Evaluating Learning-Based Compression Techniques for Radio Interferometric Visibilities Data

Xiaoran Chen, Adrien Devresse,
Dr Mathieu Salzmann, Prof. Emma Tolley

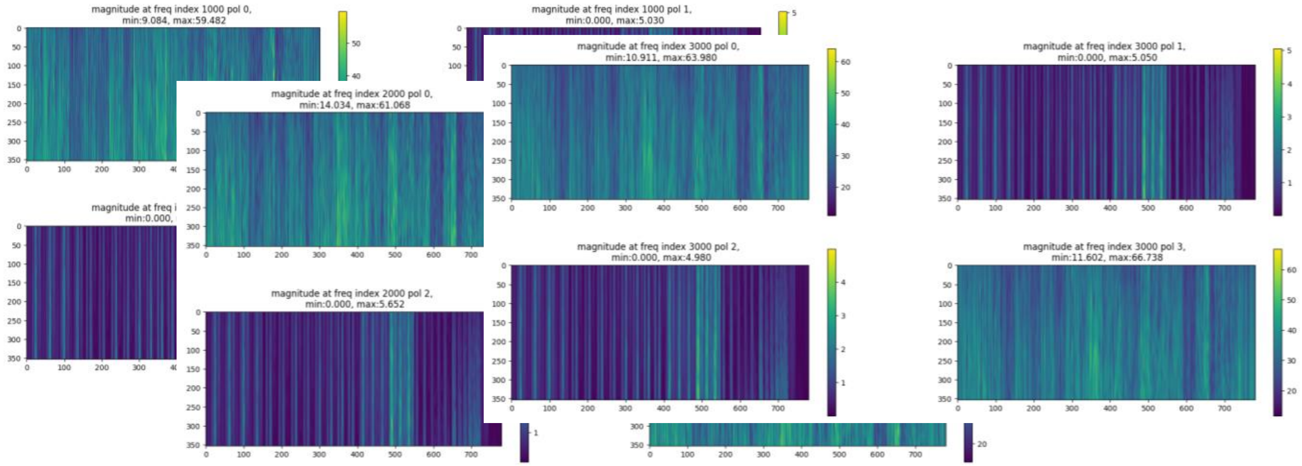
SKACH Winter Meeting 2026

Data compression for Radio Data

Visibility data by frequency channels

Illustration of Square Kilometer Array (SKA) (Wikipedia)

Data processing



Institute 1

Institute 2

Institute 3

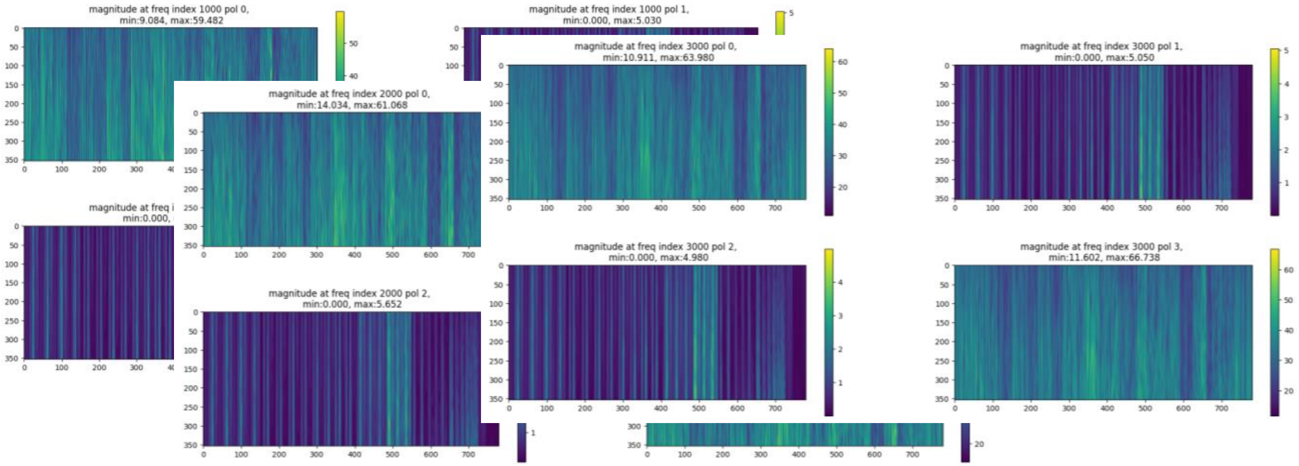
Institute 4

Data compression for Radio Data

Visibility data by frequency channels

Illustration of Square Kilometer Array (SKA) (Wikipedia)

Data processing



Transmitting large data in PB across continents

Institute 1

Institute 2

Institute 3

Institute 4

Background – Compression using DYSCO

DYSCO [Offringa, 2016] for visibility data

- Lossy compression, more preferred for noisy data,
- For complex-valued visibility data
- Two steps, normalization and quantization
- Normalization enables row (not preferred, could introduce more noise), row-frequency, antenna-frequency (preferred, lower image noise),
- Quantization, encoding the real and imaginary part using a gaussian distribution with zero mean, truncated at [-2.5σ, 2.5σ],
- Compressed into a defined bit-size, e.g. 10

Background – Compression using DYSCO

Swiss Data Science Center

- Rate-Distortion (R-D) trade-off
  - Rate (R) or bitrate: the number of bits used to represent the compressed data (bitrate);
  - Distortion (D): the loss in quality when reconstructing the original data from the compressed form (often measured using metrics like PSNR, SSIM, or perceptual loss);
  - Manually optimise R-D trade-off of DYSCO by choosing proper parameters

Background – Learning-based Models

Swiss Data Science Center

	DYSCO	Deep Learning Codecs
Optimizing R-D trade-off	Manual, via parameters	Learned, via R-D loss
Adaptive to data	No	Yes
Quantization optimization	Fixed distributions	Learned latent spaces
Practical for astronomy data?	Yes, tailored to visibility	Possible, CNN for complex valued input

AE-based → Less artifacts, faithful, but can be blurry

flow-based → Invertible transforms, more faithful, challenging training, worse results with lower bitrate

