

Efficient Deep Learning for Massive MIMO Channel State Estimation



Mason del Rosario
Doctoral Exit Seminar

September 2022

Background

Role of CSI in MIMO

CSI Estimation

Compressed Sensing

Convolutional Neural Networks

Completed Work #1: SphNet

Spherical Normalization

Completed Work #2: MarkovNet

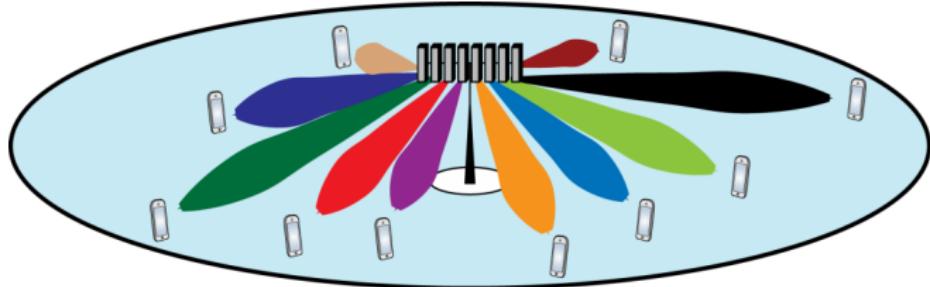
Differential Encoding

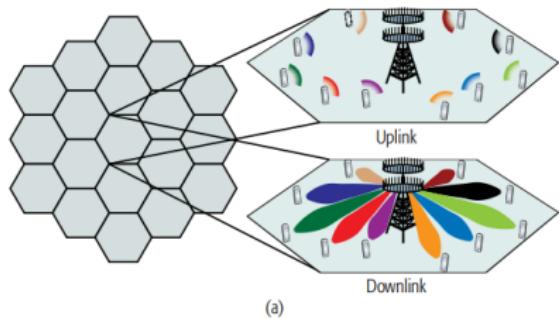
Completed Work #3: Pilots-to-delay Estimator (P2DE) and
Heterogeneous Differential Encoding

~~Current Work: Pilot Feedback and Model Re-use~~

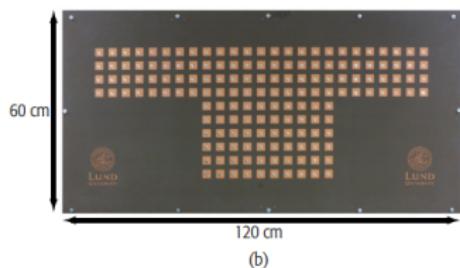
Background

Feedback-based estimation of channel state information in MIMO networks.



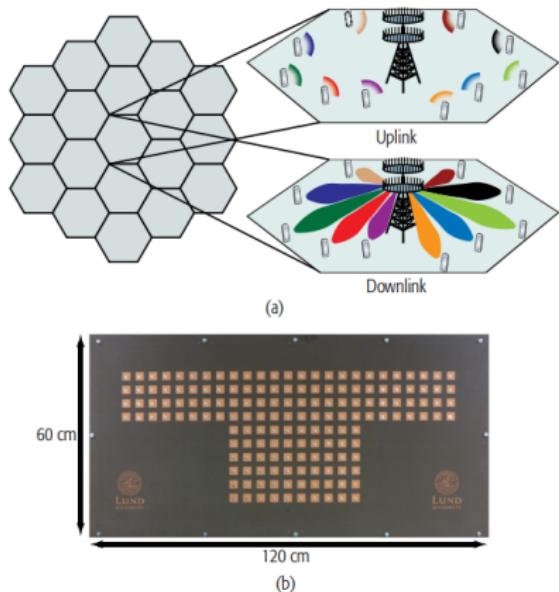


(a)



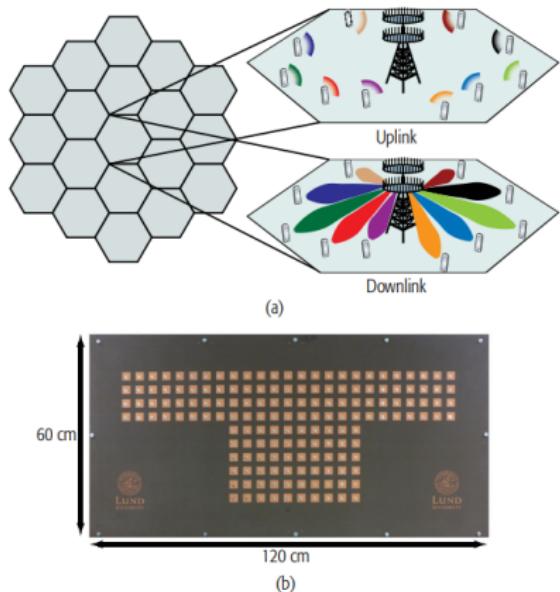
(b)

- ▶ MIMO = Multiple input multiple output



- ▶ MIMO = Multiple input multiple output
- ▶ Massive w.r.t. antenna count, not physical size.

E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten Myths and One Critical Question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016



- ▶ MIMO = Multiple input multiple output
- ▶ Massive w.r.t. antenna count, not physical size.
- ▶ Spatial diversity → **high throughput.**

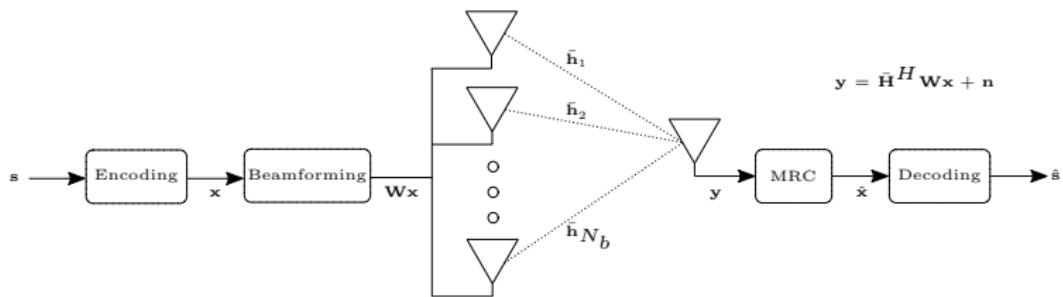


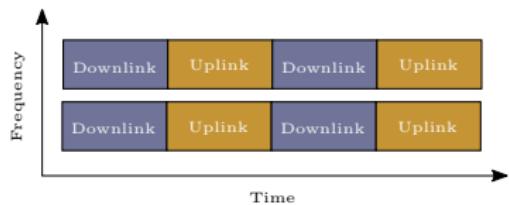
Figure: Multi-antenna transmitter (BS, gNB) and single-antenna user equipment (UE) with relevant system values.

In OFDM, the fading coefficients between Tx/Rx = **Channel State Information (CSI)**, $\bar{\mathbf{H}}$.

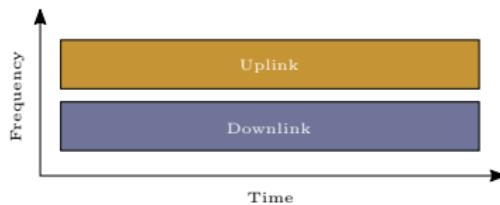
$$\bar{\mathbf{H}} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,N_f} \\ h_{2,1} & h_{2,2} & \dots & h_{2,N_f} \\ \vdots & \vdots & \vdots & \vdots \\ h_{N_b,1} & h_{N_b,2} & \dots & h_{N_b,N_f} \end{bmatrix} \in \mathbb{C}^{N_b \times N_f}$$

For N_b transmit antennas and N_f subcarriers.

Downlink-uplink reciprocity in TDD, but not in FDD.

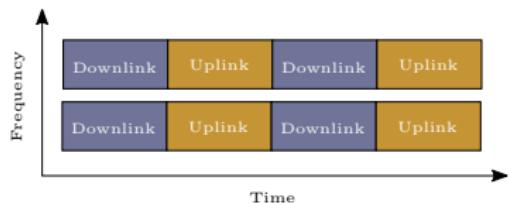


a) Time division duplex (TDD)

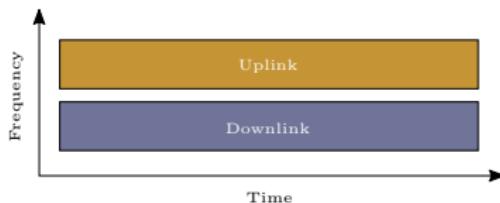


b) Frequency division duplex (FDD)

Downlink-uplink reciprocity in TDD, but not in FDD.



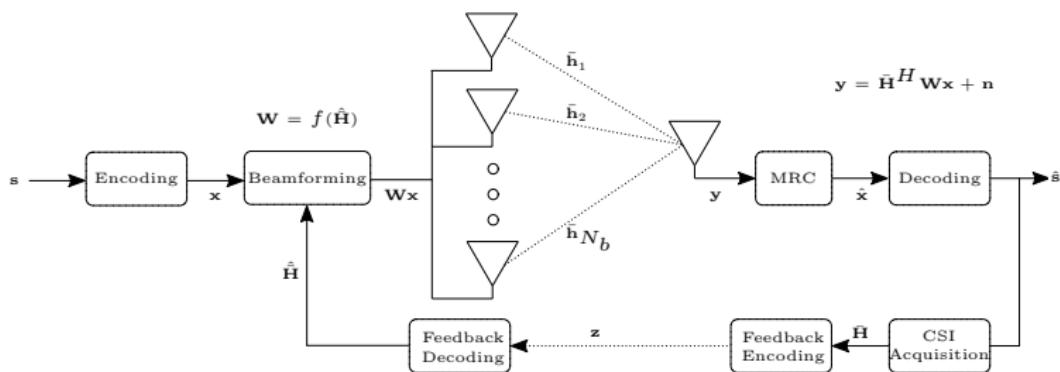
a) Time division duplex (TDD)



b) Frequency division duplex (FDD)

