Wireless 3D Point Cloud Delivery Using Deep Graph Neural Networks

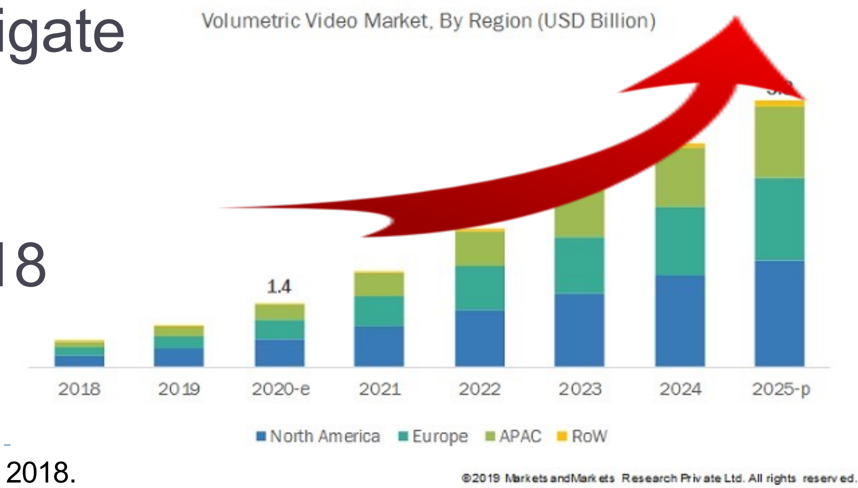
Takuya Fujihashi[1], Toshiaki Koike-Akino[2],
Siheng Chen[2], Takashi Watanabe[1],
[1]Osaka University
[2]Mitsubishi Electric Research Laboratories (MERL)

IEEE International Conference on Communications 2021
June 17, 2021

Background

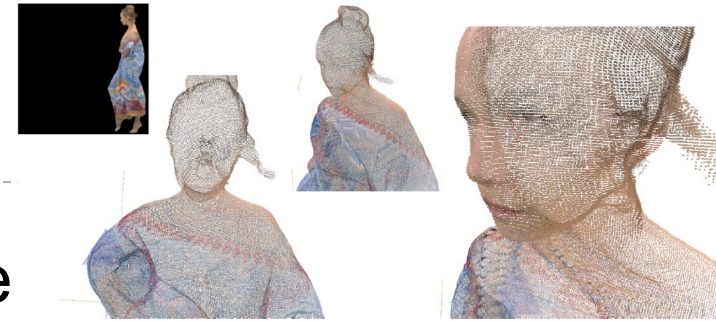
- Volumetric (3D) media
  - Reconstruct 3D scenes with full parallax and depth info.
  - Applications: entertainment, medical imaging, AR/MR

- Key applications in 5G and 5G beyond networks [1]
  - Google and Facebook investigate volumetric video streaming
  - Volumetric video market will grow from $578 million in 2018 to $2.78 billion by 2023 [2]

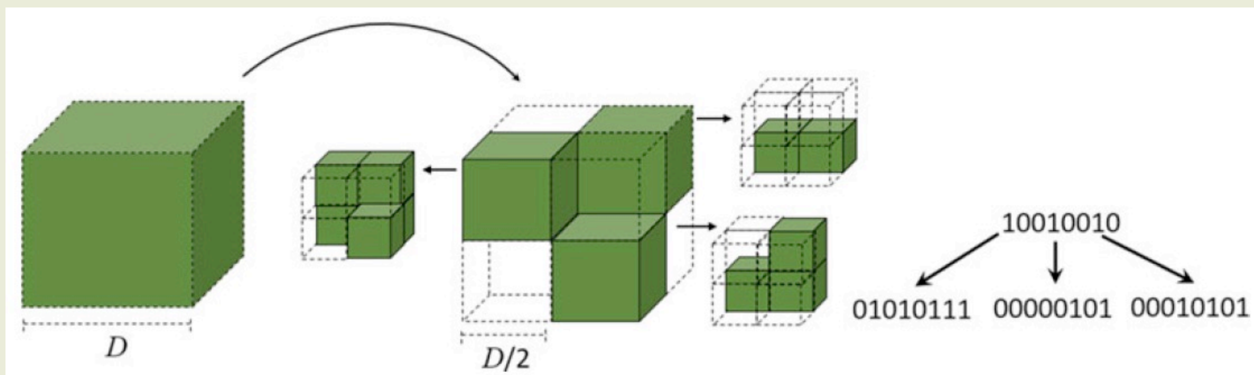


[1] AT&T Continues to Lead in Bringing 5G Experiences to Life, 2018.
[2] Volumetric Video Market by Volumetric Capture & Content Creation (Hardware (Camera & Processing Unit), Software, and Services), Application (Sports & Entertainment, Medical, Signage, Education & Training), and Geography - Global Forecast to 2023.

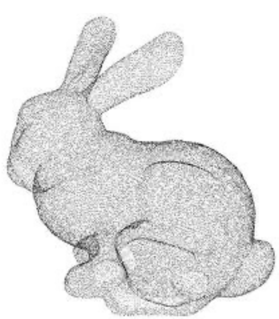
Point Cloud

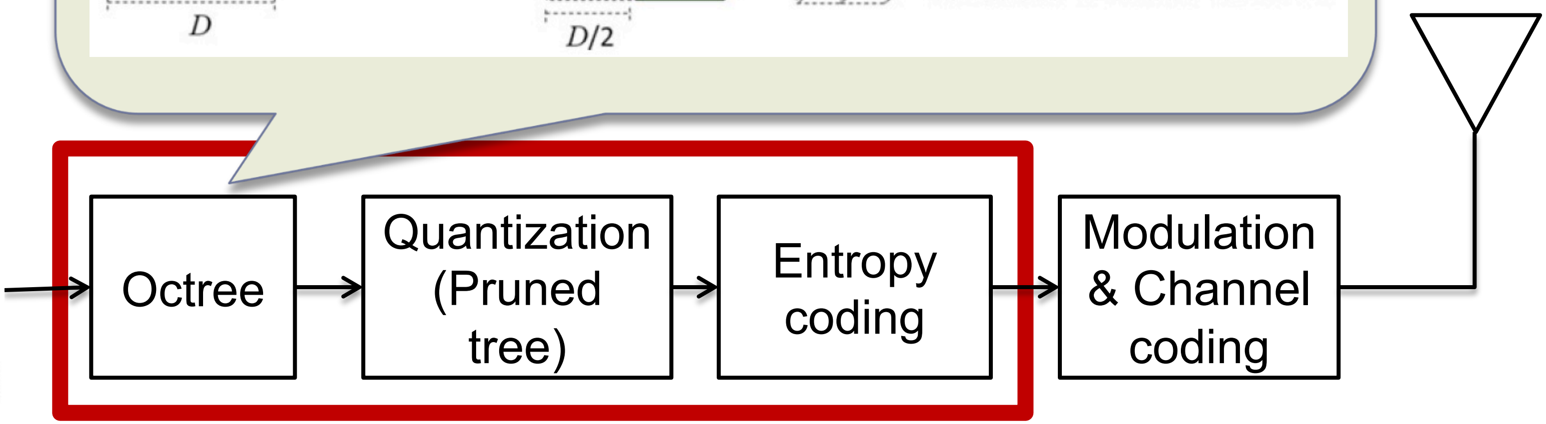


- Typical data structure for 3D scene
  - Consist of numerous and irregular structure of 3D points
  - Each point has 3D coordinate (x, y, z) information
- Existing s
  - Encoder
    wireless

