Supplementary Material

1 NOTATION

Notation is given in Table S1.

2 ADDITIONAL EXPERIMENTAL DETAILS

Table S2 lists the point clouds in our dataset and the number of points in each point cloud. The point clouds in the top part of the table are point clouds derived from meshes created in Guo et al. (2019); Meka et al. (2020), while the point clouds in the bottom part of the table are MPEG point clouds d'Eon et al. (2017); Alliez et al. (2017). The point clouds are shown in Fig. 1 in the main text and in Fig. S1 below. Point cloud visualization is performed in Meshlab Cignoni et al. (2008).

Figure S1: Point clouds *longdress*, *loot*, *redandblack*, *soldier*, *mask*, *shiva*, and *klimt*.

We implement the LVAC framework in Python using Tensorflow. The entire point cloud constitutes one batch. All configurations are trained in about 25K steps using the Adam optimizer and a learning rate of 0.01, with low bit rate configurations typically taking longer to converge. Each step takes 0.5-3.0 s on an NVIDIA P100 class GPU in eager mode with various debugging checks in place. Training loss as a function of steps for point cloud *rock* is shown in Fig. S2.

3 COORDINATE BASED NETWORKS

Figure S3 shows the RD performance of different networks: (left) *mlp(35x256x3)*, (middle) *mlp(35x64x3)*, and (right) *pa(3x32x3)* for point cloud *rock*, along with baselines. We give the corresponding plots for other point clouds in Fig. S7, Fig. S8, Fig. S9, and Fig. S10. At higher bit rates, higher target levels perform better.

4 GENERALIZATION

In the main text, we provide the RD performance on point cloud *rock* using CBNs pre-trained on point cloud *basketball*. The corresponding plots for other point clouds are given in Fig. S11 and Fig. S12.

symbol	description
$\mathbf{x}$	position vector in $\mathbb{R}^d$, typically $d = 3$ with $\mathbf{x} = (x, y, z)$
$\mathbf{y}$	attribute vector in $\mathbb{R}^r$, typically $\mathbf{y} = (R, G, B)$
$f : \mathbf{x} \mapsto \mathbf{y}$	volumetric function
$f_{\theta,\mathbf{Z}} : \mathbf{x} \mapsto \mathbf{y}$	volumetric function in family parameterized by θ, $\mathbf{Z}$
$d(f, f_{\theta,\mathbf{Z}})$	distortion between two volumetric functions
$\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_p}$	point cloud with N_p points or N_p occupied voxels in voxel grid
L	binary tree depth, or $\log_2$ of number of all voxels in voxel grid
$\ell = 0, 1, \ldots, L$	level of binary tree between root ($\ell = 0$) and leaves ($\ell = L$)
$\mathbf{n}$, $\mathbf{n}_L$, and $\mathbf{n}_R$	positions of a block at level ℓ and its left and right children at level $\ell + 1$
B, $B_{\mathbf{n}}$, or $B_{\ell,\mathbf{n}}$	block at level ℓ and position (or *offset*) $\mathbf{n}$
$N_x \times N_y \times N_z$	dimensions of a block, in voxels
w, $w_{\mathbf{n}}$, or $w_{\ell,\mathbf{n}}$	weight of block (i.e., number of point cloud points $\mathbf{x}_i$ in block)
$\mathbf{z}$, $\mathbf{z}_{\mathbf{n}}$, or $\mathbf{z}_{\ell,\mathbf{n}}$	C-dimensional latent vector for block
$f_\theta(\mathbf{x} - \mathbf{n}; \mathbf{z}_{\mathbf{n}})$	continuous function representing attributes across block at offset $\mathbf{n}$
θ	parameters of the coordinate based network (CBN) representing f_θ
$\mathbf{Z}$	$N \times C$ matrix of latent vectors, at a target level having N blocks
$\mathbf{V}$	$N \times C$ matrix of transform coefficients of $\mathbf{Z}$
$\mathbf{T}_a$	$N \times N$ analysis transform for computing $\mathbf{V} = \mathbf{T}_a\mathbf{Z}$
$\mathbf{T}_s$	$N \times N$ synthesis transform for computing $\mathbf{Z} = \mathbf{T}_s\mathbf{V}$
$\mathbf{S}$	$N \times N$ diagonal matrix of spatial normalizing weights
$\mathbf{\Delta}$	$C \times C$ diagonal matrix of quantization stepsizes
$\bar{\mathbf{V}} = \mathbf{S}^{-1}\mathbf{V}$	$N \times C$ matrix of spatially-normalized transform coefficients
$\mathbf{U} = \bar{\mathbf{V}}\mathbf{\Delta}^{-1}$	$N \times C$ matrix of stepsize-scaled spatially-normalized coefficients
$\hat{\mathbf{U}} = \lfloor \mathbf{U} \rceil$	$N \times C$ matrix of integers for entropy coding
$\hat{\bar{\mathbf{V}}} = \hat{\mathbf{U}}\mathbf{\Delta}$	$N \times C$ matrix of quantized spatially-normalized transform coefficients
$\hat{\mathbf{V}} = \mathbf{S}\hat{\bar{\mathbf{V}}}$	$N \times C$ matrix of quantized transform coefficients
$\hat{\mathbf{Z}} = \mathbf{T}_s\hat{\mathbf{V}}$	$N \times C$ matrix of quantized latent vectors
$\hat{\theta}$	quantized parameters reconstructed at decoder along with $\hat{\mathbf{Z}}$
$f_{\hat{\theta},\hat{\mathbf{Z}}} : \mathbf{x} \mapsto \mathbf{y}$	reconstructed volumetric function $f_{\hat{\theta},\hat{\mathbf{Z}}}(\mathbf{x}) = \sum_{\mathbf{n}} f_{\hat{\theta}}(\mathbf{x} - \mathbf{n}; \hat{\mathbf{z}}_{\mathbf{n}}) \mathbb{1}_{B_{\mathbf{n}}}(\mathbf{x})$
D or $D(\hat{\theta}, \hat{\mathbf{Z}})$	distortion $d(f, f_{\hat{\theta},\hat{\mathbf{Z}}})$, typically $\sum_i \|\|\mathbf{y}_i - f_{\hat{\theta},\hat{\mathbf{Z}}}(\mathbf{x}_i)\|\|^2$
R or $R(\hat{\theta}, \hat{\mathbf{Z}})$	bit rate including bits both for $\hat{\theta}$ and $\hat{\mathbf{Z}}$
J or $J(\hat{\theta}, \hat{\mathbf{Z}})$	loss function, or Lagrangian $D + \lambda R$, with Lagrange multiplier $\lambda > 0$
linear(3x3)	linear model with 3 input and 3 output channels
mlp(35x256x3)	MLP with 35=3+C input and 3 output channels, and 256 hidden units
mlp(35x64x3)	MLP with 35=3+C input and 3 output channels, and 64 hidden units
pa(3x32x3)	PA network with 3 input and 3 output channels, and 32 modulation inputs
l=21, 24, 27, 30	target levels (in binary tree where CBNs are located)
CBN	Coordinate Based Network
LVAC	Learned Volumetric Attribute Coding (our framework)
RAHT	Region Adaptive Hierarchical Transform
RLGR	adaptive Run-Length Golomb-Rice entropy coder (Malvar, 2006)
cbe	Continuous Batched Entropy model
G-PCC	MPEG Geometry-based Point Cloud Codec
TMC13	Test Model 13 (software) for MPEG G-PCC
Deep PCAC	Deep Point Cloud Attribute Coder (Sheng et al., 2021)

Table S1. Notation

Table S2. Voxelized point clouds in our dataset.

Point Cloud	# points
rock	837434
chair	791416
scooter	959388
juggling	798441
basketball	868224
basketball2	948870
jacket	805882
longdress (aka *longdress_vox10_1300*)	857966
loot (aka *loot_vox10_1200*)	805285
redandblack (aka *redandblack_vox10_1550*)	757691
soldier (aka *soldier_vox10_0690*)	1089091
mask (aka *Egyptian_mask_vox12*)	272684
shiva (aka *Shiva_00035_vox12*)	1009132
klimt (aka *Staue_Klimt_vox12*)	499660

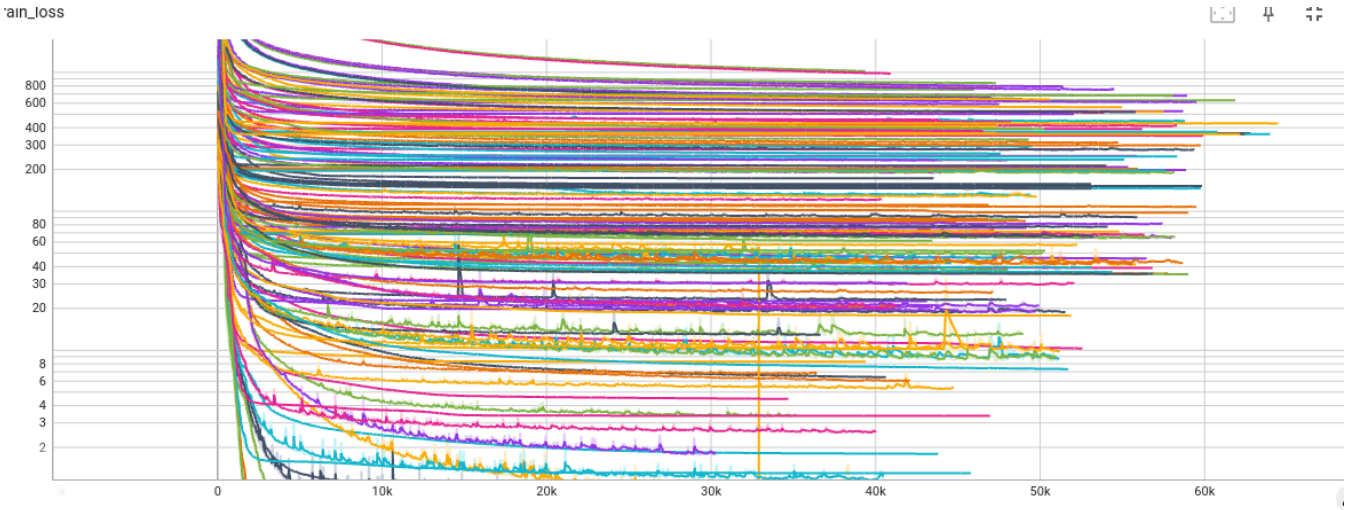


Figure S2: Training loss as a function of steps on point cloud *rock*.

