WWTF Vienna Science and Technology Fund
FWF Österreichischer Wissenschaftsfonds

Learning to Solve Wireless PHY Problems: From Structured Models to Attention-Based Methods

Rhine Summit 2026, Session 2: AI4Wireless

Associate Prof. Stefan Schwarz
in collaboration with: Kaifeng Lu, Dr. Faruk Pasic, Dr. Artan Salihu and Prof. Markus Rupp

April 2026, stefan.schwarz@tuwien.ac.at

TU WIEN Technische Universität Wien
Institute of Telecommunications

Contents

Why Combining Model and Data-Driven Approaches?

- Wireless PHY is highly structured:
  - Propagation physics, geometry, well established models

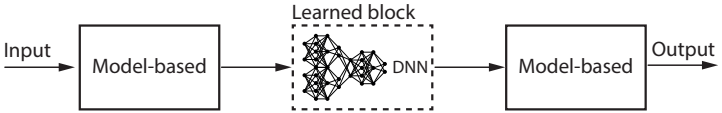
Why Combining Model and Data-Driven Approaches?

- Wireless PHY is highly structured:
  - Propagation physics, geometry, well established models
- Purely data-driven approaches:
  - Ignore domain knowledge
  - Require large datasets
  - Limited robustness under distribution shifts

Why Combining Model and Data-Driven Approaches?

- Wireless PHY is highly structured:
  - Propagation physics, geometry, well established models
- Purely data-driven approaches:
  - Ignore domain knowledge
  - Require large datasets
  - Limited robustness under distribution shifts
- Purely model-based approaches:
  - Rely on simplifying assumptions
  - Suffer from model mismatch

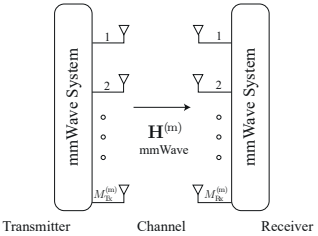
Model-based Design with Learning

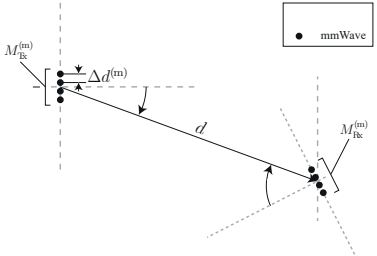


- Combine both worlds:
  - Retain structure from models
  - Learn difficult components from data
- Typical pattern:
  - Model-based pipeline
  - Replace or augment selected blocks by learned mappings

Learning complements, not replaces, signal processing models

Out-of-Band Aided Millimeter Wave Communication



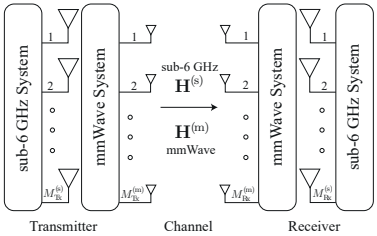


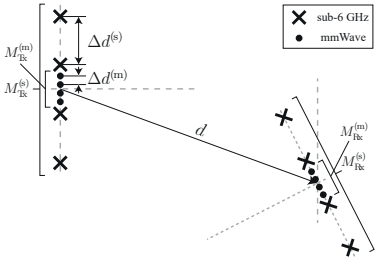
- Consider channel estimation (CE) in a mmWave MIMO system

  $\Rightarrow$ Pre-beamforming SNR is very low – pilot-based CE becomes unreliable

  $\Rightarrow$ Double-sided beam training is too slow for large arrays

Out-of-Band Aided Millimeter Wave Communication





- Consider channel estimation (CE) in a mmWave MIMO system

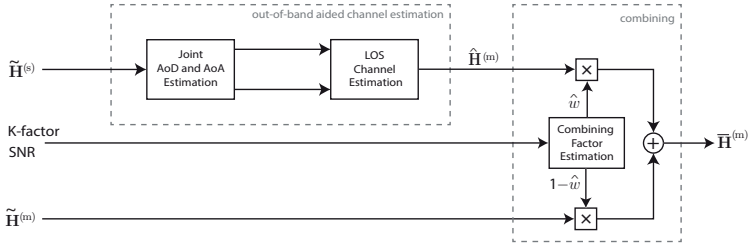
  $\Rightarrow$ Pre-beamforming SNR is very low – pilot-based CE becomes unreliable