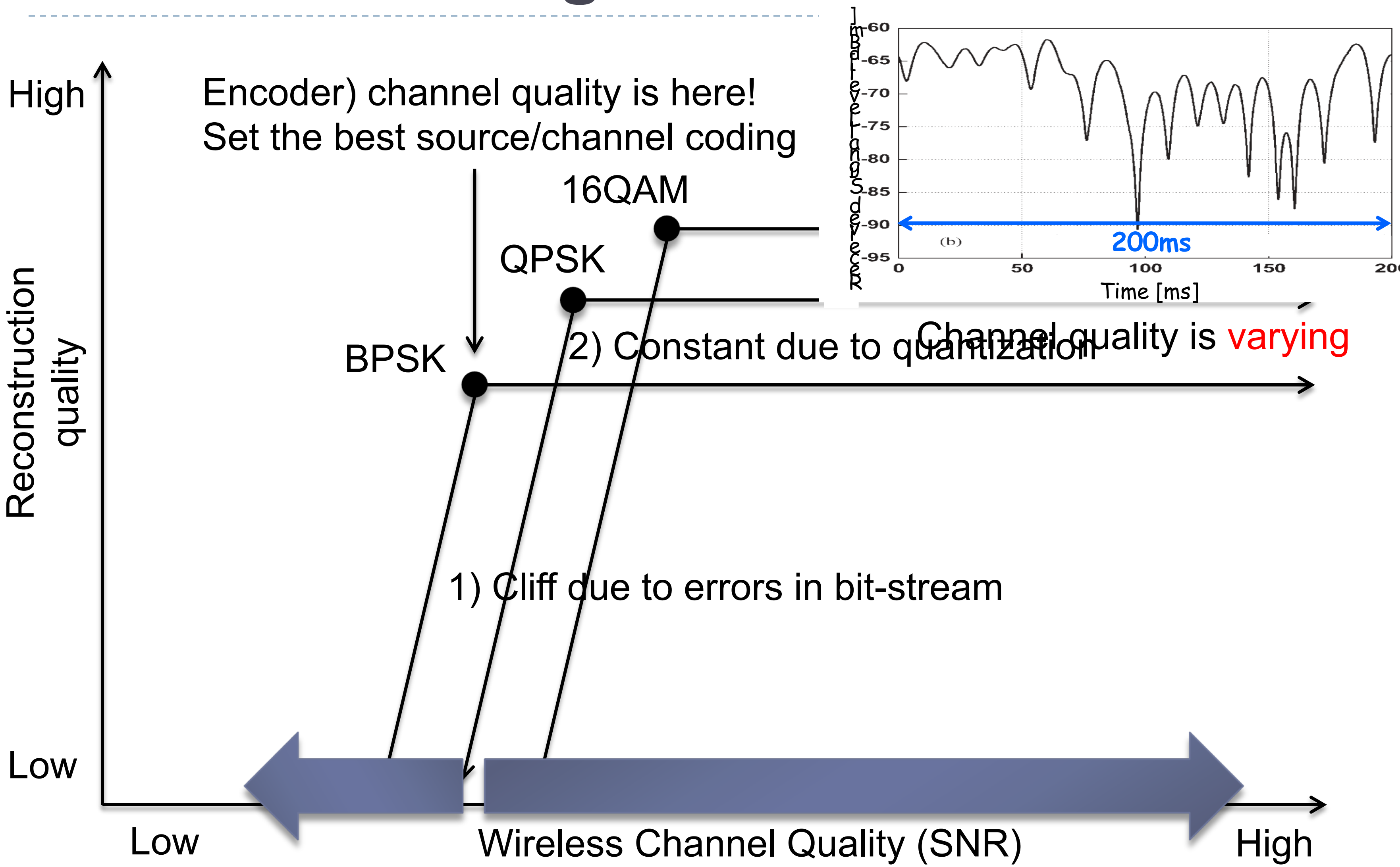


Point Cloud

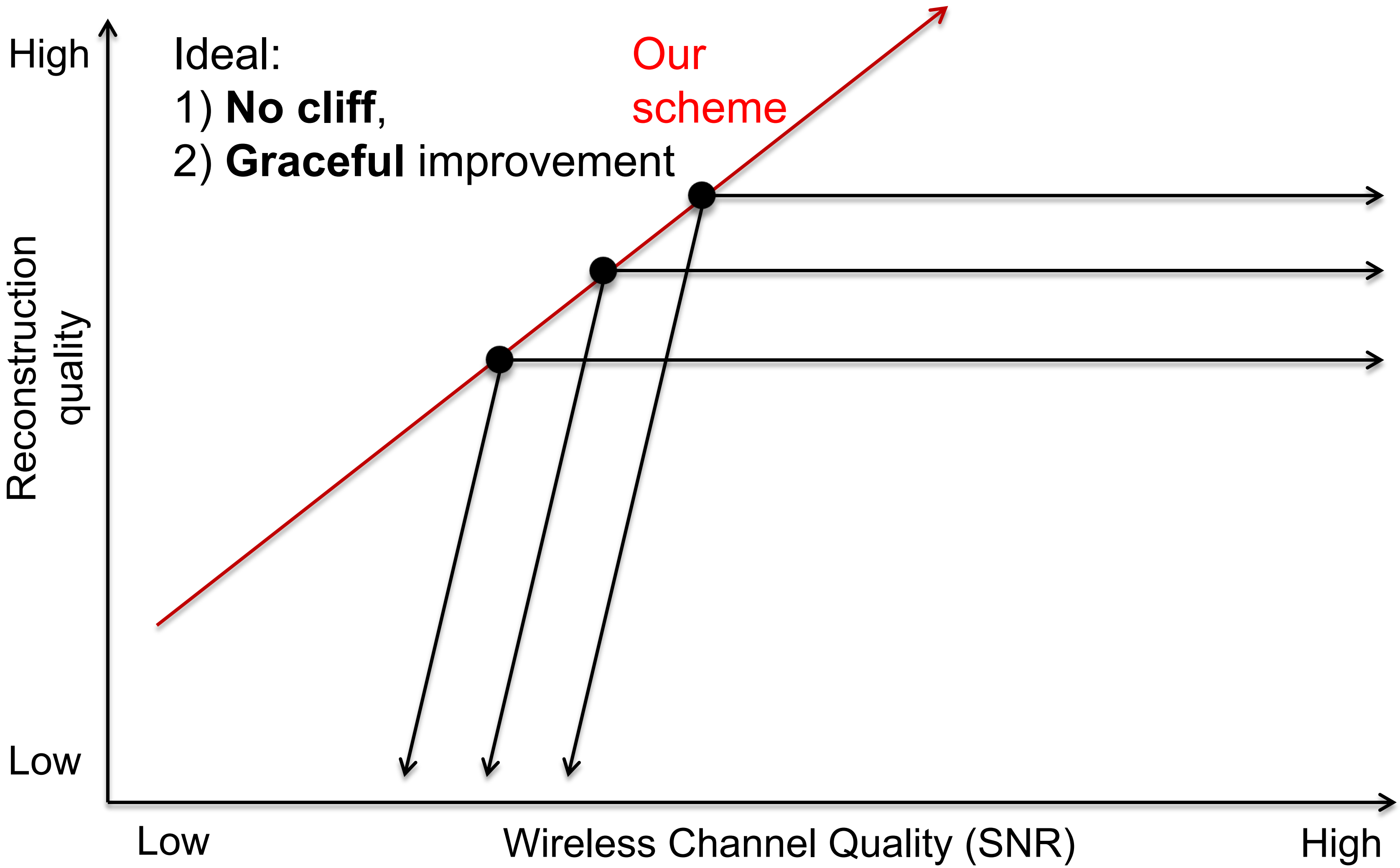




Issues of Digital-based Wireless Point Cloud Streaming



Issues of Digital-based Wireless Point Cloud Streaming

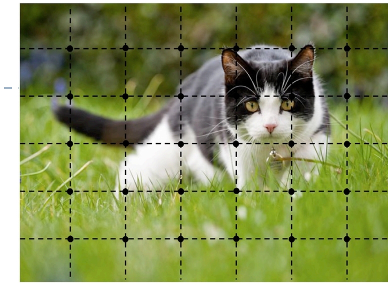


Purpose

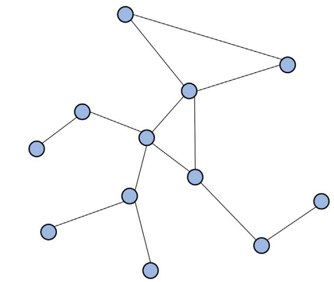
- Our study tackles the following challenging issues
  1. Cliff effect
  2. Leveling effect: Constant reconstruction quality
- Propose a novel scheme for wireless point cloud delivery
  - Regard 3D points as **graph signals** with the attributes of 3D coordinates to deal with irregular structure of holographic data formats
  - Introduce graph neural network (**GNN**)-based and multi-layer perceptron (**MLP**)-based autoencoder for point coding
  - Skip digital-based compression, instead, introduce **near-analog modulation** to realize graceful quality improvement

GNN

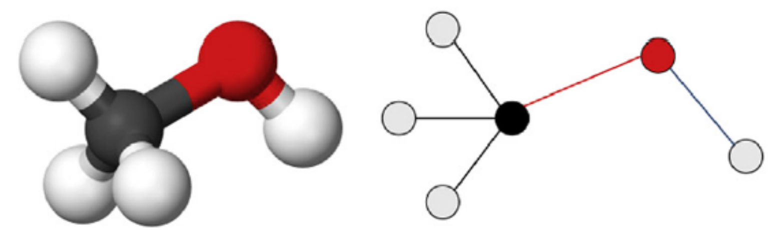
- Extend neural networks to process graph signals
  - Deal with signals in non-Euclidean space
- Applications
  - Physics, Molecule, Text, Social networks, Images
- Typical GNN model
  - Find graph structure
  - Design loss functions and computational modules

