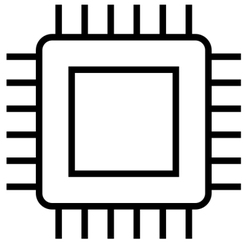
# RECENT PROGRESS IN END-TO-END LEARNING FOR THE PHYSICAL LAYER

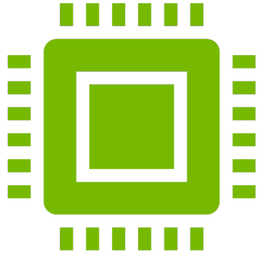
JAKOB HOYDIS, PRINCIPAL RESEARCH SCIENTIST, DECEMBER 01, 2021

# A 6G VISION

General Purpose Processors

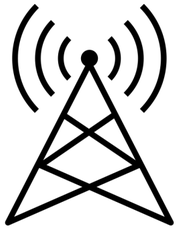


Beyond Moore's law



Domain-Specific HW Accelerators

5G/Wi-Fi



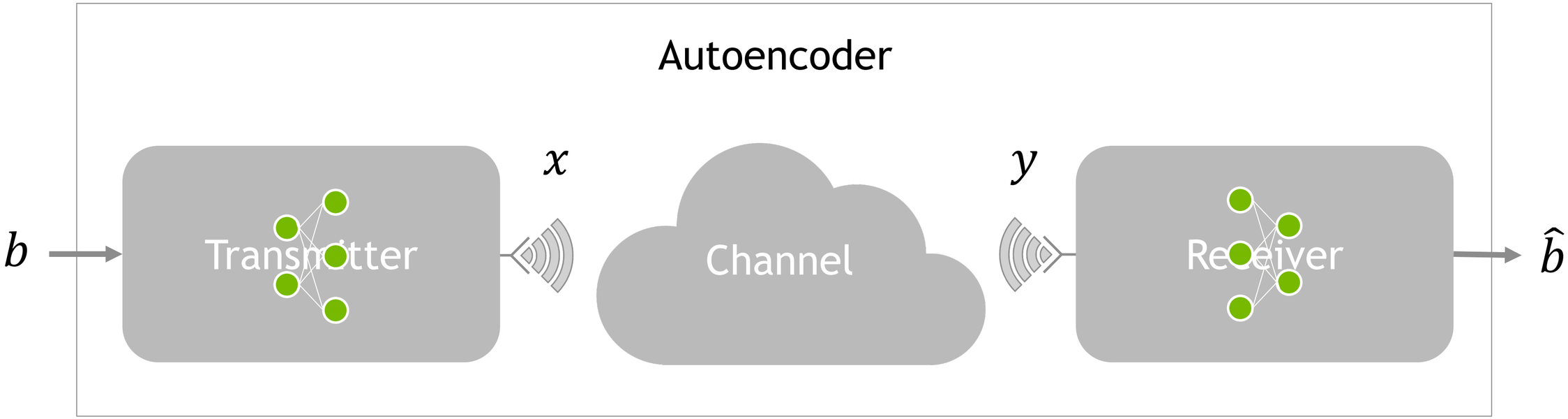
Beyond Cooper's law

6G

6G networks should be able autonomously specialize to a specific radio environment and application

*Key enablers: Software-defined RAN & Machine Learning*

# COMMUNICATION SEEN AS AN AUTOENCODER




- Channel is a penality layer (noise, distortions, limited power)
- Jointly optimize transmitter and receiver from “end-to-end”
- Universal concept that applies to any channel model & loss function

https://arxiv.org/abs/1702.00832

# WHERE ARE WE TODAY?

The entire physical layer can be learned!

- ✓ New codes achieving state-of-the-art performance
  https://arxiv.org/abs/2108.12920
- ✓ New modulation and piloting schemes
  https://arxiv.org/abs/2009.05261
- ✓ New hardware-friendly & spectrally efficient waveforms
  https://arxiv.org/abs/2109.00998
- ✓ Neural receivers outperforming traditional methods
  https://arxiv.org/abs/2005.01494
- ❑ Not much success for multi-antenna communications yet

**Computer Science > Information Theory**

*[Submitted on 2 Sep 2021]*

# Waveform Learning for Next-Generation Wireless Communication Systems

Fayçal Ait Aoudia, Jakob Hoydis

We propose a learning-based method for the joint design of a transmit and receive filter, the constellation geometry and associated bit labeling, as well as a neural network (NN)-based detector. The average power ratio (PAPR). This allows control of the tradeoff between spectral containment, peak power, and communication rate. Evaluation on an additive white Gaussian noise (AWGN) channel (RRC), without significant loss of information rate. When considering a 3rd Generation Partnership Project (3GPP) multipath channel, the learned waveform and neural receiver enable competitive or method incurs no additional complexity on the transmitter side and might be an attractive tool for waveform design of beyond-5G systems.

Subjects: **Information Theory (cs.IT)**; Machine Learning (cs.LG); Signal Processing (eess.SP)
Cite as: arXiv:2109.00998 **[cs.IT]**
(or arXiv:2109.00998v1 **[cs.IT]** for this version)