  $\Rightarrow$ Double-sided beam training is too slow for large arrays

- Utilize side information from a co-located sub-6 GHz system

  $\Rightarrow$ *Geometry of LOS (and dominant paths) is shared*

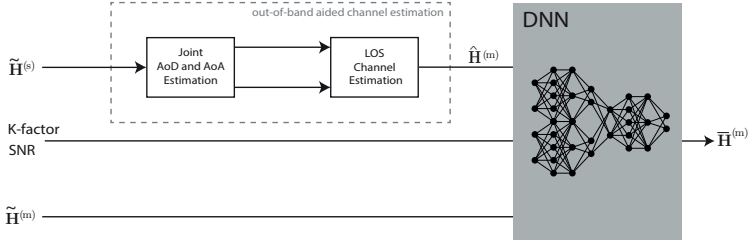
Model-Based OOB-Aided Channel Estimation



- Extract LOS angles of arrival and departure from sub-6 GHz channel estimate
- Convert to mmWave equivalent LOS channel (pathloss, array geometry)
- Convex combination with in-band estimate to obtain final estimate (SNR, K-factor)

Millimeter Wave MIMO Channel Estimation using sub-6 GHz Out-of-Band Information, F. Pasic, S. Schwarz, et al, IEEE TCOM 2024

Learning-Enhanced OOB-Aided Channel Estimation

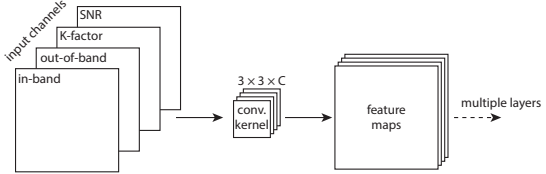


- Replace hand-crafted combination with a trainable mapping that learns:
  - Relevance of in-band vs. out-of-band CE (K-factor, SNR)
  - Optimal combination of both estimates
  - Denoising and extraction of dominant components

Learning replaces a model-based design choice

Deep Learning-based mmWave MIMO Channel Estimation using sub-6 GHz Channel Information: CNN and UNet Approaches, F. Pasic, L. Eller, S. Schwarz, et al, IEEE InfoCom 2025

DNN Design Choice: Convolutional Models



- Channel matrices exhibit spatial structure:
  - Correlations across antennas
  - Smooth variations induced by propagation geometry
- Convolutional neural networks (CNNs):
  - Apply local filters across the channel matrices
  - Share parameters across spatial locations
  - Efficient and robust for structured inputs

Single-scale processing: local spatial interactions

Extension: Multi-scale Processing (U-Net)

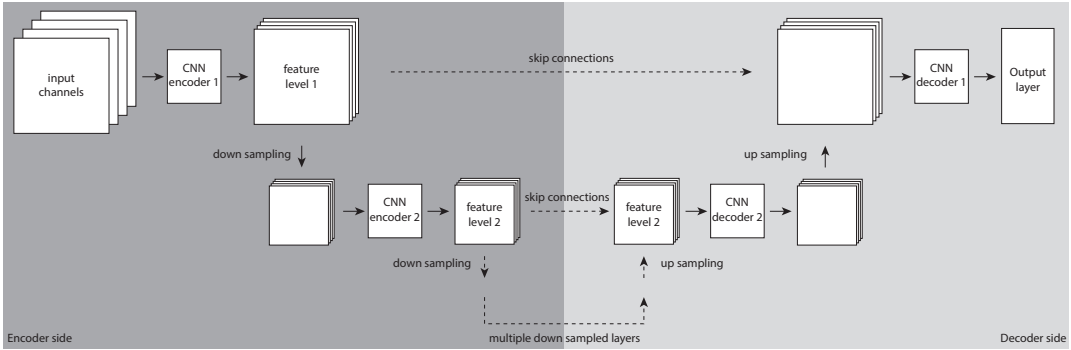
- Limitation of CNNs:
  - Interactions are local
  - Global structure captured only indirectly via DNN depth