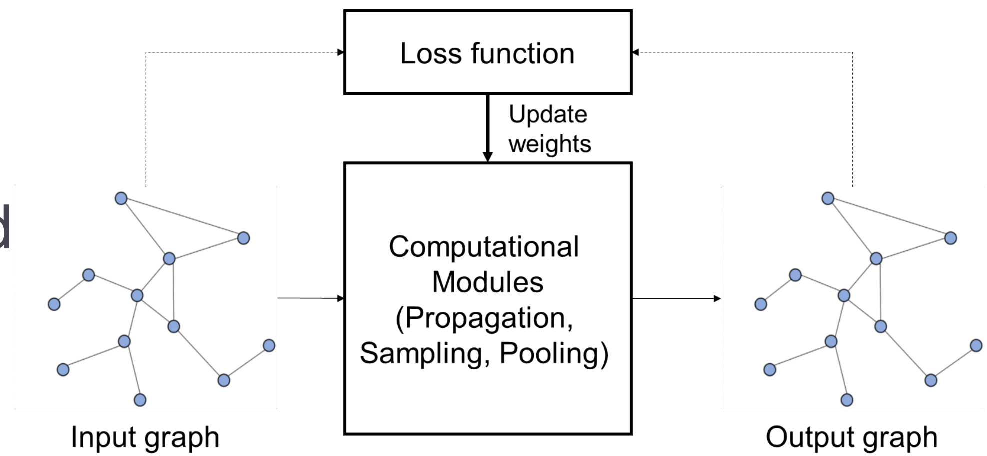


Images in Euclidean space

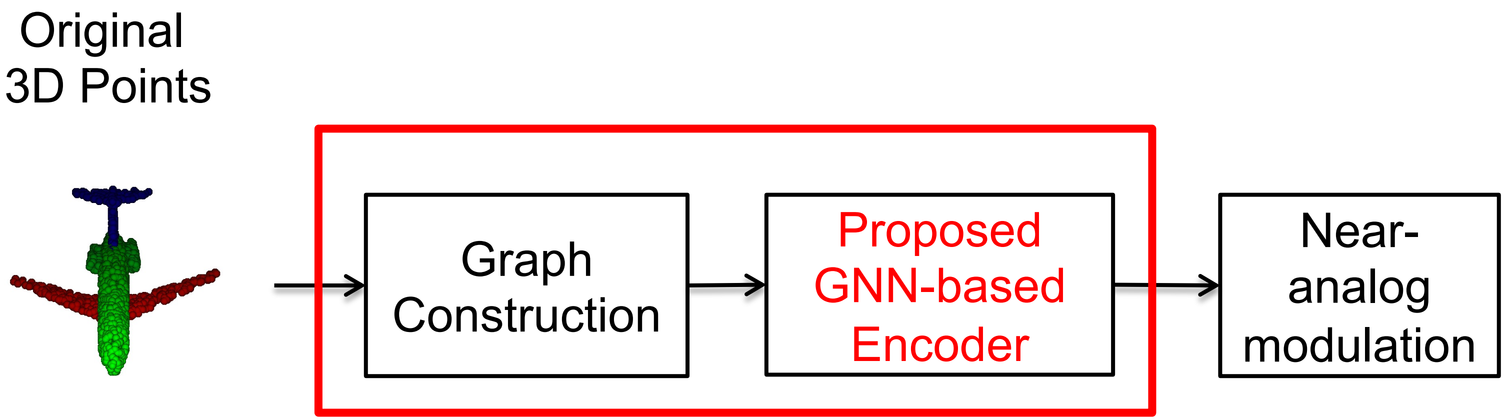


Graph in non-Euclidean space





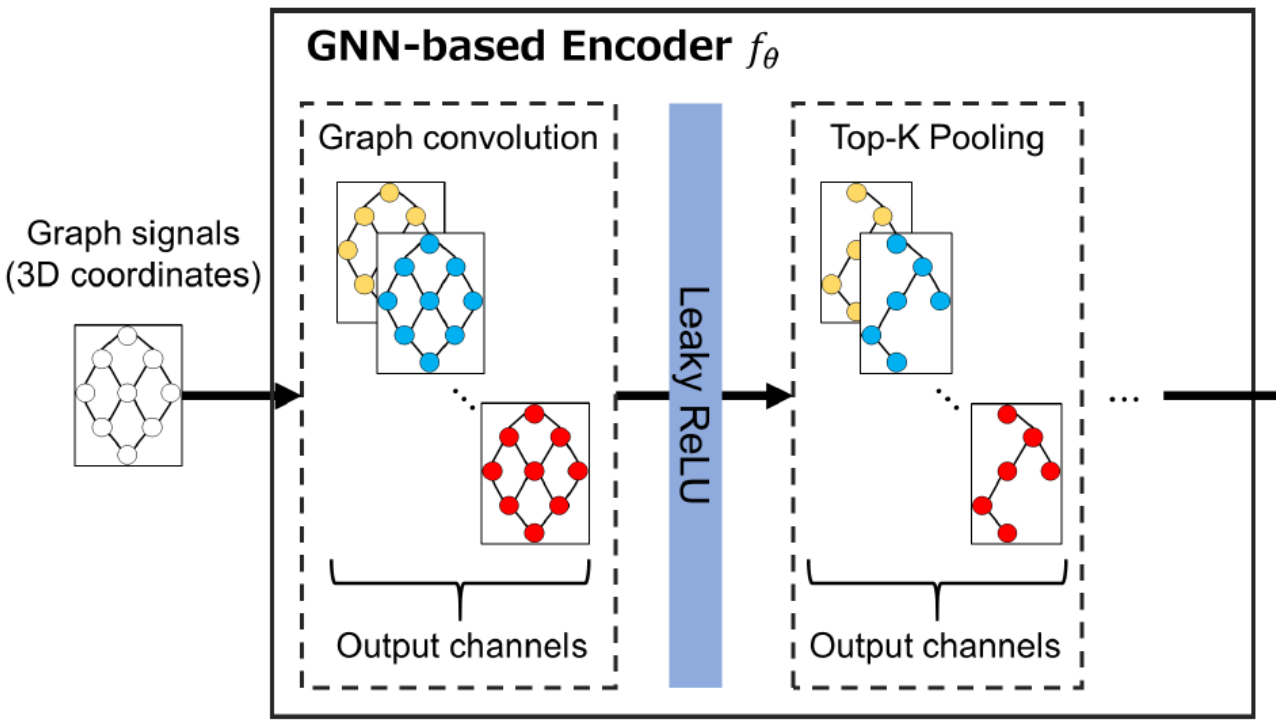
Proposed: Graph Construction



- Regard 3D points as graph $\mathcal{g} = (\boldsymbol{V}, \epsilon, \boldsymbol{W})$
  - $\boldsymbol{V}$: vertex (each 3D point), ϵ: edge
  - $\boldsymbol{W}$: adjacency matrix of positive edge weights