https://arxiv.org/abs/2109.00998

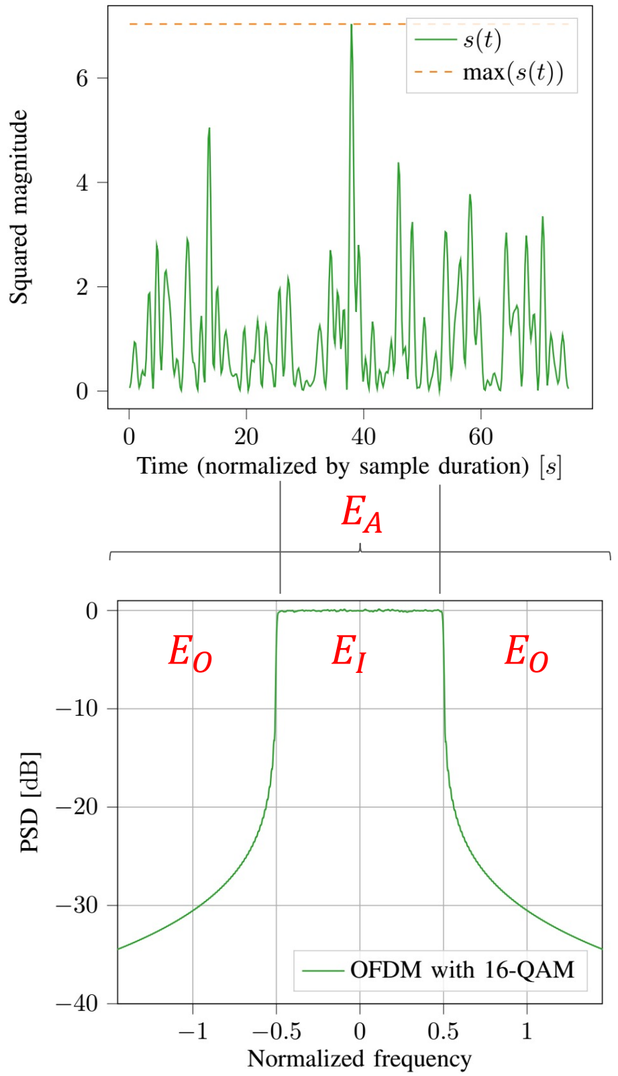
# MOTIVATION FOR WAVEFORM LEARNING

- PAPR (peak to average power ratio)

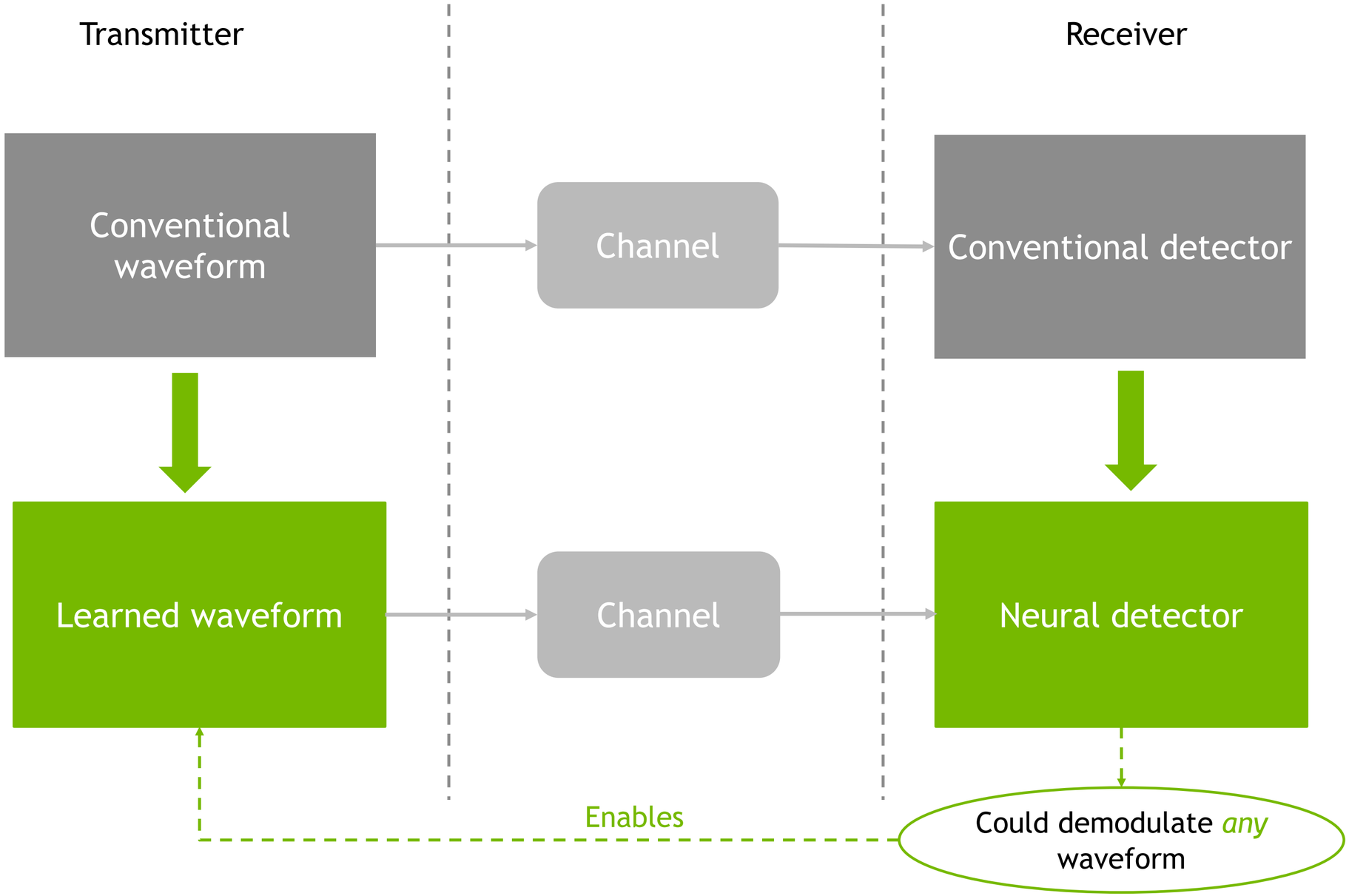
$$\text{PAPR}_\epsilon = \min e, \quad \text{s. t.} \quad P\left(\frac{|s(t)|^2}{\mathbb{E}\left[|s(t)|^2\right]} > e\right) \leq \epsilon$$

- ACLR (adjacent channel leakage ratio)

