IEEE INFOCOM

# CSI-StripeFormer: Exploiting Stripe Features for CSI Compression in Massive MIMO System

Qingyong Hu[1], Hua Kang[1], Huangxun Chen[2], Qianyi Huang[3], Qian Zhang[1], Min Cheng[4]

[1] Hong Kong University of Science and Technology
[2] Huawei
[3] Sun Yat-Sen University
[4] Noah's Ark Lab, Huawei

# Outline

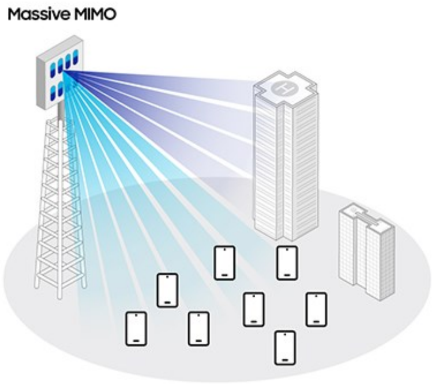
- Background
- Related Works
- Key Observations
- System Design
- Evaluations

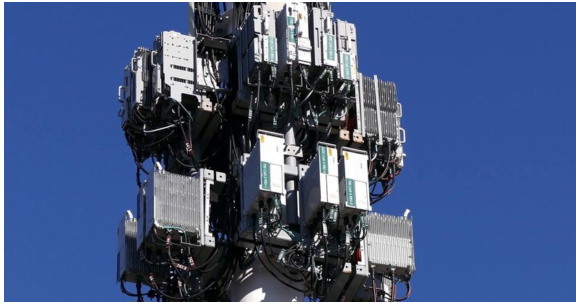
# 01

# Background

# Background

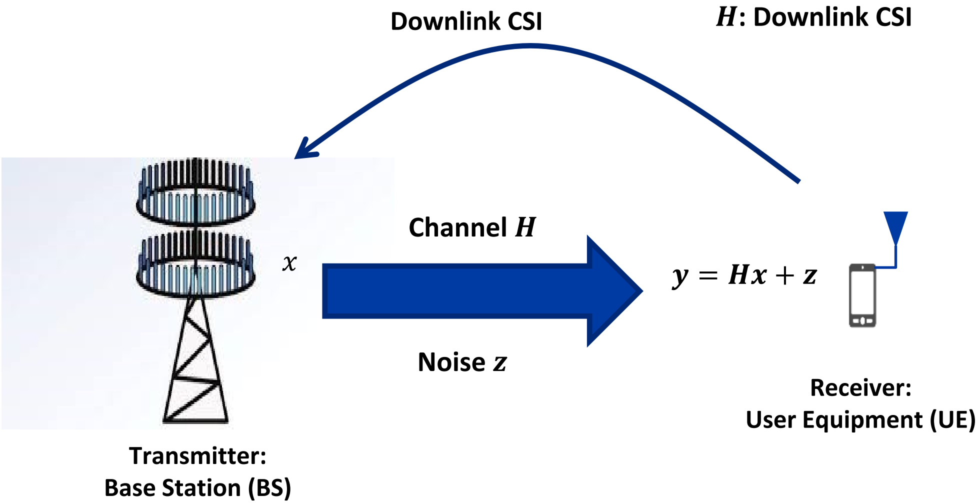
- Massive MIMO is a key enabler of improving throughput of wireless communication in 5G and beyond networks.
- It exploits spatial diversity by combining massive antennas at the base station (BS) to greatly improve the spectral efficiency.