GAN-based

Diffusion models

→ Good perceptual quality, but can hallucinate details

Background – Learning-based Models

	DYSCO	Deep Learning Codecs
Optimizing R-D trade-off	Manual, via parameters	Learned, via R-D loss
Adaptive to data	No	Yes
Quantization optimization	Fixed distributions	Learned latent spaces
Practical for astronomy data?	Yes, tailored to visibility	Possible, CNN for complex valued input

AE-based → Less artifacts, faithful, but can be blurry

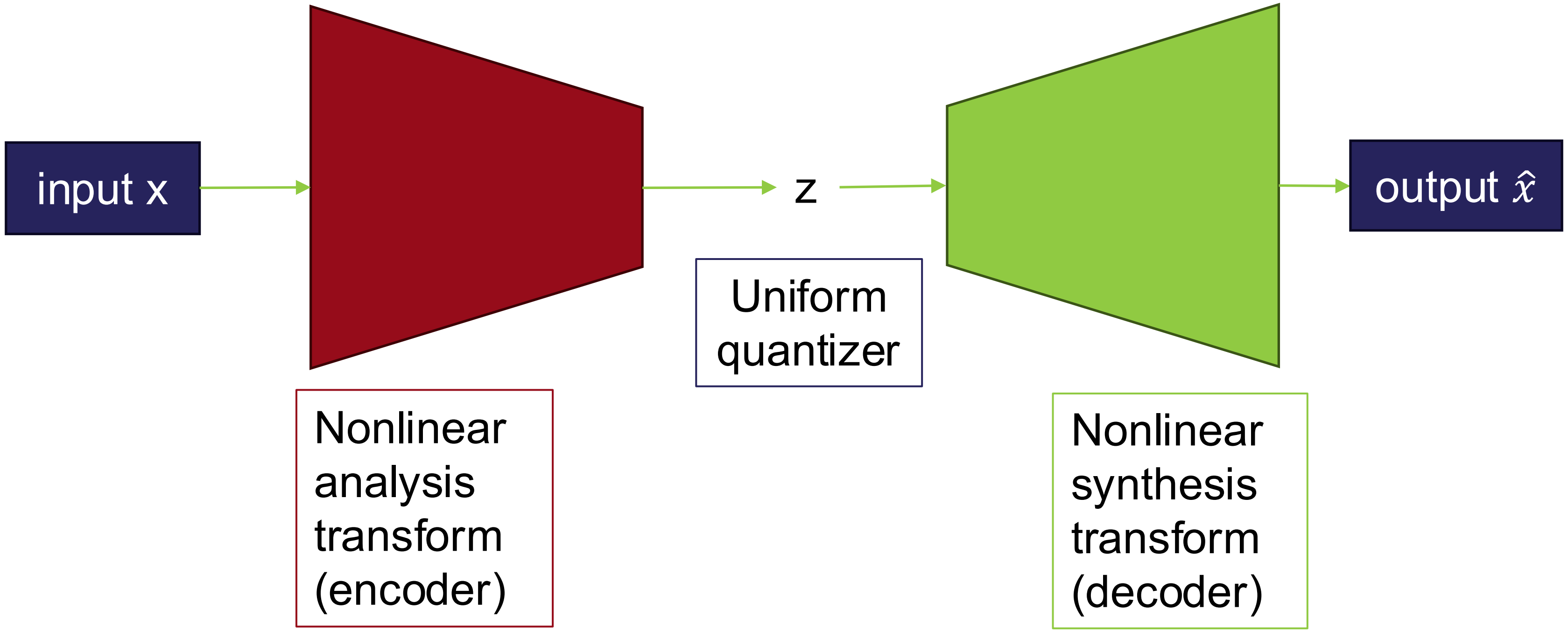
flow-based → Invertible transforms, more faithful, challenging training, worse results with lower bitrate

GAN-based, Diffusion models → Good perceptual quality, but can hallucinate details

Background – Autoencoder-based Models

- End-to-end optimized image compression (Ballé et al, 2016)
  - Proposed for image compression and aligns with VAE
  - Undifferentiable quantization handled by adding uniform noise:

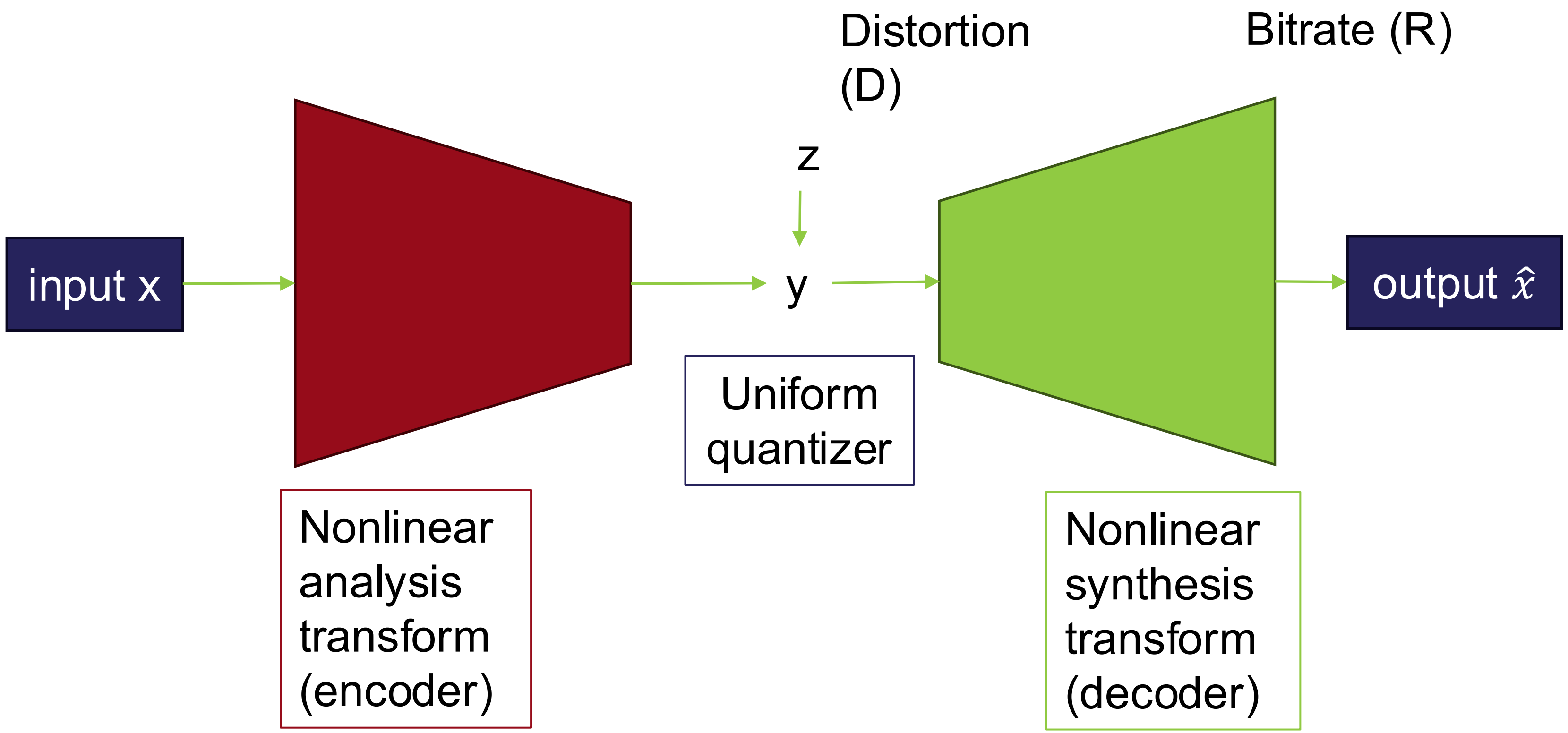
$\hat{z} = round(z)$, approximated as $\hat{z} = z + u, u \sim U(-0.5, 0.5)$



