one6G Open Lecture 1 – 6G Network AI

Learning the Physical Layer in FDD Systems: Centralized Learning of Distributed Functions

Wolfgang Utschick

Technical University of Munich
School of Computation, Information and Technology
Munich Data Science Institute

A note on UL–DL Distributional Invariance

Distributional Invariance of the Uplink & Downlink

Invariance of

- ✗ instantaneous UL & DL realizations (FDD systems ⟹ no reciprocity in general)
- ✓ UL & DL distributions

Distributional Invariance of the Uplink & Downlink:

> ***"Sampling channel state information from the same propagation environment in different frequency bands, but still with similar radio propagation characteristics, represents approximately the same underlying probability distribution."***

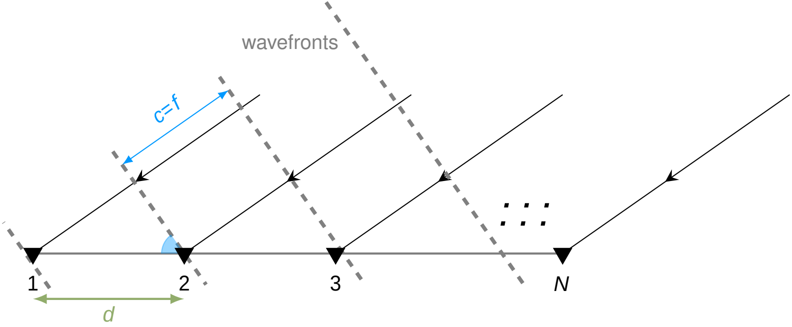
Proof:
Not a rigorous proof, but there is some evidence by statistical hypothesis testing based on two-sample tests, cf. Utschick et al, IEEE T-WC, 2022.

An intuitive explanation (1)

Consider the simple case of a ULA with N antennas.
In this case the channel vector $\boldsymbol{h}$ is such that:

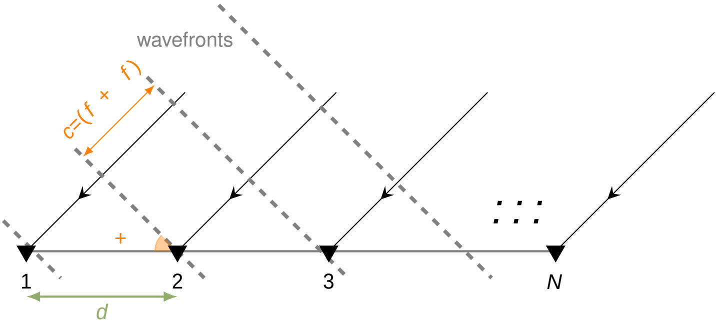
$$\boldsymbol{h} \; / \; [\;^{0}; \;^{1}; \;^{2}; \ldots; \;^{N\ \ 1}]$$

with $\; = \exp \quad j \frac{2 \;\; df}{c} \sin$

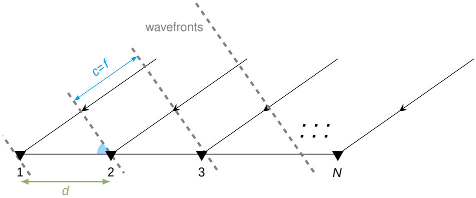


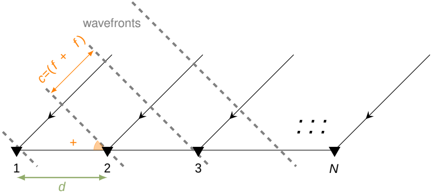
An intuitive explanation (2)

At the same time, for a different carrier frequency we have . . .