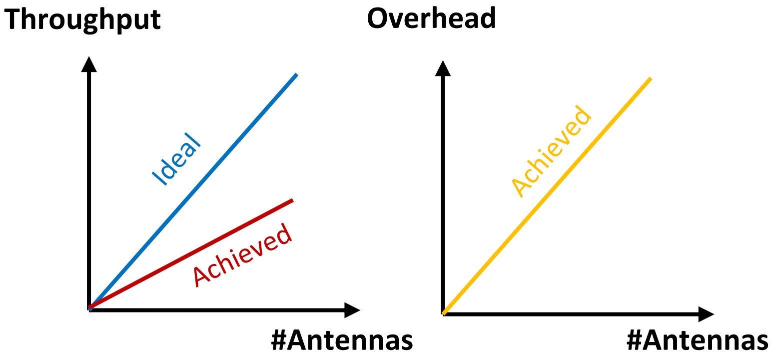


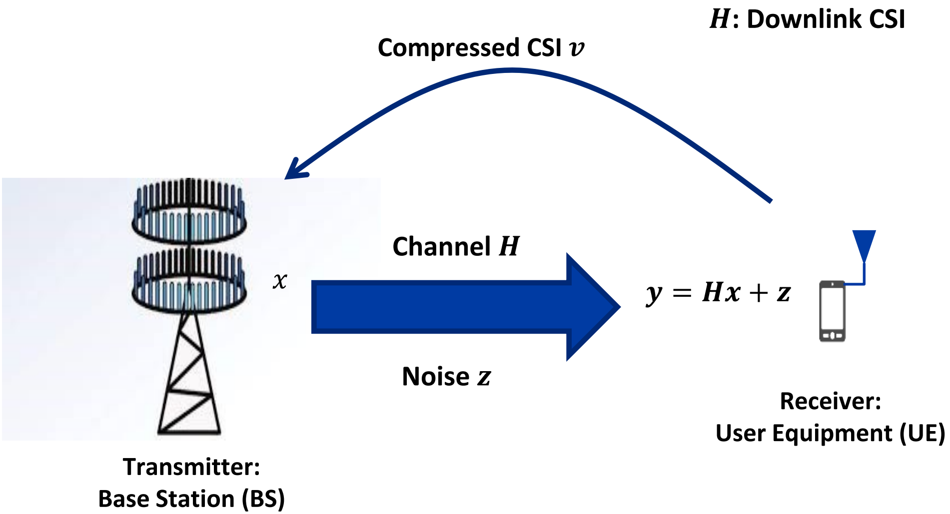
# Background

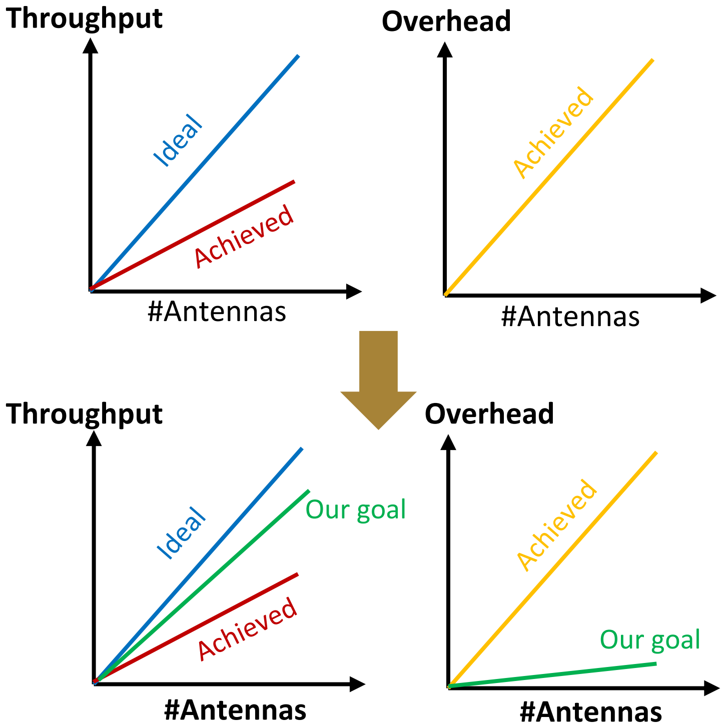
- The transmitter needs the channel state information (CSI) to design the optimal transmission scheme such as multi-user beamforming.
- However, CSI matrix grows linearly with the number of antennas and the frequency bandwidth. **More antennas indicate more feedback overhead**.



# Our goal




We can reduce the downlink CSI feedback overhead
by **compressing** it at the UE side and then **recovering** at the BS side.

02

# Related Works

# Related Works

**Compressive Sensing-based Solutions**

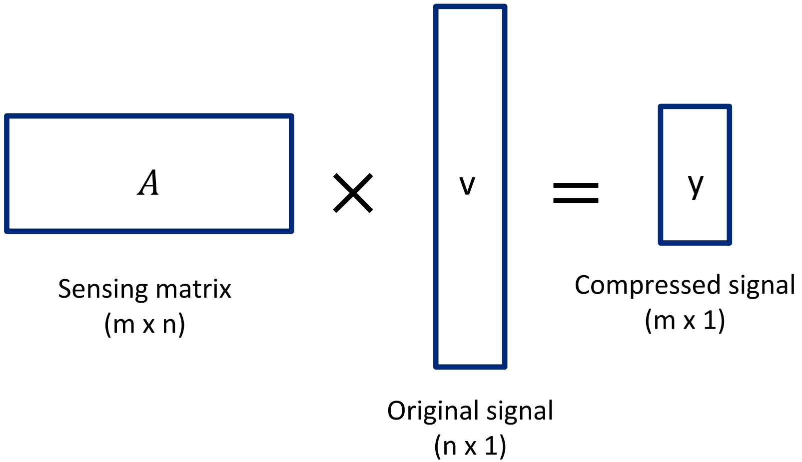
(Lasso [10], TVAL3 [11] etc.)

**Pros:**

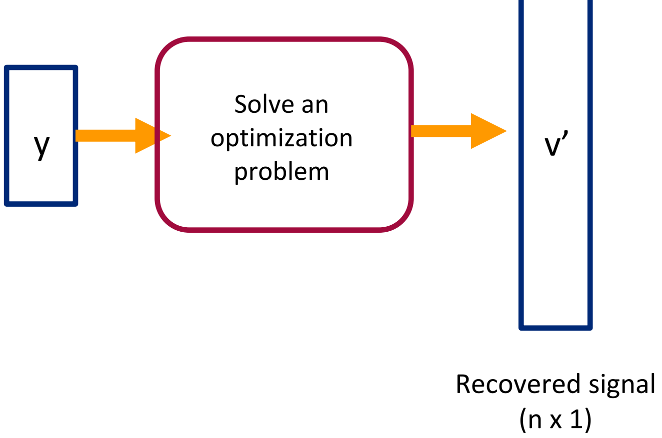
- Enjoy theoretical guarantee.

**Cons:**

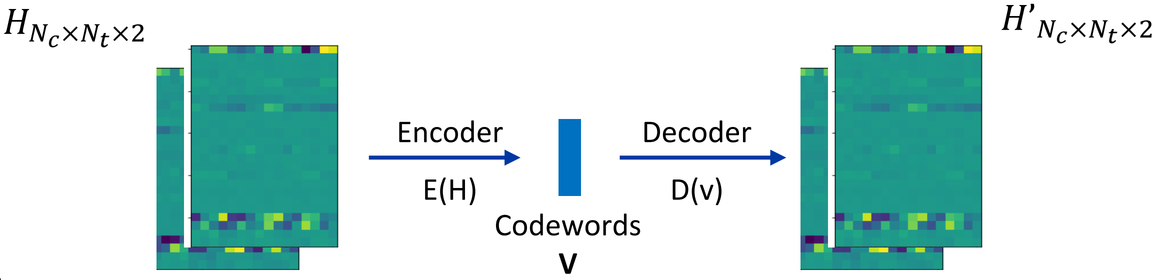
- The assumption on sparsity is too ideal.
- Hard to support high compression requirement.




$$\arg \min_{\boldsymbol{\nu}\in\mathbb{R}^{N\times 1}} \frac{1}{2}|\boldsymbol{y} - \mathbf{A}\boldsymbol{\nu}|_{l_2}^2 + \theta\sigma_\omega|\boldsymbol{\nu}|_{l_1}$$



# Related Works



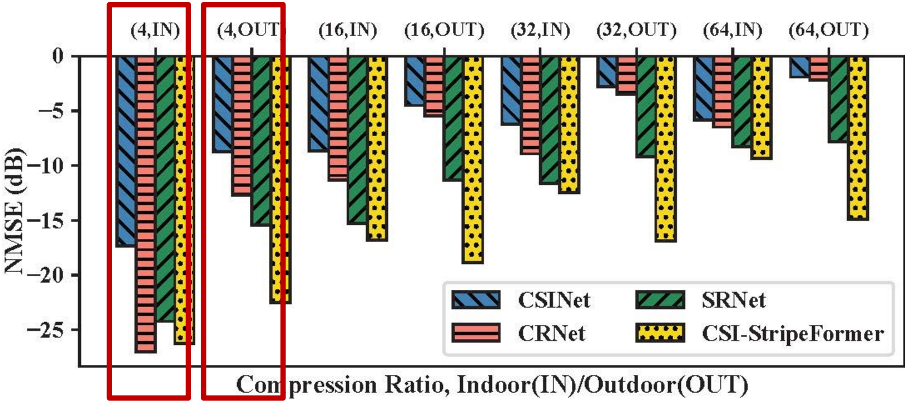

**Deep Learning-based Solutions**
(CSINet [1], CRNet [7], SRNet [8] etc.)

**Pros:**

- Relax the sparsity assumption.
- Achieve a better performance due to non-linear transformations.

