The Road Towards an AI-Native Air Interface (AI-AI) for 6G

Jakob Hoydis
Radio Systems Research & AI
Nokia Bell Labs, France
jakob.hoydis@nokia-bell-labs.com
What will future communications look like in 2030?
Creating the ‘augmented human’

Augment our Intelligence
Augment our experience
Augment our control

Learn from/with machines
Automatic security
In-body monitoring

Multi-model mixed reality telepresence
High resolution mapping
Mixed reality co-design

Domestic robots
Remote & self driving
Drone/robot swarms

6G to enable a new lifestyle at scale
The enabling foundation for that future...

Six key technologies for 6G

- AI/ML Air-Interface
- New Spectrum Technologies
- Network as a sensor
- RAN-Core Convergence & Specialization
- Extreme Connectivity
- Security and Trust
Role of ML for 5G

Slice orchestration
VM management
Anomaly detection

Random Access detection
Symbol demapping
MIMO detection
Channel estimation
MIMO User Pairing

Deployment optimization
Load balancing
Carrier Aggregation
Handover prediction
Beam Prediction
Interference Management

User localization
Trajectory prediction

Complexity
Algorithms not optimal
Lack of accurate models

No component of 5G has been designed by ML

For personal use only.
What if 6G was built so that ML could optimize parts of the PHY & MAC if needed?
Possible benefits

- Custom signaling and access schemes
- Bespoke waveforms, constellations & pilots
- Optimally adapted to hardware limitations
- Reduced standardization effort
- No heuristic parameter settings
- Faster development & deployment time

AI - AI
AI-Native Air Interface (AI-AI) for 6G

The AI-AI optimally adapts to different environments, hardware, data, and applications.

"Post Shannon": Not about reliably transmitting bits anymore, but rather serving an application with data in an optimal way.
A roadmap to an AI-Native Air Interface for 6G

1. ML replaces/enhances individual processing blocks
2. ML replaces multiple processing blocks
3. ML designs part of the PHY itself
The importance of each step in the transition

1. **ML replaces/enhances individual processing blocks**
   - Paradigm change in transceiver design & deployment
   - Online training & transfer/federated learning
   - No new signaling

2. **ML replaces multiple processing blocks**
   - Enables qualitative new features (e.g., no cyclic prefix, less pilots)
   - HW acceleration essential
   - ML-first approach without backup solution

3. **ML designs part of the PHY itself**
   - Paradigm change in communication systems design
   - Distributed & end-to-end learning
   - New signaling & procedures
Differentiable algorithms enable hybrid “ML-Expert” systems

Training on end-to-end loss:
- No need for block-wise ground-truth
- Each block optimized for end-to-end performance

Fully differentiable transceivers allows simple integration of trainable components
A universal end-to-end design goal

\[
\begin{align*}
\begin{bmatrix} b_1 \\ \vdots \\ b_M \end{bmatrix} & \xrightarrow{\text{Transmitter } f_{\theta_T}} x \\
\xrightarrow{\text{Channel } p(y|x)} y & \xrightarrow{\text{Receiver } g_{\theta_R}} \begin{bmatrix} \tilde{p}(b_1|y) \\ \vdots \\ \tilde{p}(b_M|y) \end{bmatrix}
\end{align*}
\]

Binary cross-entropy

\[
L(\theta_T, \theta_R) = \sum_{m=1}^{M} -\mathbb{E}_{b_i,y}[\log_2(\tilde{p}_\theta(b_i|y))]
\]

Achievable rate with a mismatched bit-metric decoder

\[
= M - \left( \sum_{m=1}^{M} I(b_i; y) - \sum_{m=1}^{M} \mathbb{E}_y[D_{KL}(p(b_i|y)\|\tilde{p}(b_i|y))] \right)
\]

Number of transmitted bits

Bit-metric decoding rate (depends on transmitter)

Loss due to imperfect receiver (depends on receiver)