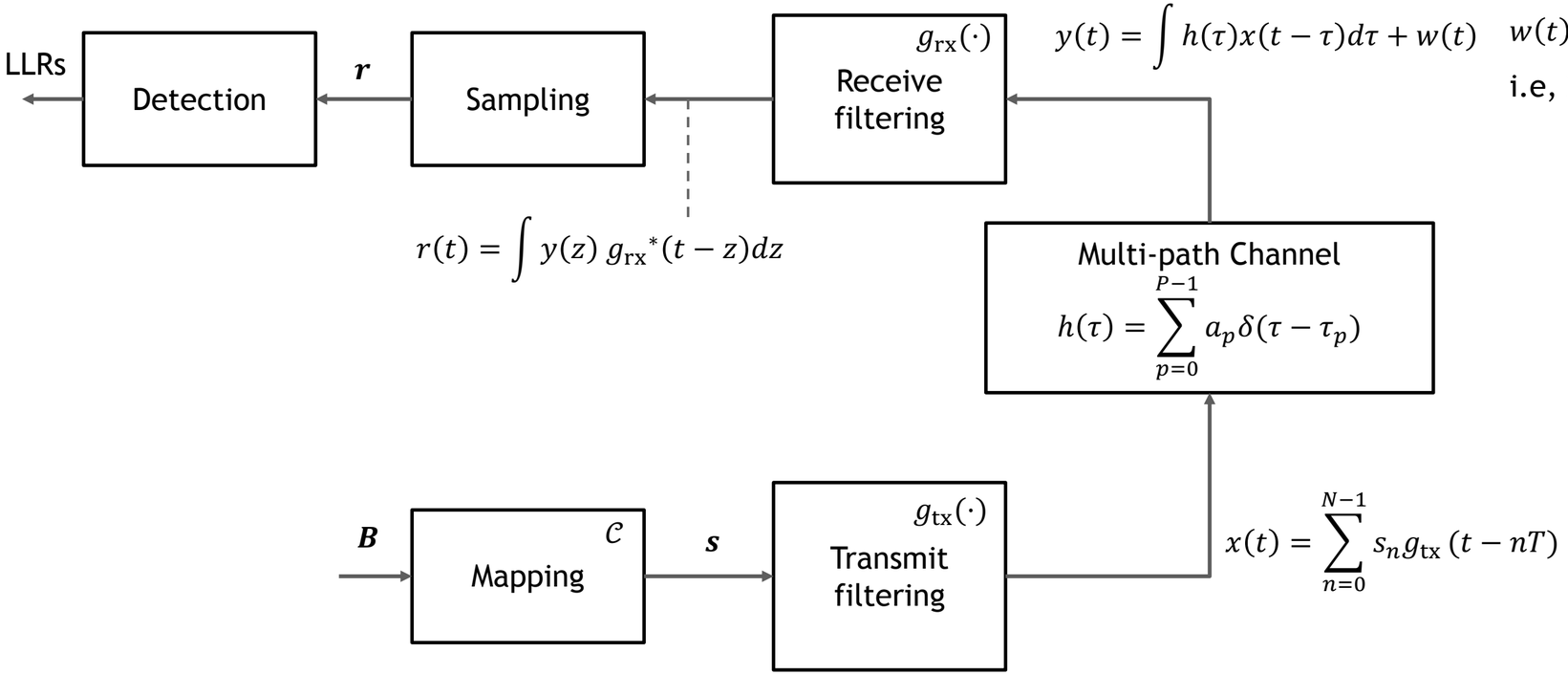
$$\text{ACLR} = \frac{\mathbb{E}_{\mathbf{x}}\left[E_O\right]}{\mathbb{E}_{\mathbf{x}}\left[E_I\right]} = \frac{\mathbb{E}_{\mathbf{x}}\left[E_A\right]}{\mathbb{E}_{\mathbf{x}}\left[E_I\right]} - 1$$