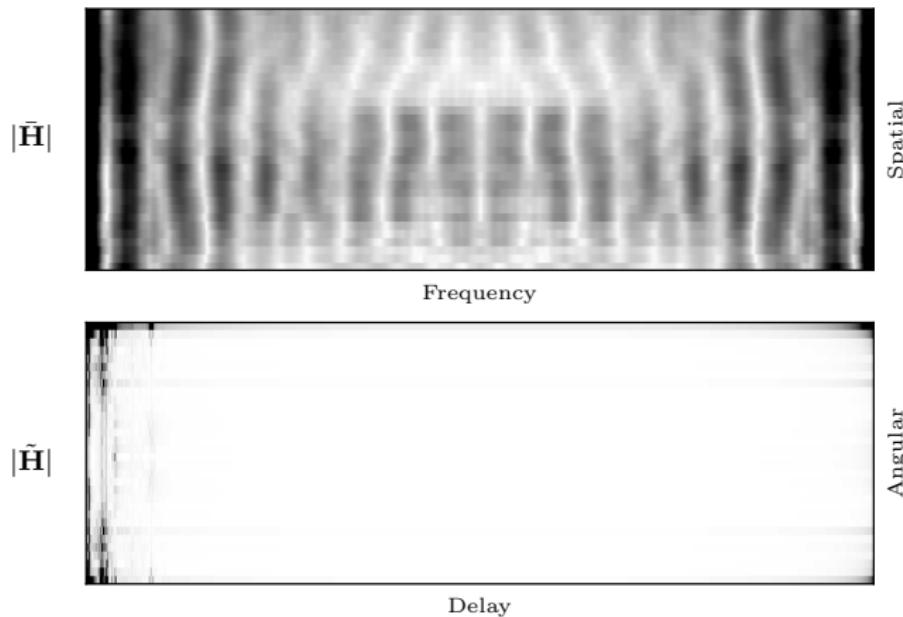
FDD requires feedback for downlink CSI estimation.

Transmitting $\bar{\mathbf{H}}$ is costly. Instead, generate estimates, $\hat{\mathbf{H}}$, based on **compressed feedback**, \mathbf{z} .

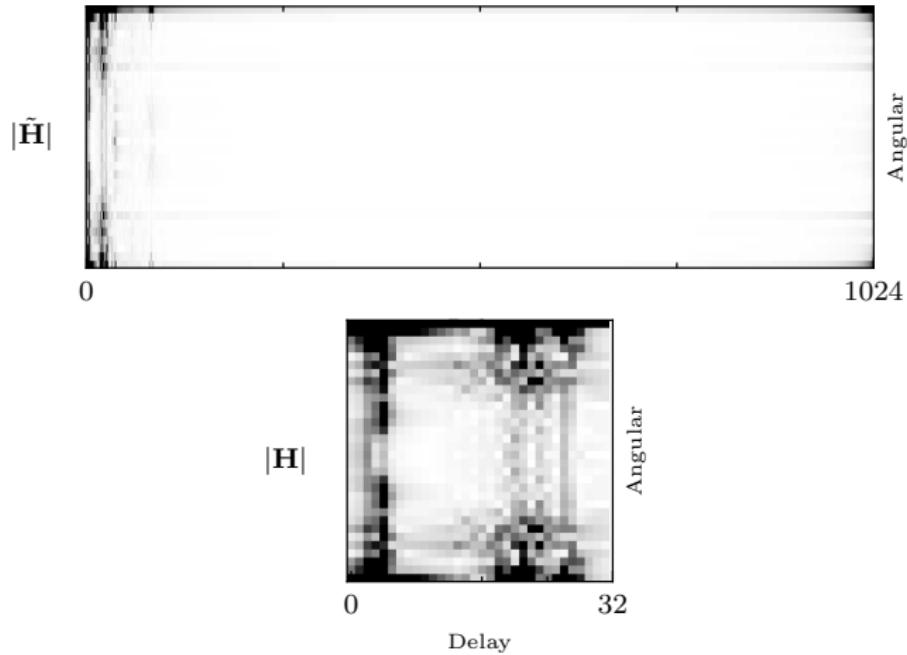


Denote 2D inverse FFT of $\bar{\mathbf{H}}$ as

$$\tilde{\mathbf{H}} = \mathbf{F}^H \bar{\mathbf{H}} \mathbf{F}.$$



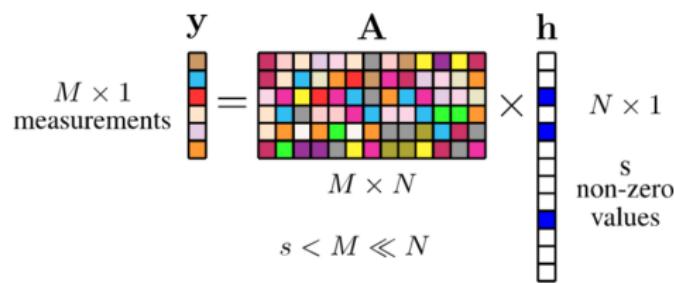
Given sparsity of $\tilde{\mathbf{H}}$, we can encode/decode a truncated version, \mathbf{H} .



1. Compressed Sensing (Conventional)
2. Convolutional Neural Networks (This proposal)

Find low-dimensional basis for sparse data, \mathbf{h} ,

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n}.$$

$$\begin{matrix} \mathbf{y} \\ M \times 1 \\ \text{measurements} \end{matrix} = \begin{matrix} \mathbf{A} \\ M \times N \\ s < M \ll N \end{matrix} \times \begin{matrix} \mathbf{h} \\ N \times 1 \\ s \text{ non-zero values} \end{matrix}$$


CS relies on the following assumptions:

1. \mathbf{h} meets a sparsity level s , number of nonzero coefficients.

CS relies on the following assumptions:

1. \mathbf{h} meets a sparsity level s , number of nonzero coefficients.
2. **Restricted Isometry Property (RIP)**. For $\delta \in [0, 1]$,

$$(1 - \delta)\|\mathbf{h}\|^2 \leq \|\mathbf{A}\mathbf{h}\|^2 \leq (1 + \delta)\|\mathbf{h}\|^2$$

for Frobenius norm $\|\cdot\|$.

CS addresses two major issues:

1. Design of **A** (stochastic or deterministic).

CS addresses two major issues:

1. Design of \mathbf{A} (stochastic or deterministic).
2. Recovery of $\hat{\mathbf{h}}$ given \mathbf{A} and \mathbf{y} , typically via convex optimization on p -norm minimization,

$$\min \|\hat{\mathbf{h}}\|_p \text{ subject to } \|\mathbf{y} - \mathbf{A}\hat{\mathbf{h}}\|_2^2 < \epsilon.$$

CS addresses two major issues:

1. Design of \mathbf{A} (stochastic or deterministic).
2. Recovery of $\hat{\mathbf{h}}$ given \mathbf{A} and \mathbf{y} , typically via convex optimization on p -norm minimization,

$$\min \|\hat{\mathbf{h}}\|_p \text{ subject to } \|\mathbf{y} - \mathbf{A}\hat{\mathbf{h}}\|_2^2 < \epsilon.$$

Problems:

- Recovery algorithms are iterative.

CS addresses two major issues:

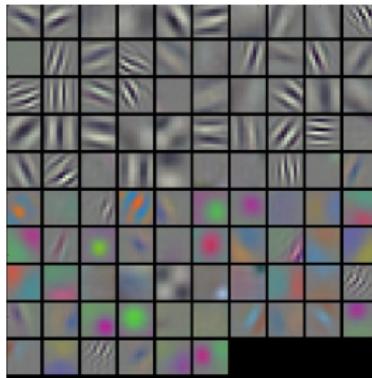
1. Design of \mathbf{A} (stochastic or deterministic).
2. Recovery of $\hat{\mathbf{h}}$ given \mathbf{A} and \mathbf{y} , typically via convex optimization on p -norm minimization,

$$\min \|\hat{\mathbf{h}}\|_p \text{ subject to } \|\mathbf{y} - \mathbf{A}\hat{\mathbf{h}}\|_2^2 < \epsilon.$$

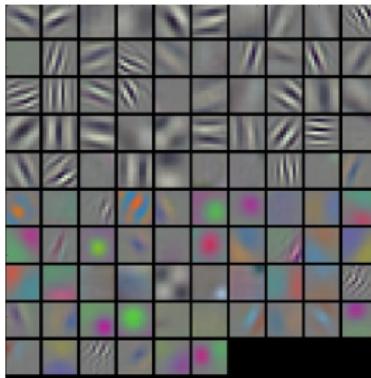
Problems:

- ▶ Recovery algorithms are iterative.
- ▶ Complexity scales with sparsity ($M \propto s$).

- ▶ Layers of trainable linear functions followed by nonlinear ‘activation’ functions.
- ▶ State-of-the art performance in image processing



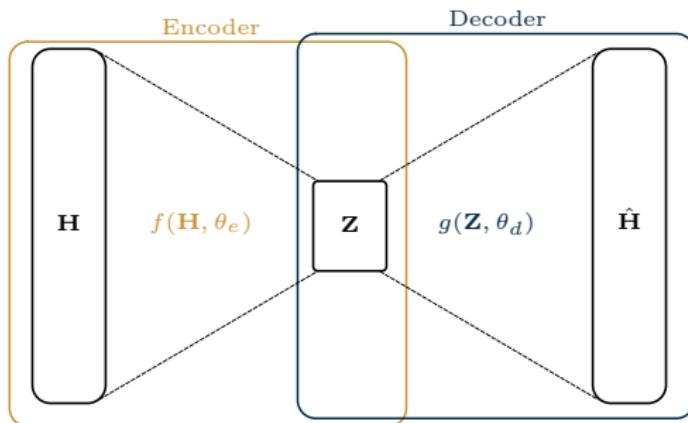
- ▶ Layers of trainable linear functions followed by nonlinear ‘activation’ functions.
- ▶ State-of-the art performance in image processing



- ▶ No assumptions on sparsity/RIP. Instantaneous decoding.