**Cons of current works:**

- **Unbalanced performance** across various scenarios: consistently much worse on the outdoor dataset than the indoor one.
- **Unsatisfactory performance** on high compression ratio requirement.
- Simply regard CSI as images without leveraging unique characteristics of CSI matrix.



## 03

# Our Observations

We aim to find the reasons of two questions:

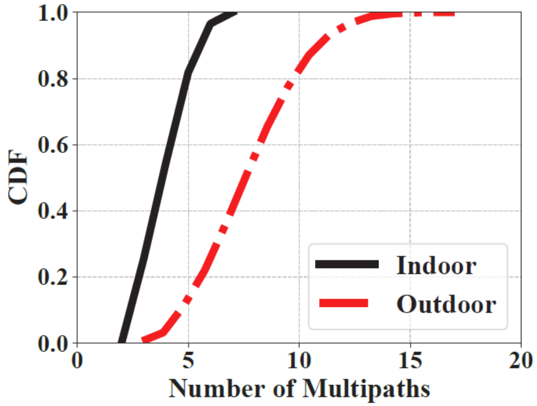
1. What is the key difference of various scenarios that influence the models' performance?
2. What is the unique feature that differs CSI matrix from images?

# Our Observations

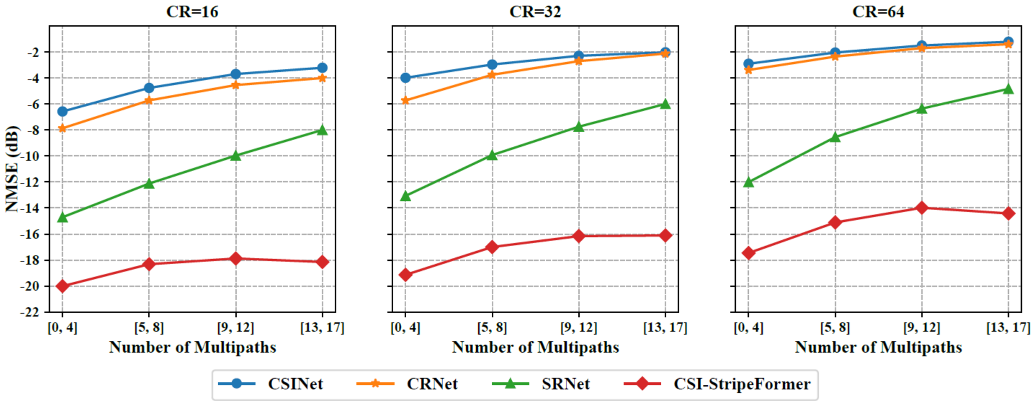
## Multipath Effects on CSI Compression

- We analyze the multipath distributions of two public COST2100 datasets with MUSIC algorithm and test the trained model on all multipath cases to validate its influence.

- The outdoor dataset contains much **more multipaths** than the indoor dataset !
- The performance of SOTA models **degrades with richer multipath effects** !




**Multipath distributions across various datasets**




**Performances of SOTA models on different subsets**

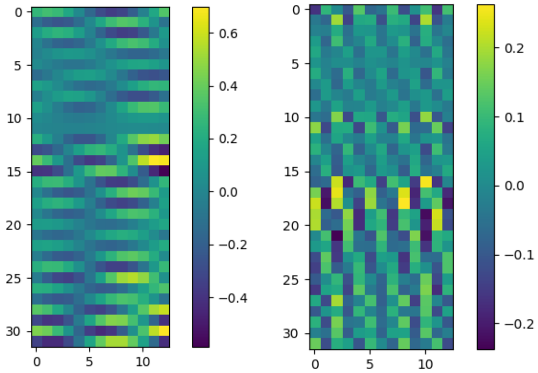
# Our Observations

## ☐ CSI Unique Feature: Stripe-based Correlation

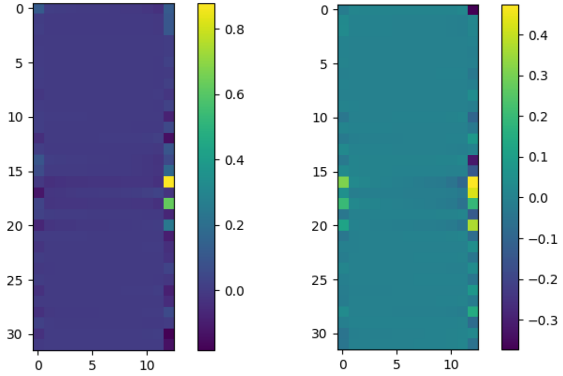
- One DL CSI signal can be determined by the frequency, propagation delay and angle-of-departure (AoD).

$$h_n(f) = \sum_{i=1}^{K} (a(f, d_i) e^{-j2\pi \frac{d_i f}{c} + j\phi(f, d_i)}) e^{-j2\pi \frac{n l cos\theta_i}{c/f}}$$

- A common method is to convert time-frequency CSI to angular-delay domain with 2D Discrete Fourier Transformation (2D DFT) for sparser representations.




Time-Frequency domain CSI




Angular-Delay domain CSI

# Our Observations

## CSI Unique Features: Stripe-based Correlation

- Ideally, one CSI corresponds to one pixel at the Angular-Delay domain.

- However, this 2D-DFT transformation is not perfect. Its resolutions are **determined by the window size**, i.e. #antennas and #subcarriers.
- This windowing effect will lead to **spectral leakage**.
- It makes the energy of one CSI element diffused in both horizontal and vertical directions of the CSI matrix and forms as ‘**Stripes**’ in both directions.