Extension: Multi-scale Processing (U-Net)

- Limitation of CNNs:
  - Interactions are local
  - Global structure captured only indirectly via DNN depth
- U-Net architecture:
  - Introduces down-sampling and up-sampling
  - Enables multi-scale representations
  - Preserves fine details via skip connections

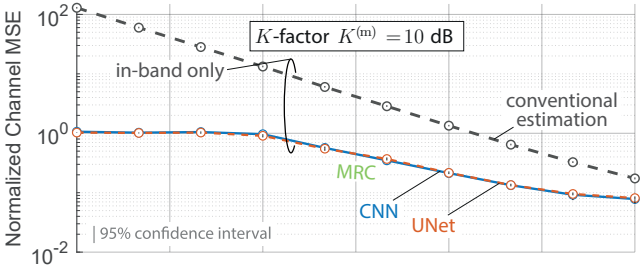
Multi-scale processing: local + global structure

U-Net Illustration



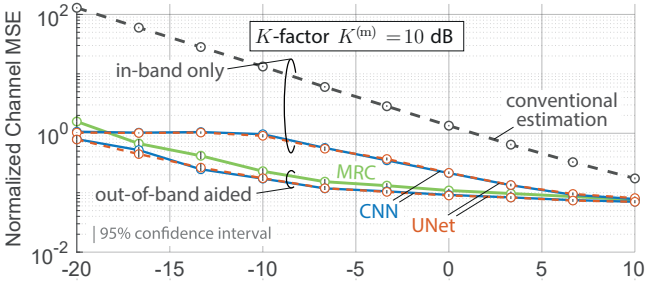
- Multiple resolution scales provide global context
- Skip connections between levels preserve details

Simulation Example



- Already in-band DNNs provide significant gain over conventional least-squares estimation
  $\Rightarrow$ Denoising through extraction of dominant LOS component (K-factor dependent)

Simulation Example



- Already in-band DNNs provide significant gain over conventional least-squares estimation
  $\Rightarrow$ Denoising through extraction of dominant LOS component (K-factor dependent)
- Out-of-band aided DNNs further improve performance due to higher SNR at sub-6 GHz
- Limited additional gain compared to simple linear combination (MRC)
- Improvements from more complex DNN architecture (U-Net) are limited
  $\Rightarrow$ **The bottleneck is not capacity, but how interactions are modeled**

From Local Processing to Interaction Modeling

- CNNs and U-Net:
  - Rely on local spatial processing
  - Capture global structure only indirectly through DNN depth

From Local Processing to Interaction Modeling

- CNNs and U-Net:
  - Rely on local spatial processing
  - Capture global structure only indirectly through DNN depth
- Many wireless problems are interaction-driven:
  - Coupling between users (interference), antennas, channels
  - Decisions depend on relative relationships

From Local Processing to Interaction Modeling

- CNNs and U-Net:
  - Rely on local spatial processing
  - Capture global structure only indirectly through DNN depth
- Many wireless problems are interaction-driven:
  - Coupling between users (interference), antennas, channels
  - Decisions depend on relative relationships
- Self-attention:
  - Directly models interactions between elements
  - Learns which parts of the input are relevant

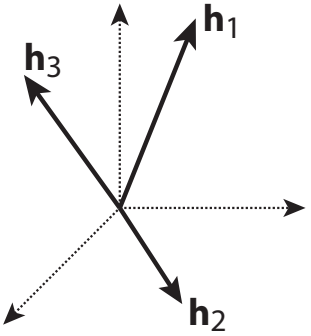
From Local Processing to Interaction Modeling

- CNNs and U-Net:
  - Rely on local spatial processing
  - Capture global structure only indirectly through DNN depth
- Many wireless problems are interaction-driven:
  - Coupling between users (interference), antennas, channels
  - Decisions depend on relative relationships
- Self-attention:
  - Directly models interactions between elements
  - Learns which parts of the input are relevant