    - $W_{i,j}$: the weight of an edge connecting vertices i and j
    - 1: vertices are connected, 0: vertices are not connected
      - Use K-nearest-neighbor graph to make the connection between the vertices

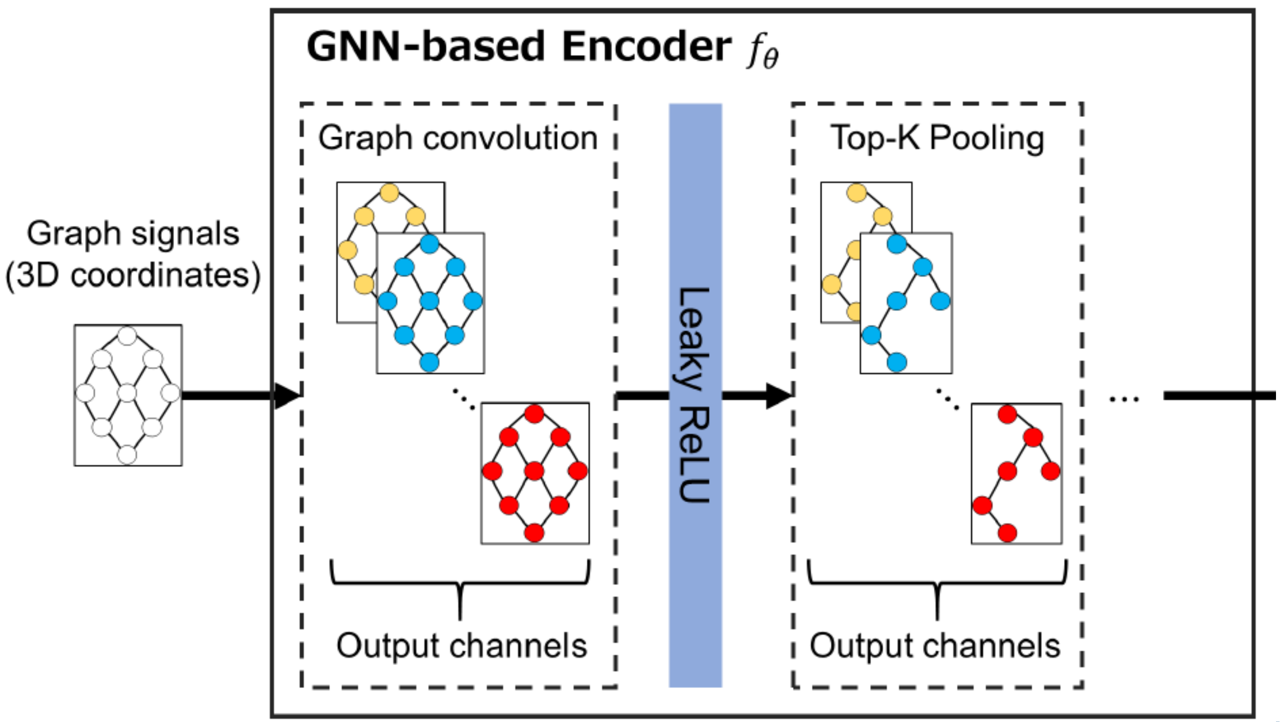
Proposed: GNN-based Encoder

- Transforms 3D points into several latent variables
  - Consists of a series of graph convolution followed by leaky rectified linear unit (ReLU) activation function, Top-K pooling, and a normalization layer