Swiss Data Science Center

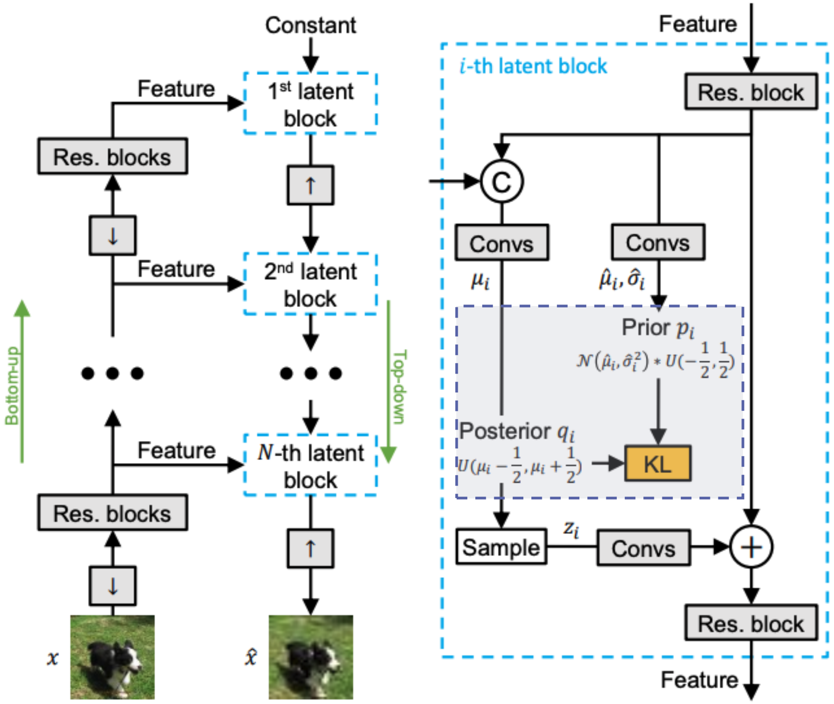
- End-to-end optimized image compression (Ballé et al, 2016)
  - Linking KL divergence with rate-distortion optimization $L = R + \lambda D$

$$D_{KL}[q||p_{z|x}] = E_{y\sim q} \log q(z|x) \; - E_{y\sim q} \log p_{x|z}(x|z) - E_{z\sim q} \log p_y(z) + const$$



Quantized Hierarchical VAE (Duan et al, WACV 2023)

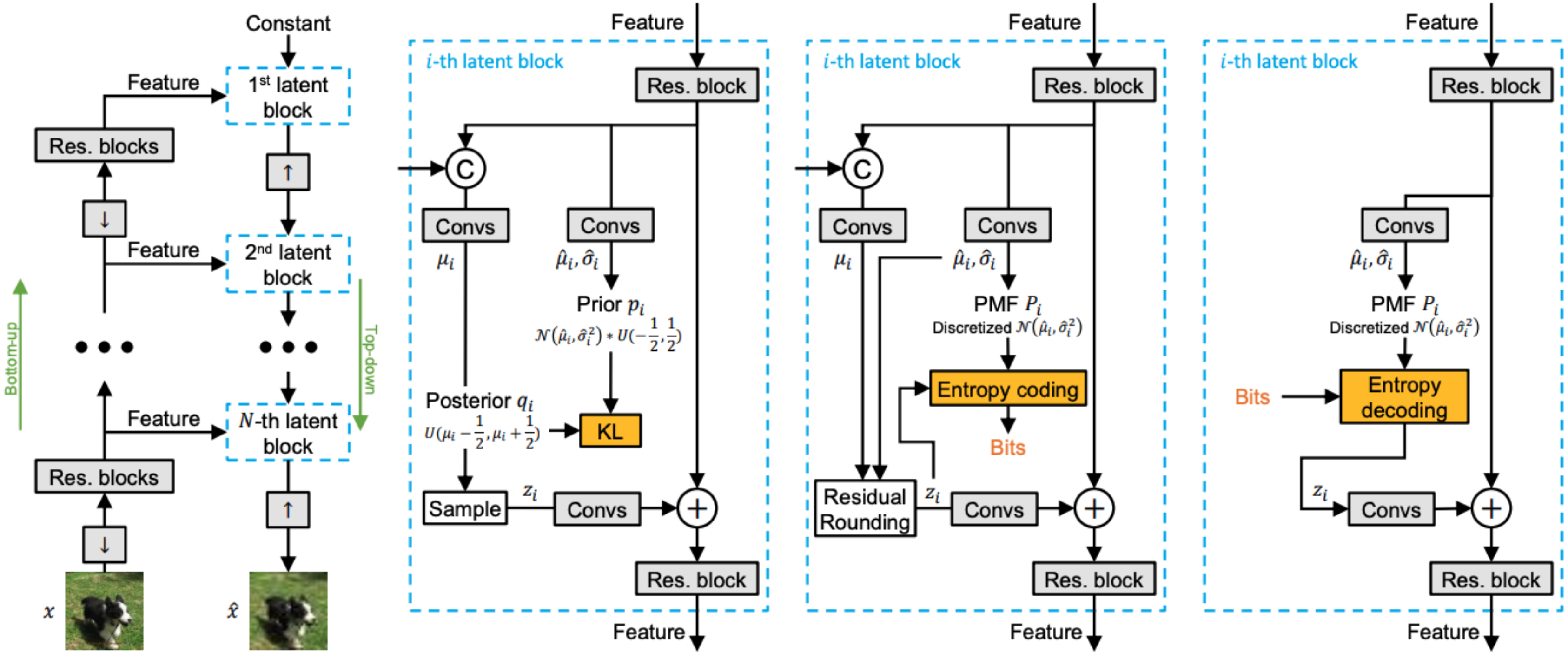
- Modelling data likelihood: $p_{X|Z}(x|z) \propto e^{-\lambda d(\hat{x},x)}$, $d(.)$ can be MSE
- Training objective: $L = \underline{D_{KL}(q_{Z|x|}||p_Z)} + E_{q_{Z|x}}[\log \frac{1}{p_{X|Z}(x|Z)}]$,