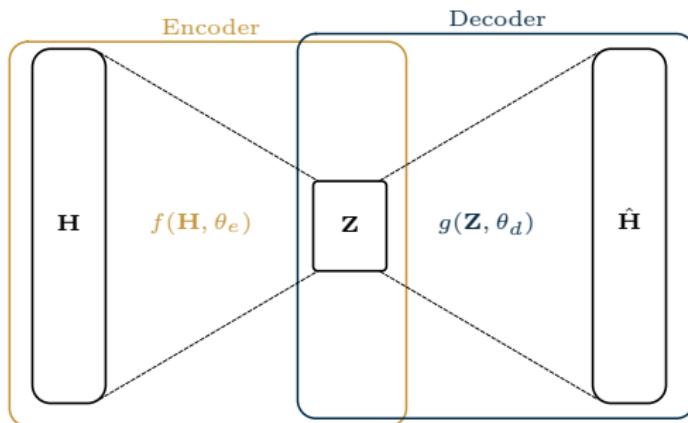
Autoencoder: Estimate $\hat{\mathbf{H}}$, latent code \mathbf{Z} with **compression ratio**,

$$\text{CR} = \frac{\dim(\mathbf{Z})}{\dim(\mathbf{H})} \text{ s.t. } \dim(\mathbf{Z}) < \dim(\mathbf{H}).$$



Autoencoder: Estimate $\hat{\mathbf{H}}$, latent code \mathbf{Z} with **compression ratio**,

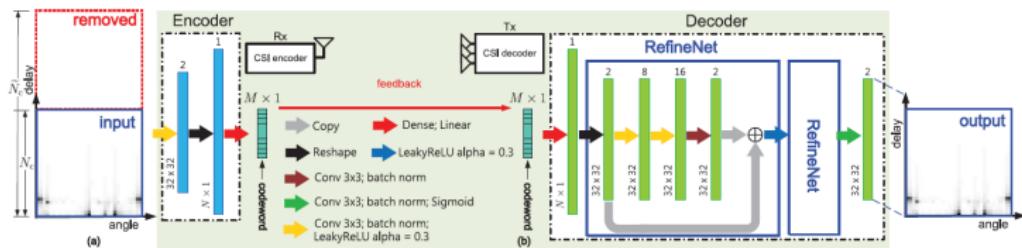
$$\text{CR} = \frac{\dim(\mathbf{Z})}{\dim(\mathbf{H})} \text{ s.t. } \dim(\mathbf{Z}) < \dim(\mathbf{H}).$$



θ_e, θ_d updated to minimize **mean-squared error (MSE)**,

$$\operatorname{argmin}_{\theta_e, \theta_d} \frac{1}{N} \sum_{i=1}^N \|\mathbf{H}_i - g(f(\mathbf{H}_i, \theta_e), \theta_d)\|^2.$$

- CNN autoencoder for learned CSI compression and feedback [3]



Metrics used:

► **Normalized Mean-squared Error**

$$\text{NMSE} = \frac{1}{N} \sum_i^N \frac{\|\mathbf{H}_i - \hat{\mathbf{H}}_i\|^2}{\|\mathbf{H}_i\|^2}$$

► **Cosine Similarity**

$$\rho = \frac{1}{NN_f} \sum_{i=1}^N \sum_{m=1}^{N_f} \frac{|\hat{\mathbf{h}}_{i,m}^H \bar{\mathbf{h}}_{i,m}|}{\|\hat{\mathbf{h}}_{i,m}\| \|\bar{\mathbf{h}}_{i,m}\|},$$

Metrics used:

► Normalized Mean-squared Error

$$\text{NMSE} = \frac{1}{N} \sum_i^N \frac{\|\mathbf{H}_i - \hat{\mathbf{H}}_i\|^2}{\|\mathbf{H}_i\|^2}$$

► Cosine Similarity

$$\rho = \frac{1}{NN_f} \sum_{i=1}^N \sum_{m=1}^{N_f} \frac{|\hat{\mathbf{h}}_{i,m}^H \bar{\mathbf{h}}_{i,m}|}{\|\hat{\mathbf{h}}_{i,m}\| \|\bar{\mathbf{h}}_{i,m}\|},$$

CNNs outperform CS at comparable compression ratios.

γ	Methods	Indoor		Outdoor	
		NMSE	ρ	NMSE	ρ
1/4	LASSO	-7.59	0.91	-5.08	0.82
	BM3D-AMP	-4.33	0.80	-1.33	0.52
	TVAL3	-14.87	0.97	-6.90	0.88
	CS-CsiNet	-11.82	0.96	-6.69	0.87
	CsiNet	-17.36	0.99	-8.75	0.91
1/16	LASSO	-2.72	0.70	-1.01	0.46
	BM3D-AMP	0.26	0.16	0.55	0.11
	TVAL3	-2.61	0.66	-0.43	0.45
	CS-CsiNet	-6.09	0.87	-2.51	0.66
	CsiNet	-8.65	0.93	-4.51	0.79
1/32	LASSO	-1.03	0.48	-0.24	0.27
	BM3D-AMP	24.72	0.04	22.66	0.04
	TVAL3	-0.27	0.33	0.46	0.28
	CS-CsiNet	-4.67	0.83	-0.52	0.37
	CsiNet	-6.24	0.89	-2.81	0.67
1/64	LASSO	-0.14	0.22	-0.06	0.12
	BM3D-AMP	0.22	0.04	25.45	0.03
	TVAL3	0.63	0.11	0.76	0.19
	CS-CsiNet	-2.46	0.68	-0.22	0.28
	CsiNet	-5.84	0.87	-1.93	0.59

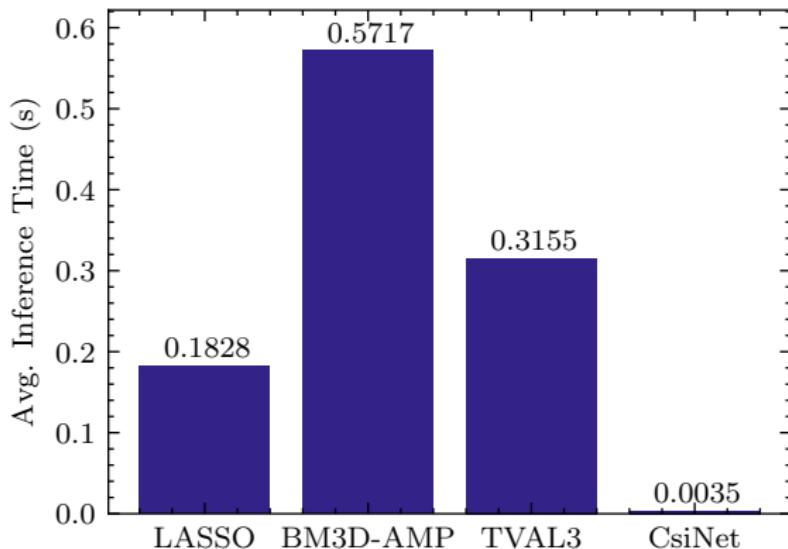


Figure: Average inference time for compressed sensing methods vs. CsiNet.

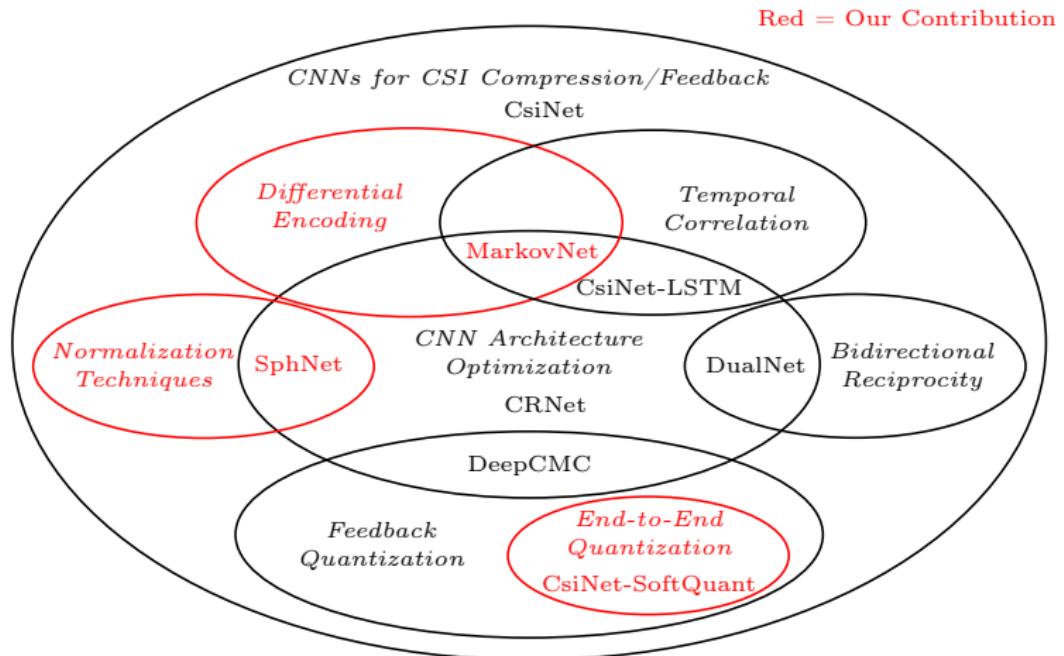
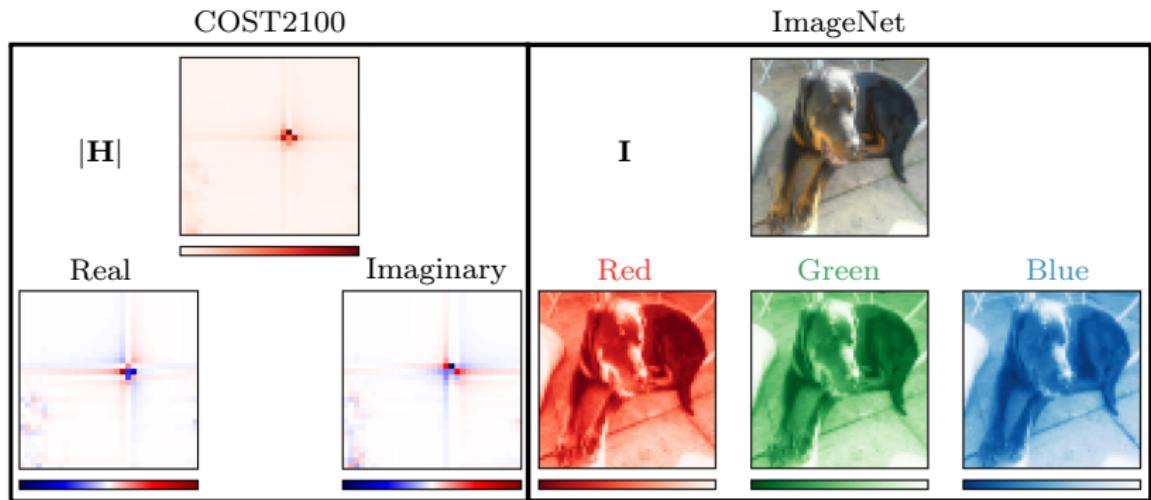
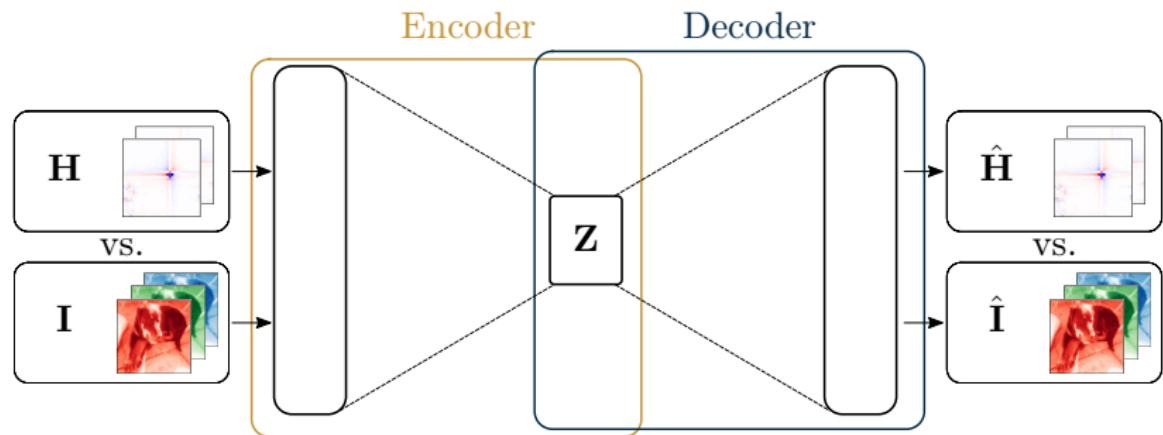