5 SIDE INFORMATION

In the main text, we show the penalty required to transmit side information, for the entropy model and for the CBNs for point cloud *rock*. The corresponding plots for other point clouds are given in Fig. S14, Fig. S15, and Fig. S16.

6 NORMALIZATION

We show the bit rate improvement due to normalization for point cloud *rock* in the main text. We give the corresponding results for other point clouds in Fig. S17 and Tab. S3.

7 SUBJECTIVE QUALITY

Figure S4 shows further results on subjective quality.

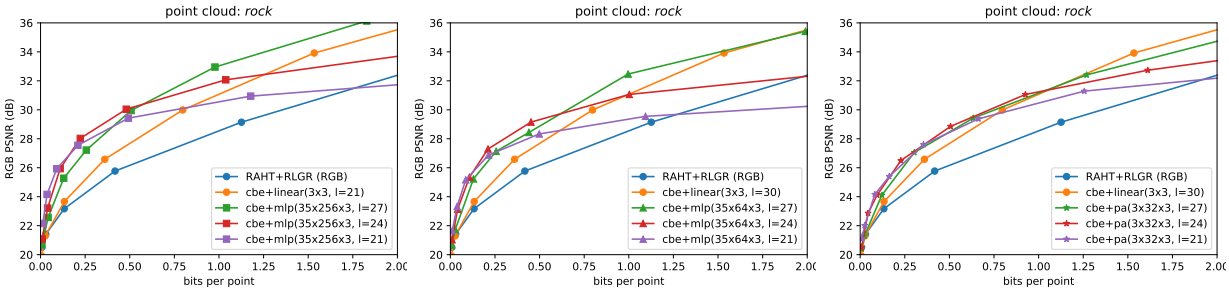


Figure S3: Coordinate Based Networks, by network. (left) *mlp(35x256x3)*. (middle) *mlp(35x64x3)*. (right) *pa(3x32x3)*.

8 ADDITIONAL POINT CLOUDS

Figure S5, Fig. S7, Fig. S9, Fig. S11, Fig. S12, Fig. S13, Fig. S14, Fig. S15, Fig. S16, Tab. S3, Fig. S17, and Fig. S18 show results for point clouds *chair*, *scooter*, *juggling*, *basketball*, *basketball2*, and *jacket* corresponding respectively to Fig. 5(a), Fig. 6, Fig. S3, Fig. 8, Fig. 9, Fig. 10, Table 1, Fig. 11, and Fig. 12 for *rock*. Figure S6, Fig. S8, and Fig. S10 show results for MPEG point clouds *longdress*, *loot*, *redandblack*, *soldier*, *mask*, *shiva*, and *klimt* corresponding respectively to Fig. 5(b), Fig. 6, and Fig. S3.

MPEG point cloud data from Fig. S8 and Fig. S10 are used in Tab. S4 to compare LVAC to Sparse PCAC (Wang and Ma, 2022), showing that LVAC and Sparse PCAC are roughly comparable. Sparse PCAC appeared while this work was in review, and Wang and Ma (2022) cite our LVAC pre-print, but we do our best to compare them here. Beware, however, that Figs. S8 and S10 — like Figs. 6 and S3 but unlike Figs. 8, 9, 10, 12, S14, S15, S16, and S18 — and hence Tab. S4, do not reflect the side information needed to completely specify the decoding. To assess this side information, please see Sec. 5.

0.129 bpp, 23.2 dB | 0.127 bpp, 24.3 dB | 0.127 bpp, 25.9 dB

0.540 bpp, 26.5 dB | 0.536 bpp, 28.2 dB | 0.531 bpp, 30.0 dB

1.005 bpp, 28.7 dB *RAHT-RLGR (RGB)* | 1.019 bpp, 31.3 dB *RAHT-RLGR (YUV)* | 1.003 bpp, 33.0 dB *LVAC mlp(35x256x3)*

Figure S4: Subjective quality of point cloud *rock*. Each row is a different bit rate. Original is shown in the main text. Zoom in to see differences.

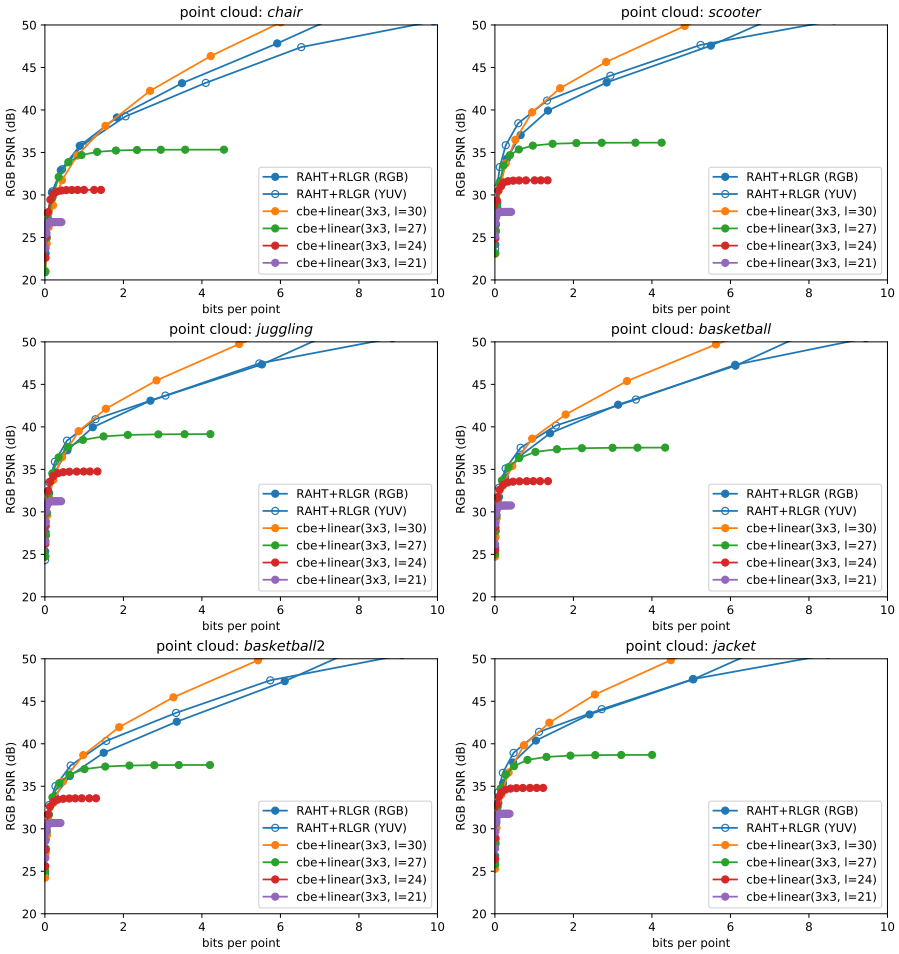


Figure S5: Baselines for six point clouds. *RAHT+RLGR (RGB)* and *(YUV)* are shown against 3×3 linear models at levels 30, 27, 24, and 21, which optimize the colorspace by minimizing $D + \lambda R$ using the *cbe* entropy model.