- Difference from images:
  - Images: strong correlations in local patch regions
  - CSI: strong correlations across stripe regions

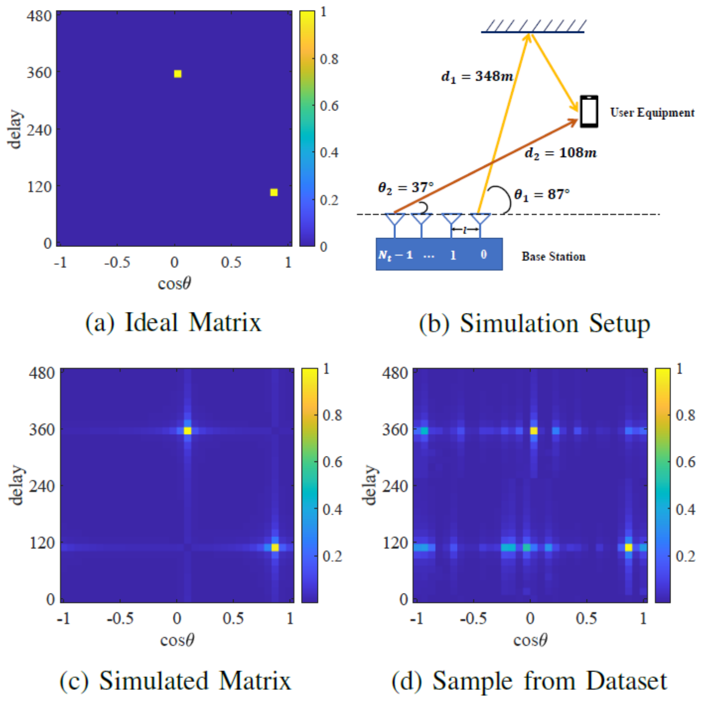



**Illustration of stripe features in angular-delay CSI**

04

# System Design

The stripe observation inspires us to tailer the deep CSI compression system for the stripe-based correlation.

# Requirements of the Deep Compression System

### Lightweight encoder

The UE is usually computation resources-limited.

### Global receptive field

The energy spreads across the whole CSI matrix.

### Ability to capture stripe-based correlation

Angular and delay dimensions have different physical characteristics.

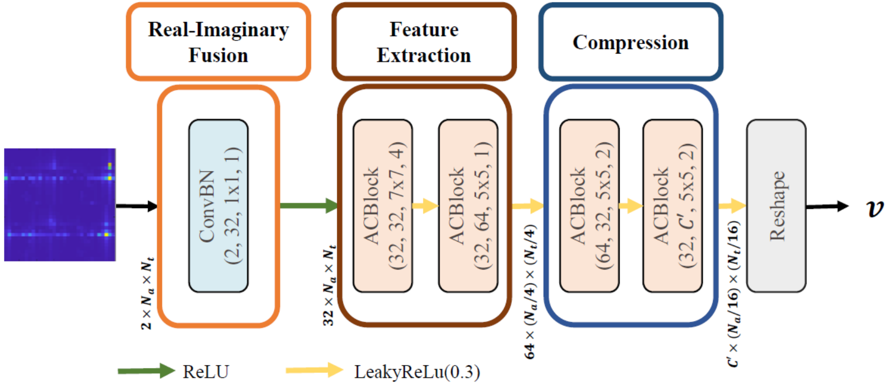
### Dynamic fusion correlations from different dimensions

Both angular and delay domains can jointly determine one CSI signal.

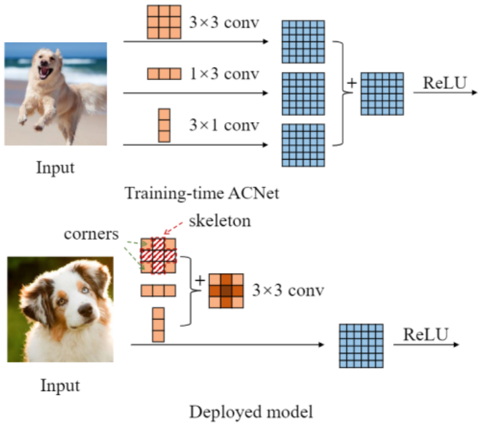
# System Design

## Lightweight Encoder

- Real-Imaginary Fusion
  - The real and imaginary part of CSI can determine the signal's amplitude and phase jointly.
- Feature Extraction
  - We leverage the asymmetric convolution block to extract features from various shapes.
- Compression
  - We use convolution blocks for adaptive compression.



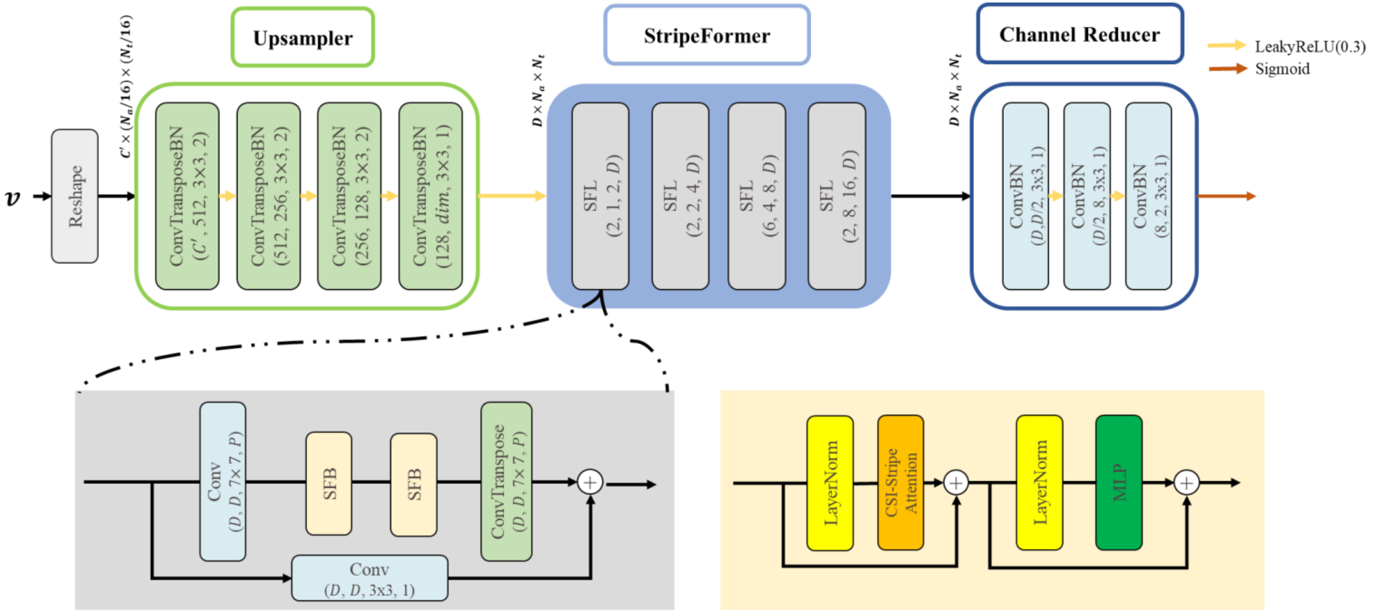

The architecture of CSI-StripeFormer Encoder




Asymmetric convolution block [2]

# System Design

## ☐ CSI-StripeFormer Decoder




(a) StripeFormer Layer (SFL)