An intuitive explanation (3)





Small changes f in the carrier frequency can be compensated by small changes of , therefore ***h*** is not changed if

$$f \sin \quad = (f + \quad f) \sin (\quad + \quad)$$

Hypothesis Testing Based On Two-Sample Tests

IEEE Trans WC 2022

Definition: Given a positive definite kernel $k(\cdot,\cdot) = \langle \varphi(\cdot), \varphi(\cdot) \rangle$ of a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ with a feature map $\varphi(\cdot) \in \mathcal{H}_k$, the maximum mean discrepancy (MMD) between two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ can be obtained by

$$\mathrm{MMD}^2(\mathbb{P}, \mathbb{Q}, k) := \mathbb{E}[k(p, p') + k(q, q') - 2k(p, q)], \tag{9}$$

with random vari[illegible]$') \sim \mathbb{P} \times \mathbb{P}$ and $(q, q') \sim \mathbb{Q} \times \mathbb{Q}$. It follows that $\mathrm{MMD}(\mathbb{P}, \mathbb{Q}, k) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$. [illegible]ther assume that we have sample sets $\mathcal{P} \sim \mathbb{P}$ and $\mathcal{Q} \sim \mathbb{Q}$ of equal sample size n, an un[illegible]imator of the squared MMD for measuring the discrepancy between $\mathbb{P}$ and $\mathbb{Q}$ can be obtai[illegible]

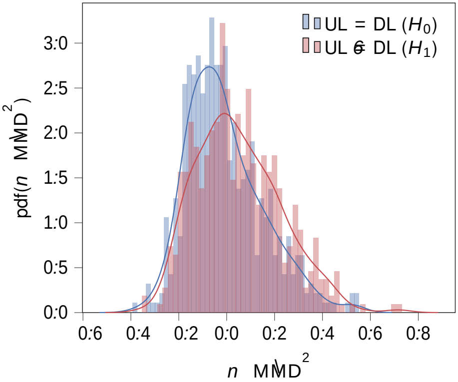
$$\widehat{\mathrm{MMD}}[illegible] := \frac{1}{n(n-1)} \sum_{i \neq j} h_{ij}, \tag{10}$$

where $h_{ij} := k(p_i, p_j) + k(q_i, q_j) - k(p_i,$[illegible]$p_j)$ with $p_i \in \mathcal{P}$ and $q_i \in \mathcal{Q}$ being the realizations of the random variables $p \sim \mathbb{P}$ and [illegible]lowing the usual kernel trick, we swap the choice of feature map $\varphi(\cdot)$ with the decision for [illegible]unction $k(\cdot,\cdot)$. The most common choice for a kernel is the Gaussian kernel, i.e.,

$$k(p, q) = \exp\left(-\frac{\|p - q\|^2}{\sigma_{50}^2}\right),$$

where $p \in \mathcal{P}$ and $q \in \mathcal{Q}$ are two samples drawn from $\mathbb{P}$ and $\mathbb{Q}$ and σ_{50} corresponds to the 50-percentile (median) distance between elements in the aggregate sample, as suggested in [28].

Hypothesis Testing Based On Two-Sample Tests



V. Rizzello, N. Turan, M. Joham, and W. Utschick. Two-sample Tests for Validating the UL-DL Conjecture in FDD systems. In Proceedings of the 17th International Symposium on Wireless Communication Systems, Berlin, Germany, September 2021.

W. Utschick, V. Rizzello, M. Joham, Z. Ma and L. Piazzi “Learning the CSI Recovery in FDD Systems”, 2021, `https://arxiv.org/abs/2104.01322`.

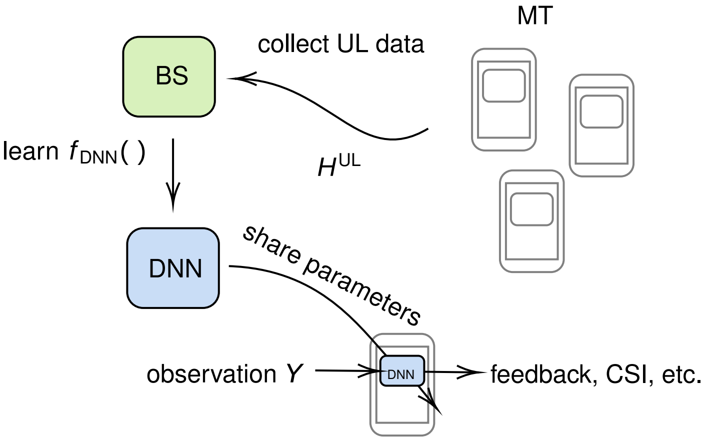
Why is this important?