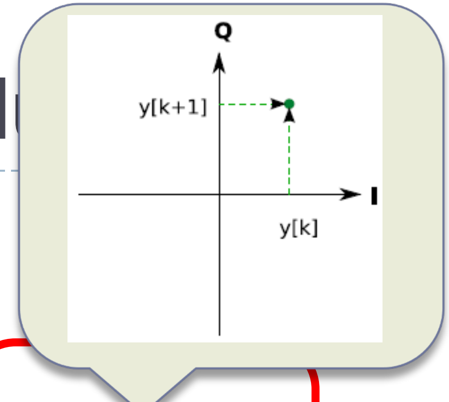


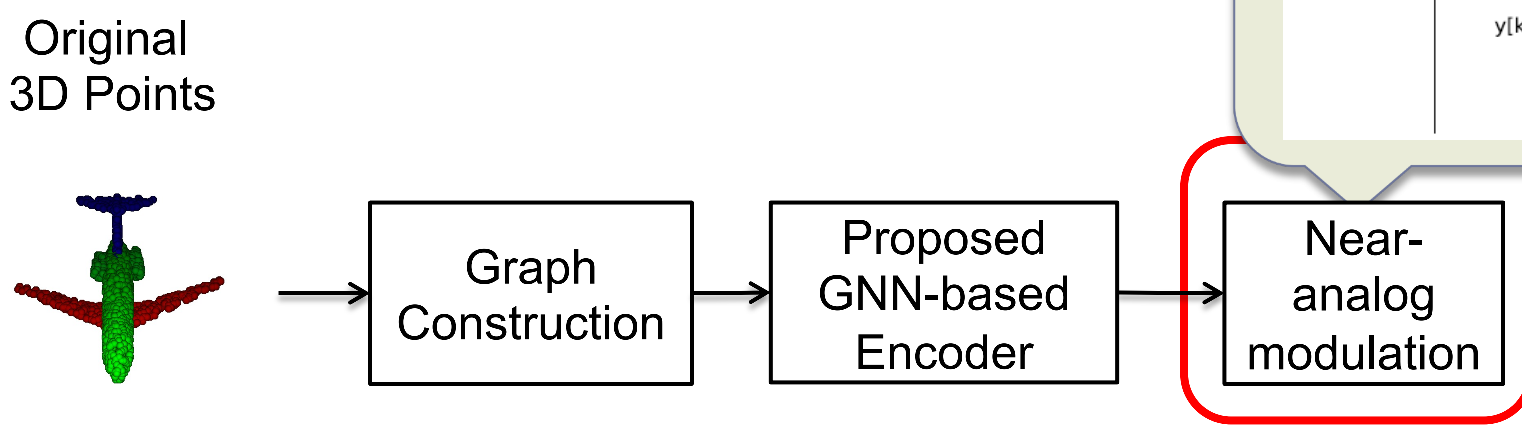
Proposed: GNN-based Encoder

- Graph Convolution: Extract the graph signal features
- Leaky ReLU: learn a mapping from the source to coded signals
- Top-K Pooling: chooses the largest values from each channel to remain important features

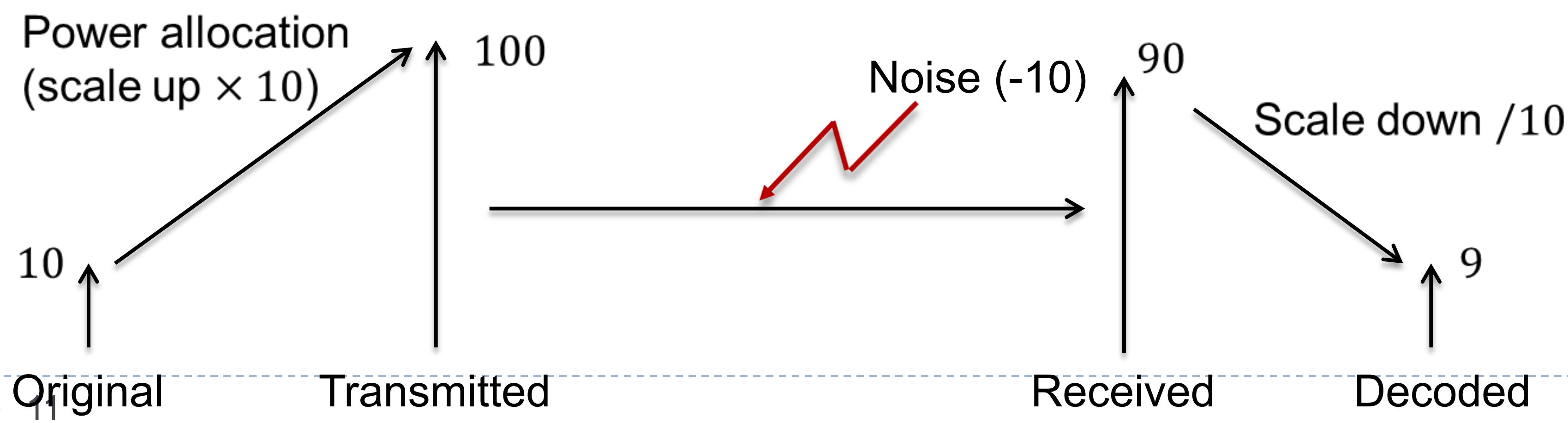


Proposed: Near-analog Mod



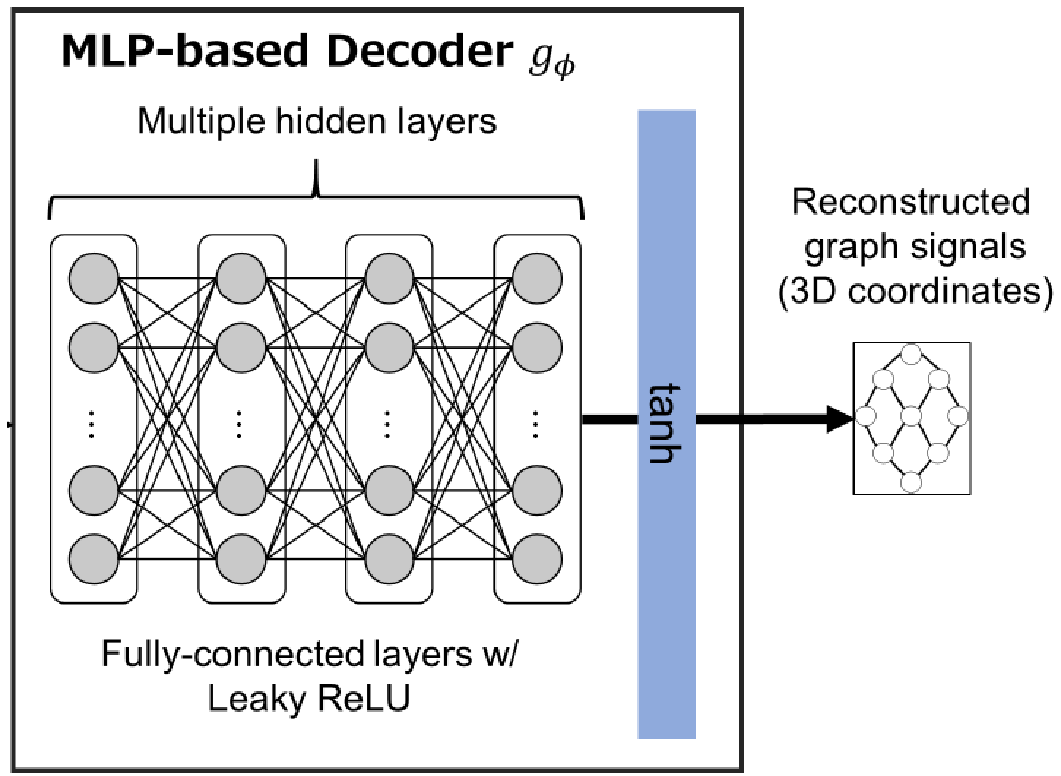


- Near-analog modulation realizes graceful quality improvement according to wireless channel quality



