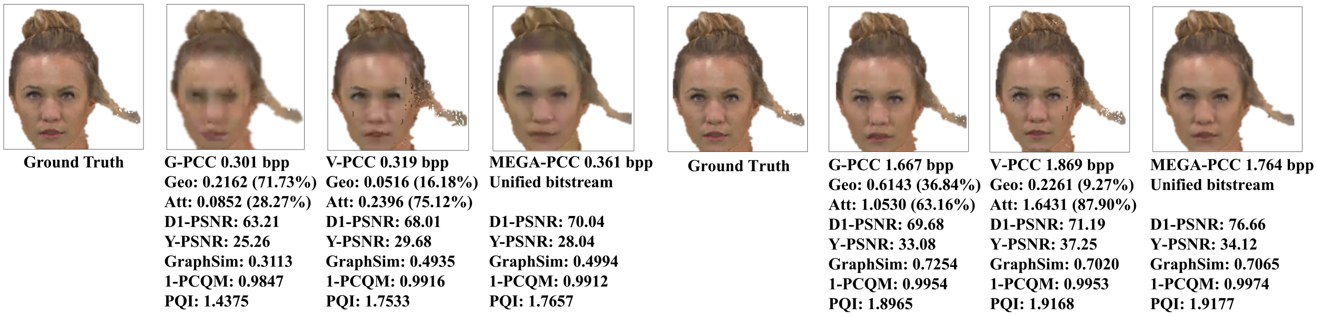

Figure 4. Visual comparison on *longdress*.

JPEG Pleno achieving superior results in attribute fidelity. This contrast highlights the importance of task-specific bit allocation to optimize both geometry and attribute reconstruction.

In comparison, the proposed MEGA-PCC method adopts a unified bitstream, where bit allocation is adaptively learned during training via a proper loss function. This approach effectively leverages the correlation between geometry and attributes, optimizing performance in a unified manner.

F. Ablation Study on Channel Flip Scanning in Mamba

Figures 7 present the R-D curves of the ablation study on the channel Flip scanning in Mamba. The results collectively support the hypothesis that capturing both spatial and channel-wise correlations is essential for enhancing representation quality and compression efficiency in point cloud processing.

G. Ablation Study on Single Decoder

To examine the effectiveness of our dual-decoder design, we implemented a single-decoder variant using the same Mamba backbone as shown in Figure 5. Unlike our proposed framework, which reconstructs geometry first and then attributes with dedicated decoders, the single-decoder attempts to jointly decode both modalities from the shared latent features in a single step.

As shown in Figure 8 and Table 4, the dual-decoder achieves substantial bit-rate reductions compared to the single-decoder across all test sequences and metrics. On average, the dual-decoder yields BD-BR savings of 40.85% (D1-PSNR), 52.05% (Y-PSNR), and 44.73% (1-PCQM), with additional reductions of 47.29% (GraphSIM) and 46.34% (PQI). The improvement is especially pronounced for attribute quality, where bit-rate savings exceed 50%, underscoring the importance of a dedicated attribute decoder. These results demonstrate that the gains of MEGA-PCC are not solely due to the Mamba backbone but primarily arise from the dual-decoder architecture. By separating the tasks, geometry decoding provides spatial guidance for attribute reconstruction, while attribute decoding benefits from a specialized pathway tailored to its higher complexity. In contrast, the single-decoder struggles to balance both tasks simultaneously, leading to higher bitrates and degraded reconstruction quality.

Table 4. BD-BR (%) of MEGA-PCC against Single-decoder.