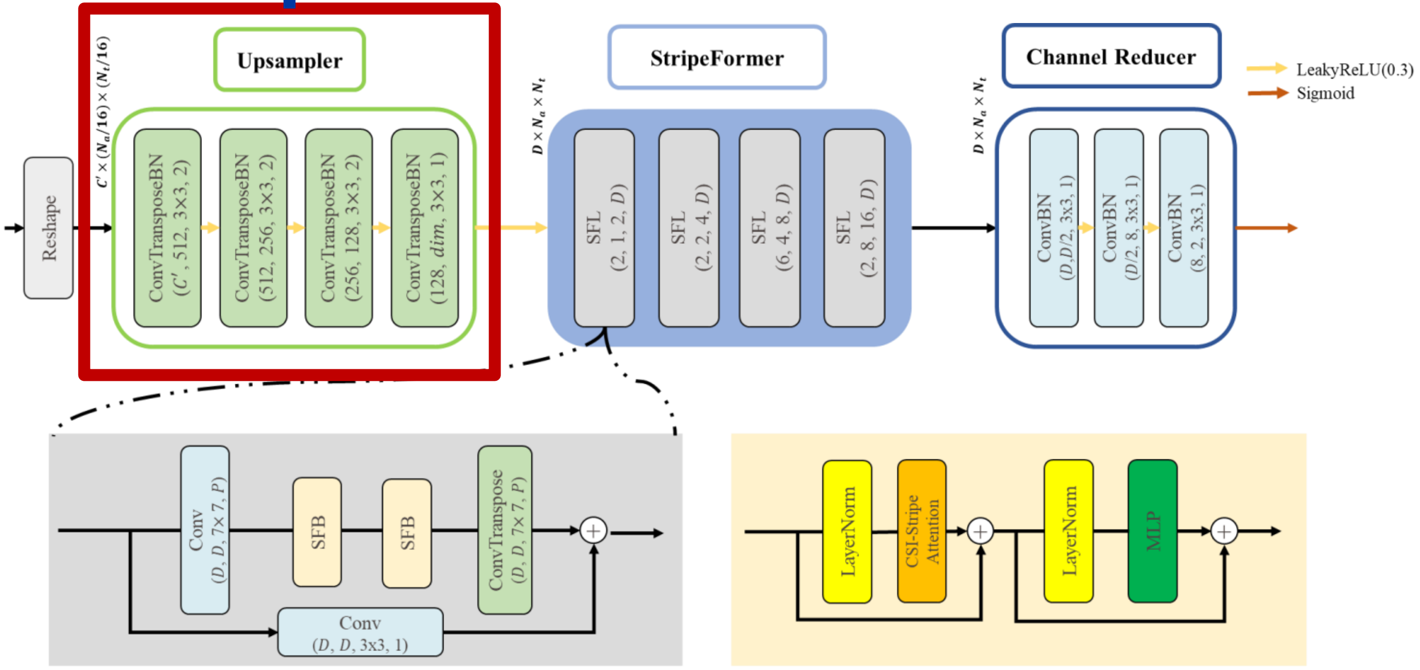
(b) StripeFormer Block (SFB)

**The architecture of CSI-StripeFormer Decoder**

# System Design

## CSI-StripeFormer Decoder

**Transposed Convs as a coarse recovery.**



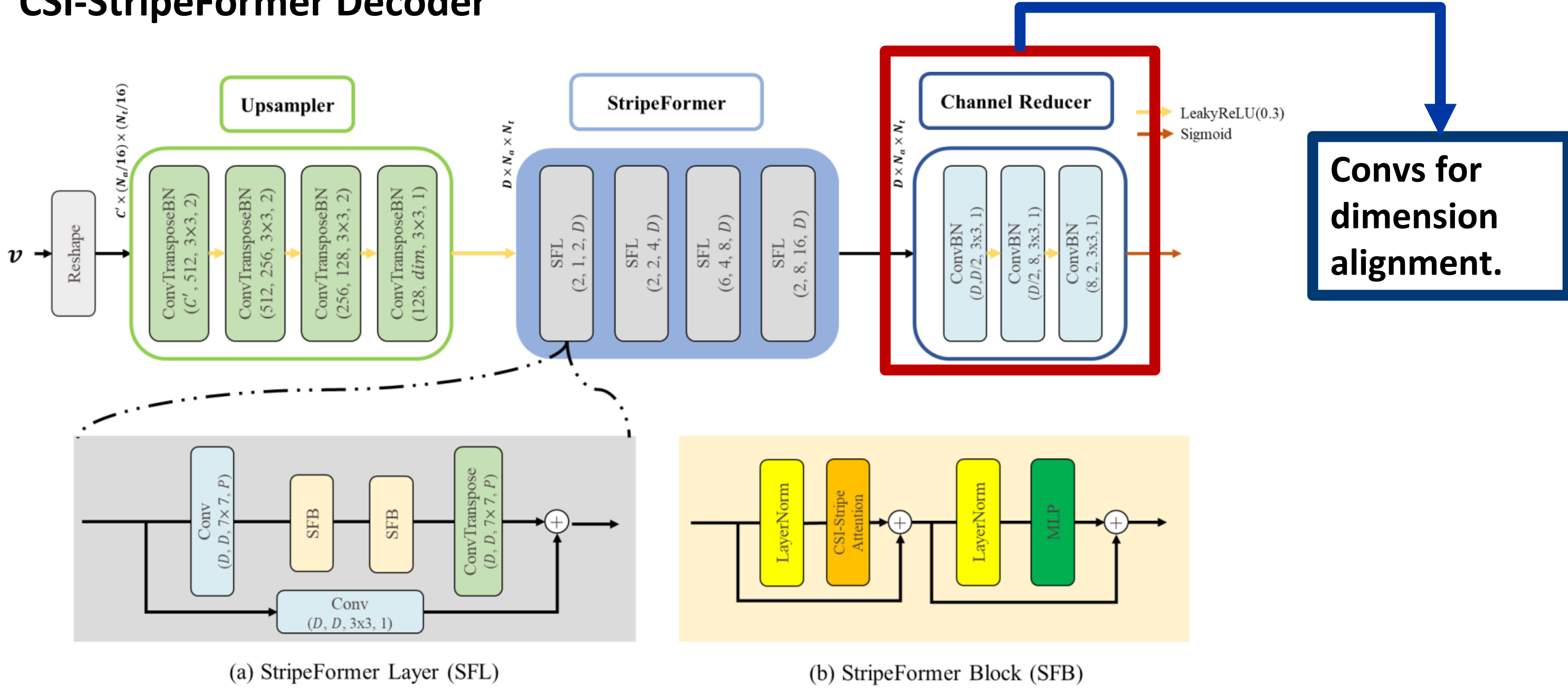

(a) StripeFormer Layer (SFL)