Figure: Areas of *domain knowledge* and corresponding CNNs.

Completed Work #1: SphNet

Power-based normalization for improved CSI reconstruction accuracy.





- **Minmax normalization** – Find minimum, maximum of channels.

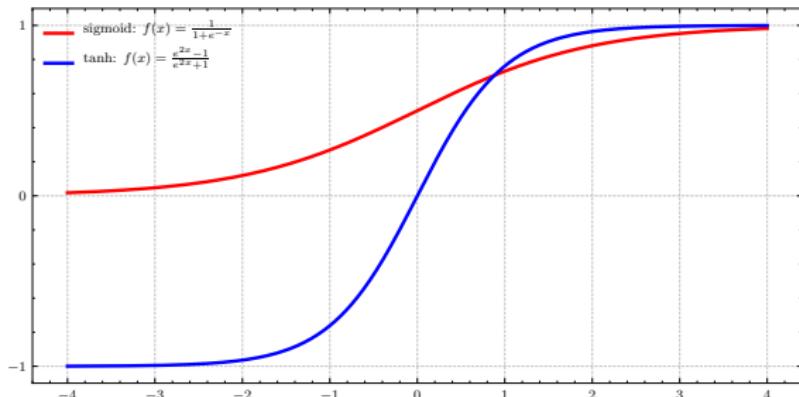
- ▶ **Minmax normalization** – Find minimum, maximum of channels.
- ▶ $H_{n,(i,j)} = (i,j)$ -th element of n -th sample

$$H_{\min\max,n,(i,j)} = \frac{H_{n,(i,j)} - H_{\min}}{H_{\max} - H_{\min}} \in [0, 1]$$

- ▶ **Minmax normalization** – Find minimum, maximum of channels.
- ▶ $H_{n,(i,j)} = (i,j)$ -th element of n -th sample

$$H_{\text{minmax},n,(i,j)} = \frac{H_{n,(i,j)} - H_{\min}}{H_{\max} - H_{\min}} \in [0, 1]$$

- ▶ Compatible with common **activation functions** (e.g., tanh, sigmoid)



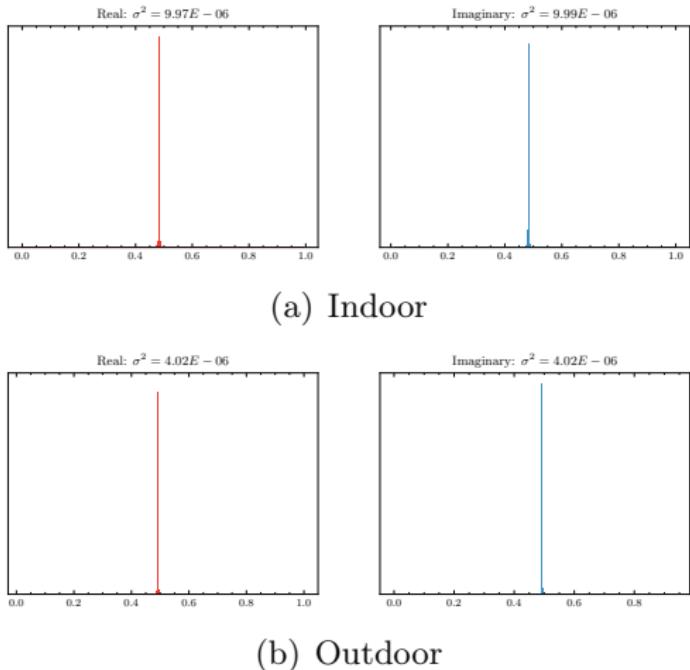


Figure: Distribution/variance of COST2100 real/imaginary channels under minmax normalization ($N = 10^5$).

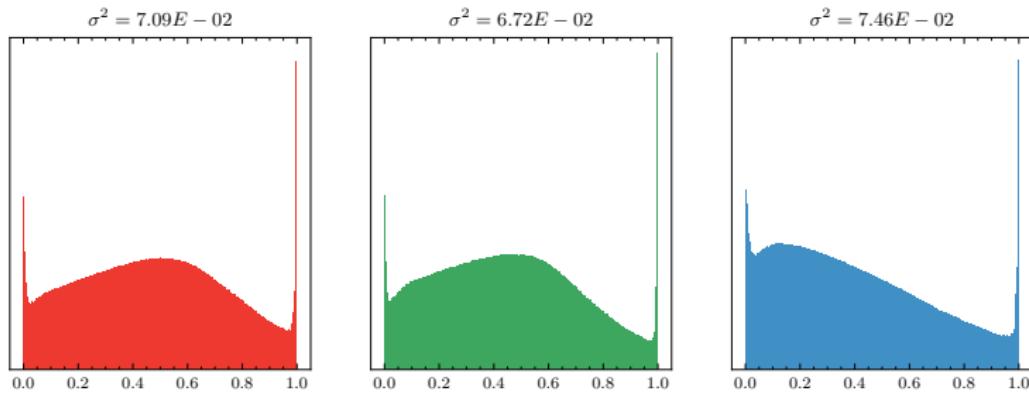


Figure: Distribution and variance of minmax-normalized ImageNet RGB channels ($N = 50000$).

Difference of four orders of magnitude.

Dataset	Env	Channels	Norm	Avg. Variance
ImageNet	-	RGB	Minmax	$7.09E^{-2}$
COST2100	Indoor	Real, Imag	Minmax	$9.98E^{-6}$
COST2100	Outdoor	Real, Imag	Minmax	$4.02E^{-6}$

Table: Minmax normalization applied to COST2100 and ImageNet dataset.

Spherical normalization – scale \mathbf{H} by power. For Frobenius norm $\|\cdot\|$,

$$\check{\mathbf{H}}^n = \frac{\mathbf{H}^n}{\|\mathbf{H}^n\|}. \quad (1)$$

Then apply minmax scaling to the entire dataset.

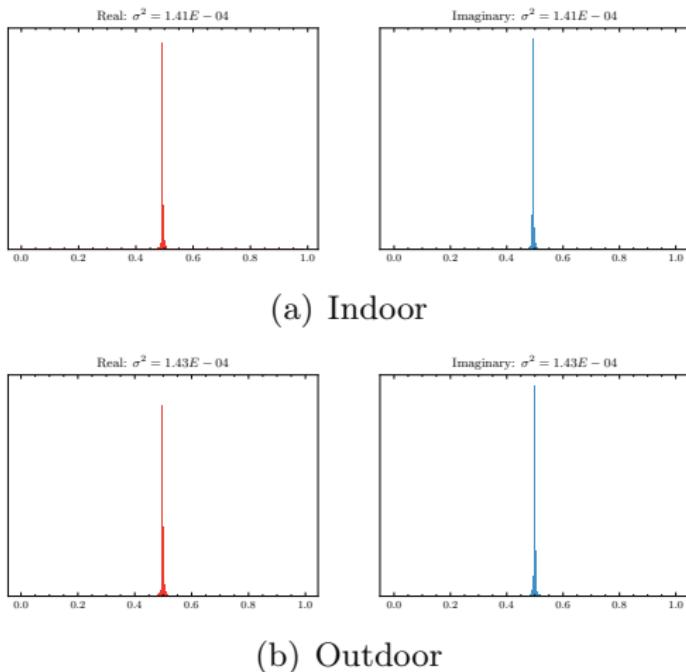


Figure: Distribution/variance of COST2100 real/imaginary channels under spherical normalization ($N = 10^5$).

Difference is now **two orders of magnitude**.