	BD-BR (%)				
Sequence	D1-PSNR	Y-PSNR	1-PCQM	GraphSIM	PQI
longdress	-29.56	-44.05	-30.66	-43.94	-40.70
loot	-42.90	-45.34	-43.53	-47.54	-43.79
redandblack	-41.42	-56.88	-56.57	-58.09	-58.66
soldier	-33.48	-41.73	-32.87	-31.85	-23.54
basketball_player	-49.84	-66.92	-55.19	-56.38	-56.24
dancer	-47.88	-57.36	-49.58	-45.92	-55.12
Average	-40.85	-52.05	-44.73	-47.29	-46.34

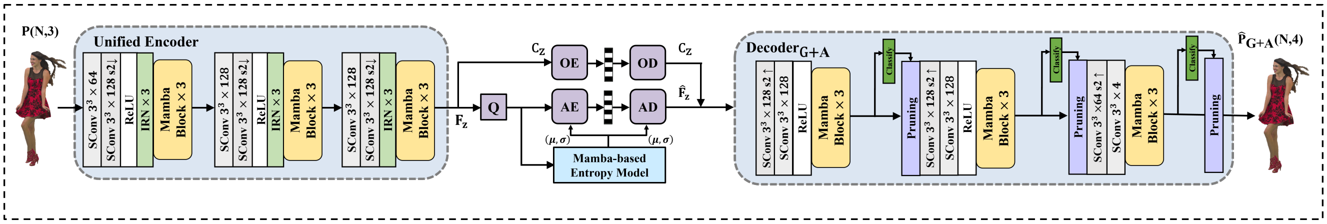

Figure 5. Overview of the Single-decoder.

Table 5. BD-BR (%) of MEGA-PCC against One-Stage training strategy.

	BD-BR (%)				
Sequence	D1-PSNR	Y-PSNR	1-PCQM	GraphSIM	PQI
longdress	-7.82	-88.57	-13.96	-91.59	-56.35
loot	-10.84	-13.53	-5.57	-13.22	-11.14
redandblack	-9.86	-42.66	-25.37	-39.66	-39.63
soldier	-9.95	-17.29	-11.67	-5.63	-7.21
basketball_player	-12.88	-60.75	-4.70	-13.23	-7.64
dancer	-12.17	-12.19	-4.69	-12.00	-12.88
Average	-10.59	-39.16	-11.09	-28.66	-21.48

H. Ablation Study on One-Stage Vs Two-Stage Training

Figure 9 and Table 5, compare MEGA-PCC trained with the proposed two-stage strategy against a one-stage baseline, where geometry and attributes are optimized jointly from the start. The results show that two-stage training consistently achieves lower bitrates across all metrics, with average BD-BR savings of 10.59% (D1-PSNR), 39.16% (Y-PSNR), 11.09% (1-PCQM), 28.66% (GraphSIM), and 21.48% (PQI). The improvements are most pronounced in attribute fidelity (Y-PSNR), indicating that attributes are particularly sensitive to training strategy. These findings confirm that decoupling training into two stages stabilizes optimization, reduces bitrate, and yields better rate–distortion performance compared to the one-stage approach.

References

[1] 3DG. V-pcc codec description. *ISO/IEC JTC 1/SC 29/WG 7 N00100*, 2022. 1, 2

[2] WG 7 and MPEG 3D Graphics Coding. G-pcc codec description v12. *ISO/IEC JTC1/SC29/WG7 N00271*, 2022. 1

[3] Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A. Chou. 8i voxelized full bodies-a voxelized point cloud dataset. *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, 2017. 1

[4] André F. R. Guarda, Nuno M. M. Rodrigues, and Fernando Pereira. The jpeg pleno learning-based point cloud coding standard: Serving man and machine. *In IEEE Access*, 13: 43289–43315, 2025. 2

[5] Jianqiang Wang, Ruixiang Xue, Jiaxin Li, Dandan Ding, Yi Lin, and Zhan Ma. A versatile point cloud compressor using universal multiscale conditional coding - part ii: Attribute. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 1, 2

[6] Jianqiang Wang, Ruixiang Xue, Jiaxin Li, Dandan Ding, and Yi Lin; Zhan Ma. A versatile point cloud compressor using universal multiscale conditional coding - part i: Geometry. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 1