# MOTIVATION FOR WAVEFORM LEARNING II



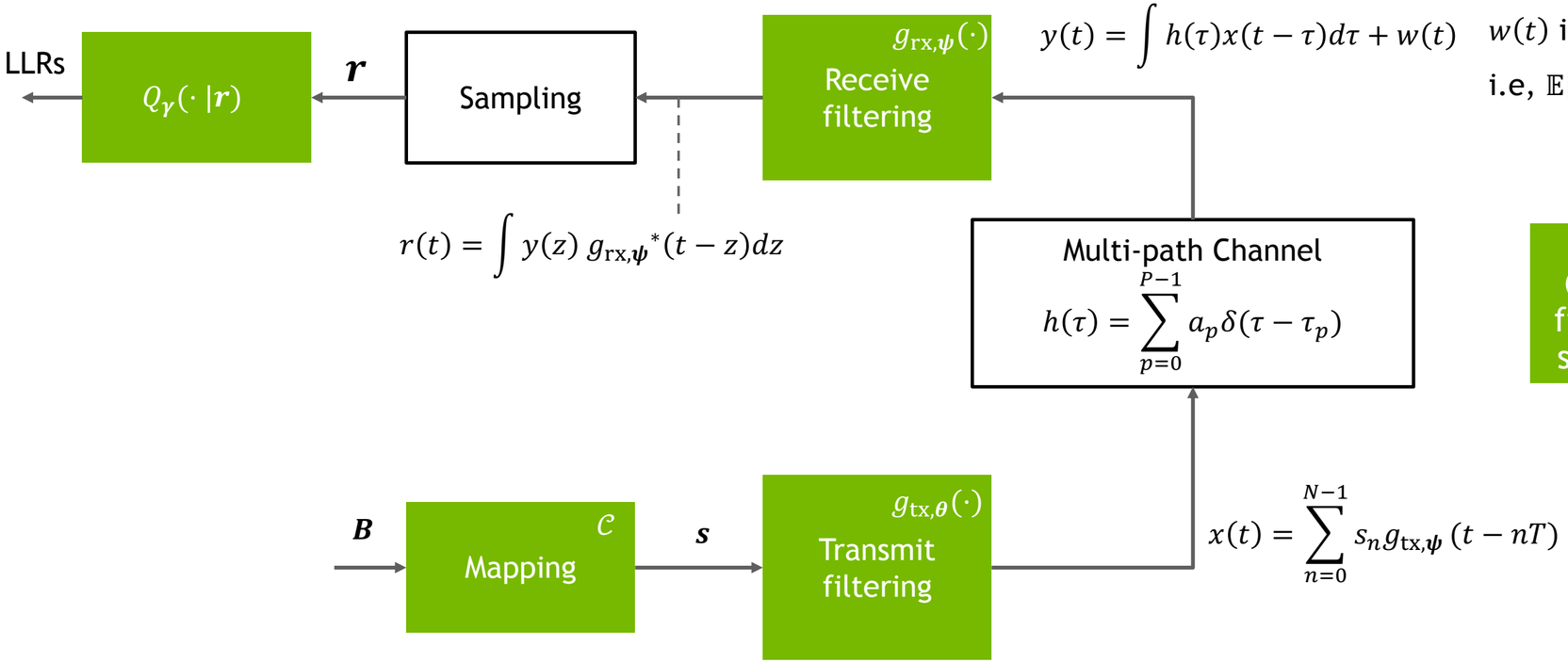
# SINGLE-CARRIER SETUP




$w(t)$ is an uncorrelated Gaussian noise,

i.e, $\mathbb{E}[w(t)w(t+\tau)^{*}] = N_0\delta(\tau)$

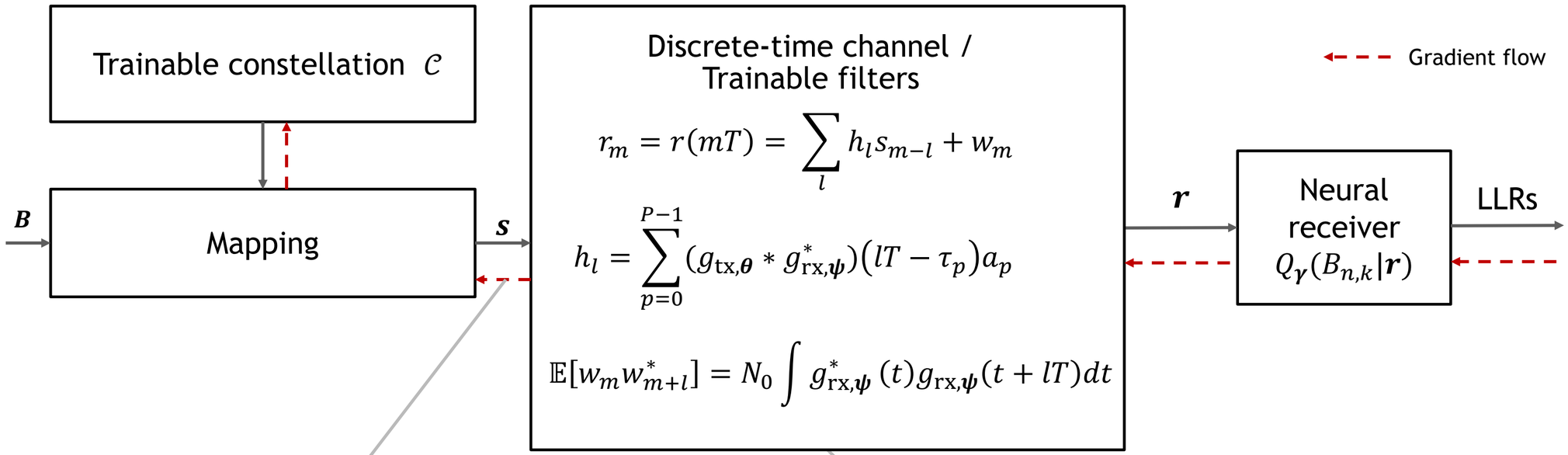
# SINGLE-CARRIER SETUP




$w(t)$ is an uncorrelated Gaussian noise,
i.e, $\mathbb{E}[w(t)w(t+\tau)^{*}] = N_0\delta(\tau)$

**Key idea:**
Optimize the transmit and receive filters jointly with the constellation shaping, bit labelling, and detector

# END-TO-END WAVEFORM LEARNING




Constraining the waveform is challenging:

- Average unit energy
- ACLR
- PAPR

Simulating the channel is challenging:

- Evaluation of the channel taps and noise correlation require integration.

# IMPLEMENTATION OF TRANSMIT AND RECEIVE FILTERS

To accurately and efficiently implement trainable transmit and receive filters, we use the functions $\{sinc(Df - s)\}_{s\in\mathbb{Z}}$, which form a basis in the frequency domain for (practical) functions time-limited to $\left(-\frac{D}{2}, \frac{D}{2}\right)$.

$$\hat{g}_{\mathrm{tx},\boldsymbol{\theta}}(f) = \sqrt{C(\boldsymbol{\theta})} \sum_{s=-S}^{S} \theta_s \mathrm{sinc}\,(Df - s) \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_S \\ \vdots \\ \theta_{-S} \end{bmatrix}$$

$$\hat{g}_{\mathrm{rx},\boldsymbol{\psi}}(f) = \sum_{s=-S}^{S} \psi_s \mathrm{sinc}\,(Df - s) \qquad \boldsymbol{\psi} = \begin{bmatrix} \psi_S \\ \vdots \\ \psi_{-S} \end{bmatrix}$$

$C(\boldsymbol{\theta})$ is a normalization constant that ensures the transmit filter has unit energy:

$$C(\boldsymbol{\theta}) = \frac{1}{\int_{-D/2}^{D/2} \left| g_{\mathrm{tx},\boldsymbol{\theta}}(t) \right|^2 dt}$$

This leads to the time-domain expressions

$$g_{\mathrm{tx},\boldsymbol{\theta}}(t) = \frac{\sqrt{C(\boldsymbol{\theta})}}{D} \mathrm{rect}\left(\frac{t}{D}\right) \sum_{s=-S}^{S} \theta_s e^{\frac{j2\pi s}{D}t}$$

$$g_{\mathrm{rx},\boldsymbol{\psi}}(t) = \frac{1}{D} \mathrm{rect}\left(\frac{t}{D}\right) \sum_{s=-S}^{S} \psi_s e^{\frac{j2\pi s}{D}t}$$