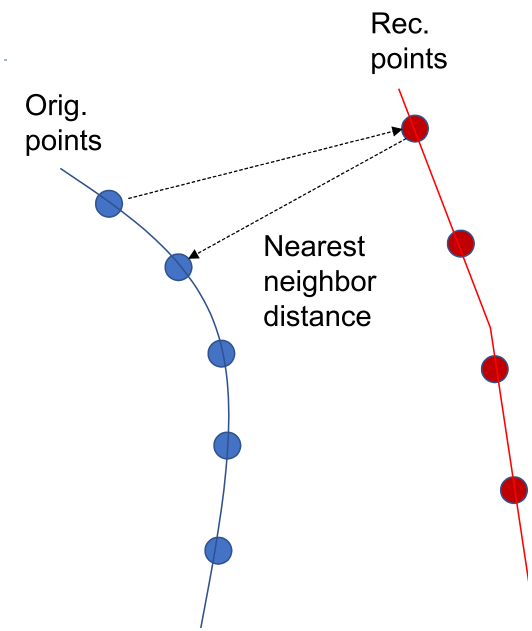
Proposed: MLP-based Decoding

- Latent variables z_i are impaired according to a channel transfer function with pre/post equalizations
  - $\eta_{\text{preeq}} = |h_i| z_i + n_i$
  - $\eta_{\text{posteq}} = z_i + n_i / h_i$
    - h_i: multiplicative fading coefficient
    - n_i: effective noise with a variance of σ^2
- Decoder
  - Consists of a series of fully-connected layers and leaky ReLU



Proposed: Loss Function

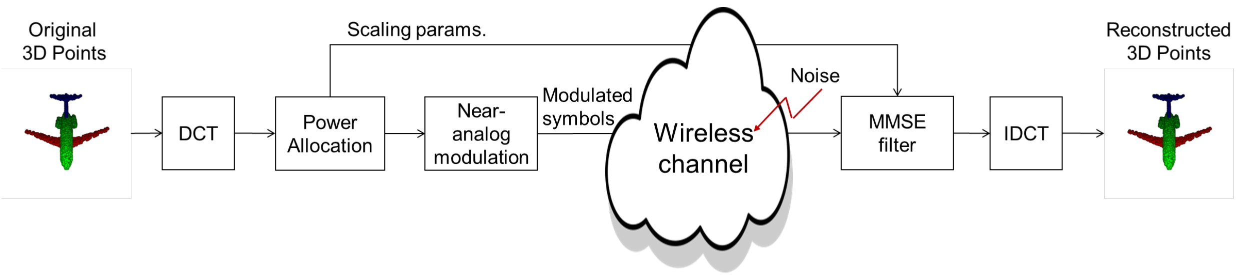
- GNN-based encoding and MLP-based decoding functions are trained to minimize a loss function
  - $(\theta, \phi) = \arg\min_{\theta,\phi} \mathrm{E}[\, d(\boldsymbol{p}, \widehat{\boldsymbol{p}}_{\theta,\phi})]$
  - $\widehat{\boldsymbol{p}}_{\theta,\phi}$: reconstructed 3D coordinates via the proposed encoder and decoder with parameter sets of θ and ϕ



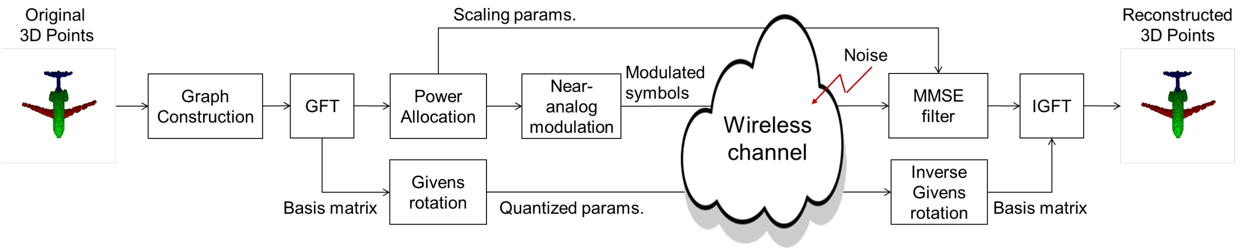
- Consider augmented Chamfer distance for loss function
  - $\max\left\{\frac{1}{|S|}\sum_{\boldsymbol{p}\in S}\min_{\widehat{\boldsymbol{p}}\in\hat{S}}\left|\left|\boldsymbol{p}-\widehat{\boldsymbol{p}}\right|\right|_2, \frac{1}{|\hat{S}|}\sum_{\widehat{\boldsymbol{p}}\in\hat{S}}\min_{\boldsymbol{p}\in S}\left|\left|\boldsymbol{p}-\widehat{\boldsymbol{p}}\right|\right|_2\right\}$
  - If Chamfer distance is even small, the original and reconstructed 3D coordinates are close each other

Evaluation

- Comparative schemes
  - SoftCast [3]: DCT-based
  - HoloCast [4]: GFT-based
  - Proposed: GNN