[7] Qi Yang, Zhan Ma, Yiling Xu, Zhu Li, and Jun Sun. Inferring point cloud quality via graph similarity. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2

[8] Junteng Zhang, Tong Chen, Dandan Ding, and Zhan Ma. Yoga: Yet another geometry-based point cloud compressor. *In Proceedings of the ACM International Conference on Multimedia*, pages 9070–9081, 2023. 1, 2

[9] Junzhe Zhang, Gexin Liu, Junteng Zhang, Dandan Ding, and Zhan Ma. Deeppcc: Learned lossy point cloud compression. *In IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 18–37, 2024. 1, 2

[10] Junzhe Zhang, Tong Chen, Dandan Ding, and Zhan Ma. Revisit point cloud quality assessment: Current advances and a multiscale-inspired approach. *In IEEE Transactions on Visualization and Computer Graphics*, 2025. 1, 2

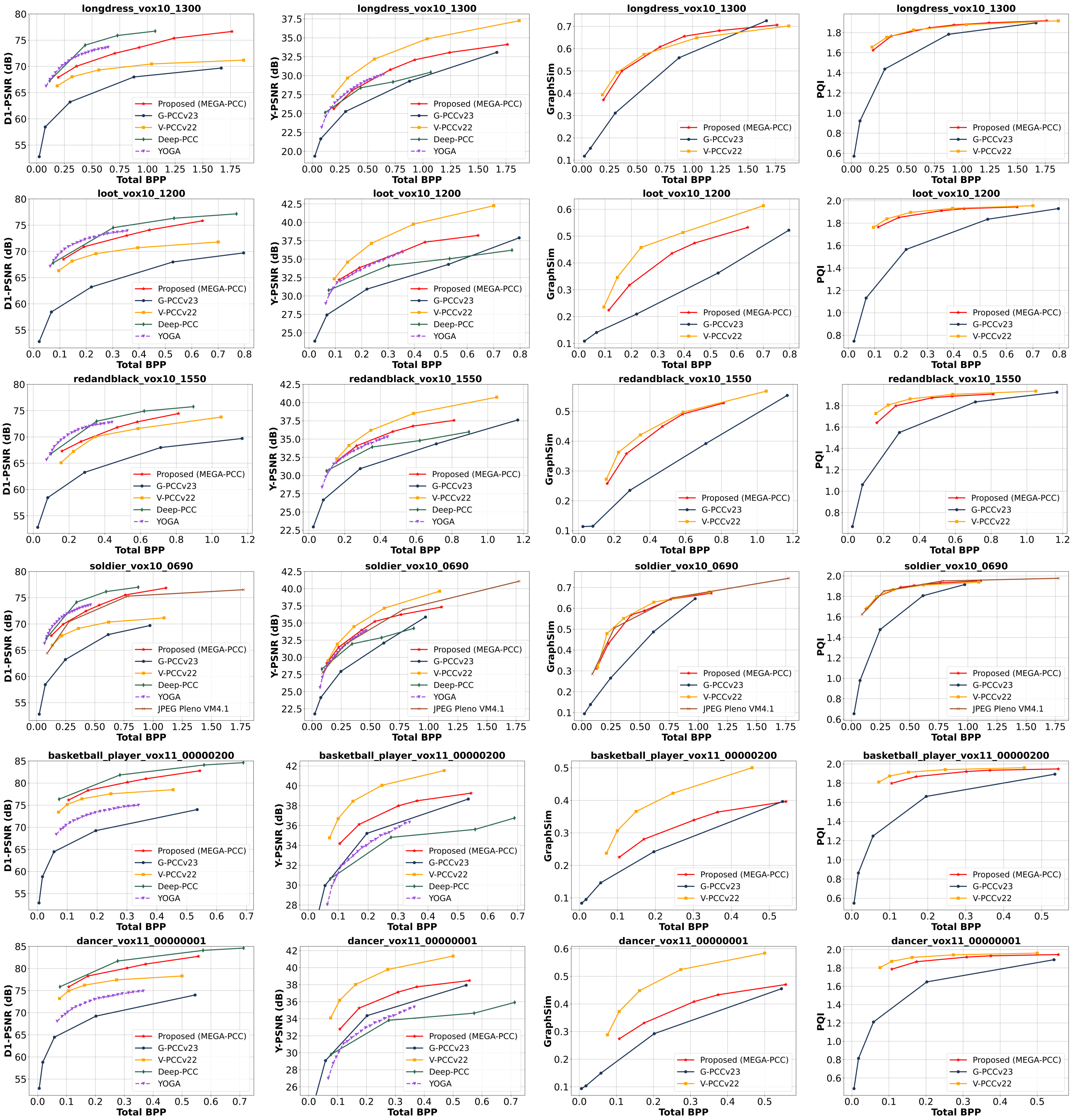

Figure 6. Rate-distortion performance in terms of various quality metric.