(a) General framework (b) Latent block (training)

Swiss Data Science Center

Quantized Hierarchical VAE (Duan et al, WACV 2023)
- Compress: $z \leftarrow \hat{\mu}_i + nearest\_round(\mu_i - \hat{\mu}_i)$,
- Encode into bits: $P_i(n) = p_i(\hat{\mu}_i + n|Z_{<i})$ with rANS
- Decompress: compute $P_i(.)$ for each latent block, add decoded z_i into feature maps



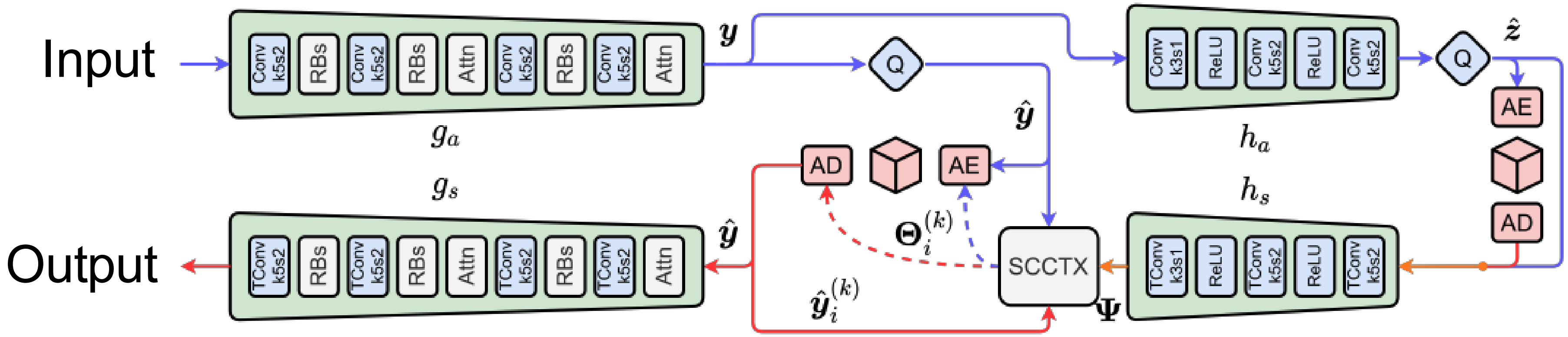
(a) General framework (b) Latent block (training) (c) Latent block (compress) (d) Latent block (decompress)

Background – Autoencoder-based Models

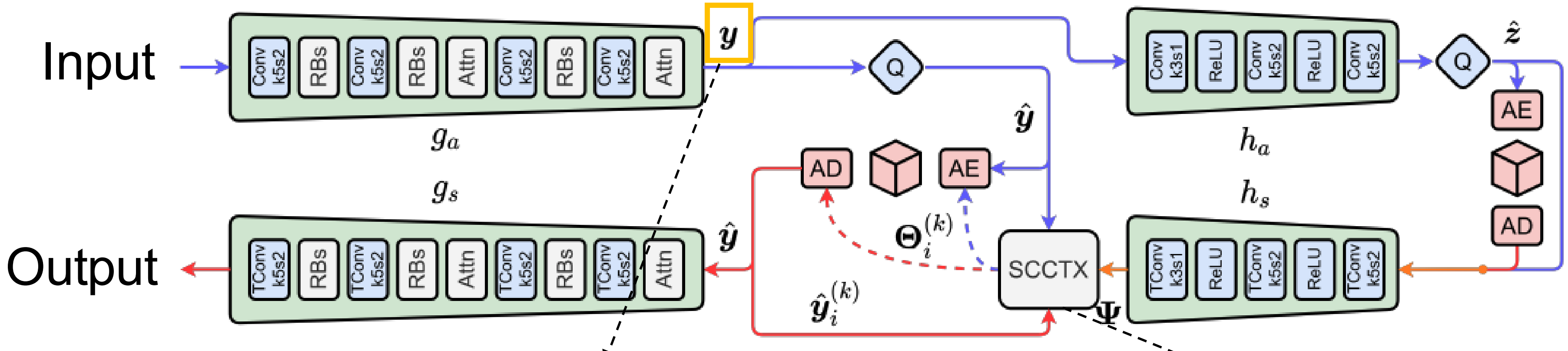
Swiss Data Science Center

Efficient Learned Image Compression (He et al, CVPR 2022)