From local filters to learned interactions

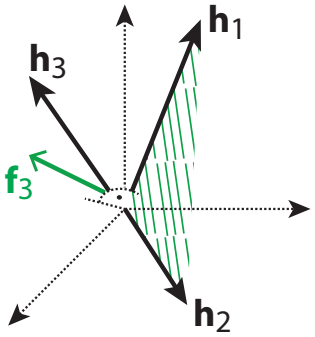
Example: Interactions Across Users

- Consider multi-antenna channel vectors of multiple users at a given subcarrier



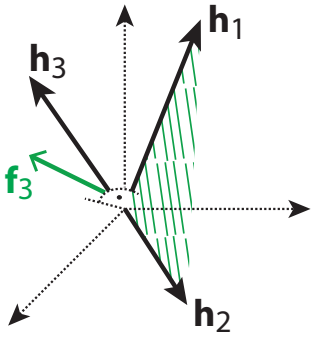
Example: Interactions Across Users

- Consider multi-antenna channel vectors of multiple users at a given subcarrier
- System performance depends on interactions between users:
  - Interference depends on channel similarity
  - Beamforming and power allocation are coupled



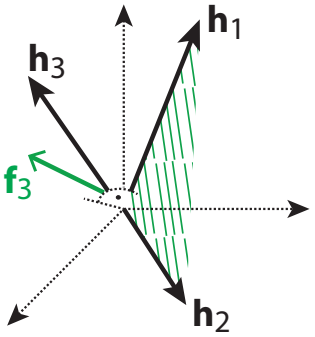
Example: Interactions Across Users

- Consider multi-antenna channel vectors of multiple users at a given subcarrier
- System performance depends on interactions between users:
  - Interference depends on channel similarity
  - Beamforming and power allocation are coupled
- Limitation of local processing (CNN):
  - No natural notion of locality across users
  - Interactions are not explicitly modeled



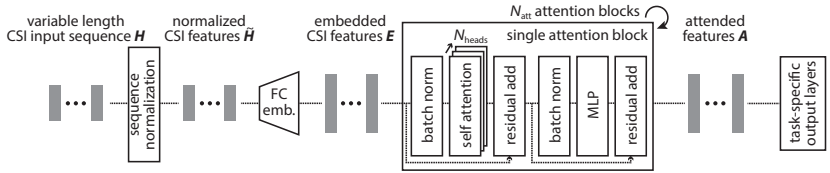
Example: Interactions Across Users

- Consider multi-antenna channel vectors of multiple users at a given subcarrier
- System performance depends on interactions between users:
  - Interference depends on channel similarity
  - Beamforming and power allocation are coupled
- Limitation of local processing (CNN):
  - No natural notion of locality across users
  - Interactions are not explicitly modeled
- Self-attention:
  - Models pairwise interactions across users
  - Identifies relevant users for each decision



Multi-user systems are inherently interaction-driven – captured by self-attention

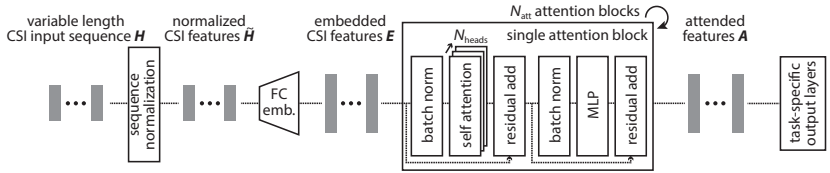
Wireless Transformer Architecture



- Transformers operate on sets or sequences of elements:
  - Natural language: sequences of word embeddings (tokens)
  - Wireless systems: collections of channel vectors or matrices

Attention Aided CSI Wireless Localization, A. Salihu, S. Schwarz and M. Rupp, IEEE SPAWC 2022

Wireless Transformer Architecture



- Transformers operate on sets or sequences of elements:
  - Natural language: sequences of word embeddings (tokens)
  - Wireless systems: collections of channel vectors or matrices
- Flexible definition of sequence dimension:
  - Users, subcarriers, symbols, or combinations thereof (multi-dimensional transformers)

$$\boldsymbol{H} = \{\ldots, \mathbf{h}_u[k], \ldots\}, \quad \mathbf{h}_u[k] = \left[\Re\left(\bar{\mathbf{h}}_u[k]\right), \Im\left(\bar{\mathbf{h}}_u[k]\right), 10\log_{10}\left(\frac{\|\mathbf{h}_u[k]\|^2}{\sigma_n^2}\right)\right]$$

Attention Aided CSI Wireless Localization, A. Salihu, S. Schwarz and M. Rupp, IEEE SPAWC 2022

Learning to Optimize System Objectives

- Many wireless problems are naturally formulated as optimization problems:
  - Beamforming, power allocation, scheduling,...