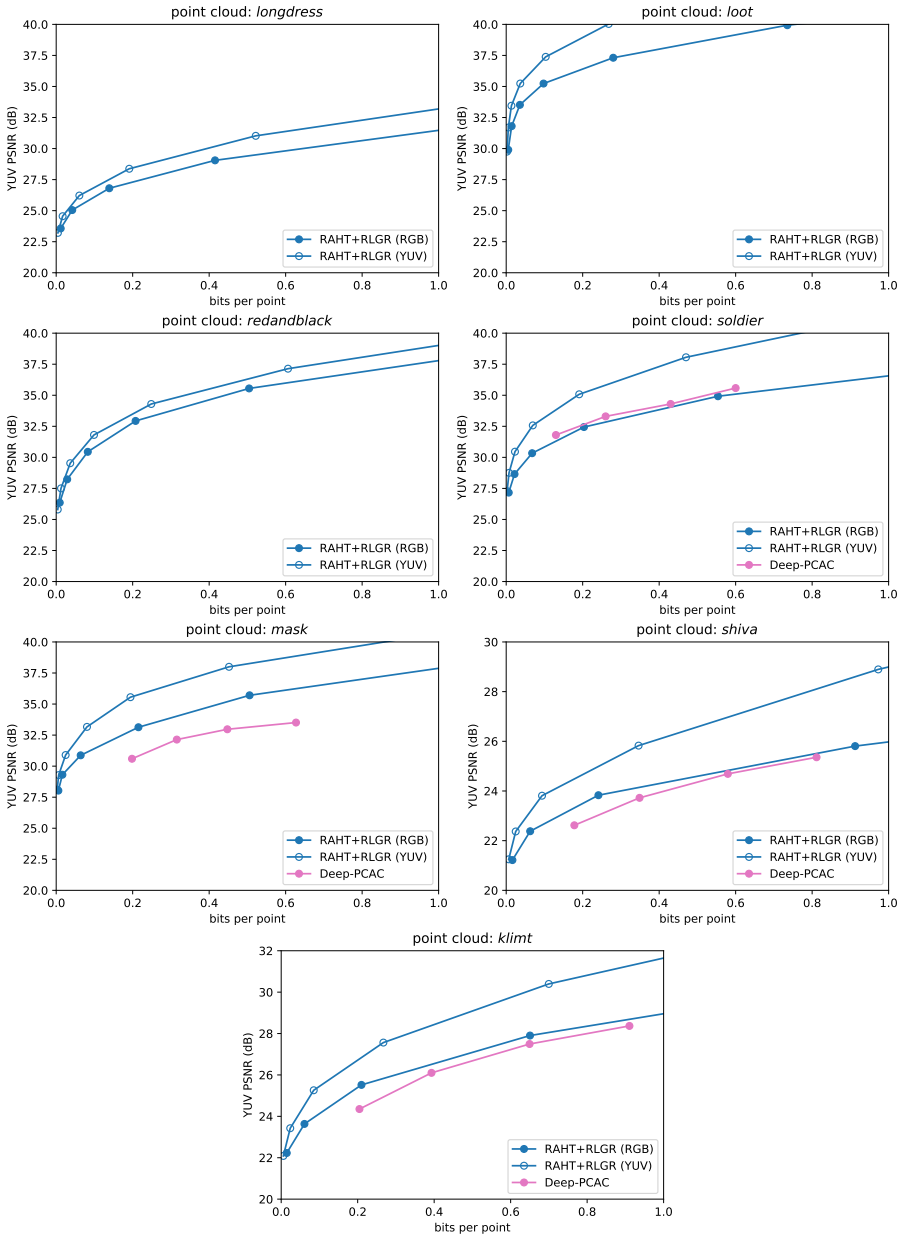


Figure S6: Baselines for MPEG point clouds: *RAHT+RLGR (RGB)*, *RAHT+RLGR (YUV)*, and (for some point clouds) Deep-PCAC Sheng et al. (2021). Baselines for other point clouds are shown in Fig. S5.

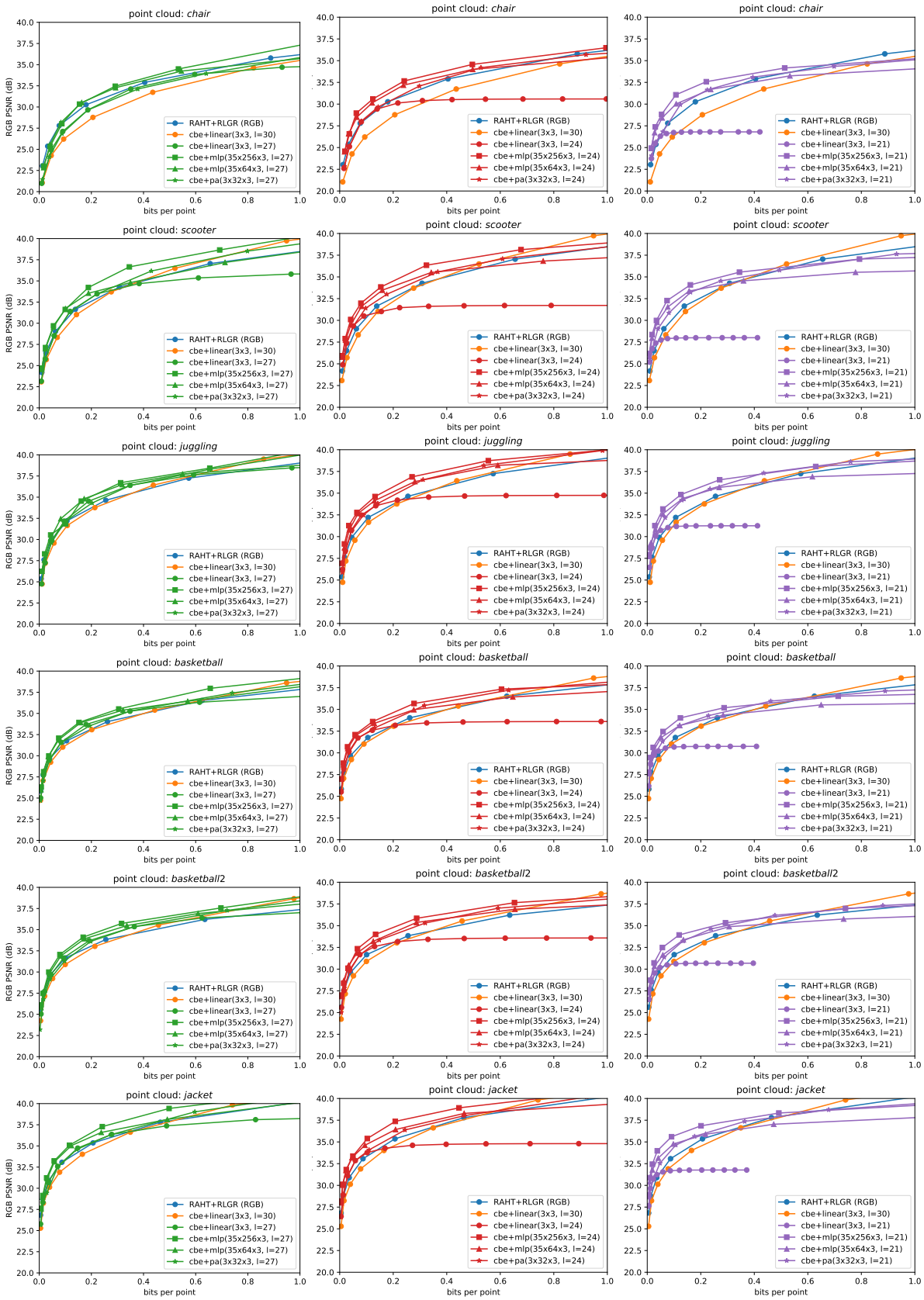


Figure S7: Coordinate Based Networks, by target level. Each row is a different point cloud. Left, middle, right columns each show *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)* CBNs, along with baselines, at levels 27, 24, 21.See Fig. S8 for other point clouds.

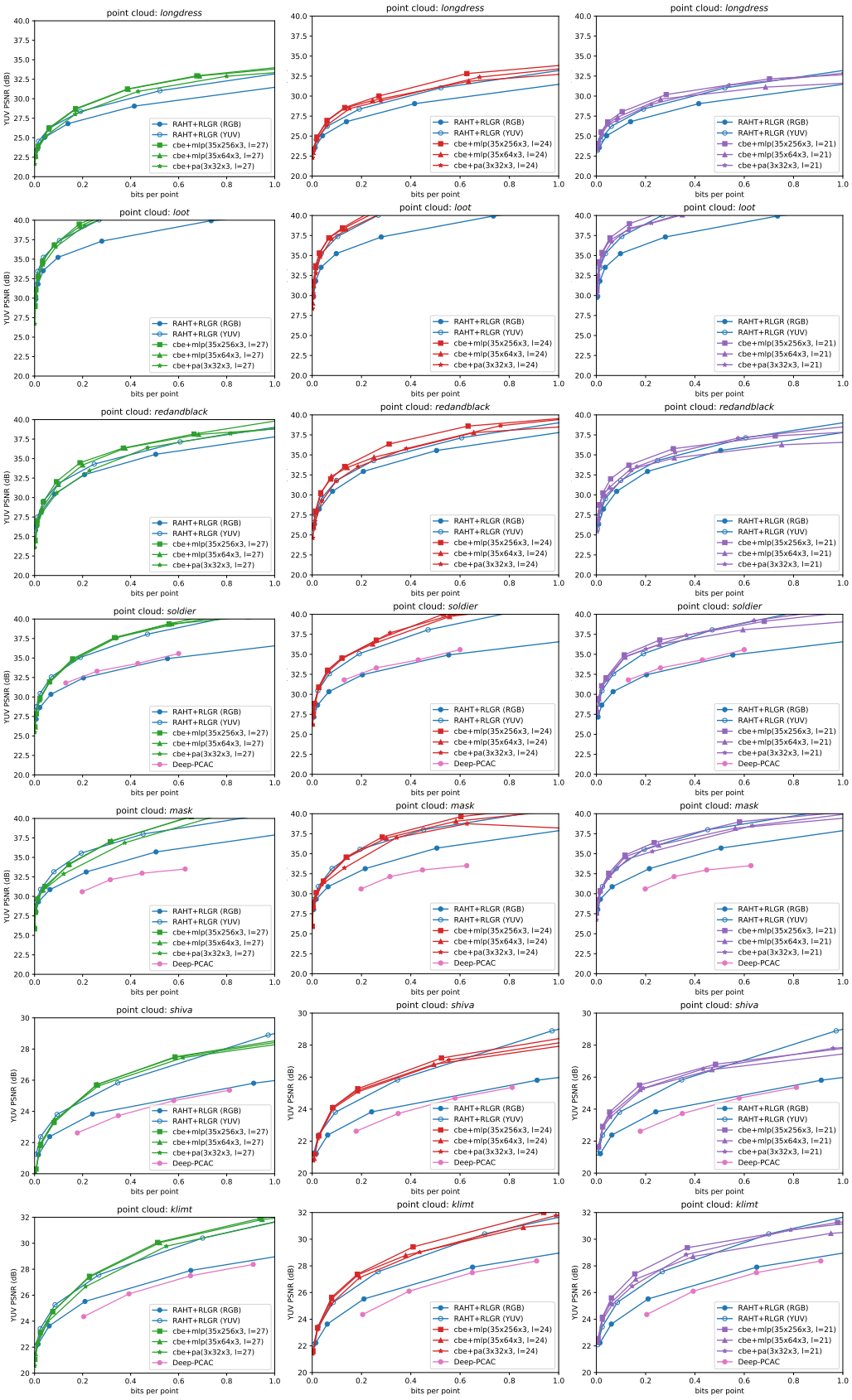


Figure S8: Coordinate Based Networks, by target level, for MPEG point clouds. Each row is a different point cloud. Left, middle, right columns each show *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)* CBNs, along with baselines, at levels 27, 24, 21. See Fig. S7 for other point clouds.