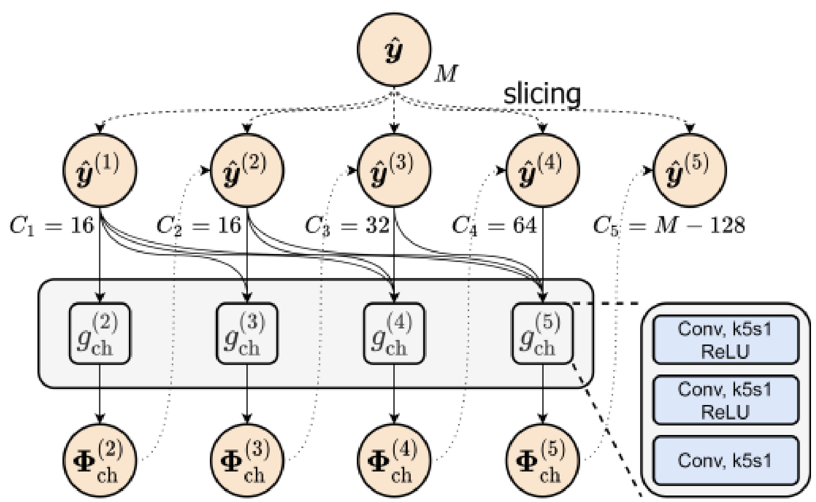
- Training objective: $L = R(\hat{y}) + \lambda D(x, g_s(\hat{y}))$, where $R(\hat{y}) = -E[\log p_{\hat{u}}(\hat{y})]$;
- Modelling latent codes: $p_{y|z}(y|z) = [N(\mu, \sigma^2) * U(-0.5, 0.5)]$;

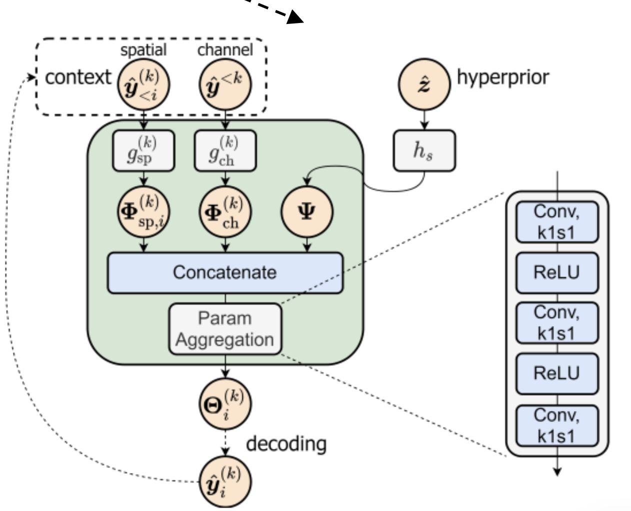


Efficient Learned Image Compression (He et al, CVPR 2022)