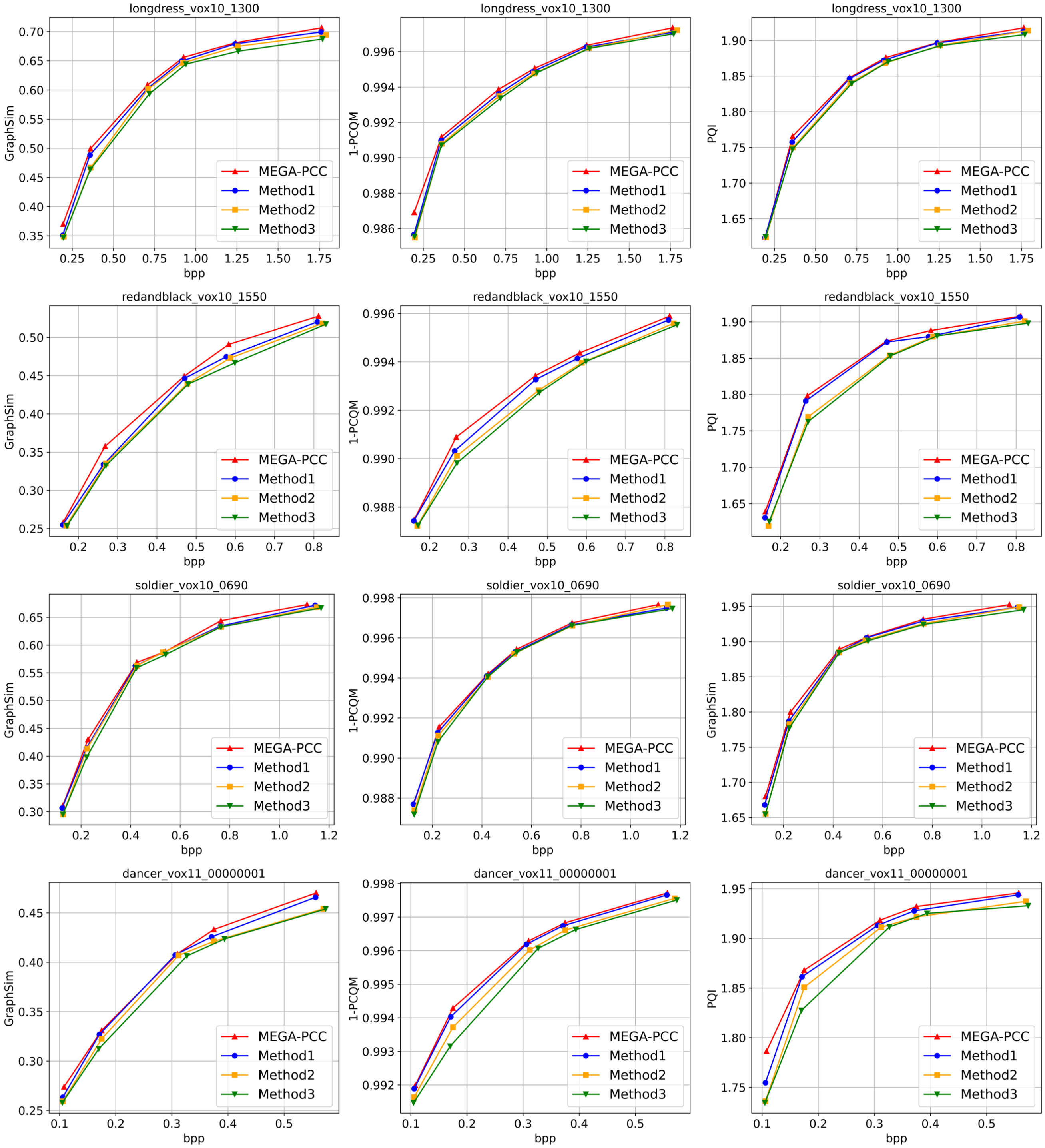

Figure 7. Ablation study on Channel Flip Scanning in Mamba.