Dataset	Env	Channels	Norm	Avg. Variance
ImageNet	-	RGB	Minmax	$7.09E^{-2}$
COST2100	Indoor	Real, Imag	Spherical	$1.41E^{-4}$
COST2100	Outdoor	Real, Imag	Spherical	$1.43E^{-4}$
COST2100	Indoor	Real, Imag	Minmax	$9.98E^{-6}$
COST2100	Outdoor	Real, Imag	Minmax	$4.02E^{-6}$

Table: Minmax vs. spherical normalization applied to COST2100 datasets compared with ImageNet.

Spherical normalization → MSE equivalent to NMSE.

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2, \quad \text{NMSE} = \frac{1}{N} \sum_{k=1}^N \frac{\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2}{\|\mathbf{H}_k\|}$$

Spherical normalization → MSE equivalent to NMSE.

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^N \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2, \quad \text{NMSE} = \frac{1}{N} \sum_{k=1}^N \frac{\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2}{\|\mathbf{H}_k\|}$$

MSE of spherically normalized estimator yields,

$$\begin{aligned}\text{MSE}_{\text{Sph}} &= \frac{1}{N} \sum_{k=1}^N \|\check{\mathbf{H}}_k - \hat{\check{\mathbf{H}}}_k\|^2 \\ &= \frac{1}{N} \sum_{k=1}^N \left\| \frac{\mathbf{H}_k}{\|\mathbf{H}_k\|} - \frac{\hat{\mathbf{H}}_k}{\|\mathbf{H}_k\|} \right\|^2 \\ &= \frac{1}{N} \sum_{k=1}^N \frac{\|\mathbf{H}_k - \hat{\mathbf{H}}_k\|^2}{\|\mathbf{H}_k\|}.\end{aligned}$$

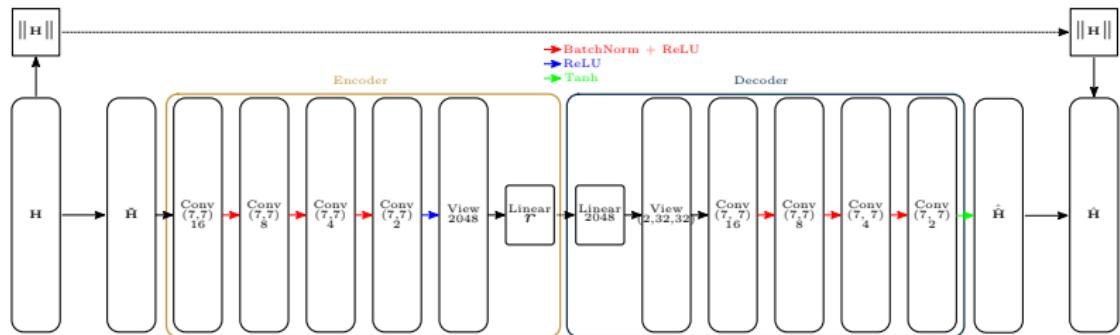


Figure: SphNet – CsiNetPro architecture with Spherical Normalization.

Z. Liu, M. del Rosario, X. Liang, L. Zhang, and Z. Ding, “Spherical Normalization for Learned Compressive Feedback in Massive MIMO CSI Acquisition,” in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020

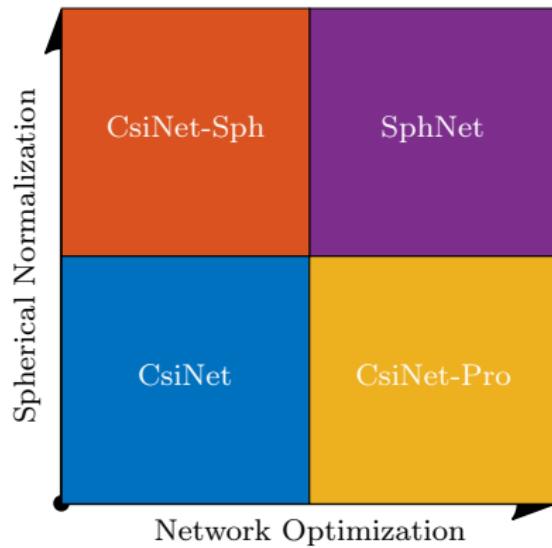
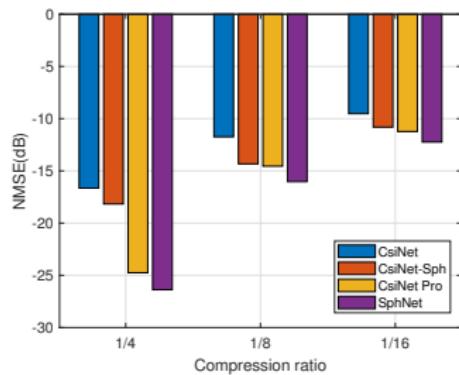


Figure: Illustration of techniques used in different models.

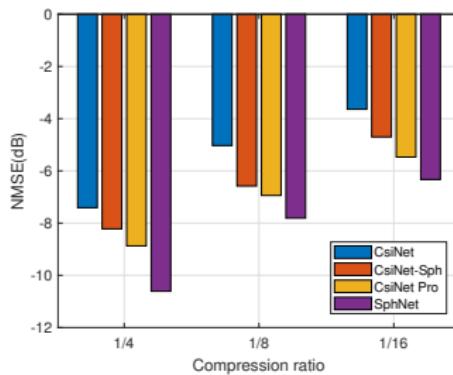
Z. Liu, M. del Rosario, X. Liang, L. Zhang, and Z. Ding, “Spherical Normalization for Learned Compressive Feedback in Massive MIMO CSI Acquisition,” in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020

Table: Parameters for COST2100 model in this work.

Environment	Indoor	Outdoor
Num. gNB Antennas (N_b)		32
Num. Subcarriers (N_f)		1024
Truncation Value (R_d)		32
Carrier Frequency	5.3 GHz	300 MHz
UE Mobility	0.001 m/s	1 m/s
UE Starting Position	20×20 m	400×400 m
Num. Channel Samples (N)		10^5
Training/Validation Split		70%/30%
Feedback interval		40 ms



(a) Indoor



(b) Outdoor

Figure: Ablation study for CsiNet-Pro and spherical normalization [4] (lower NMSE is better).

Z. Liu, M. del Rosario, X. Liang, L. Zhang, and Z. Ding, “Spherical Normalization for Learned Compressive Feedback in Massive MIMO CSI Acquisition,” in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020

Completed Work #2: MarkovNet

A deep differential autoencoder for efficient temporal learning.

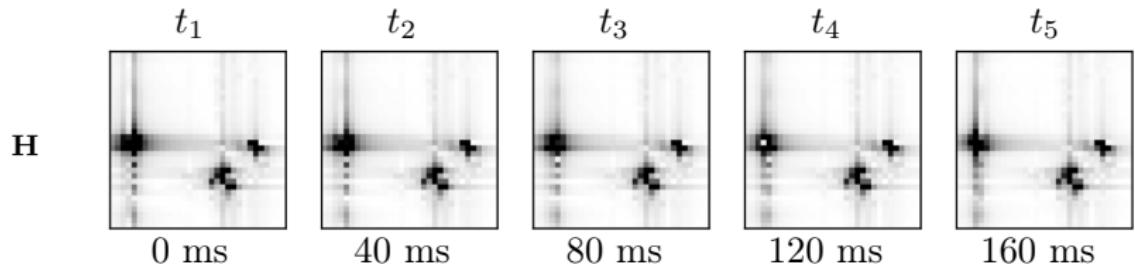


Figure: Ground truth CSI (**H**) for five timeslots (T_1 through T_5) on one outdoor sample from the validation set.

Recurrent neural networks (RNNs) contain trainable long short-term memory (LSTM) cells which learn temporal relationships.

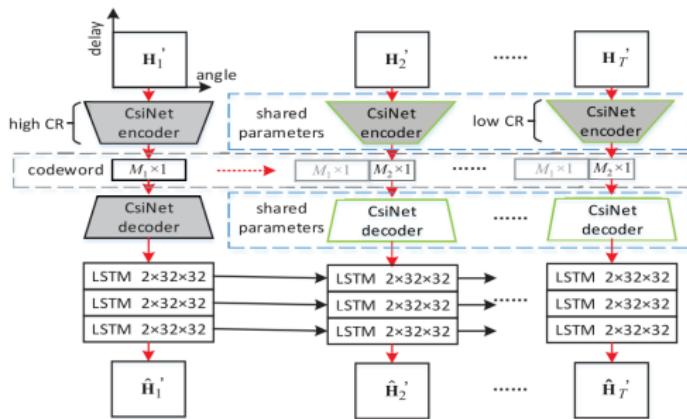


Figure: CsiNet-LSTM network architecture [6].

LSTMs improve NMSE at smaller compression ratios.

	CR	LASSO	BM3D-AMP	TVAL3	CsiNet	CsiNet-LSTM
Indoor	1/16	-2.96	0.25	-3.20	-10.59	-23.06
	1/32	-1.18	20.85	-0.46	-7.35	-22.33
	1/64	-0.18	26.66	0.60	-6.09	-21.24
	ρ	1/16	0.72	0.29	0.73	0.99
		1/32	0.53	0.17	0.45	0.99
		1/64	0.30	0.16	0.24	0.99
	runtime	1/16	0.2471	0.3454	0.3148	0.0001
		1/32	0.2137	0.5556	0.3148	0.0001
		1/64	0.2479	0.6047	0.2860	0.0001
	NMSE↓	1/16-1/64	94%	105	1.19	42% 8%
Outdoor	1/16	-1.09	0.40	-0.53	-3.60	-9.86
	1/32	-0.27	18.99	0.42	-2.14	-9.18
	1/64	-0.06	24.42	0.74	-1.65	-8.83
	ρ	1/16	0.49	0.23	0.46	0.95
		1/32	0.32	0.16	0.28	0.94
		1/64	0.19	0.16	0.19	0.93
	runtime	1/16	0.2122	0.4210	0.3145	0.0001
		1/32	0.2409	0.6031	0.2985	0.0001
		1/64	0.0166	0.5980	0.2850	0.0001
	NMSE↓	1/16-1/64	94%	60	2.40	54% 10%

Problem: Number of parameters/FLOPs for RNNs is large.

Table: Model size/computational complexity per timeslot for CsiNet-LSTM and CsiNet. M: million.