This implementation of the transmit and receive filters enables exact and low complexity computation of $C(\boldsymbol{\theta})$ and of the

- channel taps
- noise covariance
- ACLR

*Details available in our paper.*

# IMPLEMENTATION OF THE NEURAL RECEIVER

The neural receiver is implemented by a residual convolutional neural network


Residual block


Neural receiver for the AWGN channel

# LOSS FUNCTION

Our goal is to optimize the set of trainable parameters $(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma})$ to <u>maximize the BMD rate</u>

Mutual information between bit $B_{n,k}$ and received samples $\boldsymbol{r}$

True posterior distribution on $B_{n,k}$ given $\boldsymbol{r}$

Posterior distribution on $B_{n,k}$ given $\boldsymbol{r}$ approximated by the neural receiver

$$R(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} I(B_{n,k}; \boldsymbol{r} | \boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}) - \mathbb{E}_{\boldsymbol{r}}\left[\mathrm{D}_{\mathrm{KL}}\left(P\left(B_{n,k}|\boldsymbol{r}\right)||\mathrm{Q}_{\boldsymbol{\gamma}}(B_{n,k}|\boldsymbol{r})\right)\right]$$

Achievable information rate assuming an ideal receiver

Rate loss due to mismatched receiver

By training on the BMD rate, optimization of the constellation geometry is performed while considering the labelling of each point. This leads to the joint optimization of the bit labelling and constellation geometry.

# LOSS FUNCTION

A practically useful observation is that the BMD rate $R$ is closely related to the binary cross-entropy:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma}) = K - R(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma})$$

where

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma}) \coloneqq -\frac{1}{N}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\mathbb{E}\left[\log_2\left(Q_{\boldsymbol{\gamma}}\left(B_{n,k}|\boldsymbol{r}\right)\right)|\mathcal{C}, \boldsymbol{\theta}, \boldsymbol{\psi}\right] \approx -\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\log_2\left(Q_{\boldsymbol{\gamma}}\left(B_{n,k}^{[m]}|\boldsymbol{r}^{[m]}\right)\right)$$

Maximizing $R$ is equivalent to minimizing $\mathcal{L}$

$R(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathcal{C}, \boldsymbol{\gamma})$ is known to be an achievable rate for practical systems that rely on BICM and BMD detection [1,2]

[2] G. Böcherer, “Achievable Rates for Probabilistic Shaping,” preprint arXiv:1707.01134, 2017.

[3] F. Ait Aoudia and J. Hoydis, "End-to-end Learning for OFDM: From Neural Receivers to Pilotless Communication," in IEEE Transactions on Wireless Communications.

# PROBLEM STATEMENT

$$\min_{\boldsymbol{\theta},\mathcal{C},\boldsymbol{\psi},\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\theta},\boldsymbol{\psi},\mathcal{C},\boldsymbol{\gamma})$$

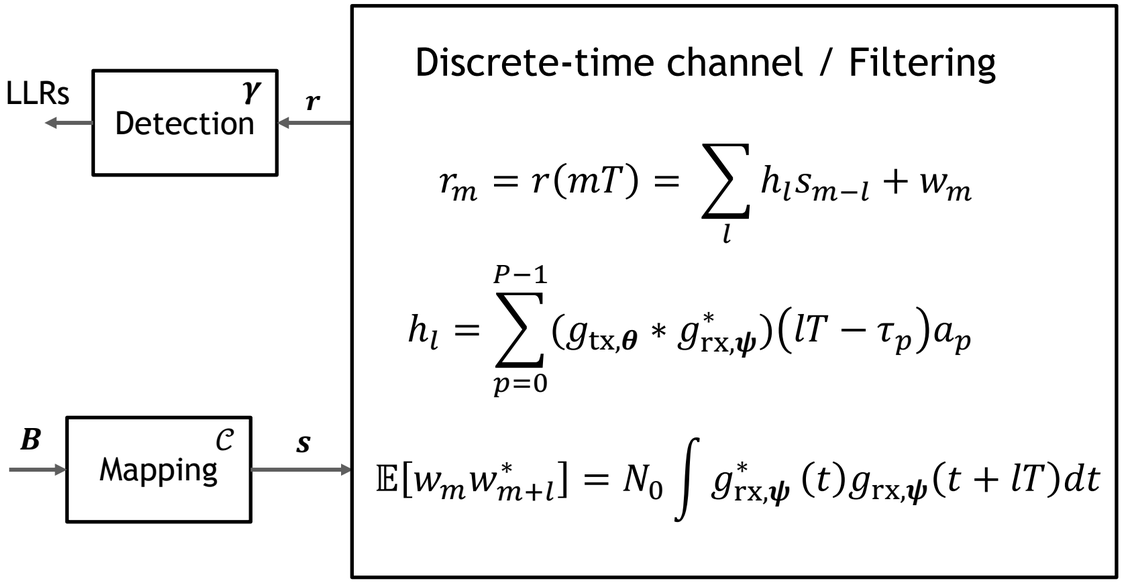
subject to

$$\int_{-D/2}^{D/2} \left|g_{\mathrm{tx},\boldsymbol{\theta}}(t)\right|^2 dt = 1$$

$$\mathbb{E}_{c\sim\mathcal{C}}[|c|^2] = 1$$

$$\mathbb{E}_{t\sim\mathcal{U}\left(-\frac{T}{2},\frac{T}{2}\right)}\left[max\left(\frac{|x(t)|^2}{\overline{|x|^2}} - \epsilon_V, 0\right)\right] = 0$$