SoftCast



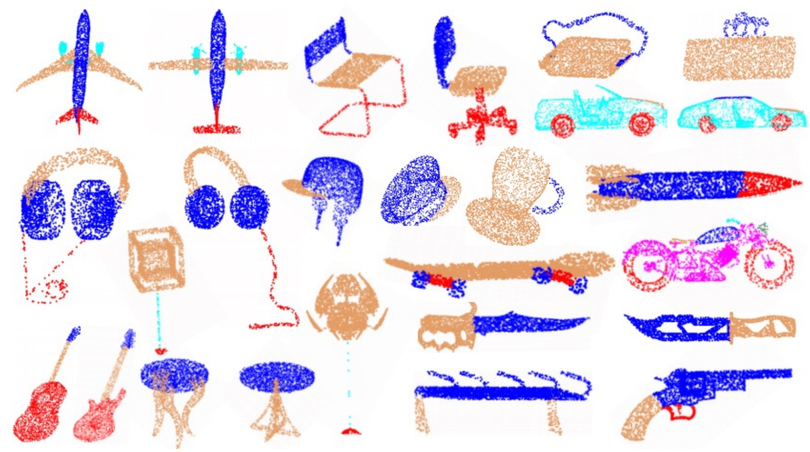
HoloCast

- Reference point cloud
  - ShapeNet: dataset of 3D points

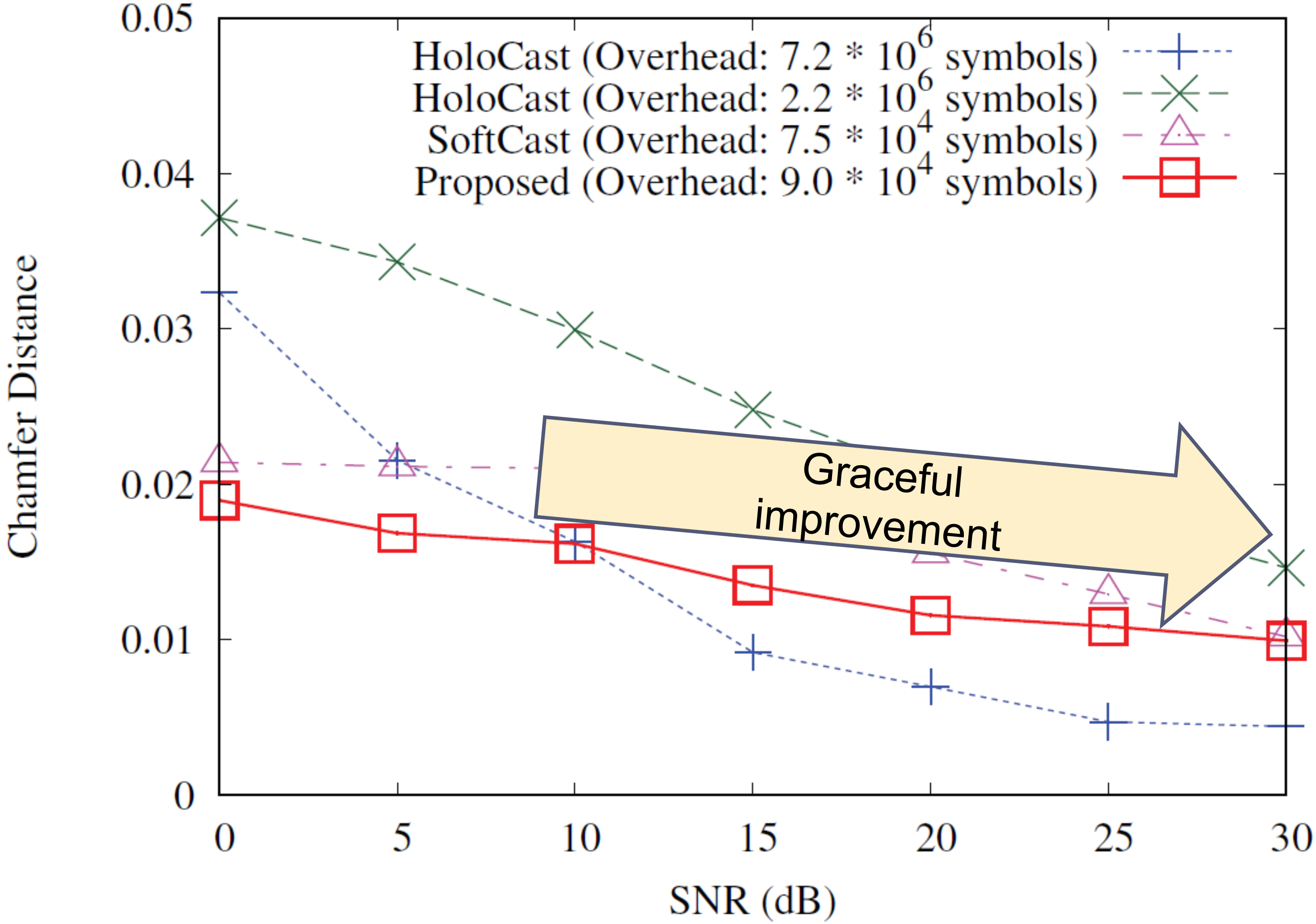
[3] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in ACM Annual International Conference on Mobile Computing and Networking, Sep. 2011, pp. 289–300. [4] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. Orlik, "HoloCast:Graph signal processing for graceful point cloud delivery," in IEEE International Conference on Communications, 2019, pp. 1–7

Evaluation

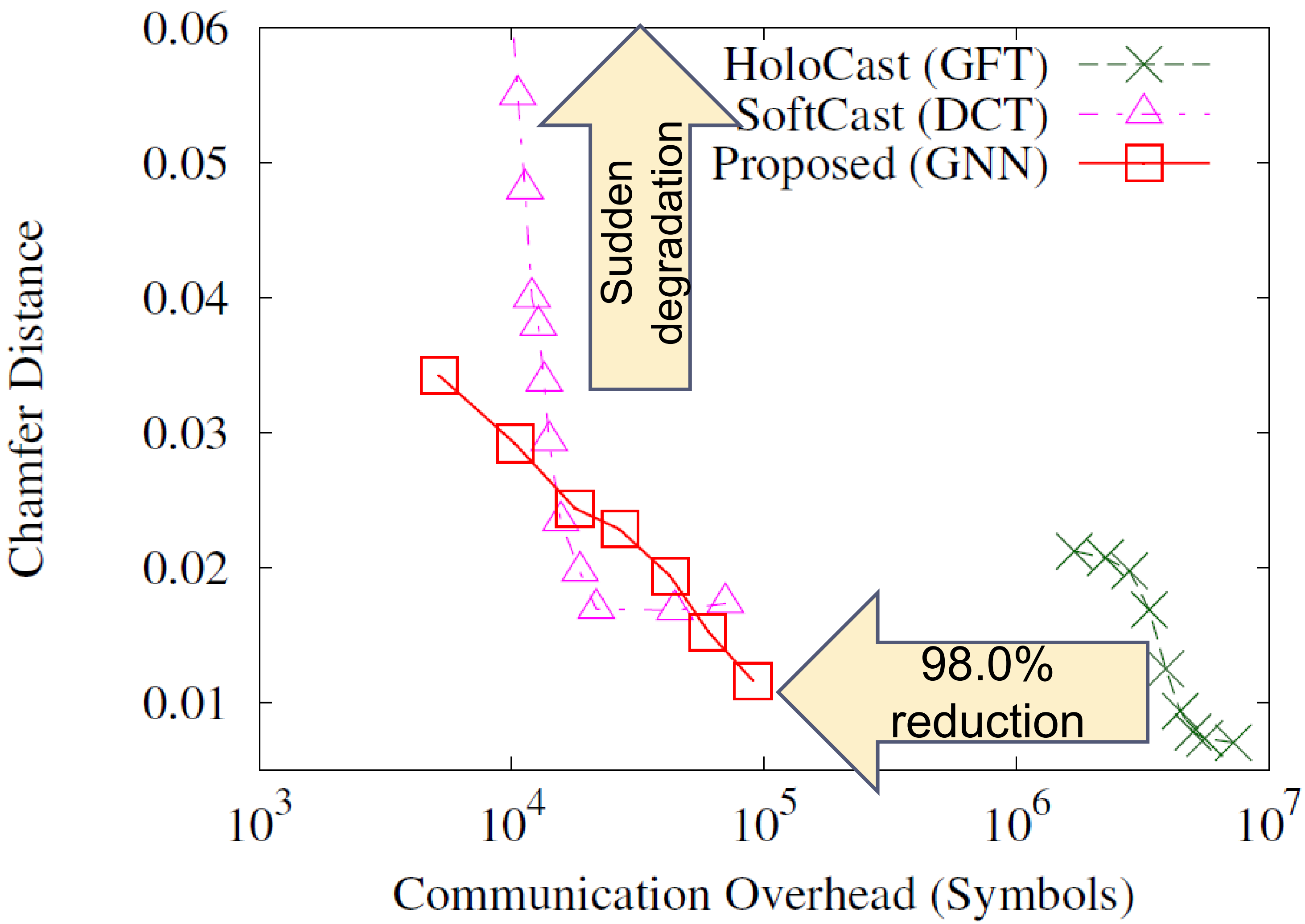
- Reference point cloud
  - ShapeNet: dataset of 3D points
    - Category: Airplane
    - 2115 point clouds for training, 234 point clouds for testing