Because what really matters in machine learning is the distribution of data!

Because what really matters in machine learning is the ~~distribution~~ of data!
ensembles

Novel Design Options for FDD PHY Layer Functions

PHY Layer Functions Soley Based on UL-Data

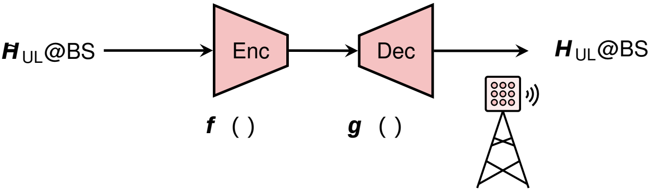


Application:

Channel Compression & Reconstruction

Channel Compression & Reconstruction

1. Train an autoencoder solely based on UL CSI



$\boldsymbol{f}$: encoder neural net

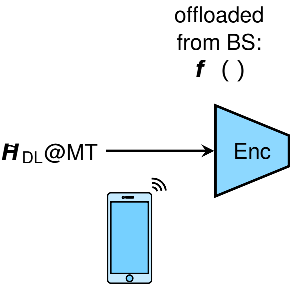
$\boldsymbol{g}$: decoder neural net

$$\boldsymbol{z}_{UL} = \boldsymbol{f}(\tilde{\boldsymbol{H}}_{UL})$$

$$\boldsymbol{H}_{UL} \approx \hat{\boldsymbol{H}}_{UL} = \boldsymbol{g}(\boldsymbol{z}_{UL}).$$

Channel Compression & Reconstruction

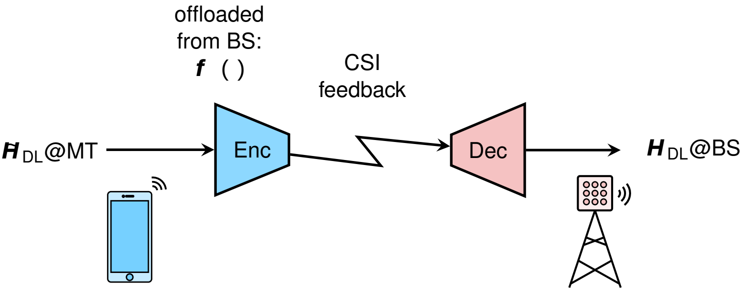
2. Offload the UL-trained encoder to each MTs in the cell
3. Compress the DL CSI at the MT



$$\boldsymbol{z}_{\mathrm{DL}} = \boldsymbol{f}\,(\boldsymbol{H}_{\mathrm{DL}})$$

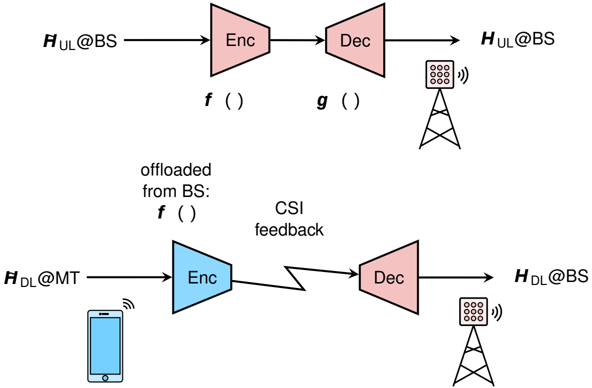
Channel Compression & Reconstruction

4. Feed the encoded DL CSI back to the BS
5. Decode the DL CSI at the BS with the UL-trained decoder



$$\boldsymbol{H}_{\mathrm{DL}} \approx \hat{\boldsymbol{H}}_{\mathrm{DL}} = \boldsymbol{g}(\boldsymbol{z}_{\mathrm{DL}})$$