Learning to Optimize System Objectives

- Many wireless problems are naturally formulated as optimization problems:
  - Beamforming, power allocation, scheduling,...
- Instead of supervised learning:
  - Learn directly from system objectives
  - No need for ground-truth labels

Learning to Optimize System Objectives

- Many wireless problems are naturally formulated as optimization problems:
  - Beamforming, power allocation, scheduling,...
- Instead of supervised learning:
  - Learn directly from system objectives
  - No need for ground-truth labels
- Optimization-based learning:
  - Train model to maximize a performance metric
  - Use differentiable objective as training signal

$$\textit{Learning signal} = \textit{system objective}$$

- Sum-rate maximization for multi-user beamforming:

$$\max_{\{\mathbf{f}_u\}} \sum_{u=1}^{U} \log_2\left(1 + \frac{|\mathbf{h}_u^H \mathbf{f}_u|^2}{\sum_{v \neq u} |\mathbf{h}_u^H \mathbf{f}_v|^2 + \sigma^2}\right)$$

$$\text{s.t.} \sum_{u=1}^{U} \|\mathbf{f}_u\|^2 \leq P$$

- Sum-rate maximization for multi-user beamforming:

$$\max_{\{\mathbf{f}_u\}} \sum_{u=1}^{U} \log_2\left(1 + \frac{|\mathbf{h}_u^H \mathbf{f}_u|^2}{\sum_{v \neq u} |\mathbf{h}_u^H \mathbf{f}_v|^2 + \sigma^2}\right)$$

$$\text{s.t.} \sum_{u=1}^{U} \|\mathbf{f}_u\|^2 \leq P$$

- Adjustable fairness via Jain's index:

$$\frac{\left(\sum_{u=1}^{U} R_u\right)^2}{U \sum_{u=1}^{U} R_u^2} \geq J_{LB}$$

- Sum-rate maximization for multi-user beamforming:

$$\max_{\{\mathbf{f}_u\}} \sum_{u=1}^{U} \log_2\left(1 + \frac{|\mathbf{h}_u^H \mathbf{f}_u|^2}{\sum_{v \neq u} |\mathbf{h}_u^H \mathbf{f}_v|^2 + \sigma^2}\right)$$

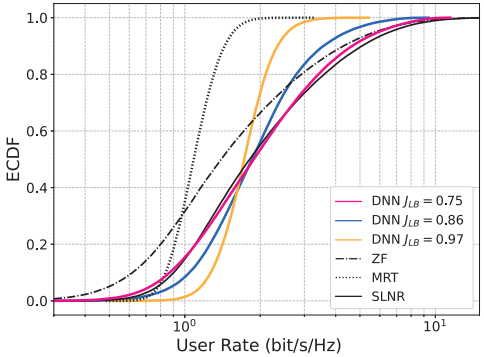
$$\text{s.t.} \sum_{u=1}^{U} \|\mathbf{f}_u\|^2 \leq P$$

- Adjustable fairness via Jain's index:

$$\frac{\left(\sum_{u=1}^{U} R_u\right)^2}{U \sum_{u=1}^{U} R_u^2} \geq J_{LB}$$

- Challenging optimization problem:
  - Non-convex due to inter-user interference and fairness constraint
- Optimization-based learning:
  - Beamformers can be learned directly via gradient-based optimization (WiT)