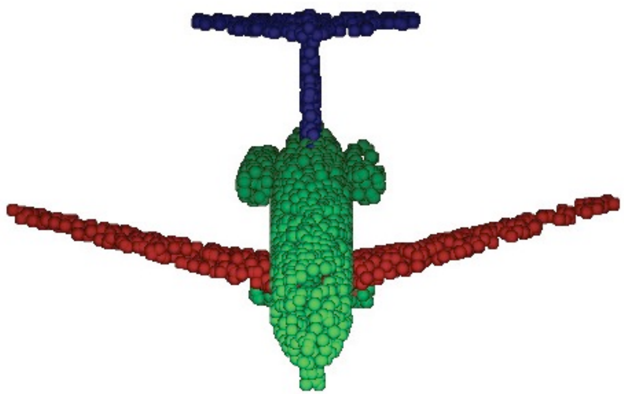
Reconstruction Quality vs. Wireless Channel Quality



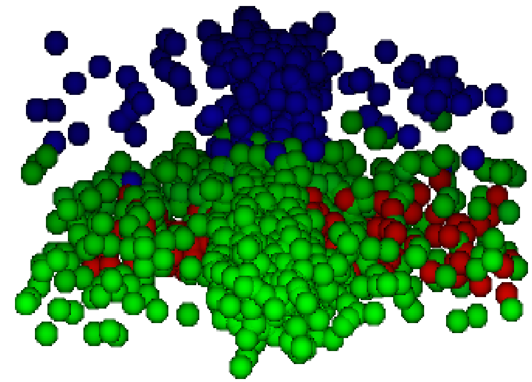
Reconstruction Quality vs Traffic



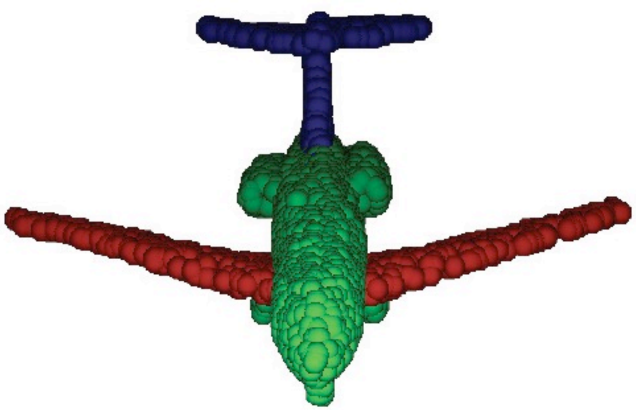
Visual Quality of Point Cloud Reconstruction



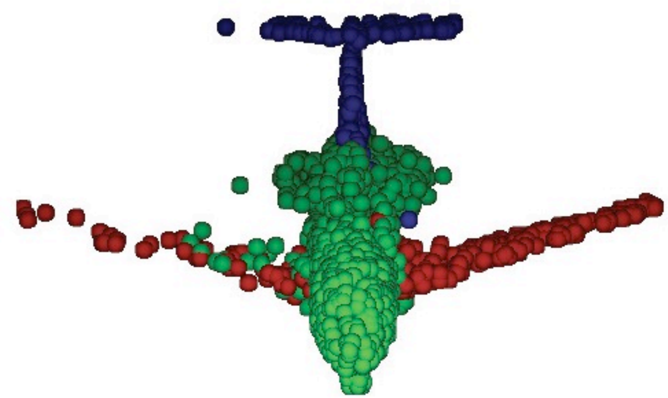
Original



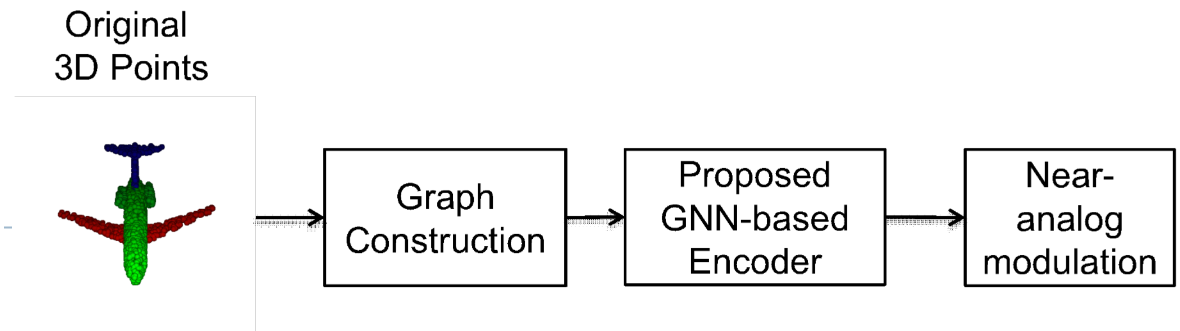
SoftCast
Chamfer distance: 0.020



HoloCast
Chamfer distance: 0.003



Proposed
Chamfer distance: 0.012



Conclusion

- Proposed a novel scheme for wireless point cloud delivery
  - Regard 3D points as **graph signals** with the attributes of 3D coordinates and color components to deal with irregular structure of holographic data formats
  - Introduce graph neural network (**GNN**)-based and multi-layer perceptron (**MLP**)-based autoencoder for point coding
  - Skip digital-based compression, instead, introduce **near-analog modulation** to realize graceful quality improvement
- Validated the advantage
  - Gracefully improve reconstruction quality with the improvement of wireless channel quality
  - Better reconstruction quality with a limited amount of traffic

Q&A

- Please send questions and comments

fujihashi.takuya@ist.osaka-u.ac.jp