CR	Parameters		FLOPs	
	CsiNet-LSTM	CsiNet	CsiNet-LSTM	CsiNet
1/4	132.7 M	2.1 M	412.9 M	7.8 M
1/8	123.2 M	1.1 M	410.8 M	5.7 M
1/16	118.5 M	0.5 M	409.8 M	4.7 M
1/32	116.1 M	0.3 M	409.2 M	4.1 M
1/64	115.0 M	0.1 M	409.0 M	3.9 M

For short enough feedback interval, CSI data form a Markov chain,

$$\mathbf{H}_t = \gamma \mathbf{H}_{t-1} + \mathbf{V}_t,$$

with $\gamma \in \mathbb{R}^+$ and i.i.d $\mathbf{V}_t \sim \mathcal{CN}(\mathbf{0}, \Sigma_V)$.

Z. Liu †, M. del Rosario †, and Z. Ding, “A Markovian Model-Driven Deep Learning Framework for Massive MIMO CSI Feedback,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1214–1228, 2022 († equal contribution)

The ordinary least-squares solution, γ , is given as

$$\gamma = \frac{\text{Trace}(\mathbb{E} [\mathbf{H}_{t-1}^H \mathbf{H}_t])}{\mathbb{E} \|\mathbf{H}_t^H \mathbf{H}_t\|^2}.$$

The ordinary least-squares solution, γ , is given as

$$\gamma = \frac{\text{Trace}(\mathbb{E} [\mathbf{H}_{t-1}^H \mathbf{H}_t])}{\mathbb{E} \|\mathbf{H}_t^H \mathbf{H}_t\|^2}.$$

Utilize estimator, $\hat{\gamma}$, based on the sample statistics,

$$\hat{\gamma} = \frac{\sum_{i=1}^N \text{Trace}([\mathbf{H}_{t-1}^H(i) \mathbf{H}_t(i)])}{\sum_{i=1}^N \|\mathbf{H}_t^H(i) \mathbf{H}_t(i)\|^2},$$

for training set of size N .

Using $\hat{\gamma}$, train encoder on estimation error as

$$\begin{aligned}\mathbf{E}_t &= \mathbf{H}_t - \hat{\gamma} \hat{\mathbf{H}}_{t-1} \\ \mathbf{z}_t &= f_{e,t}(\mathbf{E}_t).\end{aligned}$$

Jointly train a decoder,

$$\begin{aligned}\hat{\mathbf{E}}_t &= f_{d,t}(\mathbf{z}_t) \\ \hat{\mathbf{H}}_t &= \hat{\mathbf{E}}_t + \hat{\gamma} \hat{\mathbf{H}}_{t-1}.\end{aligned}$$

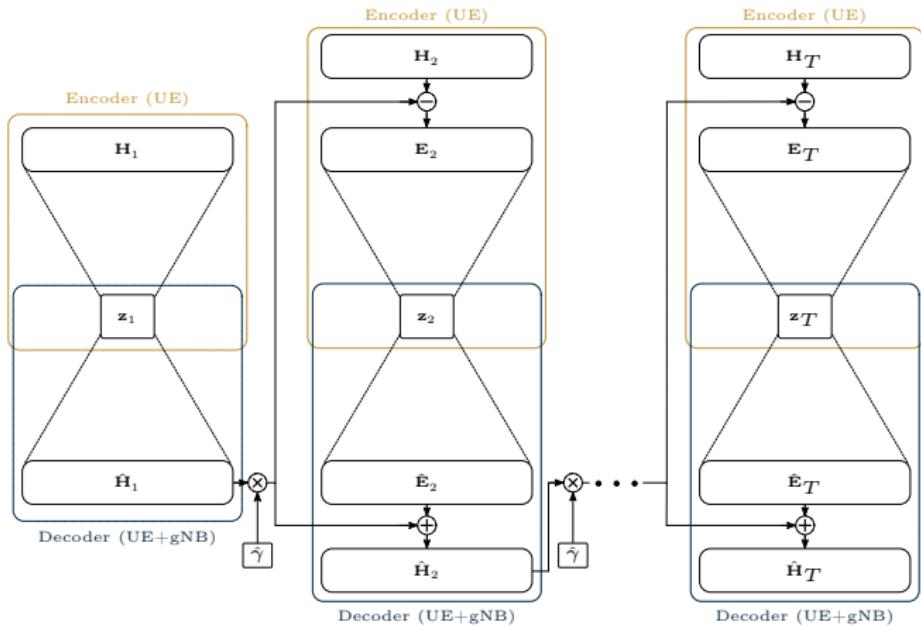
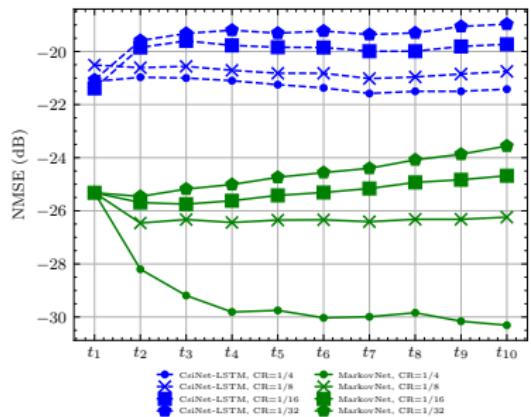


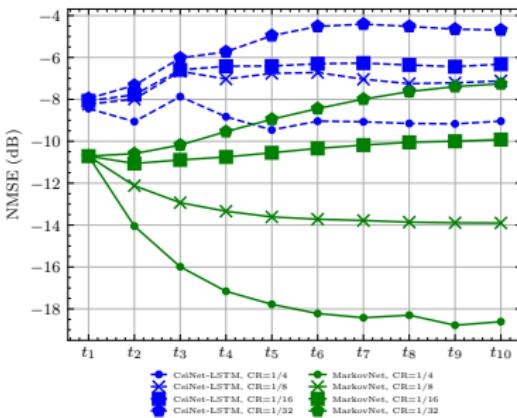
Figure: MarkovNet architecture. Networks at $t \geq 2$ predict estimation error, $\hat{\mathbf{E}}_t$.

MarkovNet Results – NMSE Performance

45



(a) Indoor



(b) Outdoor

Figure: NMSE (lower is better) comparison of MarkovNet and CsiNet-LSTM at multiple CRs.

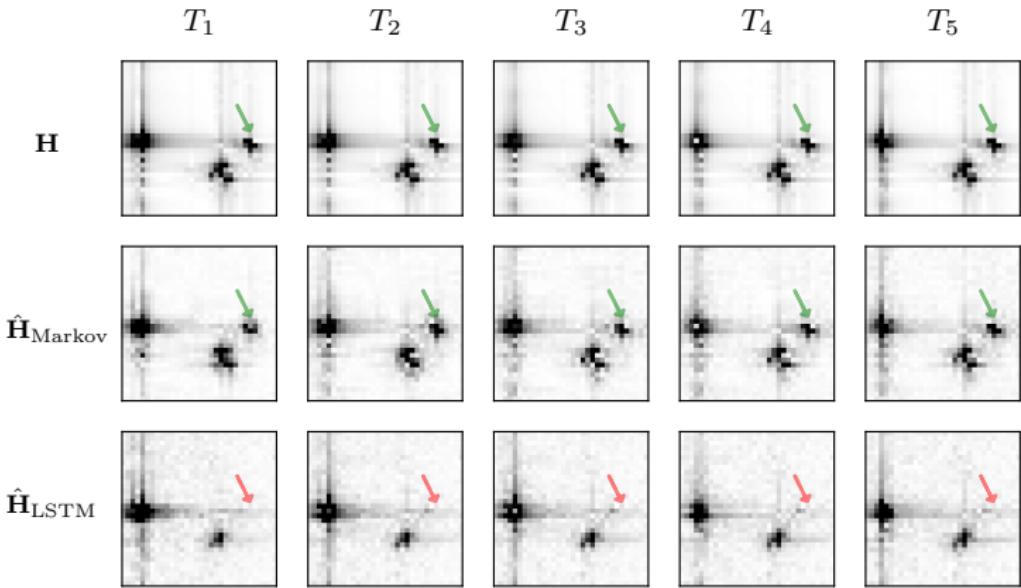


Figure: Ground truth (\mathbf{H}), MarkovNet estimates ($\hat{\mathbf{H}}_{\text{Markov}}$), and CsiNet-LSTM estimates ($\hat{\mathbf{H}}_{\text{LSTM}}$) on from outdoor test set ($\text{CR} = \frac{1}{4}$).

Table: Model size/computational complexity of tested temporal networks (CsiNet-LSTM, MarkovNet) and comparable non-temporal network (CsiNet). M: million.

	Parameters		
	CsiNet-LSTM	MarkovNet	CsiNet
CR=1/4	132.7 M	2.1 M	2.1 M
CR=1/8	123.2 M	1.1 M	1.1 M
CR=1/16	118.5 M	0.5 M	0.5 M
CR=1/32	116.1 M	0.3 M	0.3 M
CR=1/64	115.0 M	0.1 M	0.1 M
	FLOPs		
	CsiNet-LSTM	MarkovNet	CsiNet
CR=1/4	412.9 M	44.5 M	7.8 M
CR=1/8	410.8 M	42.4 M	5.7 M
CR=1/16	409.8 M	41.3 M	4.7 M
CR=1/32	409.2 M	40.8 M	4.1 M
CR=1/64	409.0 M	40.5 M	3.9 M

Completed Work #3: Pilots-to-delay Estimator (P2DE) and Heterogeneous Differential Encoding

Acquiring delay domain CSI under practical pilot placement. Improving differential encoding.

Current Work: Pilot Feedback and Model Re-use

UE-focused reduction of complexity of CSI feedback.

Problem: Computation due to encoder at UE. Can we reduce this?