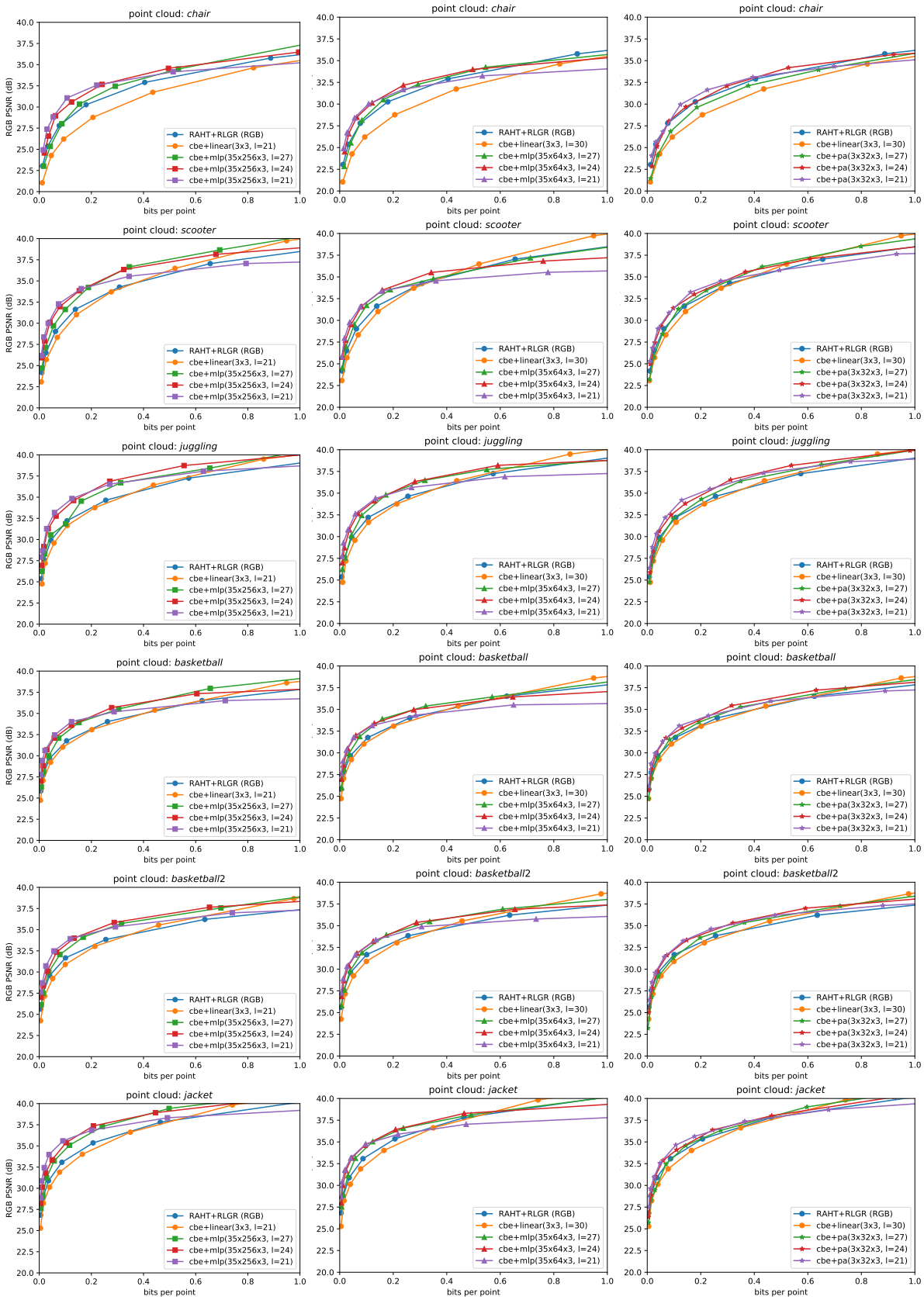


Figure S9: Coordinate Based Networks, by network. Each row is a different point cloud. Left, middle, right columns each show levels 27, 24, and 21, along with baselines, for CBNs *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)*. See Fig. S3 for point cloud *rock* and Fig. S10 for other point clouds.

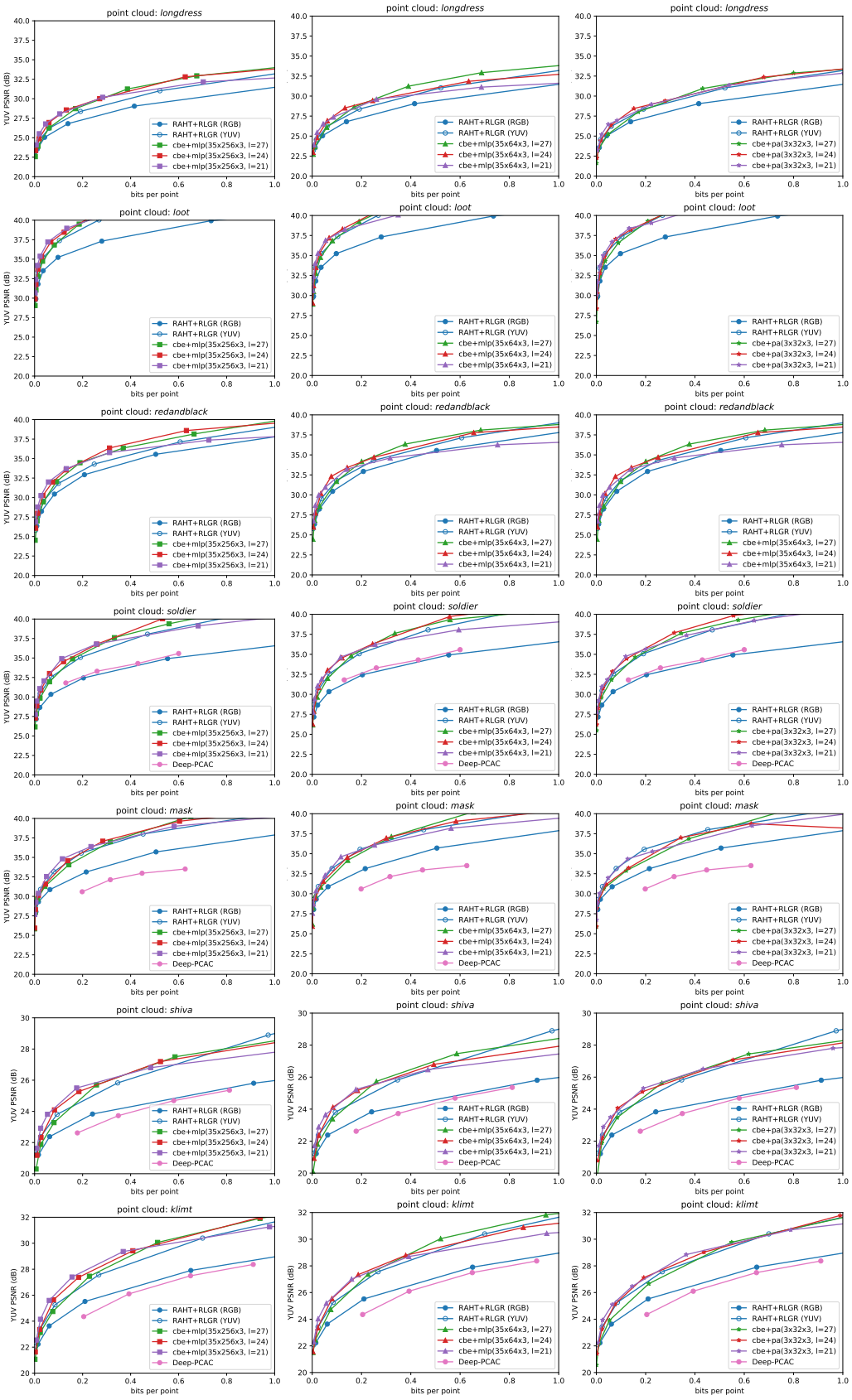


Figure S10: Coordinate Based Networks, by network, for MPEG point clouds. Each row is a different point cloud. Left, middle, right columns each show levels 27, 24, and 21, along with baselines, for CBNs *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)*. See Fig. S3 and Fig. S9 for other point clouds.