(b) StripeFormer Block (SFB)

**The architecture of CSI-StripeFormer Decoder**

# System Design

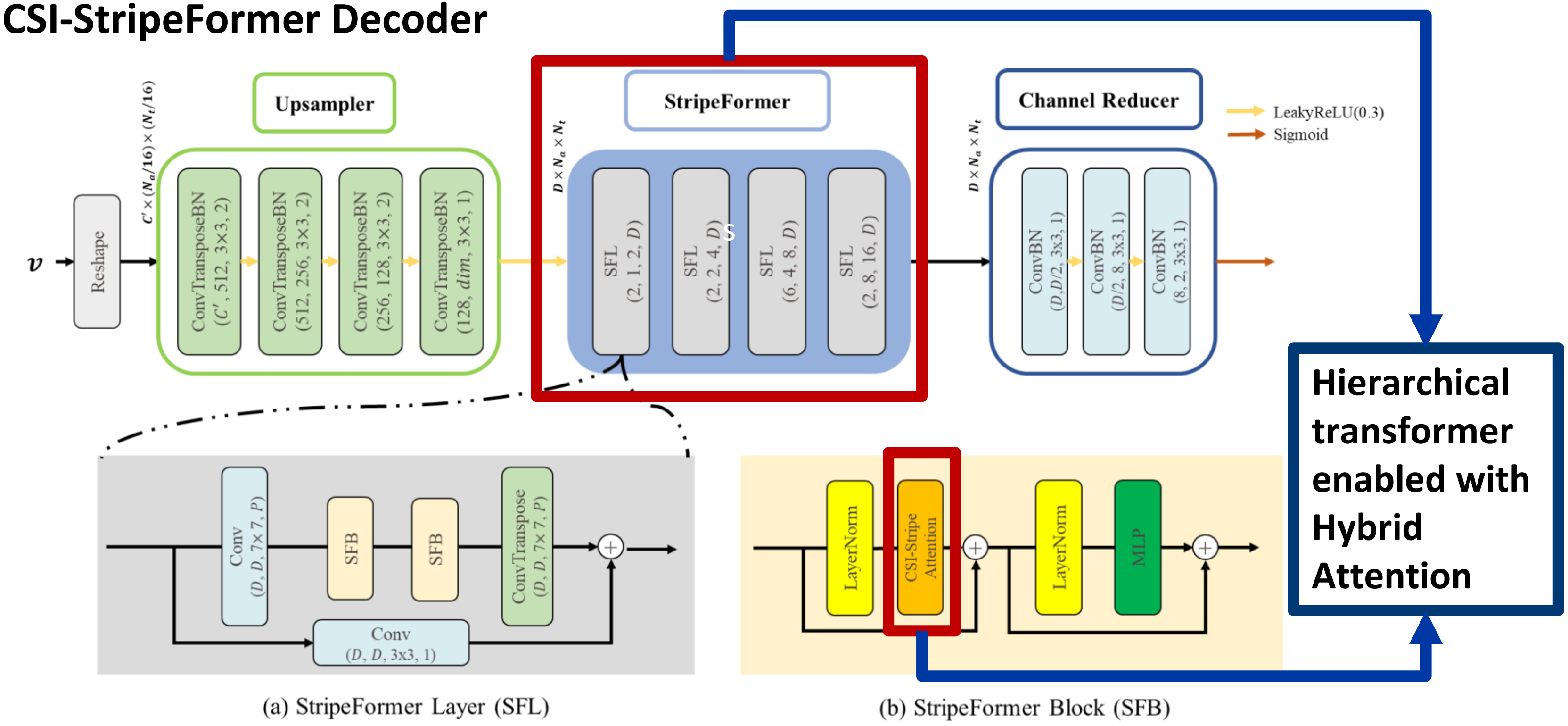
## ☐ CSI-StripeFormer Decoder




**The architecture of CSI-StripeFormer Decoder**

# System Design

- CSI-StripeFormer Decoder




The architecture of CSI-StripeFormer Decoder

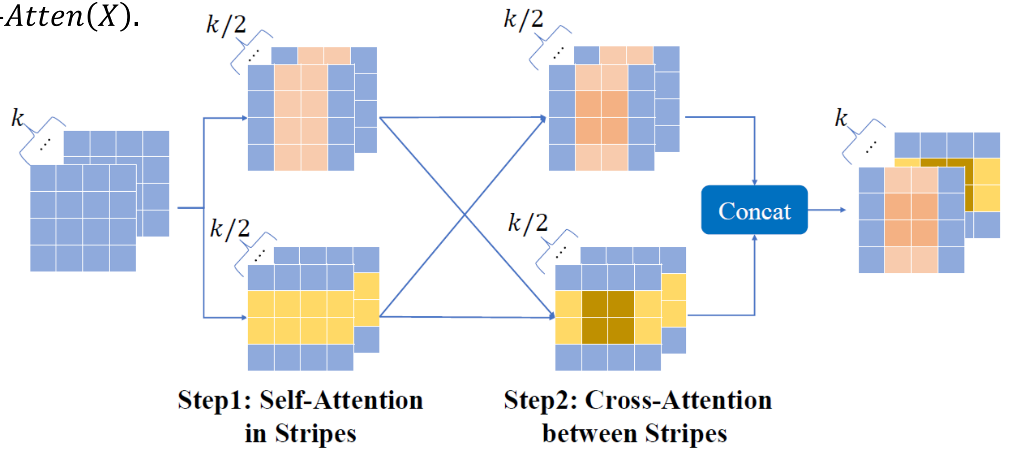
# CSI-StripeFormer Decoder

## Hybrid Attention

- Step 1: Angular-Delay Self-Attention

We leverage CSWin Attention [4] for calculating Angular-Delay Self-Attention.
Suppose the input is $X$, the output of the horizontal stripes is $H\text{-}Atten(X)$.
Similar operations on vertical stripes generate the output $V\text{-}Atten(X)$.

$$
\begin{aligned}
[H_1, H_2, ..., H_M] &= \mathrm{Split}(X), \\
[Q_i^k, K_i^k, V_i^k] &= [H_i W_Q^k, H_i W_K^k, H_i W_V^k], \\
A_i^k &= \mathrm{Softmax}\left[\frac{Q_i^k (K_i^k)^T}{\sqrt{d_k}}\right], \\
\mathrm{LePE}(V_i^k) &= \mathrm{Conv}(V_i^k), \\
O_i^k &= A_i^k V_i^k + \mathrm{LePE}(V_i^k), \\
\text{H-Atten}^k(X) &= [O_1^k, O_2^k, ..., O_M^k], \\
\text{H-Atten}(X) &= [\text{H-Atten}^1(X), ..., \text{H-Atten}^N(X)].
\end{aligned}
$$