- ▶ Z. Liu, **M. del Rosario**, X. Liang, L. Zhang, and Z. Ding, “Spherical Normalization for Learned Compressive Feedback in Massive MIMO CSI Acquisition,” in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020
- ▶ Z. Liu †, **M. del Rosario** †, and Z. Ding, “A Markovian Model-Driven Deep Learning Framework for Massive MIMO CSI Feedback,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1214–1228, 2022
- ▶ **del Rosario, Mason** and Z. Ding, “Learning-Based MIMO Channel Estimation under Spectrum Efficient Pilot Allocation and Feedback,” *IEEE Transactions on Wireless Communications*, 2022. Accepted in IEEE TWC.

- ▶ Thesis committee

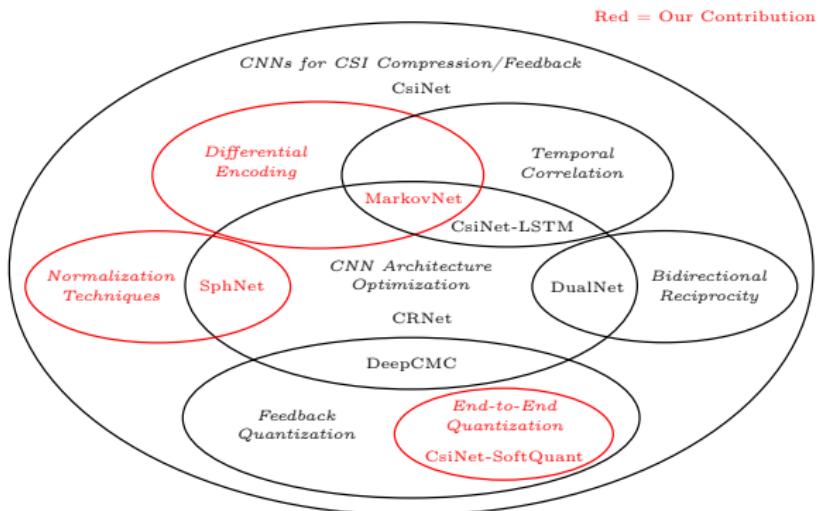
- ▶ Thesis committee
- ▶ Prof. Ding, lab mates, collaborators

- ▶ Thesis committee
- ▶ Prof. Ding, lab mates, collaborators
- ▶ My parents, my brother

- ▶ Thesis committee
- ▶ Prof. Ding, lab mates, collaborators
- ▶ My parents, my brother
- ▶ My SO

Questions?

mdeirosa@ucdavis.edu



-
- [1] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten Myths and One Critical Question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.
 - [2] E. Crespo Marques, N. Maciel, L. Naviner, H. Cai, and J. Yang, "A Review of Sparse Recovery Algorithms," *IEEE Access*, vol. 7, pp. 1300–1322, 2019.
 - [3] C. Wen, W. Shih, and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," *IEEE Wireless Communications Letters*, vol. 7, pp. 748–751, Oct 2018.
 - [4] Z. Liu, M. del Rosario, X. Liang, L. Zhang, and Z. Ding, "Spherical Normalization for Learned Compressive Feedback in Massive MIMO CSI Acquisition," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2020.
 - [5] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Qutin, F. Tufvesson, and P. D. Doncker, "The COST 2100 MIMO Channel Model," *IEEE Wireless Communications*, vol. 19, pp. 92–99, December 2012.
 - [6] T. Wang, C. Wen, S. Jin, and G. Y. Li, "Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels," *IEEE Wireless Comm. Letters*, vol. 8, pp. 416–419, April 2019.
 - [7] Z. Liu †, M. del Rosario †, and Z. Ding, "A Markovian Model-Driven Deep Learning Framework for Massive MIMO CSI Feedback," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1214–1228, 2022.
 - [8] del Rosario, Mason and Z. Ding, "Learning-Based MIMO Channel Estimation under Spectrum Efficient Pilot Allocation and Feedback," *IEEE Transactions on Wireless Communications*, 2022.

Accepted in IEEE TWC.

Appendix

SphNet (and benchmark networks)

- ▶ **Epochs:** 1000
- ▶ **Optimizer:** Adam with learning rate 10^{-3}

MarkovNet

- ▶ **Epochs (t_1):** 1000
- ▶ **Epochs (t_2, \dots, t_T):** 150
- ▶ **Optimizer:** Adam with learning rate 10^{-3}
- ▶ Each timeslot is initialized with weights from previous timeslot.

CsiNet-LSTM

- ▶ **Epochs:** 1000 (pretraining CsiNet), 500 (CsiNet-LSTM)
- ▶ **Optimizer:** Adam with learning rate 10^{-3}

CsiNet-SoftQuant has three stages of training:

1. **Autoencoder pretraining:** No latent quantization (1000 epochs). MSE objective function.
2. **Center pretraining:** Train soft quantization layer to initialize centers, \mathbf{C} (1000 epochs). Using θ_e from stage 1, train on $\mathbf{Z} = f(\mathbf{H}, \theta_e)$ to minimize the cluster energy, $\operatorname{argmin}_{\mathbf{C}} \sum_{i=1}^N \|\tilde{\mathbf{Z}} - Q(\tilde{\mathbf{Z}})\|^2$.
3. **SHVQ finetuning:** Using the results of stage 1 and 2, finetune autoencoder and soft quantization layer (50 epochs). Loss function is the entropy-regularized MSE. Use different β to achieve different rates.

D-AMP = Denoising approximate message passing. Initialize $x^0 = \mathbf{0}$, and alternate between:

$$x^{t+1} = D_{\hat{\sigma}^t}(x^t + \mathbf{A}^* z^t)$$

$$z^t = y - \mathbf{A}x^t + z^{t-1} \frac{\text{div} D_{\hat{\sigma}^{t-1}}(x^{t-1} + \mathbf{A}^* z^{t-1})}{m}$$

where $\hat{\sigma}^t = \text{Var}(x^t + \mathbf{A}^* z^t)$, $D_{\hat{\sigma}_t}$ = denoising algorithm.

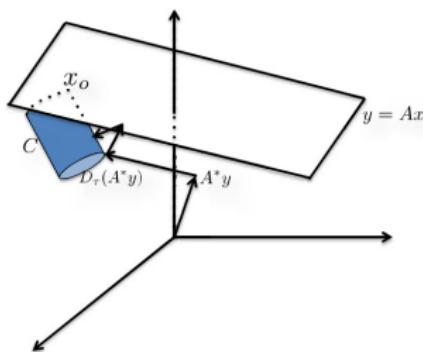


Figure: Subspaces of interest in D-AMP.

BM3D-AMP = D-AMP with *block matching 3D collaborative filtering* (BM3D).

- ▶ Combination of non-local means (NLM) and wavelet thresholding.
- ▶ Procedure:
 1. Compare patches of pixels in images
 2. Group similar patches
 3. 2D (DCT or Bior Wavelet) + 1D Haar wavelet transforms on group
 4. Shrink coefficients in groups ($N \rightarrow M$)
 5. Perform inverse transform by 1) hard thresholding and 2) Wiener filter ($M \rightarrow N$)

Given mean μ , standard deviation σ w.r.t \mathbf{H} ,

$$H_{\text{tanh}}(i, j) = \tanh\left(\frac{H(i, j) - \mu}{2\nu\sigma}\right) + 1.$$

Scale parameter ν chosen by designer.

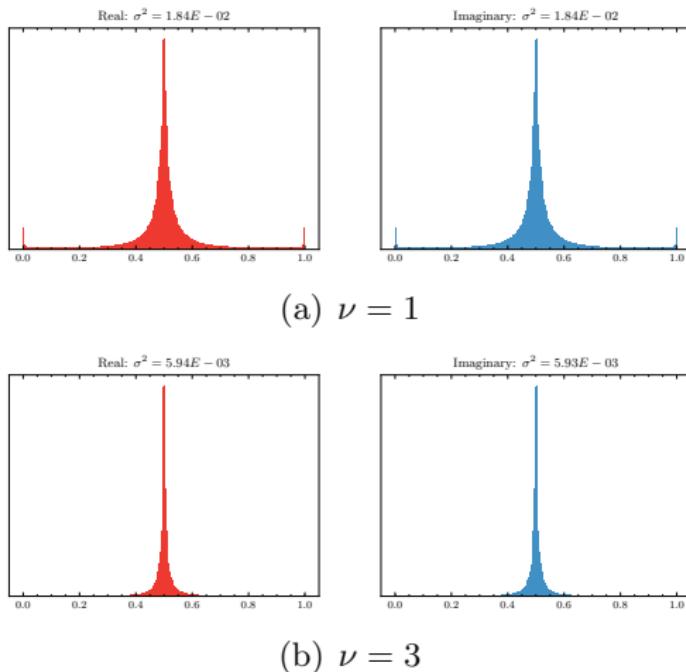


Figure: Distribution/variance of indoor COST2100 real/imaginary channels under tanh normalization ($N = 9.910^5$).

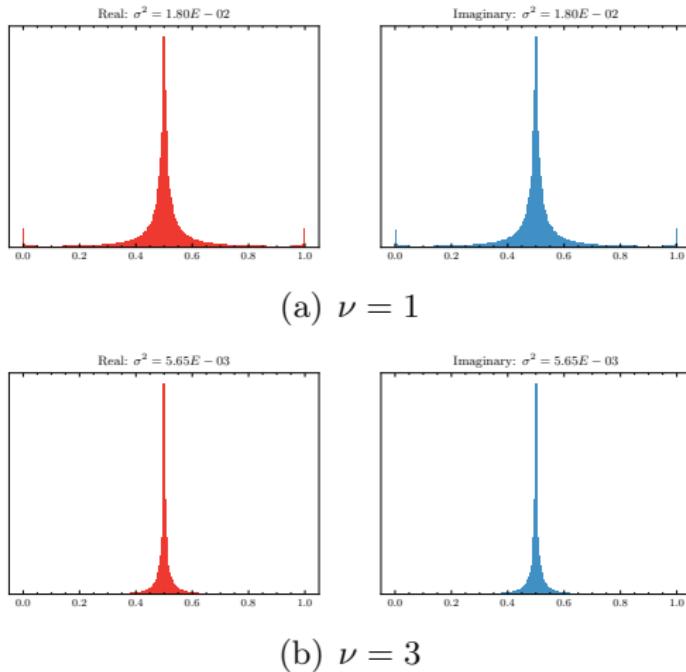


Figure: Distribution/variance of outdoor COST2100 real/imaginary channels under tanh normalization ($N = 10^5$).