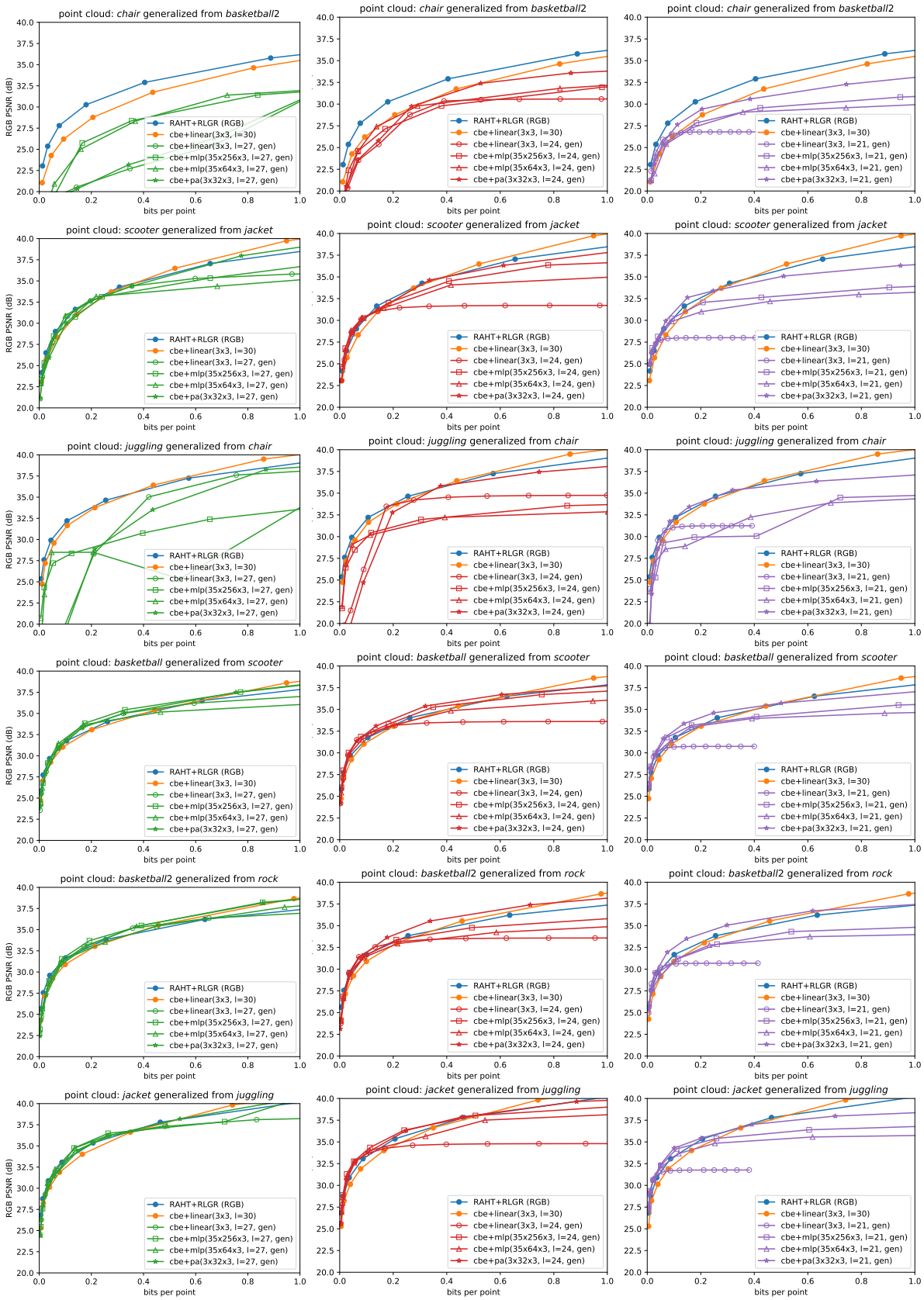


Figure S11: Coordinate Based Networks with generalization, by target level. Each row is a different point cloud. Left, middle, right columns each show *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)* CBNs, along with baselines, at levels 27, 24, 21.

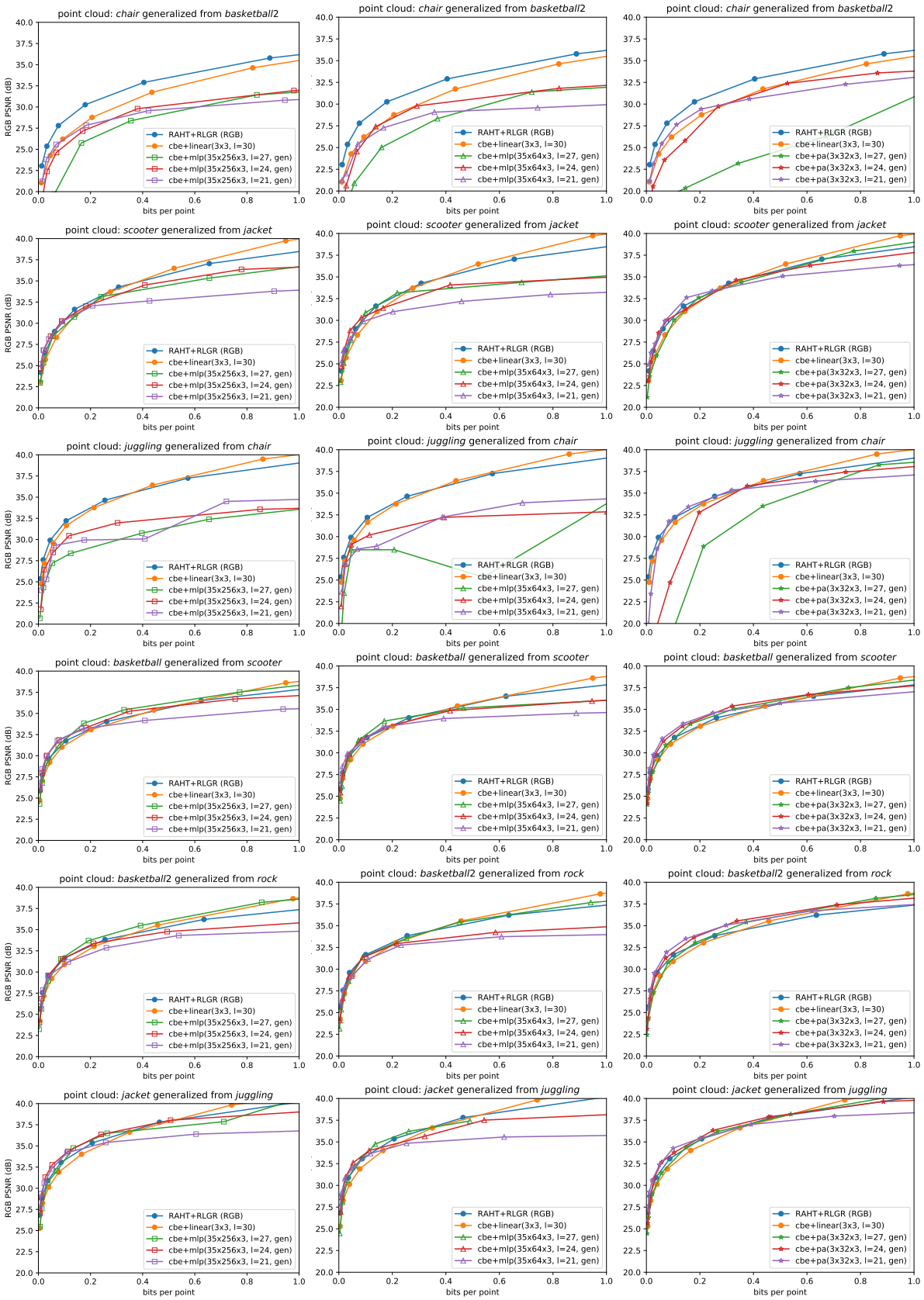


Figure S12: Coordinate Based Networks with generalization, by network. Each row is a different point cloud. Left, middle, right columns each show levels 27, 24, and 21, along with baselines, for CBNs *mlp(35x256x3)*, *mlp(35x64x3)*, and *pa(3x32x3)*.

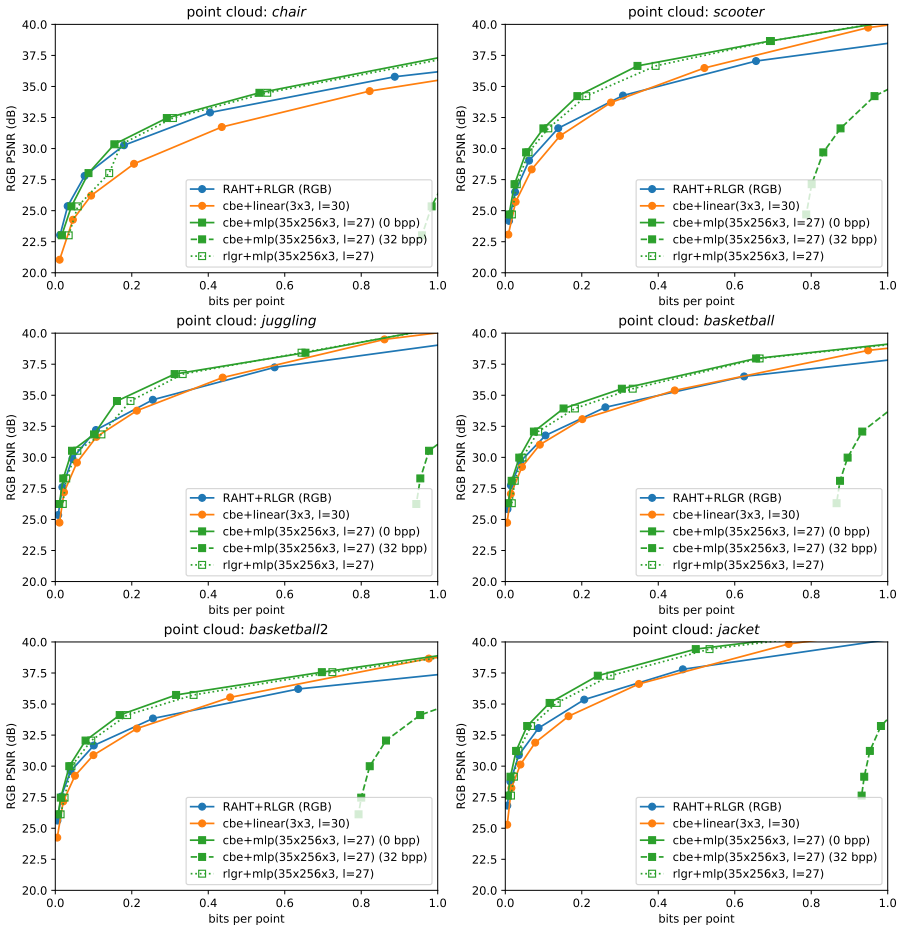


Figure S13: Side information for entropy model. Sending 32 bits per parameter for the *cbe* entropy model would reduce RD performance from solid to dashed green lines. But the backward-adaptive *RLGR* entropy coder (dotted, unfilled) obviates the need to send side information with almost no loss in performance.