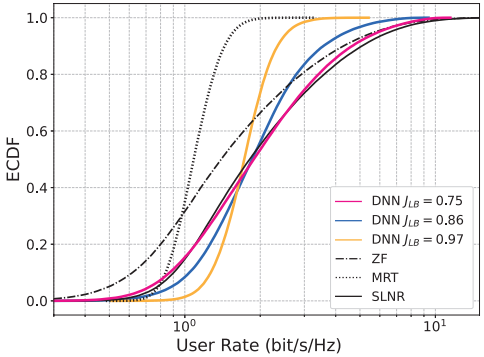
Simulation Example

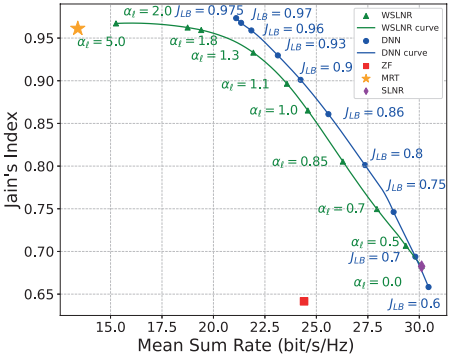


- System with $N_t = 16$ antennas serving $N_u = 12$ users

DNN-Enabled Multi-User Beamforming for Throughput Maximization under Adjustable Fairness, K. Lu, M. Rupp and S. Schwarz, IEEE ICMLCN 2026

Simulation Example

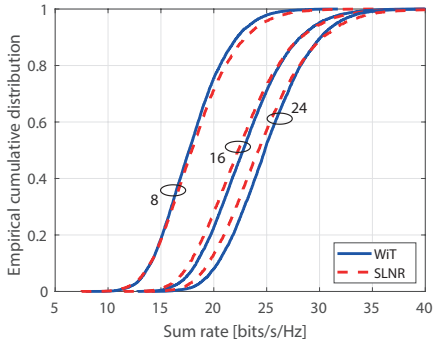




- System with $N_t = 16$ antennas serving $N_u = 12$ users
- With tight fairness constraint, significant gains over weighted SLNR benchmark

DNN-Enabled Multi-User Beamforming for Throughput Maximization under Adjustable Fairness, K. Lu, M. Rupp and S. Schwarz, IEEE ICMLCN 2026

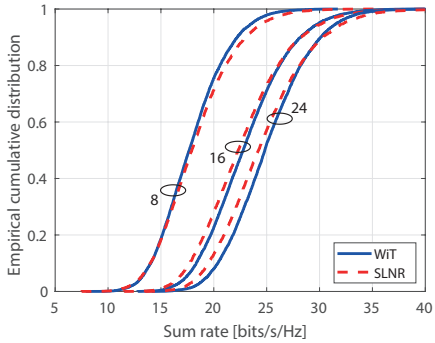
Simulation Example

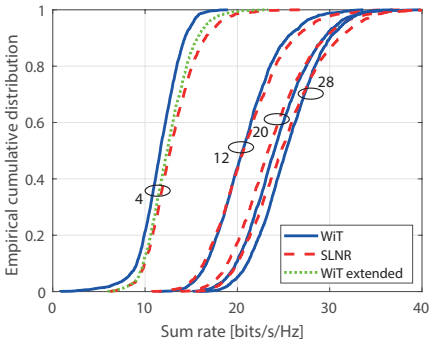


- Transformers can handle varying input sequence lengths

Transformer DNNs for Versatile CSI-Based Wireless Communication and Localization, S. Schwarz, K. Lu, and M. Rupp, IEEE WPMC 2025

Simulation Example





- Transformers can handle varying input sequence lengths
- Interpolation and limited extrapolation beyond the training set are supported

Transformer DNNs for Versatile CSI-Based Wireless Communication and Localization, S. Schwarz, K. Lu, and M. Rupp, IEEE WPMC 2025

Learning from Data without Labels

- Supervised learning:
  - Requires large amounts of labeled data
  - Labels are costly or unavailable in wireless systems (e.g., localization)

Learning from Data without Labels

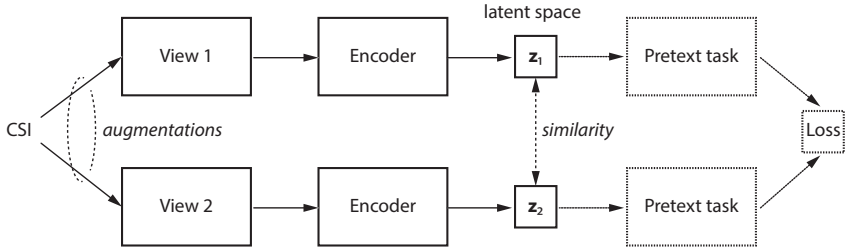
- Supervised learning:
  - Requires large amounts of labeled data
  - Labels are costly or unavailable in wireless systems (e.g., localization)
- Optimization-based learning:
  - Requires well-defined system objectives; often non-convex (local optimum)
  - Not always applicable (e.g., throughput vs Shannon rate)