Channel Compression & Reconstruction

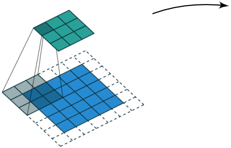


not necessary to collect DL CSI data for training

robust against Gaussian noise

V. Rizzello, W. Utschick, “Learning the CSI Denoising and Feedback Without Supervision”, 2021, `https://arxiv.org/abs/2104.05002`.

Channel Compression & Reconstruction



Layer type	Output shape			#Parameters $\boldsymbol{\theta}$
Input	64	160	2	0
Conv2D, strides=2	32	80	8	152
Batch normalization	32	80	8	32
ReLU	32	80	8	0
Conv2D, strides=2	16	40	16	1168
Batch normalization	16	40	16	64
ReLU	16	40	16	0
Conv2D, strides=2	8	20	32	4640
Batch normalization	8	20	32	128
ReLU	8	20	32	0
Conv2D, strides=2	4	10	64	18496
Batch normalization	4	10	64	256
ReLU	4	10	64	0
Conv2D, strides=2	2	5	128	73856
Batch normalization	2	5	128	512
ReLU	2	5	128	0
Flatten		1280		0
Fully-connected		256		327936
Tanh		256		0

autoencoder is based on convolutional layers instead of fully connected layers

$\Longrightarrow$ architecture can be easily scaled to high dimensional CSI

Channel Compression & Reconstruction

QuaDRiGa channel simulator

Urban Microcell NLoS

58 paths

training samples = 48 K

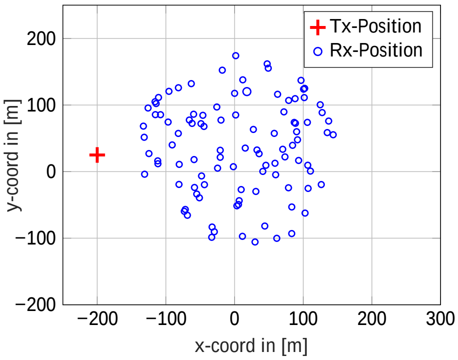
validation/test samples = 2   6 K

number of antennas = 64

number of carriers = 160

center frequency = 2.5 GHz

frequency gap = 120 and 480 MHz



Channel Compression & Reconstruction

QuaDRiGa channel simulator

Urban Microcell NLoS

58 paths

training samples = 48 K

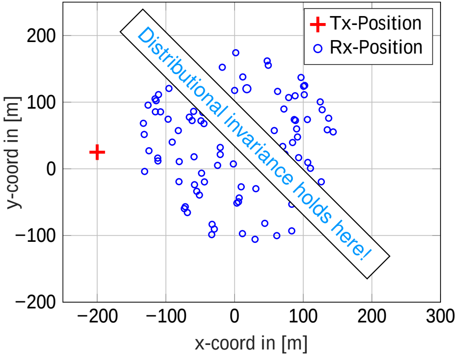