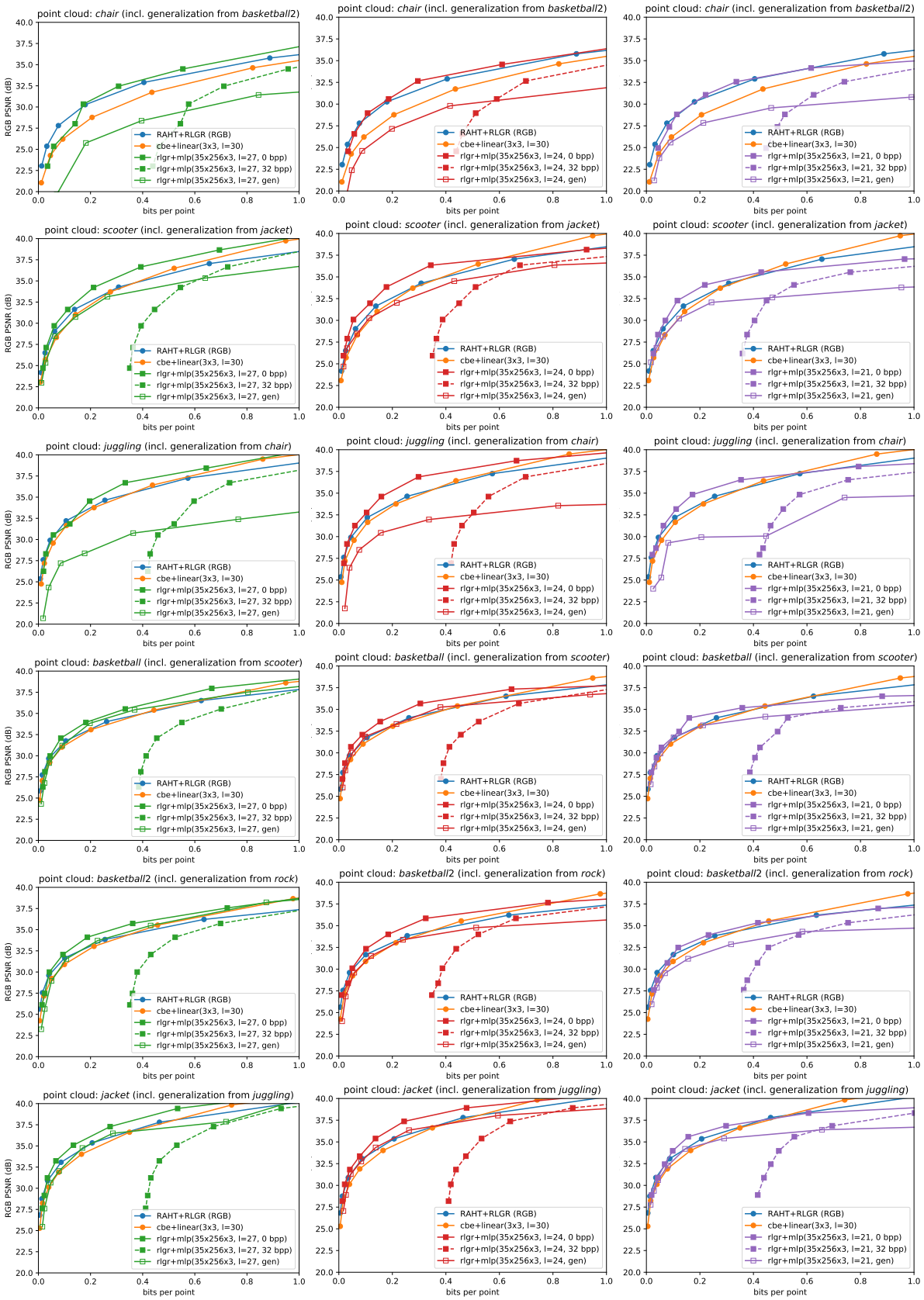


Figure S14: Effect of side information for coordinate based network *mlp(35x256x3)* at levels 27 (left), 24 (middle), and 21 (right). Each row is a different point cloud.

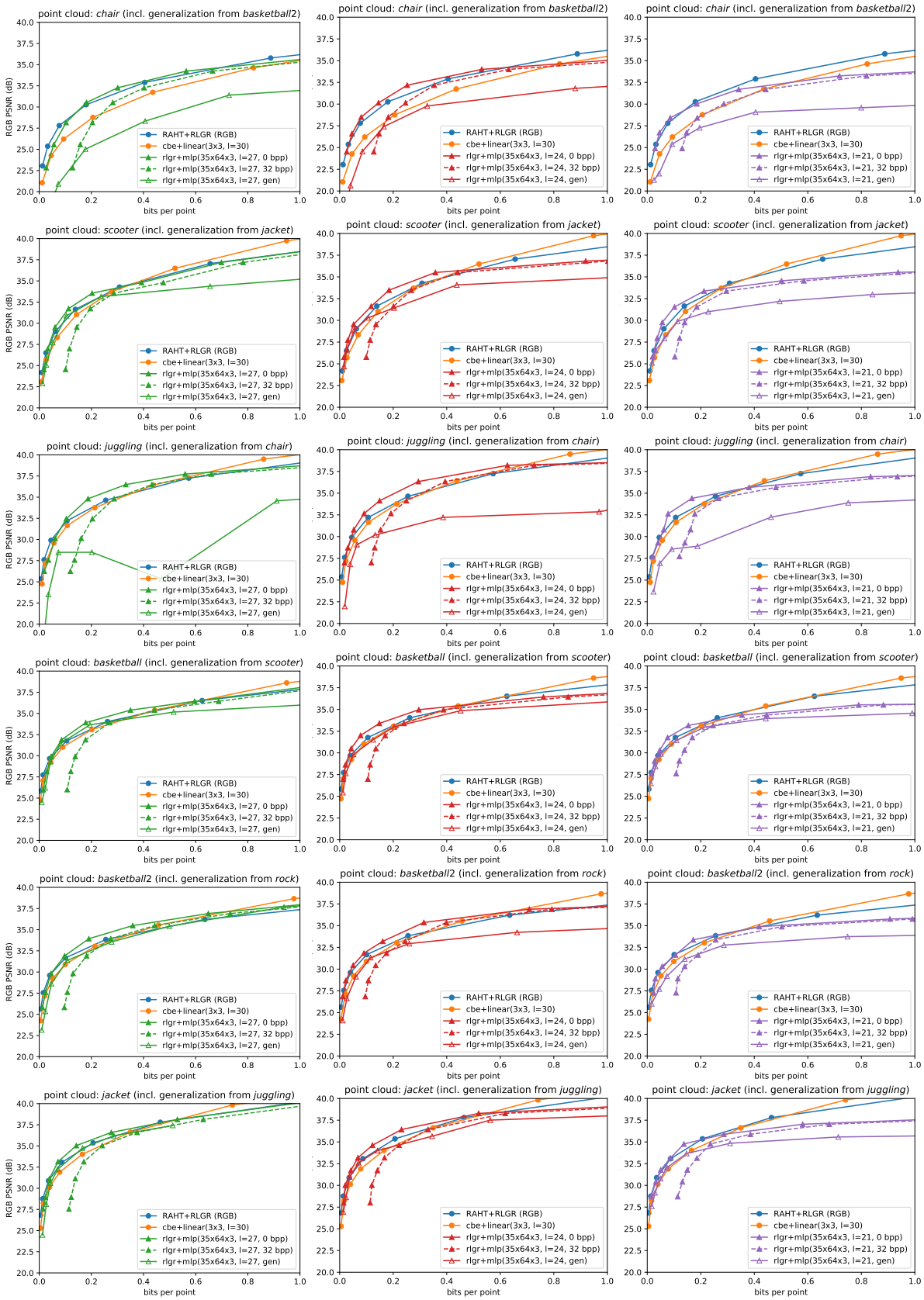


Figure S15: Effect of side information for coordinate based network *mlp(35x64x3)* at levels 27 (left), 24 (middle), and 21 (right). Each row is a different point cloud.