Uneven grouped context to speed up encoding



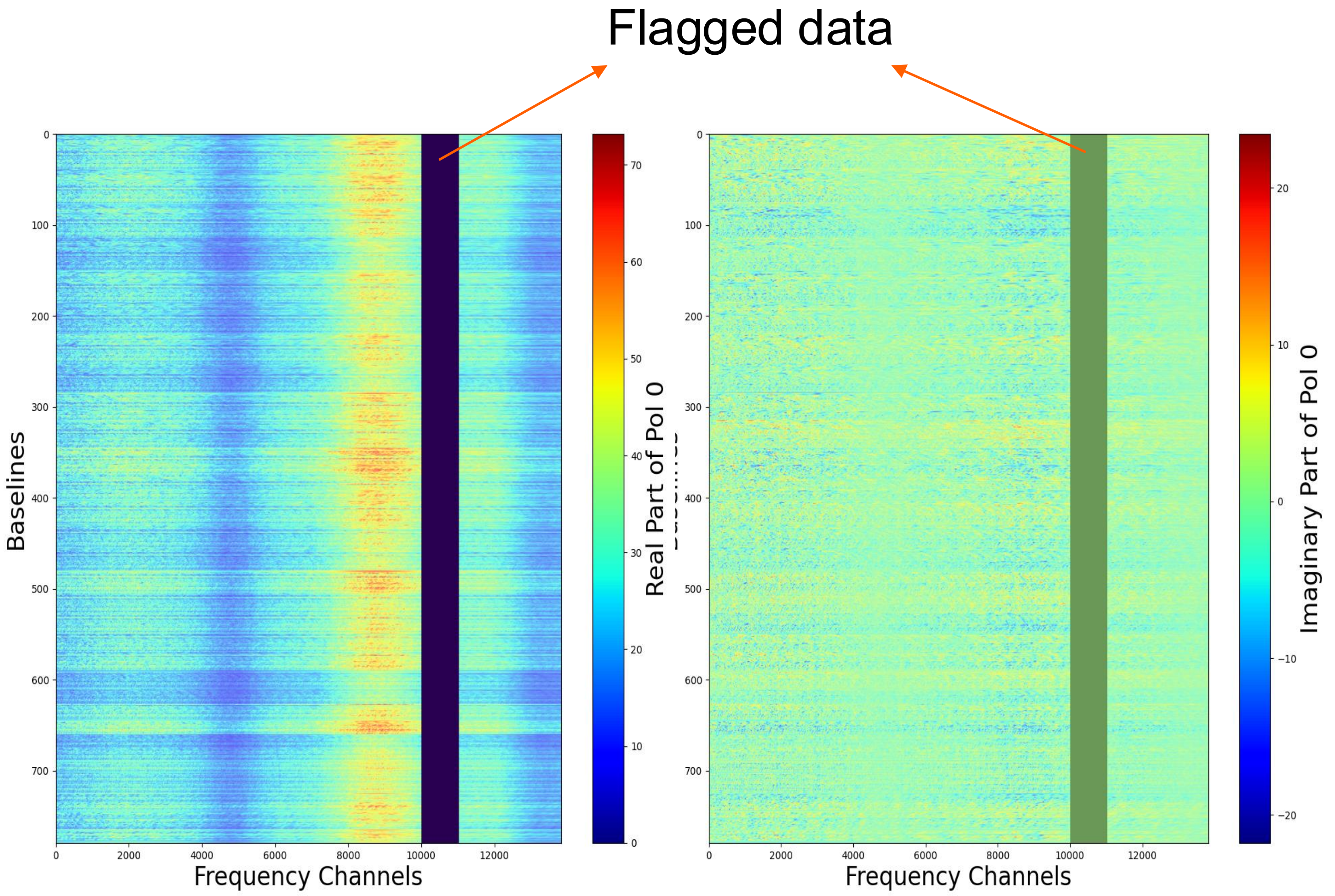


Parallel backward adaption coding in the spatial and channel axes

Datasets

Synthetic data

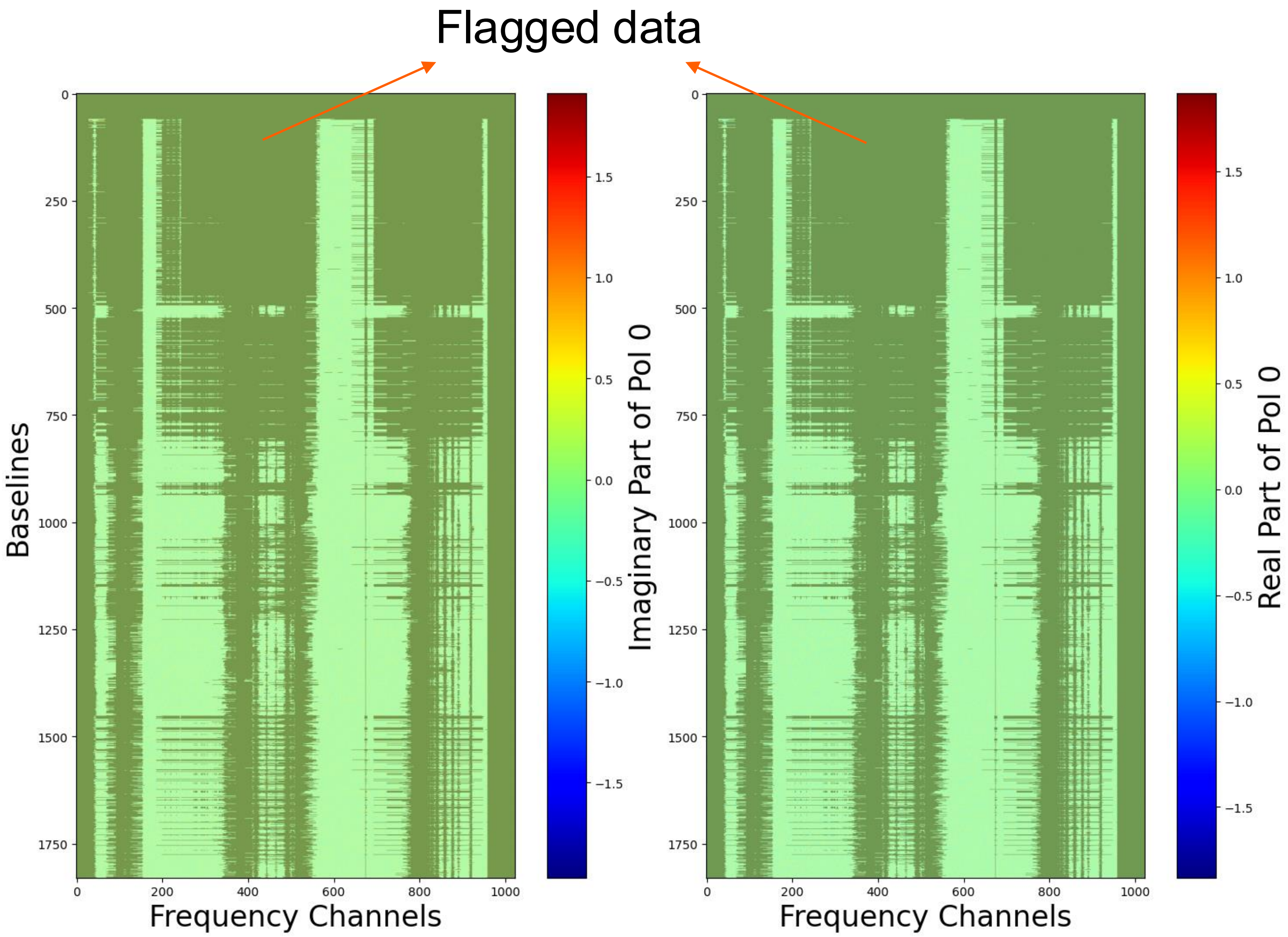
- Test dataset for the SKA SDP pipeline for SKA LOW (PI26 - 2025)
- Size: 875GB
- 13824 frequencies in the range of [1.000e+08Hz,1.750e+08Hz];
- 353 time integration of 1 s;
- 780 baselines;
- 7.41% flagged data among time integrations



Swiss Data Science Center

Real data

- MeerKAT dataset associated to the RATT PARROT publication
  - https://doi.org/10.1093/mnras/stae303
- Size: 411GiB
- 1024 frequencies in the range of [8.563e+08Hz,1.711e+09Hz];
- 3644 time integrations of 8s;
- 1830 baselines;
- 63.86% flagged data among time integrations



Applying Autoencoder-based Models to Radio Data

Swiss Data Science Center

Data preparation

- Concatenate complex values as real and imaginary parts → data with 8 channels
- Explore spatial and temporal correlation → data shape of (n_time, n_baseline) across frequencies;
- Replace flagged data as 0+0i;
- Random split into 70% as training data, 20% as validation data and 10% as test data.

Training objectives

- Additional loss for magnitude and phases to preserve relations between real and imaginary parts:

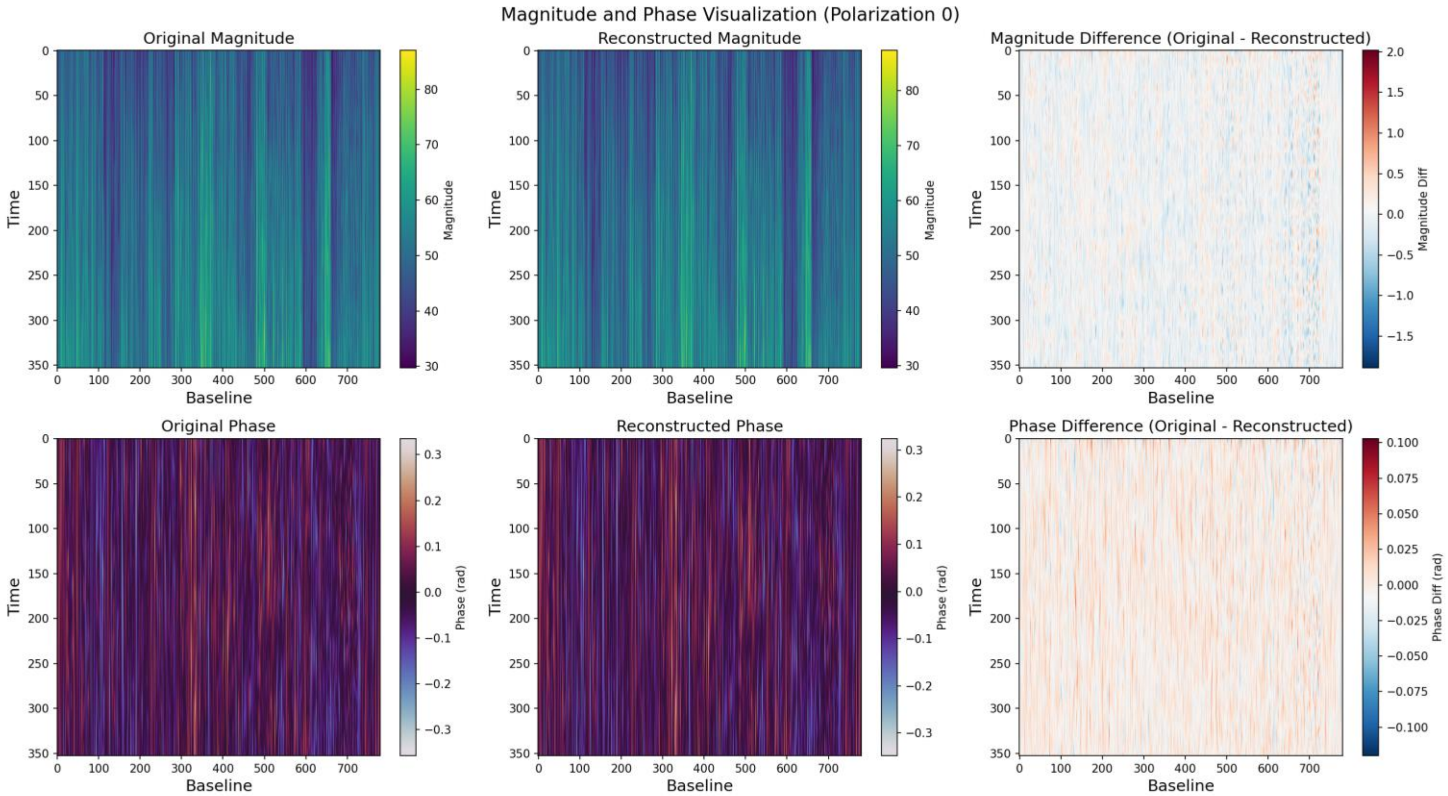
$$L = L_{model} + \lambda_1 \left| \mathrm{Mag}_{\mathrm{input}} - \mathrm{Mag}_{\mathrm{output}} \right| + \lambda_2 wrap\left(|\phi_{gt} - \phi_{rec}| \right)$$

Infrastructure

- Data preprocessing: CPU nodes on CSCS Eiger, with x86_64 architecture;
- Model training: A100 GPU on CSCS Daint, with aarch64 architecture

Comparative Results – Synthetic Data

	MSE	BPP	Mag. bias	Phase bias	Decoding time (s/freq)
ELiC	0.108	0.271	0.107	0.047	2.605
QH-VAE	0.127	0.280	0.113	0.061	2.790