Hybrid Attention in StripeFormer Block

# CSI-StripeFormer Decoder

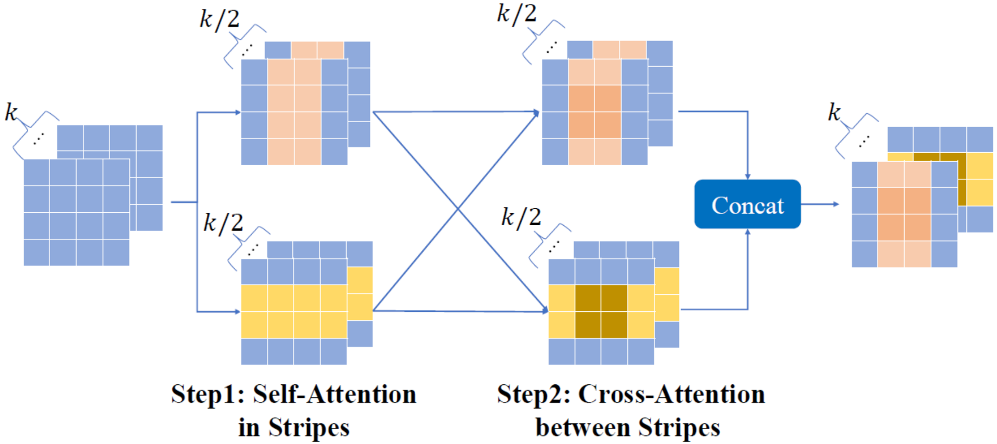
## ☐ Hybrid Attention

- Step 2: Angular-Delay Cross-Attention

We use residual cross attention [5] to fuse the correlations of angular and delay domains by using one single domain as (query, key), and the other domain as value.

$$[Q_{X_2}, K_{X_2}, V_{X_1}] = [X_2 W_Q, X_2 W_K, X_1 W_V],$$
$$\mathrm{CrxAtten}(X1, X2) = [\mathrm{Softmax}(Q_{X_2} K_{X_2}^T)] V_{X_1} + X1.$$

$$\text{H-V Atten} = \mathrm{CrxAtten}(\text{H-Atten(X)}, \text{V-Atten(X)}),$$
$$\text{V-H Atten} = \mathrm{CrxAtten}(\text{V-Atten(X)}, \text{H-Atten(X)}),$$
$$Y = \mathrm{Concat}[\text{H-V Atten}, \text{V-H Atten}] W^O,$$



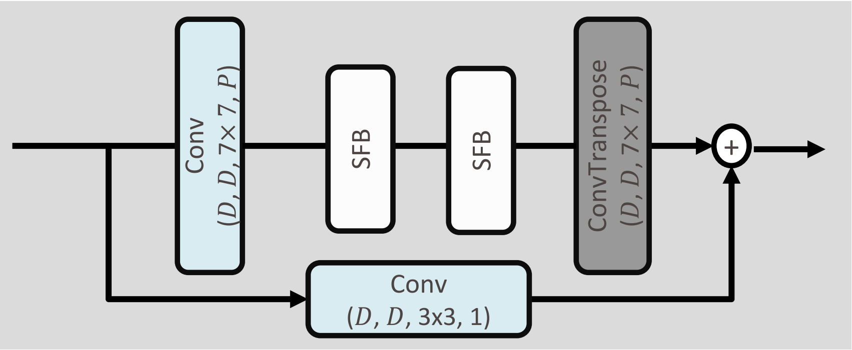

Hybrid Attention in StripeFormer Block

# CSI-StripeFormer Decoder

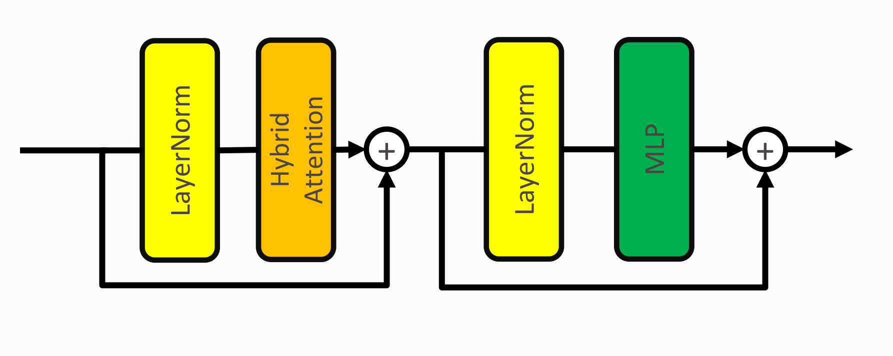
## Putting all together

$$\hat{X}^l = \text{Hybrid-Attention}(\text{LN}(X^{l-1})) + X^{l-1},$$
$$X^l = \text{MLP}(\text{LN}(\hat{X}^l)) + \hat{X}^l.$$

- **StripeFormer Block (SFB)**
  - SFB consists of hybrid attention block, layer normalization (LN) block and multilayer perceptron (MLP) block. $X^l$ is the output of the $l$-th SFB or the output of convolutional embeddings.
- **StripeFormer Layer (SFL)**
  - SFL is a two-branch structure to jointly consider stripe features and potential patch features. The stripe branch utilizes convolution kernels to assist SFBs with computation reduction and transposed convolution kernels to recover the output of SFBs.




**StripeFormer Layer (SFL)**




**StripeFormer Block (SFB)**

05

# Evaluations on COST2100 and ALAR Datasets

# Evaluation Setup

- COST2100 [3]

- Evaluated on two public benchmark datasets with a bandwidth of 20 MHz
- Indoor picocellular scenario at 5.3 GHz
- Outdoor rural scenario at 300 MHz
- BS has $N_t = 32$ uniform linear array antennas, UE has $N_r = 1$ antenna.
- Evaluation metric: Normalized Mean Square Error (NMSE)

$$\mathrm{NMSE} = \mathbb{E}\left[\frac{\| \boldsymbol{H_a} - \hat{\boldsymbol{H}}_{\boldsymbol{a}} \|^2}{\| \boldsymbol{H_a} \|^2}\right]$$

- Baseline:
  - Deep Neural Network (DNN) Based: CSINet [1], CRNet [7], TransNet [9], ACCsiNet [12]
  - Hybrid Model (HM) Based: SRNet [8], IdasNet [13]