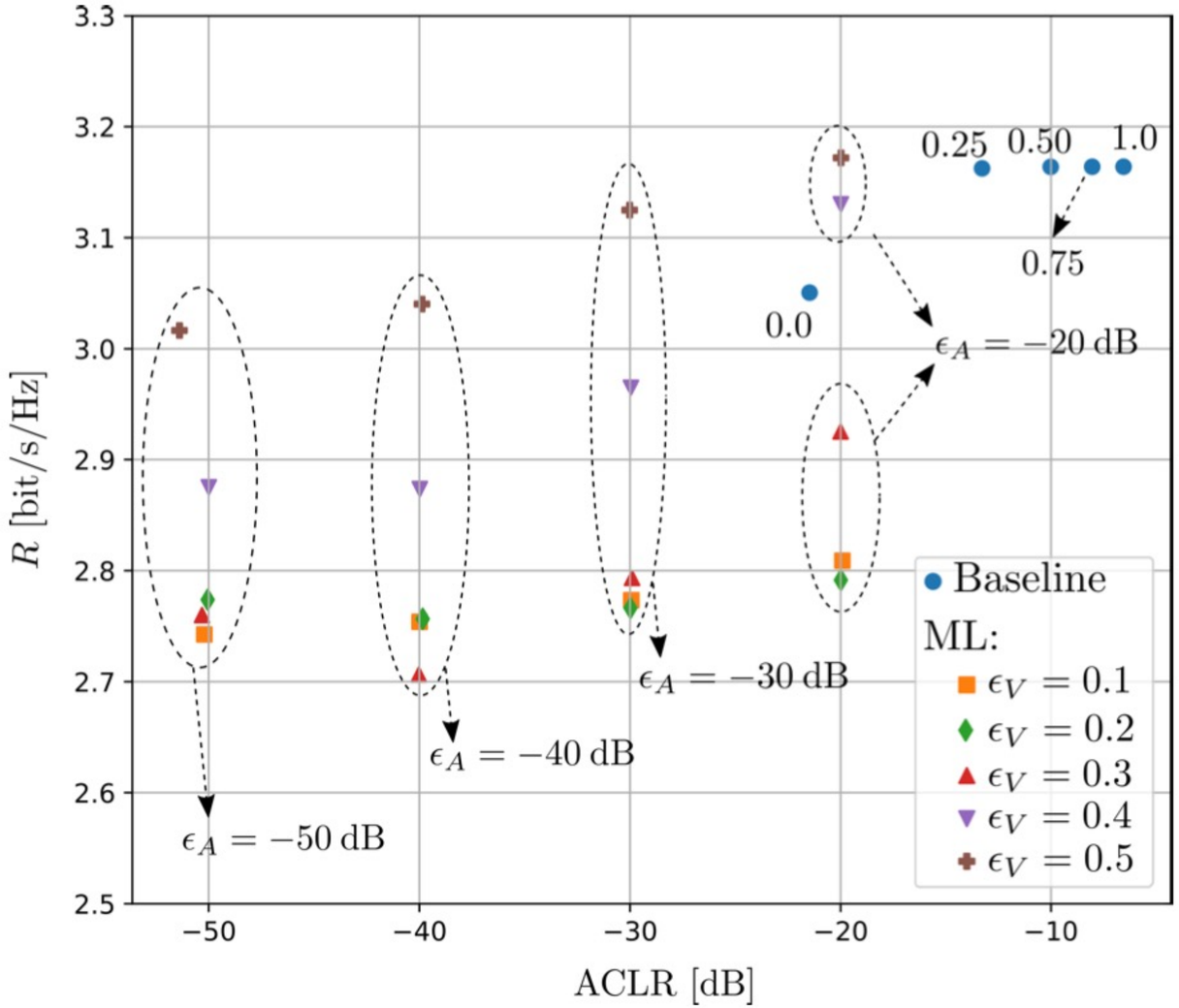
$$\mathrm{ACLR}(\boldsymbol{\theta}) \leq \epsilon_A$$




Training is performed using an augmented Lagrangian based approach.