Learning from Data without Labels

- Supervised learning:
  - Requires large amounts of labeled data
  - Labels are costly or unavailable in wireless systems (e.g., localization)
- Optimization-based learning:
  - Requires well-defined system objectives; often non-convex (local optimum)
  - Not always applicable (e.g., throughput vs Shannon rate)
- Desired capability:
  - Learn generalizable representations from data
  - Adapt efficiently to different downstream tasks
- Key question: Can we learn directly from the structure of wireless data?

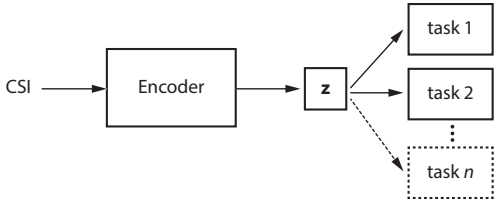
Learning signal derived from data itself – CSI is an omnipresent data source

Self-Supervised Learning Principle



- Learn representations without external labels
  - Define pretext tasks using inherent data structure
  - Learn invariances and meaningful features
- Learned representation can efficiently be adapted to various downstream tasks

Self-Supervised Learning Principle



- Learn representations without external labels
  - Define pretext tasks using inherent data structure
  - Learn invariances and meaningful features
- Learned representation can efficiently be adapted to various downstream tasks
- Once trained, the encoder is connected to task-specific lightweight heads
  $\Rightarrow$ Quick adaptation to previously unseen tasks

Self-Supervision for Wireless CSI: Constructing Views

- Wireless CSI provides multiple observations of the same channel:
  - Across antennas, subcarriers, and time
- View construction:
  - Subsampling and masking of CSI
  - Perturbations (noise, phase noise, Doppler)
  - Domain transformations (time-frequency vs delay-Doppler)
  - Multi-band CSI (sub-6 GHz to mmWaves)

Self-Supervision for Wireless CSI: Constructing Views

- Wireless CSI provides multiple observations of the same channel:
  - Across antennas, subcarriers, and time
- View construction:
  - Subsampling and masking of CSI
  - Perturbations (noise, phase noise, Doppler)
  - Domain transformations (time-frequency vs delay-Doppler)
  - Multi-band CSI (sub-6 GHz to mmWaves)
- Key idea:
  - Views share the same underlying propagation geometry

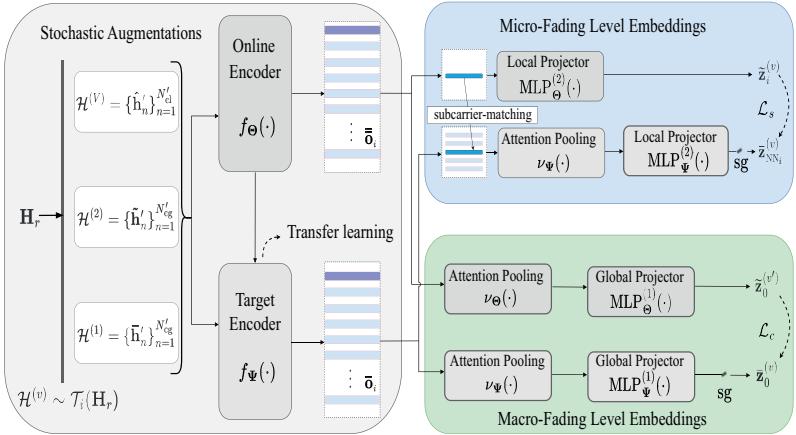
What is Learned from CSI?

- Learning objective:
  - Consistent views $\rightarrow$ similar representations
  - Different channels $\rightarrow$ distinguishable
- Learned representation:
  - Captures invariances across antennas, subcarriers, and time
  - Reflects structure induced by propagation

What is Learned from CSI?