validation/test samples = 2   6 K

number of antennas = 64

number of carriers = 160

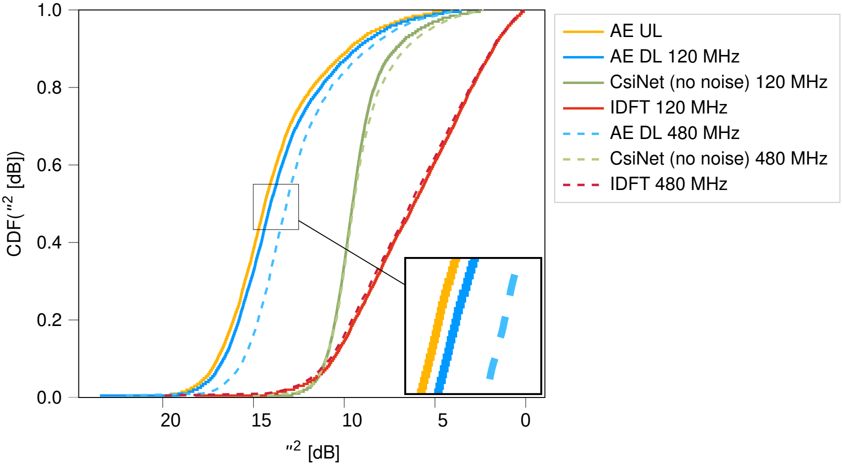
center frequency = 2.5 GHz

frequency gap = 120 and 480 MHz

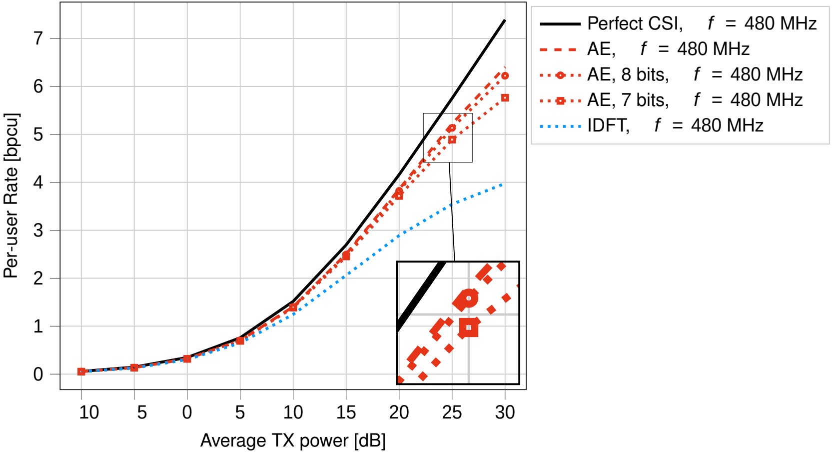


Channel Compression & Reconstruction

NMSE of different methods for SNR=10 dB.



Channel Compression & Reconstruction



achievable per-user rate in a multiuser scenario (8 user)

zero-forcing precoding based on recovered DL CSI

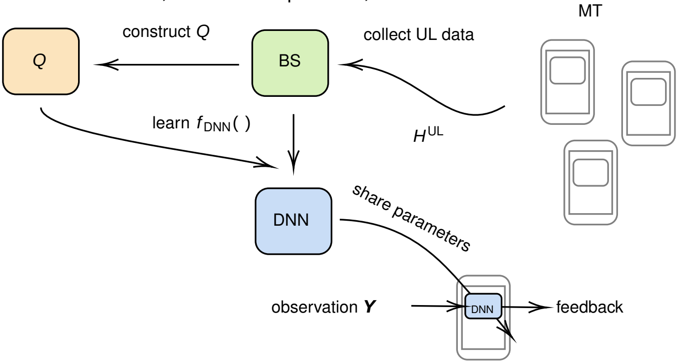
the hyperbolic tangent as activation function in the latent space makes easy to quantize the CSI

Application:

Codebook Construction & Feedback Generation

Codebook Construction & Feedback Generation

Proposed concept: Training based soley on uplink (UL) data at the base station (@BS) and subsequent offloading of trained deep neural networks to the mobile terminals (MTs) for feedback generation, channel estimation, channel compression, etc.



kind of distributed implementation of AI-aided physical layer functions

offloading deep functionalities anywhere and anytime in the networks for increasing performance

N. Turan, M. Koller, S. Bazzi, W. Xu and W. Utschick, „Unsupervised Learning of Adaptive Codebooks for Deep Feedback Encoding in FDD Systems." , 2021, `https://arxiv.org/abs/2105.09125`.

Urban Macrocell Scenario

3GPP 38.901 UMa, single carrier scenario:

Urban Macrocell Scenario

3GPP 38.901 UMa, single carrier scenario:

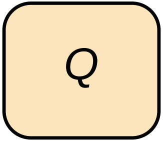
Distributional invariance holds here!