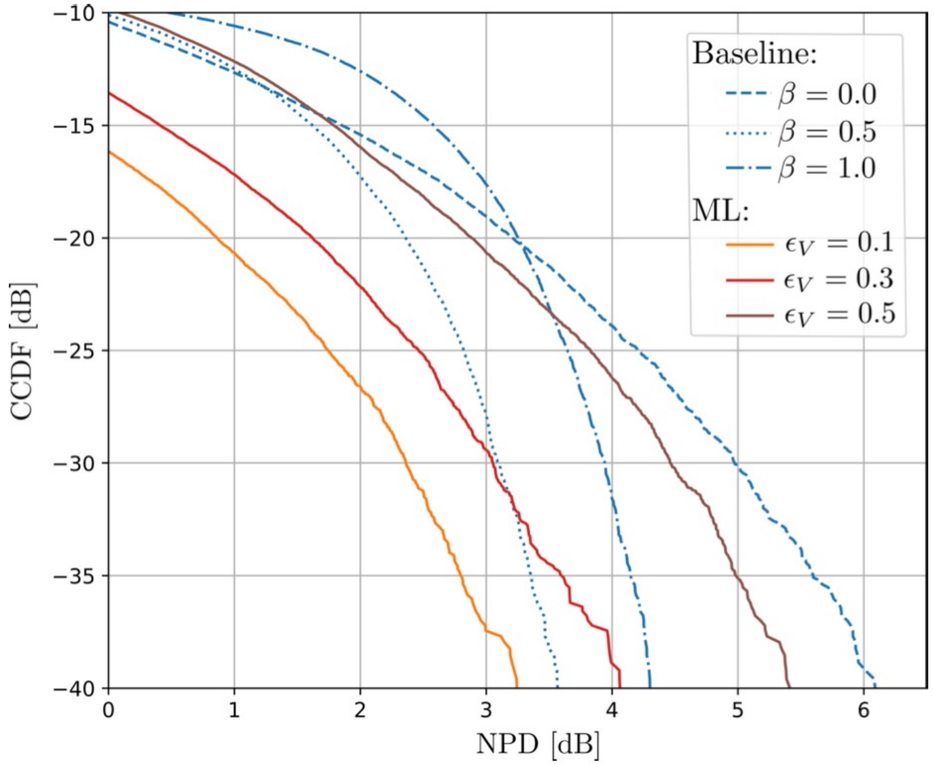
*Details available in our paper.*

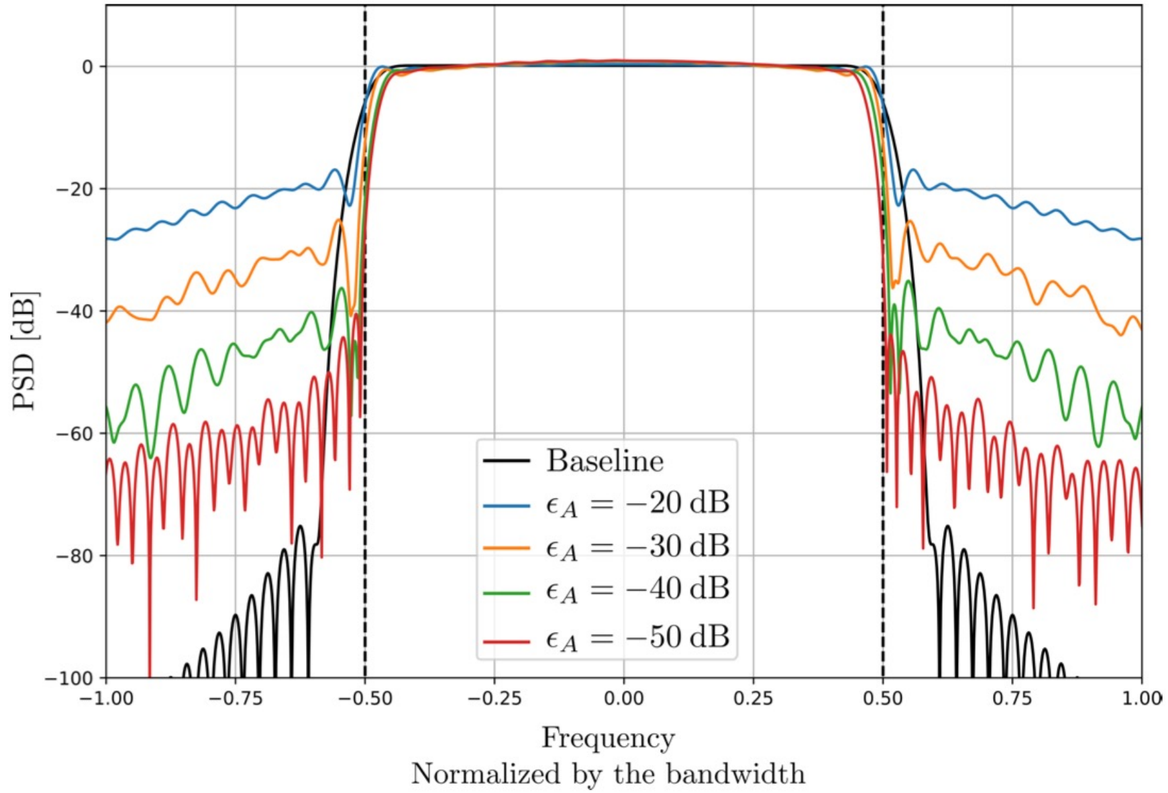
# RESULTS FOR AWGN CHANNEL



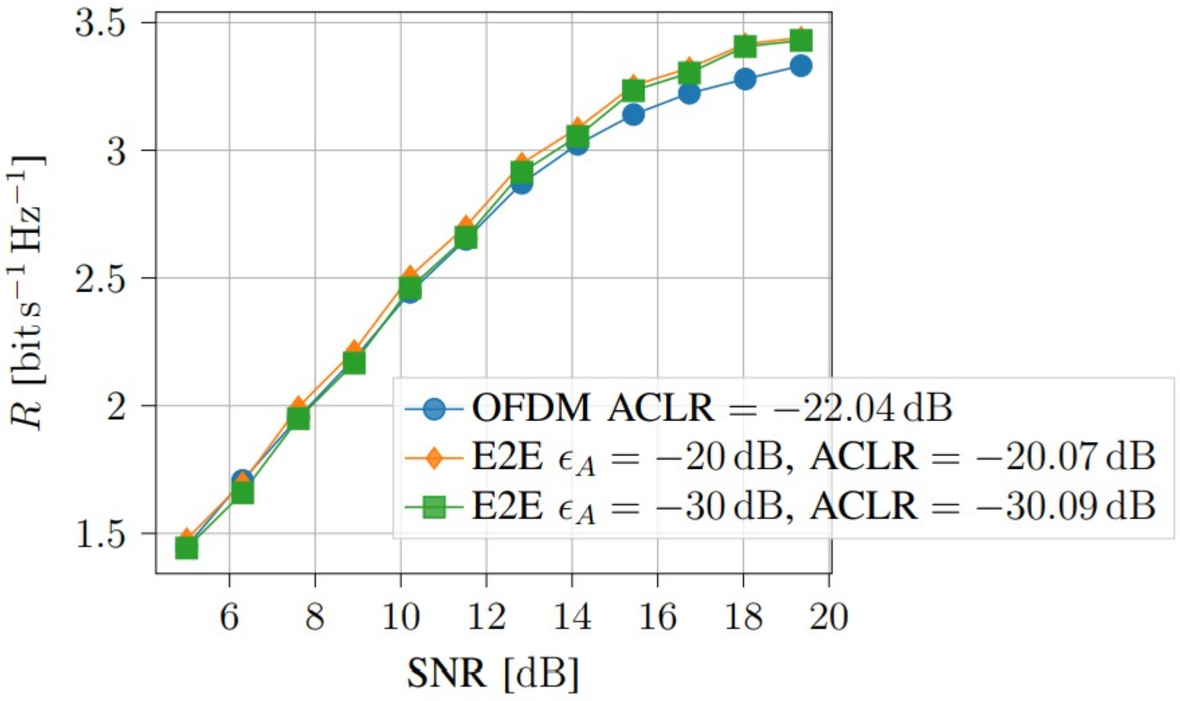

- Baseline:
  - 16-QAM
  - RRC filter with Blackman window
  - Roll-off factors $\beta = \{0.0, 0.25, 0.50, 0.75, 1.0\}$
- E2E Approach:
  - $\epsilon_V \in \{0.1,0.2,0.3,0.4,0.5\}$
  - $\epsilon_A \in \{-50, -40, -30, -20\}$dB
- Observations:
  - Accurate ACLR control possible!
  - PAPR constraint incurs the strongest impact on rate