Comparative Results – Synthetic Data

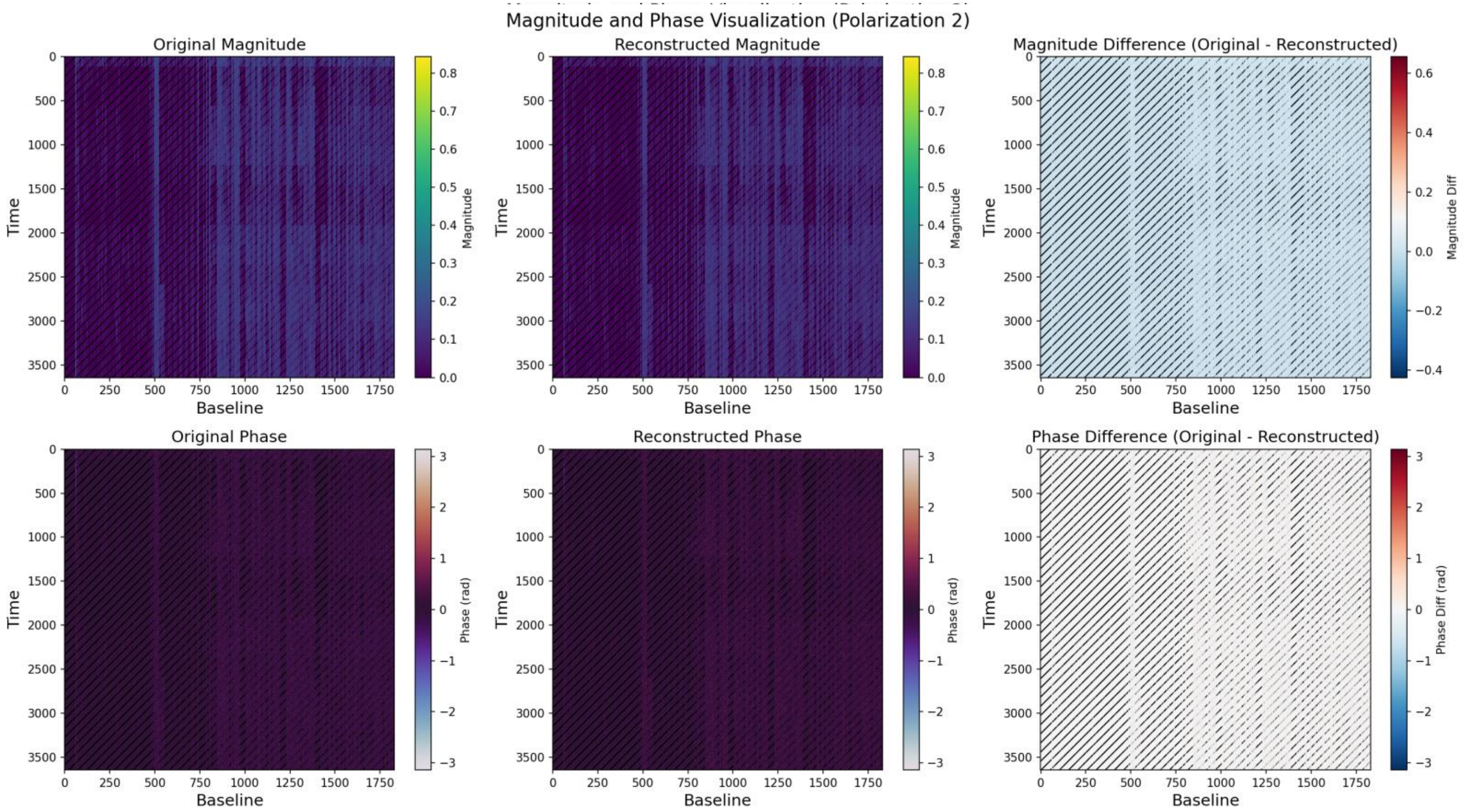
Swiss Data Science Center

Effect of the additional loss with varied λ_1 and λ_2

	MSE	BPP	Mag. bias	Phase bias
No Additional Loss	0.274	0.121	0.341	0.147
λ_1=1, λ_2=1	0.201	0.139	0.267	0.135
λ_1=5, λ_2=5	0.231	0.196	0.131	0.102
λ_1=10, λ_2=10	0.108	0.271	0.107	0.047

Comparative Results – Real Data

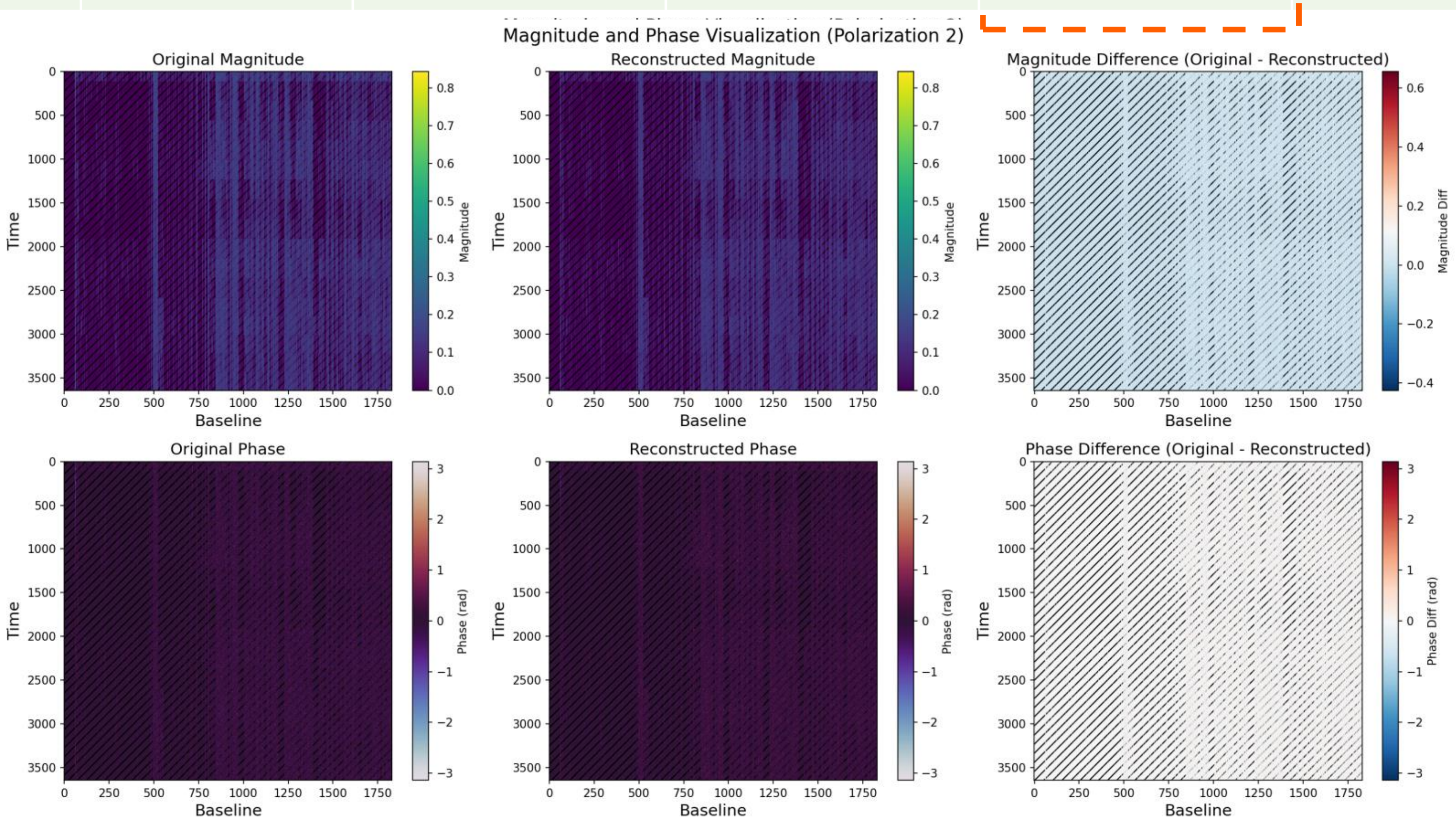
	MSE	BPP	Mag. bias	Phase bias	Decoding time (s/freq)
ELiC	0.039	13.631	0.084	0.234	15.17
QH-VAE	0.047	13.872	0.093	0.252	14.37



58.126% masked data

Comparative Results – Real Data

	MSE	BPP	Mag. bias	Phase bias	Decoding time (s/freq)
ELiC	0.039	13.631	0.084	0.234	15.17
QH-VAE	0.047	13.872	0.093	0.252	14.37



58.126% masked data

Thank you!

References

1. Offringa, A. R. "Compression of interferometric radio-astronomical data." *Astronomy & Astrophysics* 595 (2016): A99.
2. Ballé, Johannes, Valero Laparra, and Eero P. Simoncelli. "End-to-end optimized image compression." arXiv preprint arXiv:1611.01704 (2016).
3. He, Dailan, et al. "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.
4. Duan, Zhihao, et al. "Lossy image compression with quantized hierarchical vaes." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2023.