- Learning objective:
  - Consistent views $\rightarrow$ similar representations
  - Different channels $\rightarrow$ distinguishable
- Learned representation:
  - Captures invariances across antennas, subcarriers, and time
  - Reflects structure induced by propagation
- Outcome:
  - Compact embedding adaptable to downstream tasks
- Perspective:
  - Physics-informed constraints could further guide learning

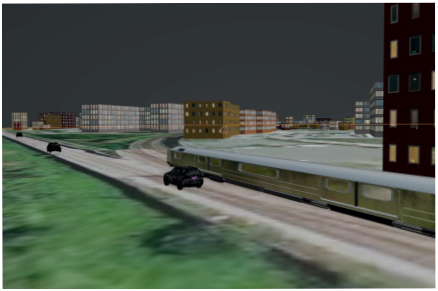
SSL for CSI-Fingerprinting based Wireless Localization



- Two SSL branches: one focusing on microscopic fading the other on macroscopic properties

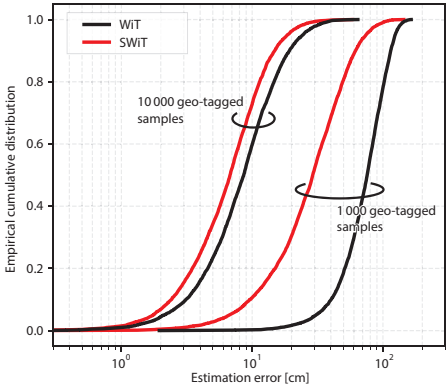
Self-Supervised and Invariant Representations for Wireless Localization, A. Salihu, S. Schwarz, et al, IEEE TWC 2024

CSI-Fingerprinting based Wireless Localization – Simulation Scenario



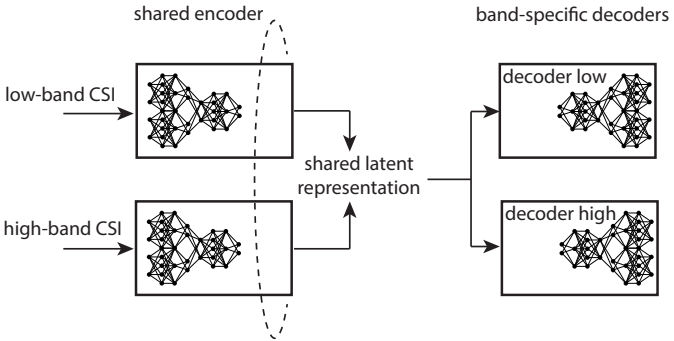
- Ray-tracing generated channel data for a rail road track in Vienna
- $N_a = 64$ antennas and $N_{\text{sc}} = 32$ subcarriers sampled over 20 MHz bandwidth
- Randomness is introduced through moving objects (vehicles, trains) and variations in material parameters

Localization Performance



- Self-supervised pre-training allows to reduce required number of geo-tagged samples
- WiT is a generalizable architecture that support various tasks – same as used for beamforming

Future Work: Towards Shared Representations Across Frequency Bands



- Learn a shared latent representation of multi-band CSI
  - Capture common geometric structure (LOS, dominant scatterers)
  - Decouple representation from specific frequency band

Summary and Conclusions

- Model-based and data-driven approaches are complementary in wireless PHY
- Local processing is insufficient: interaction modeling is key (self-attention)
- Optimization-based learning enables direct training from system objectives
- Self-supervised learning provides generalizable representations from CSI

WWTF Vienna Science and Technology Fund
FWF Österreichischer Wissenschaftsfonds

Learning to Solve Wireless PHY Problems: From Structured Models to Attention-Based Methods

Rhine Summit 2026, Session 2: AI4Wireless

Associate Prof. Stefan Schwarz
in collaboration with: Kaifeng Lu, Dr. Faruk Pasic, Dr. Artan Salihu and Prof. Markus Rupp

April 2026, stefan.schwarz@tuwien.ac.at

TU WIEN Technische Universität Wien
Institute of Telecommunications