Urban Macrocell Scenario

3GPP 38.901 UMa, single carrier scenario:

- QuaDRiGa channel simulator
- MIMO channels: (BS antennas, MT antennas) = (16,4) or (32,16),
- non-line-of-sight (NLOS), line-of-sight (LOS) and outdoor-to-indoor (O2I)
- UL carrier frequency = 2.53 GHz,
- DL carrier frequency = 2.73 GHz,
- BS-ULA with “3GPP-3D” antennas,
- MT-ULA with ”omni- directional” antennas,
- BS placed at a height of 25 m with a sector of 120 ,
- minimum distance of the MT location to the BS is 35 m,
- maximum distance to the BS is 500 m.
- 10^4 training, 2.5 10^3 validation and 5 10^3 test samples.

Codebook Construction



Codebook Construction

unsupervised codebook design

using the k-means algorithm based on the achievable date rate metric

solely based on UL channel state information (CSI)

1. Divide the training set H into K clusters $V_k^{(i)}$:

$$V_k^{(i)} = \{ \boldsymbol{H} \in H \mid r(\boldsymbol{H};\boldsymbol{Q}_k^{(i)}) \quad r(\boldsymbol{H};\boldsymbol{Q}_j^{(i)}); k \neq j \}.$$

2. Find new covariance matrices or update the so called “cluster centers”:

$$\boldsymbol{Q}_k^{(i+1)} = \underset{\boldsymbol{Q} \quad 0}{\text{argmax}} \frac{1}{|V_k^{(i)}|} \sum_{\boldsymbol{H} \in V_k^{(i)}} r(\boldsymbol{H};\boldsymbol{Q})$$

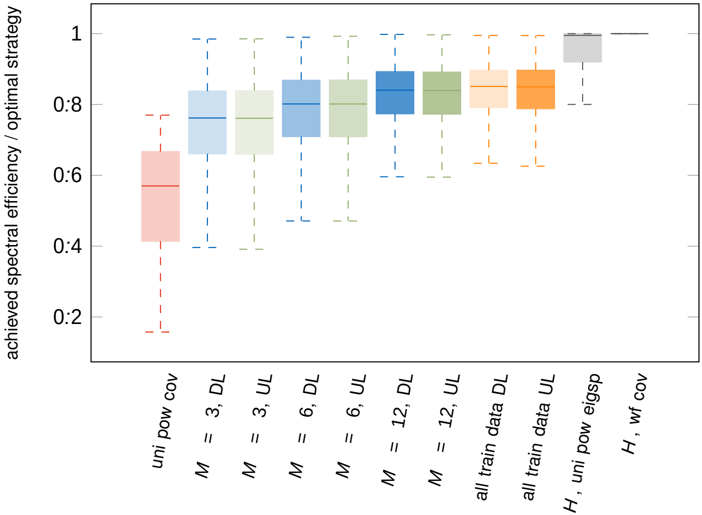
$$\text{subject to} \quad \text{trace}(\boldsymbol{Q}) \qquad \text{and} \quad \text{rank}\, \boldsymbol{Q} \quad N_{\text{rx}}.$$

Applying a projected gradient algorithm on the candidate set of precoding covariances:

$$\boldsymbol{g}_{\boldsymbol{Q}} = \frac{1}{{}^{2}_{n} \ln(2)} \sum_{\boldsymbol{H} \in V_k^{(i)}} \boldsymbol{H}^{\text{H}} \quad \boldsymbol{I} + \frac{1}{{}^{2}_{n}} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\text{H}} \quad {}^{1} \boldsymbol{H};$$

$$\boldsymbol{Q} \qquad \boldsymbol{Q} + \quad \boldsymbol{g}_{\boldsymbol{Q}}.$$