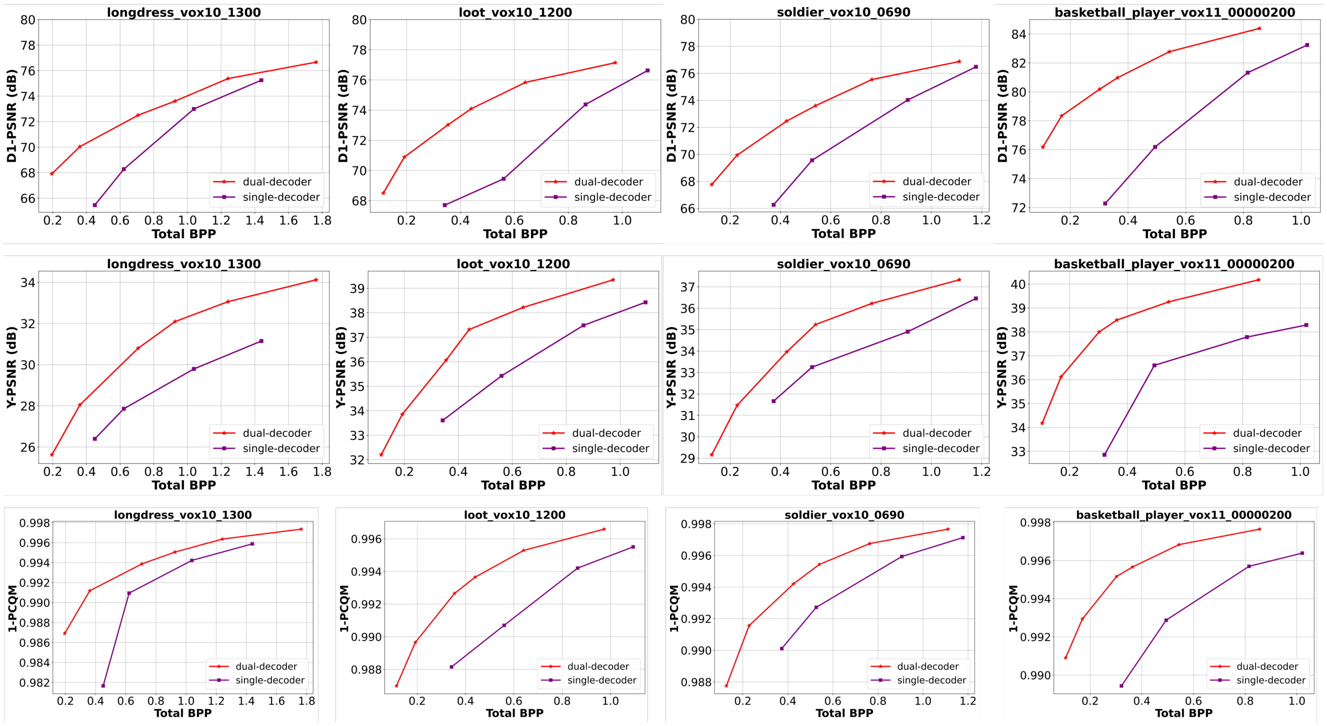

Figure 8. Ablation study on Single-decoder and Dual-decoder.