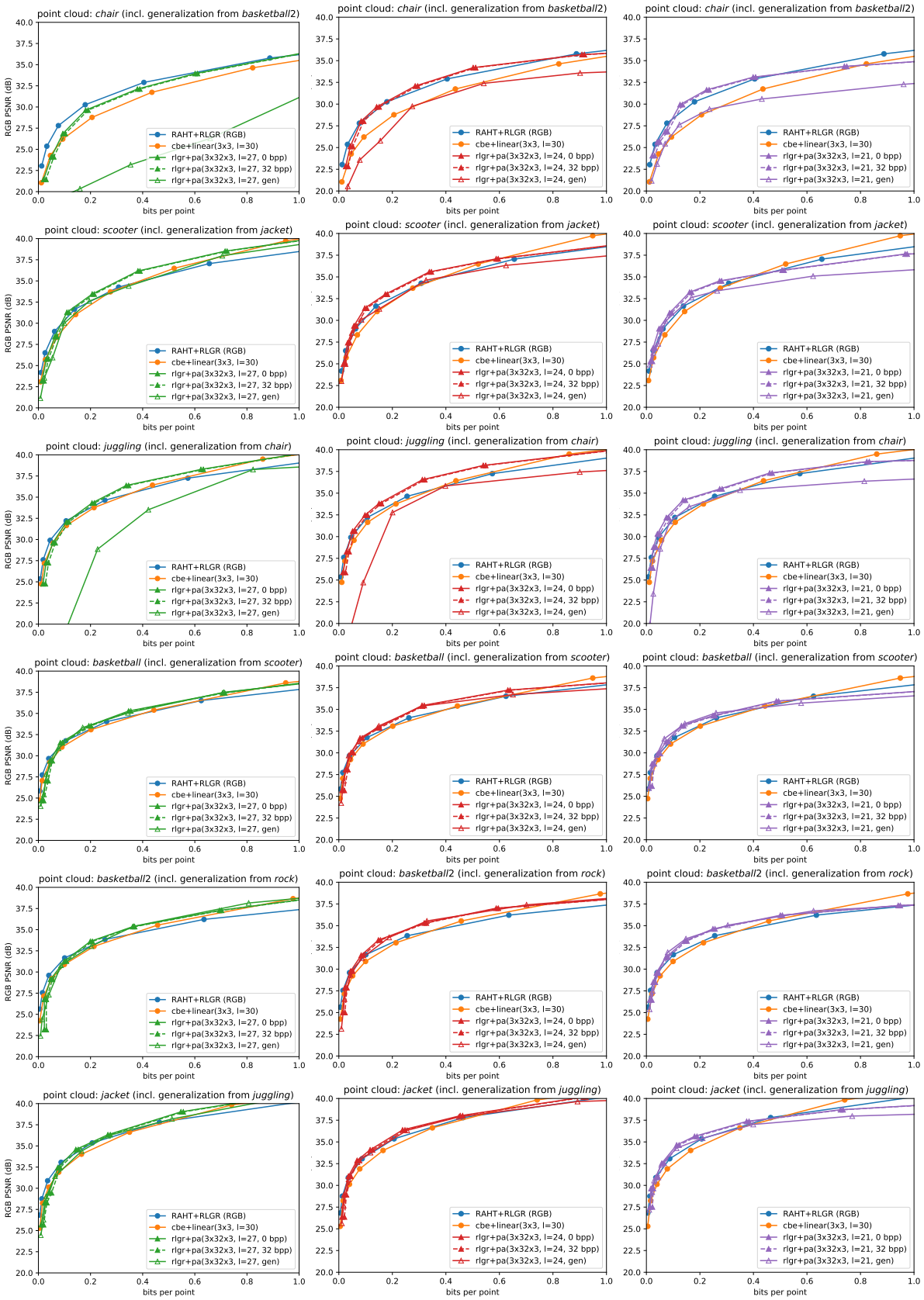
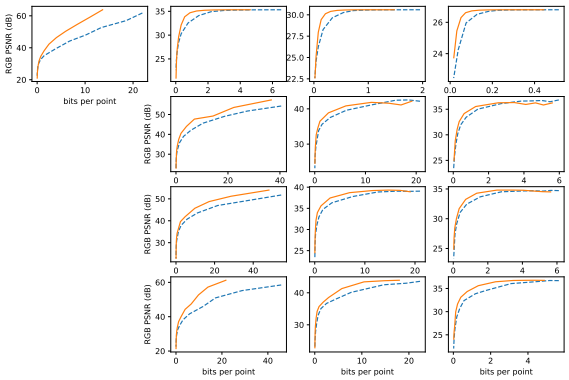


Figure S16: Effect of side information for coordinate based network *pa(3x32x3)* at levels 27 (left), 24 (middle), and 21 (right). Each row is a different point cloud.

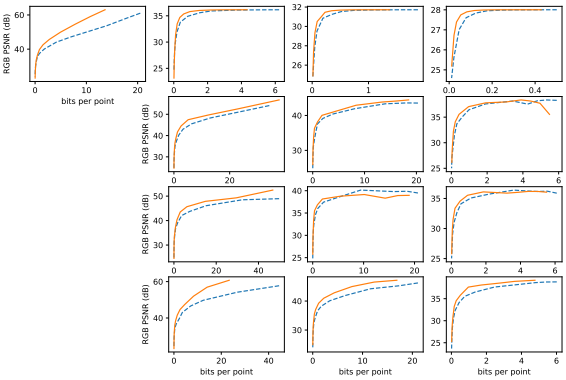
Table S3. BD-Rate reductions due to normalization.

rock	level			
CBN	30	27	24	21
linear(3x3)	-36.5%	-26.1%	-29.6%	-35.2%
mlp(35x256x3)	N/A	-35.8%	-39.4%	-28.7%
mlp(35x64x3)	N/A	-32.7%	-34.4%	-27.8%
pa(3x32x3)	N/A	-41.6%	-39.8%	-37.5%
chair	level			
CBN	30	27	24	21
linear(3x3)	-36.8%	-26.5%	-34.1%	-47.4%
mlp(35x256x3)	N/A	-34.8%	-39.5%	-28.0%
mlp(35x64x3)	N/A	-34.3%	-47.5%	-38.4%
pa(3x32x3)	N/A	-45.2%	-44.7%	-45.6%
scooter	level			
CBN	30	27	24	21
linear(3x3)	-32.6%	-19.5%	-34.6%	-41.8%
mlp(35x256x3)	N/A	-32.1%	-38.1%	-20.6%
mlp(35x64x3)	N/A	-27.1%	-30.2%	-39.1%
pa(3x32x3)	N/A	-41.9%	-42.2%	-40.6%
juggling	level			
CBN	30	27	24	21
linear(3x3)	-26.9%	-8.5%	-23.4%	-32.4%
mlp(35x256x3)	N/A	-22.7%	-35.0%	-27.7%
mlp(35x64x3)	N/A	-16.3%	-35.2%	-30.5%
pa(3x32x3)	N/A	-47.7%	-47.6%	-39.1%
basketball	level			
CBN	30	27	24	21
linear(3x3)	-30.0%	-19.7%	-27.7%	-28.9%
mlp(35x256x3)	N/A	-20.9%	-29.2%	-42.3%
mlp(35x64x3)	N/A	-5.8%	-28.4%	-32.8%
pa(3x32x3)	N/A	-31.1%	-27.7%	-24.5%
basketball2	level			
CBN	30	27	24	21
linear(3x3)	-29.2%	-14.6%	-20.8%	-36.5%
mlp(35x256x3)	N/A	-34.5%	-24.0%	-15.9%
mlp(35x64x3)	N/A	-28.7%	-27.8%	-24.1%
pa(3x32x3)	N/A	-41.2%	-42.5%	-41.7%
jacket	level			
CBN	30	27	24	21
linear(3x3)	-29.3%	-15.2%	-27.7%	-41.9%
mlp(35x256x3)	N/A	-28.8%	-35.1%	-28.4%
mlp(35x64x3)	N/A	-22.0%	-21.3%	-25.1%
pa(3x32x3)	N/A	-39.1%	-38.0%	-41.7%