# Evaluation Results on COST2100

## Overall Performance

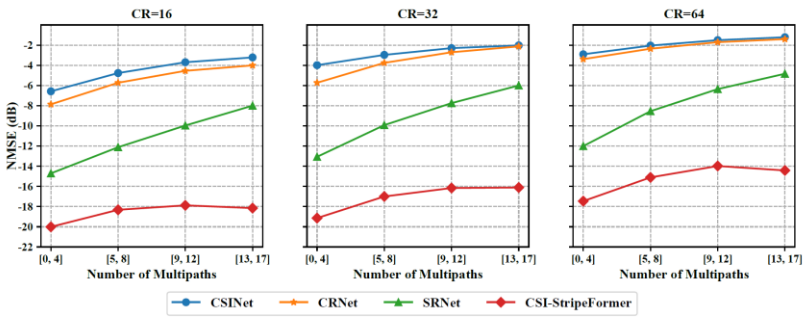
- Our model achieves best performance on various scenarios and compression ratios.
- Our model can significantly reduce NMSE by over 7dB compared with the best baseline SRNet [8]. Its performance under CR=64 is even comparable with that of SRNet [8] under CR=4 in the multipath-rich outdoor scenario.

| Category | Model | CR=4 | | CR=8 | | CR=16 | | CR=32 | | CR=64 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Indoor | Outdoor | Indoor | Outdoor | Indoor | Outdoor | Indoor | Outdoor | Indoor | Outdoor |
| HM | SRNet | -24.23 | <u>-15.43</u> | -19.26 | <u>-13.47</u> | <u>-15.26</u> | <u>-11.31</u> | <u>-11.61</u> | <u>-9.17</u> | -8.27 | <u>-7.80</u> |
| | IdasNet | / | / | -18.87 | -10.34 | -13.51 | -6.15 | -10.13 | -5.03 | <u>-9.34</u> | -3.63 |
| DNN | CSINet | -17.36 | -8.75 | -12.70 | -7.61 | -8.65 | -4.51 | -6.24 | -2.81 | -5.84 | -1.93 |
| | CRNet | <u>-26.99</u> | -12.71 | -16.01 | -8.04 | -11.35 | -5.44 | -8.93 | -3.51 | -6.49 | -2.22 |
| | ACCsiNet | / | / | / | / | -14.81 | -11.76 | -11.00 | -9.14 | -7.46 | -7.11 |
| | TransNet | **-32.38** | -14.86 | **-22.91** | -9.99 | -15.00 | -7.82 | -10.49 | -4.13 | -6.08 | -2.62 |
| | Ours | -26.24 | **-22.50** | <u>-22.29</u> | **-20.35** | **-16.80** | **-18.86** | **-12.48** | **-16.86** | **-9.37** | **-14.89** |

# Evaluation Results on COST2100

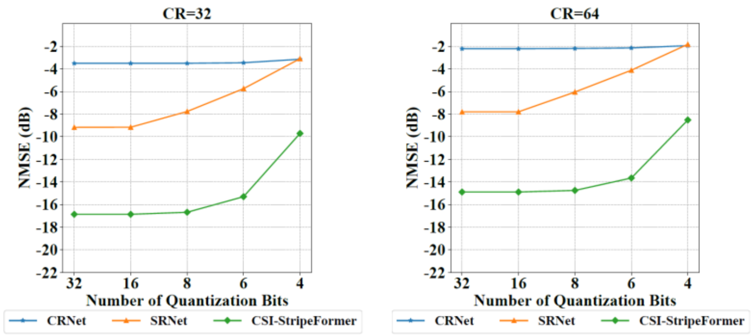
## Multipath Effects on CSI-StripeFormer

- Our model achieves better performance across various conditions.
- Our model's performance is more consistent and is more robust to multipath effects.




## Quantization Influence on CSI-StripeFormer

- We perform post training uniform quantization for evaluation.
- Our model has little degradation when the quantization bits are only 6 bits.
- It validates that our model is more robust to quantization influence.




(a) CR=32 (b) CR=64

# Ablation Study on COST2100

## Validation of Hybrid Attention

- Hyrbid Attention outperforms several SOTA attention mechanisms. Comparing CSWin [4] with Swin [21] can further validate stripe corrleations in CSI.

| Transformer Block Type | NMSE↓ (dB) |
|---|---|
| None | -4.73 |
| CSWin [15] | <u>-12.19</u> |
| Swin [25] | -11.66 |
| Our SFB | **-14.89** |

## Impact of embedded dimension

- Our model is scalable when we increase the dimension of the decoder.
- NMSE can be further reduced to -25.11 dB for CR=64 when D=256.
- Given D=256, the NMSE of our model at CR=64 is around 10 dB lower than baselines at CR=4.

| $D$ | 32 | 64 | 128 | 256 |
|---|---|---|---|---|
| NMSE↓ (dB) | -6.07 | -7.62 | <u>-14.89</u> | **-25.11** |

## Impact of SFL configurations

- The performance is mainly determined by the number of layers for multi-resolution feature extraction.

| CFG | Layers | Split Width | NMSE↓ (dB) |
|---|---|---|---|
| 1 | [2, 2, 6, 2] | [1, 2, 4, 8] | <u>-14.89</u> |
| 2 | [3, 3, 3, 3] | [1, 2, 4, 8] | **-14.99** |
| 3 | [1, 1, 3, 1] | [1, 2, 4, 8] | -13.65 |
| 4 | [2, 2, 2] | [1, 2, 4] | -12.12 |
| 5 | [2, 2] | [1, 2] | -9.85 |

# Evaluations on COST2100

- Computation Complexity

- Our model offloads the complexity from UE to BS for a better performance.

| Model | CR | UE | | BS | | NMSE↓ (dB) |
|---|---|---|---|---|---|---|
| | | Params | FLOPs | Params | FLOPs | |
| CRNet | 32 | 131K | 383K | **136K** | **3.23M** | -3.51 |
| TransNet | 32 | 271K | 16.91M | 276K | 16.97M | -4.13 |
| SRNet | 32 | **58K** | **238K** | 2.07M | 658M | -9.17 |
| Ours | 32 | 165K | 7.50M | 11.38M | 5.76G | **-16.86** |

# Concluding Remarks

- Multipath effects hinders the practice of CSI compression systems.
- The stripe feature differs CSI from images. This can be a potential aspect for further dedicated CSI-based sensing / communication system design.
- We design CSI-StripeFormer, a stripe-aware encoder-decoder framework to explicitly exploit stripe features with a hierarchical hybrid Transformer-based architecture.
- CSI-StripeFormer can achieve a NMSE loss of -25.11 dB at a high compression ratio of 64, which is much better than best baselines at compression ratio of 4.

# Thank you!