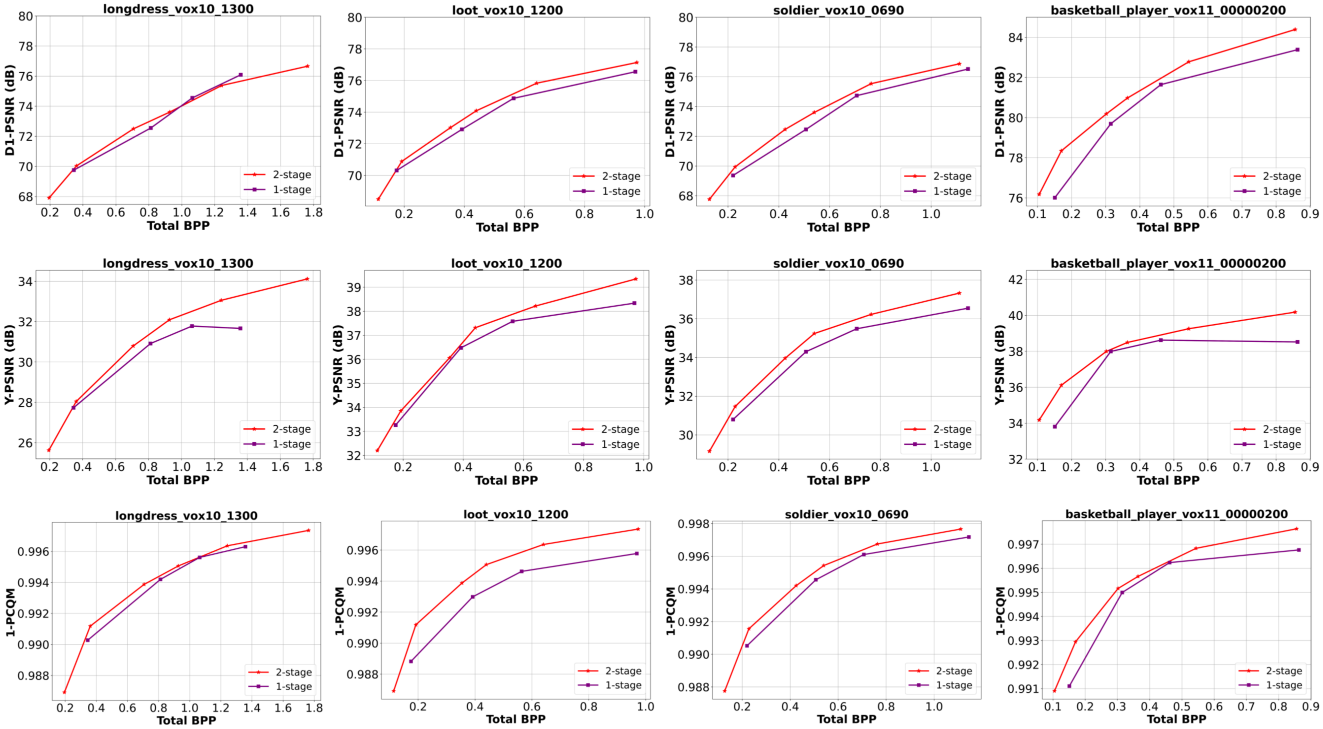

Figure 9. Ablation study on one-stage and two-stage training