Codebook Construction

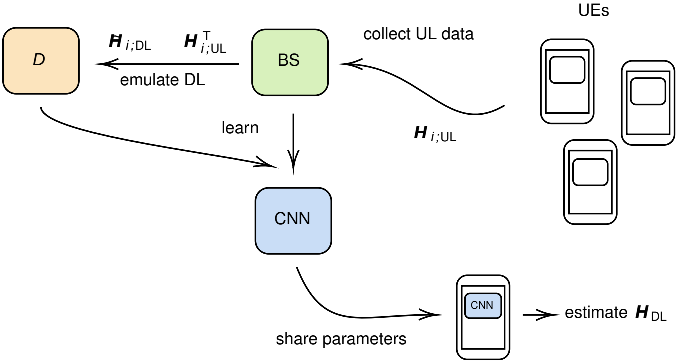


$N_{tx} = 16$; $N_{rx} = 4$, with different codebook sizes (2^M)

Application:

Channel Estimation

Channel Estimation



Downlink observations @MT:

$$\boldsymbol{Y}_{i;\mathrm{DL}} = \boldsymbol{H}_{i;\mathrm{DL}}\boldsymbol{X} + \boldsymbol{Z}_i$$

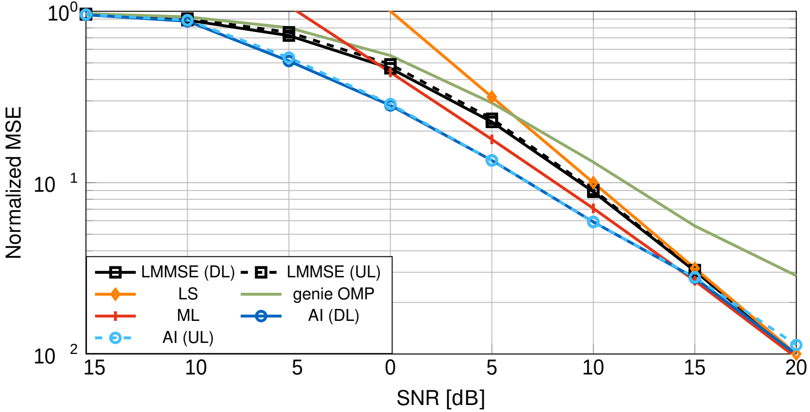
Emulated downlink system @BS:

$$\tilde{\boldsymbol{Y}}_{i;\mathrm{DL}} = \boldsymbol{H}^{\mathrm{T}}_{i;\mathrm{UL}}\boldsymbol{X} + \boldsymbol{Z}_i$$

B. Fesl, N. Turan, M. Koller, M. Joham, and W. Utschick, „Centralized Learning of the Distributed Downlink Channel Estimators in FDD Systems using Uplink Data.“ , 2021, https://arxiv.org/abs/2105.10746.

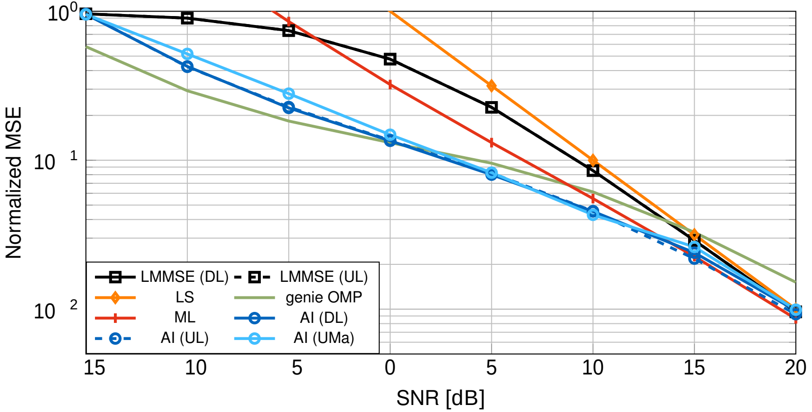
Channel Estimation

$N_{BS} = 64$, $N_{MS} = 4$, $N_P = 64$, mixed NLOS/LOS scenario, 20K training samples



Channel Estimation

$N_{BS} = 64$, $N_{MS} = 4$, $N_P = 64$, mainly LOS scenario, 20K training samples



Thank You!