Rather than scalar $\hat{\gamma} \in \mathbb{R}^+$, we can derive a multivariate p -step predictor, $\mathbf{W}_1, \dots, \mathbf{W}_p$. Given p prior CSI samples, the mean-square optimal predictor \hat{H}_t is a linear combination of these the prior CSI samples,

$$\hat{\mathbf{H}}_t = \mathbf{H}_{t-1}\mathbf{W}_1 + \cdots + \mathbf{H}_{t-p}\mathbf{W}_p + \mathbf{E}_t. \quad (2)$$

Error terms are uncorrelated with the CSI samples (i.e. $\mathbf{H}_{t-i}^H \mathbf{E}_t = 0$ for all $i \in [0, \dots, p]$), and we pre-multiply by \mathbf{H}_{t-i}^H ,

$$\begin{aligned}\mathbf{H}_{t-i}^H \hat{\mathbf{H}}_t &= \mathbf{H}_{t-i}^H \mathbf{H}_{t-1} \mathbf{W}_1 + \cdots + \mathbf{H}_{t-i}^H \mathbf{H}_{t-p} \mathbf{W}_p + \mathbf{H}_{t-i}^H \mathbf{E}_t \\ &= \mathbf{H}_{t-i}^H \mathbf{H}_{t-1} \mathbf{W}_1 + \cdots + \mathbf{H}_{t-i}^H \mathbf{H}_{t-p} \mathbf{W}_p.\end{aligned}\tag{3}$$

Denote the correlation matrix $\mathbf{R}_i = \mathbb{E}[\mathbf{H}_{t-i}^H \mathbf{H}_t]$. Presume CSI matrices arise from a stationary process, implying the following properties:

1. $\mathbf{R}_i = \mathbb{E}[\mathbf{H}_{t-i}^H \mathbf{H}_t] = \mathbb{E}[\mathbf{H}_t^H \mathbf{H}_{t+i}]$
2. $\mathbf{R}_i = \mathbf{R}_{-i}^H$

Taking the expectation, write (3) as a linear combination of \mathbf{R} ,

$$\mathbf{R}_{i+1} = \mathbf{R}_i \mathbf{W}_1 + \cdots + \mathbf{R}_{i-p+1} \mathbf{W}_p.$$

For p CSI samples, write a system of p equations, admitting the following,

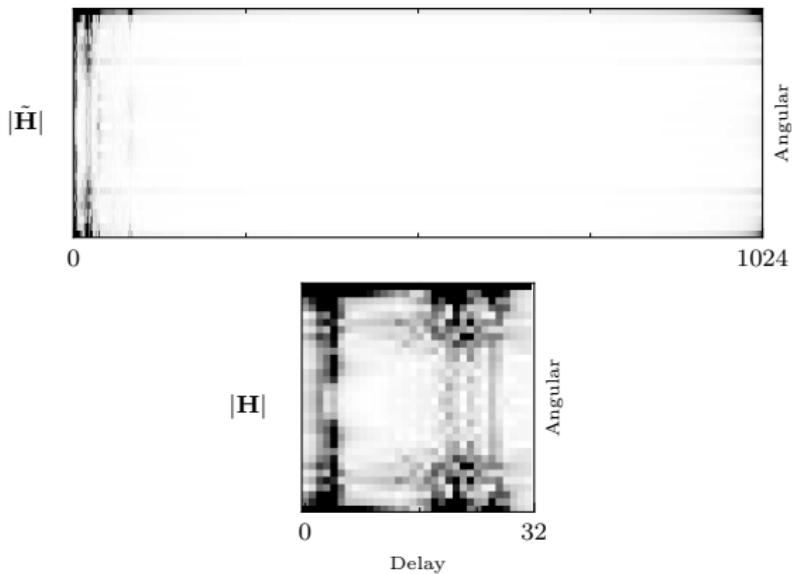
$$\begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_p \end{bmatrix} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1^H & \cdots & \mathbf{R}_{p-1}^H \\ \mathbf{R}_1 & \mathbf{R}_0 & \cdots & \mathbf{R}_{p-2}^H \\ \vdots & & \ddots & \vdots \\ \mathbf{R}_{p-1} & \mathbf{R}_{p-2} & \cdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \cdots \\ \mathbf{W}_p \end{bmatrix}.$$

Solving for the coefficient matrices admits the solution

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_p \end{bmatrix} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1^H & \cdots & \mathbf{R}_{p-1}^H \\ \mathbf{R}_1 & \mathbf{R}_0 & \cdots & \mathbf{R}_{p-2}^H \\ \vdots & & \ddots & \vdots \\ \mathbf{R}_{p-1} & \mathbf{R}_{p-2} & \cdots & \mathbf{R}_0 \end{bmatrix}^+ \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_p \end{bmatrix}, \quad (4)$$

where $[\cdot]^+$ denotes the Moore-Penrose pseudoinverse.

$$\text{NMSE}_{\text{all}} = \frac{1}{N} \sum_i^N \frac{\|\tilde{\mathbf{H}}_i - \hat{\mathbf{H}}_i\|^2}{\|\tilde{\mathbf{H}}_i\|^2}, \quad \text{NMSE}_{\text{truncate}} = \frac{1}{N} \sum_i^N \frac{\|\mathbf{H}_i - \hat{\mathbf{H}}_i\|^2}{\|\mathbf{H}_i\|^2},$$



		MarkovNet		CsiNet-LSTM	
Env	CR	NMSE _{truncate}	NMSE _{all}	NMSE _{truncate}	NMSE _{all}
Indoor	$\frac{1}{4}$	-29.26	-20.81	-21.28	-18.4
	$\frac{1}{8}$	-26.25	-20.26	-20.76	-18.12
	$\frac{1}{16}$	-25.27	-19.99	-19.96	-17.67
	$\frac{1}{32}$	-24.62	-19.78	-19.41	-17.34
Outdoor	$\frac{1}{4}$	-16.8	-12.4	-8.89	-7.99
	$\frac{1}{8}$	-13.19	-10.86	-7.17	-6.60
	$\frac{1}{16}$	-10.45	-9.13	-6.65	-6.15
	$\frac{1}{32}$	-8.87	-7.92	-5.33	-4.99

Table: NMSE of truncated vs. full CSI matrices.

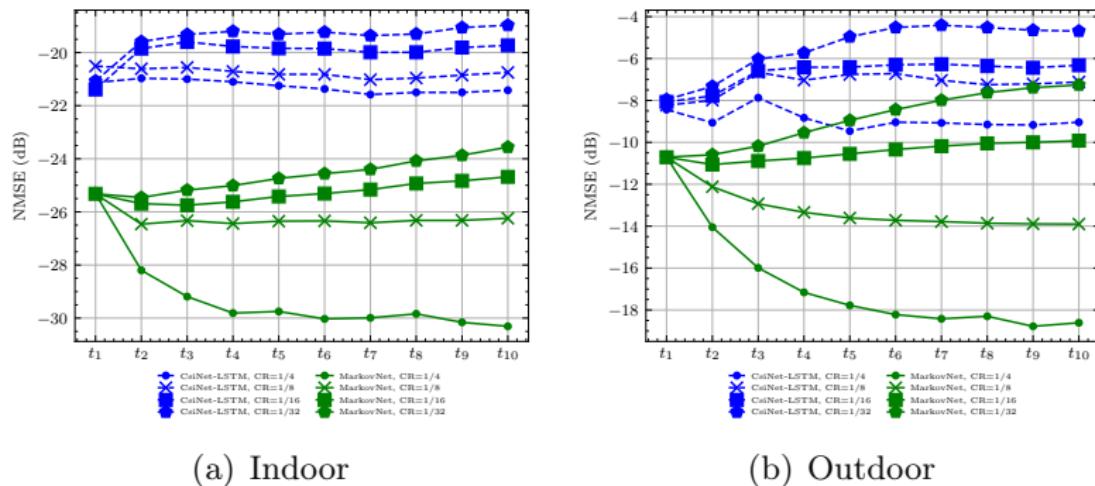


Figure: NMSE_{truncated} comparison of MarkovNet and CsiNet-LSTM at various compression ratios (CR).

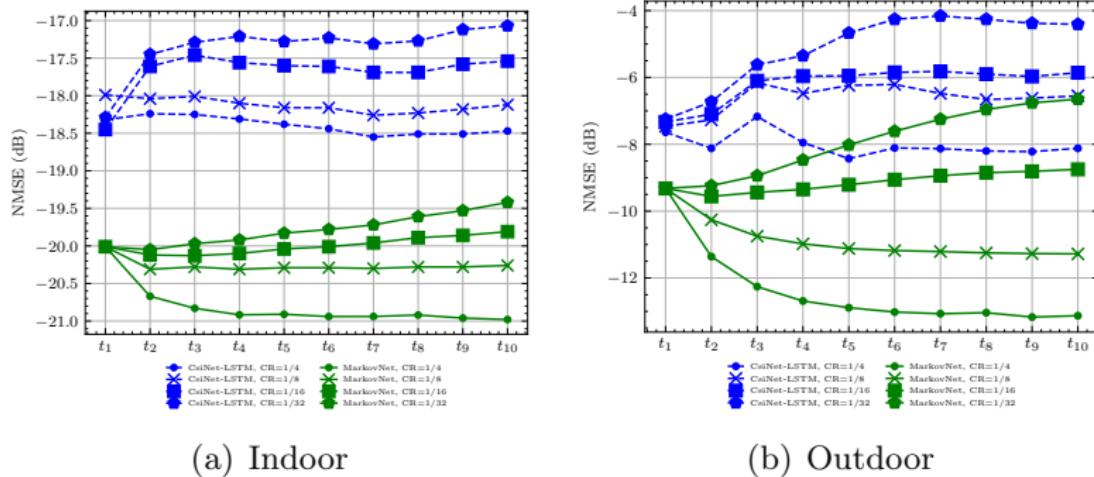


Figure: NMSE_{all} comparison of MarkovNet and CsiNet-LSTM at various compression ratios (CR).