Minimizing the binary cross-entropy maximizes an achievable rate with a practical decoder
Understanding end-to-end learning on an AWGN channel

Constellation determines bit-metric decoding rate

Decision regions determine loss w.r.t. MAP

Constellation

First Bit Decision Region

Second Bit Decision Region

Third Bit Decision Region

Fourth Bit Decision Region

Achievable rate (Bit / complex channel use)

EsNo (dB)

Achievable rate (Bit / complex channel use)

EsNo (dB)

Autoencoder

Autoencoder MAP RX

16QAM Bit-wise MI

Shannon Capacity

Unachievable Region

Gain: -1.682 bit/ccu
Case study:
From Neural Receivers to Pilotless Transmissions
SISO doubly-selective channel

\[ Y = H \circ X + N \]
\[ \text{vec}(H) \sim \mathcal{CN}(0, R_F \otimes R_T) \]

Spectral correlation
- \([R_F]_{i,k} = \sum_{l=1}^{L} S_l e^{j2\pi l D_s \Delta F (i-k)}\)
- Subcarrier spacing \(\Delta F = 30 \text{ kHz}\)
- Delay spread \(D_s = 100 \text{ ns}\)
- TDL-A power delay profile

Temporal correlation
- \([R_T]_{i,k} = J_0 \left( 2\pi \frac{v}{c} f_c \Delta T (i-k) \right)\)
- Carrier frequency \(f_c = 3.5 \text{ GHz}\)
- Speed \(v = 50 \text{ km/h}\)

Modulation & Coding
- 64 QAM
- 5G code \(n=1024, r=2/3\)
Baseline receiver

- Least-squares channel estimation at pilot positions
- Equalization using the nearest pilot
- Exact LLR computation assuming a Gaussian post-equalized channel
- Textbook sum-product BP decoder with 40 iterations
Potential performance enhancements

![Graph showing BER vs. E_b/N_0 for Baseline and Perfect CSI]
Deficits of the baseline

- Imperfect channel estimation & channel aging lead to
  - Mismatched LLR computation
  - SNR degradation
Neural demapper (symbol-wise)

Channel Estimation → Equalization → Decoding

\[ \text{Re}\{\hat{X}_{i,j}\}, \text{Im}\{\hat{X}_{i,j}\}, SNR_{i,j}, \text{Pos. enc. } i, \text{Pos. enc. } j \rightarrow LLR_{i,j}^{(1)}, \cdots, LLR_{i,j}^{(6)} \]

Learns grid position-dependent statistics for better LLR computation
Neural demapper (grid-wise)

Channel Estimation → Equalization → Decoding

\[ \text{Re}\{\hat{X}\}, \text{Im}\{\hat{X}\}, \text{SNR} \] → \(72 \times 14 \times 3\) → \(LLR \) → \(72 \times 14 \times 6\)

Leverages pilots and data to compensate for channel aging and mismatched LLR computation.
Neural network architecture is key to success

Fully convolutional ResNet

Dilated separable convolutions


Each output value has a receptive field spanning the entire resource grid
Data-aided channel estimation, equalization, and demapping for unprecedented performance
End-to-end learning with Neural receiver

$$\mathcal{C} = \frac{1}{64} \sum_{c \in \mathcal{C}} c$$

64\times2 trainable weights

Zero-mean unit energy constellation

End-to-end learning enables pilotless transmissions without performance loss
Learned constellation for pilotless communication

F. Ait Aoudia, J. Hoydis, “End-to-end Learning for OFDM: From Neural Receivers to Pilotless Communication”, arXiv2009.05261

How could this be standardized?
Important research topics for end-to-end learning

• New waveforms for new spectrum
• Learning for systems with (extreme) hardware constraints
• Joined communications + X
• Signals conveying a few bits of information
• Application-specific end-to-end learning
• Semantic communications
• Decentralized & federated learning
• Transfer & meta learning
The next frontier:
Protocol learning
Emerging a RAN protocol

3GPP Way

ML Way


Can we learn a MAC protocol?
Thank you!