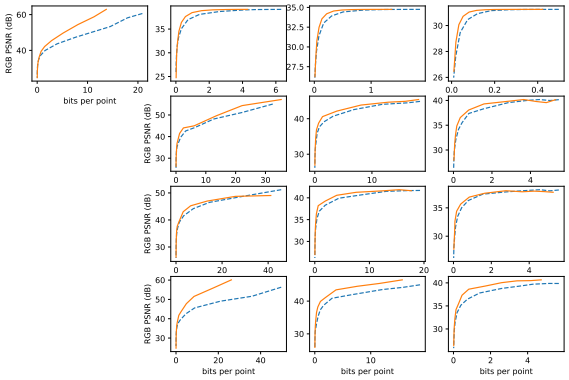
point cloud: *chair*



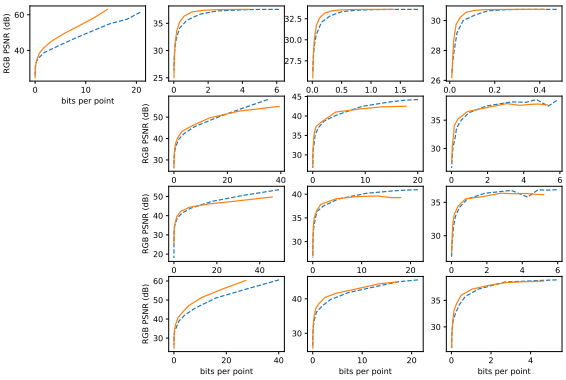
point cloud: *scooter*



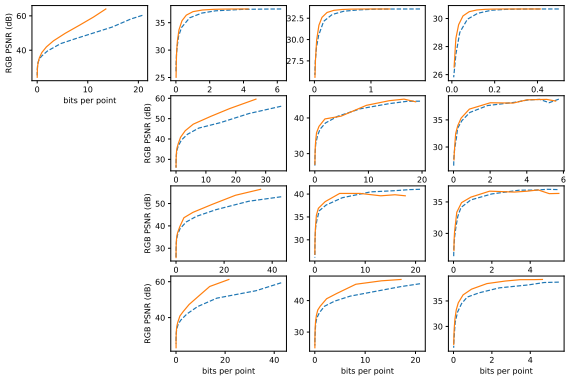
point cloud: *juggling*



point cloud: *basketball*



point cloud: *basketball2*



point cloud: *jacket*

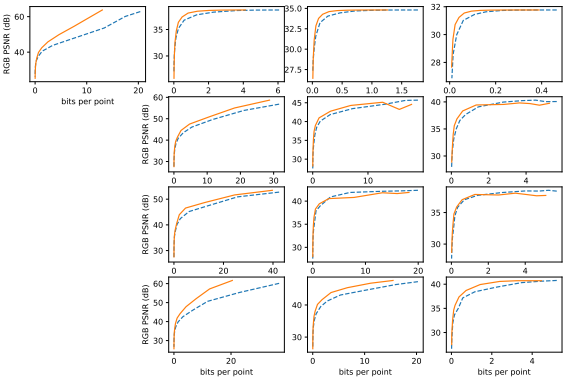


Figure S17: RD performance (RGB PSNR vs. bit rate) improvement due to normalization, corresponding to entries in Tab. S3.

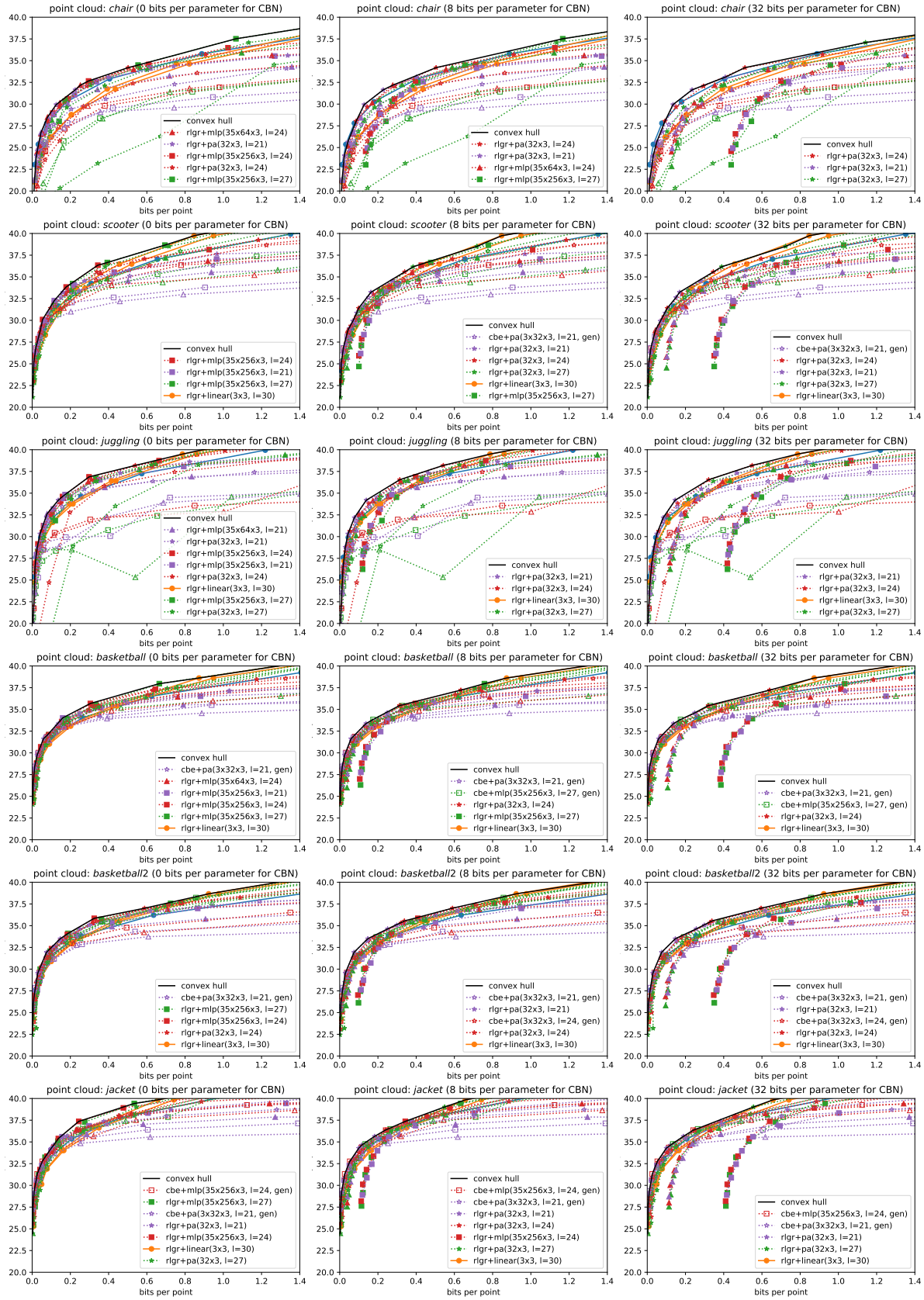


Figure S18: Convex hull (solid black line) of RD performances of all CBN configurations across all levels, including side information using 0 (left), 8 (middle), and 32 (right) bits per CBN parameter. Each row is a different point cloud.

Table S4. Comparison to Sparse PCAC (Wang and Ma (2022)). Sparse PCAC (left) and LVAC (right) are each compared to references TMC13v14 and RAHT, in terms of BD-BD (average % bitrate change over YUV PSNR range) and BD-PSNR (average YUV PSNR dB change over bitrate range) relative to the reference. Negative/positive BD-BR means better/worse than the reference. Positive/negative BD-PSNR means better/worse than the reference. The references are the best-performing versions: TMC13v14 (YUV, predRAHT) and RAHT-RLGR (YUV). Both Sparse PCAC and LVAC are better than RAHT, not as good as TMC13v14, and about as good as each other. Sparse PCAC data taken from (Wang and Ma, 2022, Table 1). LVAC data computed from convex hulls of LVAC curves in Fig. S8 and Fig. S10. Bold values indicate which one of LVAC and Sparse PCAC performs better.

Point Cloud	Sparse PCAC (Wang and Ma (2022))				LVAC (ours)			
	vs. TMC13v14		vs. RAHT		vs. TMC13v14		vs. RAHT	
	BD-BR	BD-PSNR	BD-BR	BD-PSNR	BD-BR	BD-PSNR	BD-BR	BD-PSNR
longdress	**+27 %**	**-0.81 dB**	**-42 %**	**+1.72 dB**	+54 %	-1.25 dB	-33 %	+1.00 dB
loot	+142 %	-2.50 dB	+6 %	-0.10 dB	**+61 %**	**-1.75 dB**	**-30 %**	**+0.31 dB**
red&black	+55 %	-1.28 dB	-36 %	+1.30 dB	**+41 %**	**-0.98 dB**	**-41 %**	**+1.40 dB**
soldier	+66 %	-1.60 dB	-25 %	+0.75 dB	**+59 %**	**-1.17 dB**	**-29 %**	**+1.27 dB**
Average	+72 %	-1.55 dB	-24 %	+0.92 dB	**+54 %**	**-1.29 dB**	**-33 %**	**+1.00 dB**

REFERENCES

Alliez, P., Forge, F., De Luca, L., Pierrot-Deseilligny, M., and Preda, M. (2017). Culture 3D Cloud: A Cloud Computing Platform for 3D Scanning, Documentation, Preservation and Dissemination of Cultural Heritage. *ERCIM News* , 64

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. (2008). MeshLab: an Open-Source Mesh Processing Tool. In *Eurographics Italian Chapter Conference*, eds. V. Scarano, R. D. Chiara, and U. Erra (The Eurographics Association). doi:10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136

d'Eon, E., Harrison, B., Meyers, T., and Chou, P. A. (2017). *8i Voxelized Full Bodies — A Voxelized Point Cloud Dataset*. input document M74006 & m42914, ISO/IEC JTC1/SC29 WG1 & WG11 (JPEG & MPEG), Ljubljana, Slovenia

Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., et al. (2019). The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Trans. Graph.* 38. doi:10.1145/3355089.3356571

Malvar, H. (2006). Adaptive run-length/golomb-rice encoding of quantized generalized gaussian sources with unknown statistics. In *Data Compression Conference (DCC'06)*. 23–32

Meka, A., Pandey, R., Haene, C., Orts-Escolano, S., Barnum, P., Davidson, P., et al. (2020). Deep relightable textures - volumetric performance capture with neural rendering. vol. 39. doi:10.1145/3414685.3417814

Sheng, X., Li, L., Liu, D., Xiong, Z., Li, Z., and Wu, F. (2021). Deep-pcac: An end-to-end deep lossy compression framework for point cloud attributes. *IEEE Transactions on Multimedia* , 1–1doi:10.1109/TMM.2021.3086711

Wang, J. and Ma, Z. (2022). Sparse tensor-based point cloud attribute compression. In *2022 IEEE Int'l Conf. Multimedia Information Processing and Retrieval (MIPR)*