Links to Recent Work

- Channel Estimation, Prediction, and Extrapolation:

  Learning The CSI Recovery in FDD Systems (IEEE T-WC 2022) (https://arxiv.org/abs/2104.01322)

  Centralized Learning of the Distributed Downlink Channel Estimators in FDD Systems using Uplink Data (WSA 2021) (https://arxiv.org/abs/2105.10746)

- Codebook Design and Feedback Generation in Multiuser MIMO Systems:

  Learning The CSI Denoising and Feedback Without Supervision (SPAWC 2021) (https://arxiv.org/abs/2104.05002)

  Unsupervised Learning of Adaptive Codebooks for Deep Feedback Encoding in FDD Systems (Asilomar 2021) (https://arxiv.org/abs/2105.09125)

Appendix

Quantitative Analysis:

More About Two-Sample Tests

MMD definition

The two-sample tests are based on the maximum mean discrepancy metric.
The maximum mean discrepancy is the distance between feature means:

$$\mathrm{MMD}^2(\mathbb{P}, \mathbb{Q}; k) := \mathbb{E}[k(p, p') + k(q, q') - 2k(p, q)], \quad (1)$$

where:

- $k(\cdot, \cdot) = \langle \varphi(\cdot), \varphi(\cdot) \rangle$: positive definite kernel of a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ with a feature map $\varphi(\cdot) \in \mathcal{H}_k$
- $(p, p') \sim \mathbb{P} \times \mathbb{P}$
- $(q, q') \sim \mathbb{Q} \times \mathbb{Q}$

Hence, $\mathrm{MMD}(\mathbb{P}, \mathbb{Q}; k) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$.

Kernels descriptions

"1-percentile": = 1-percentile of the distance between points in the aggregate sample

$$k(a;b) = \exp\left(\frac{ka \quad bk^2}{{}^{2}_{1}} \right)$$

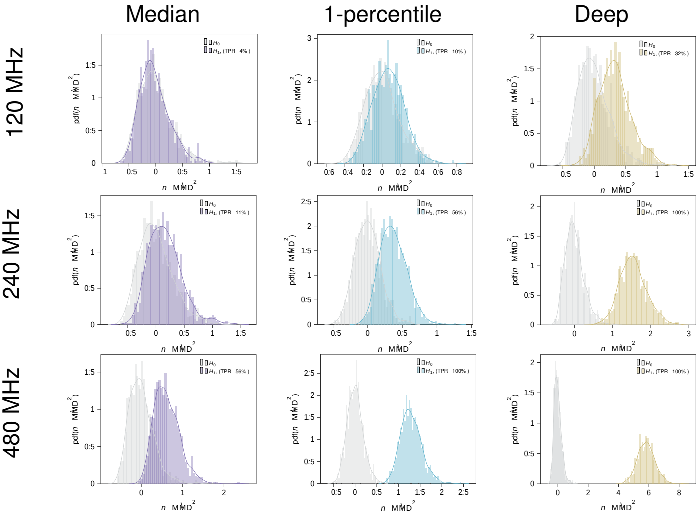
"median": = 50-percentile of the distance between points in the aggregate sample

$$k(a;b) = \exp\left(\frac{ka \quad bk^2}{{}^{2}_{50}} \right)$$

"deep": = 50-percentile of the distance between points in the aggregate sample of the latent space of an autoencoder

$$k(a;b) = \exp\left(\frac{kf \; (a) \quad f \; (b)k^2}{{}^{2}_{50;\mathrm{latent}}} \right)$$

Different frequency gaps in a UMi LOS scenario



The larger the frequency gaps, the larger is the discrepancy between the UL & DL distributions.

Over different UMi NLOS ($\Delta f = 120$ MHz)