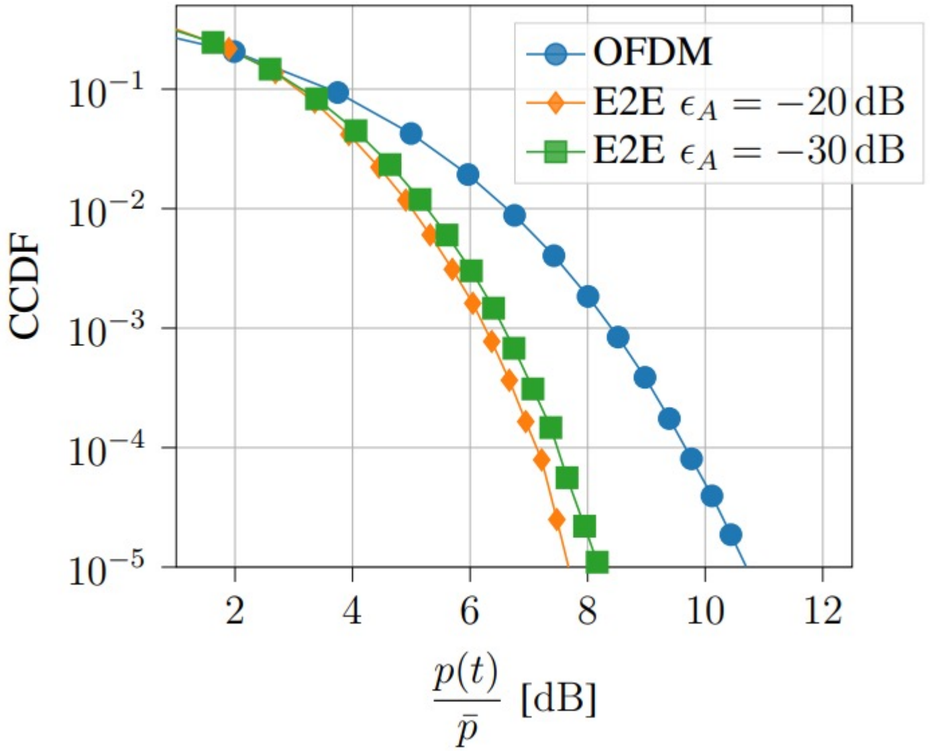
# RESULTS FOR AWGN CHANNEL II



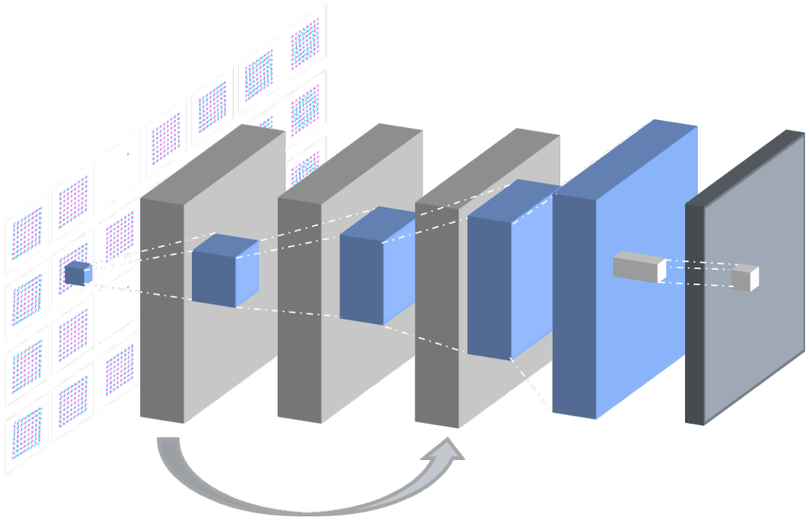
# CAN WE BEAT OFDM WITH CYCLIC PREFIX?



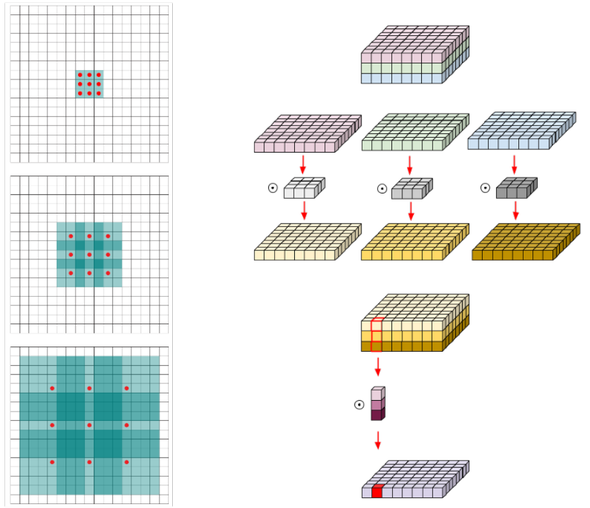




- 3GPP 38.901 UMi line-of-sight (LoS) and non-line-of-sight (NLoS) channel models are considered
- Training and evaluation were performed on two separate datasets
- Baseline is Gray labelled 16QAM with OFDM waveform, with 166 subcarriers, 30 kHz subcarrier spacing, and a 5G NR pilot pattern

# THE NEED FOR INTERDISCIPLINARY RESEARCH



Fully convolutional residual networks

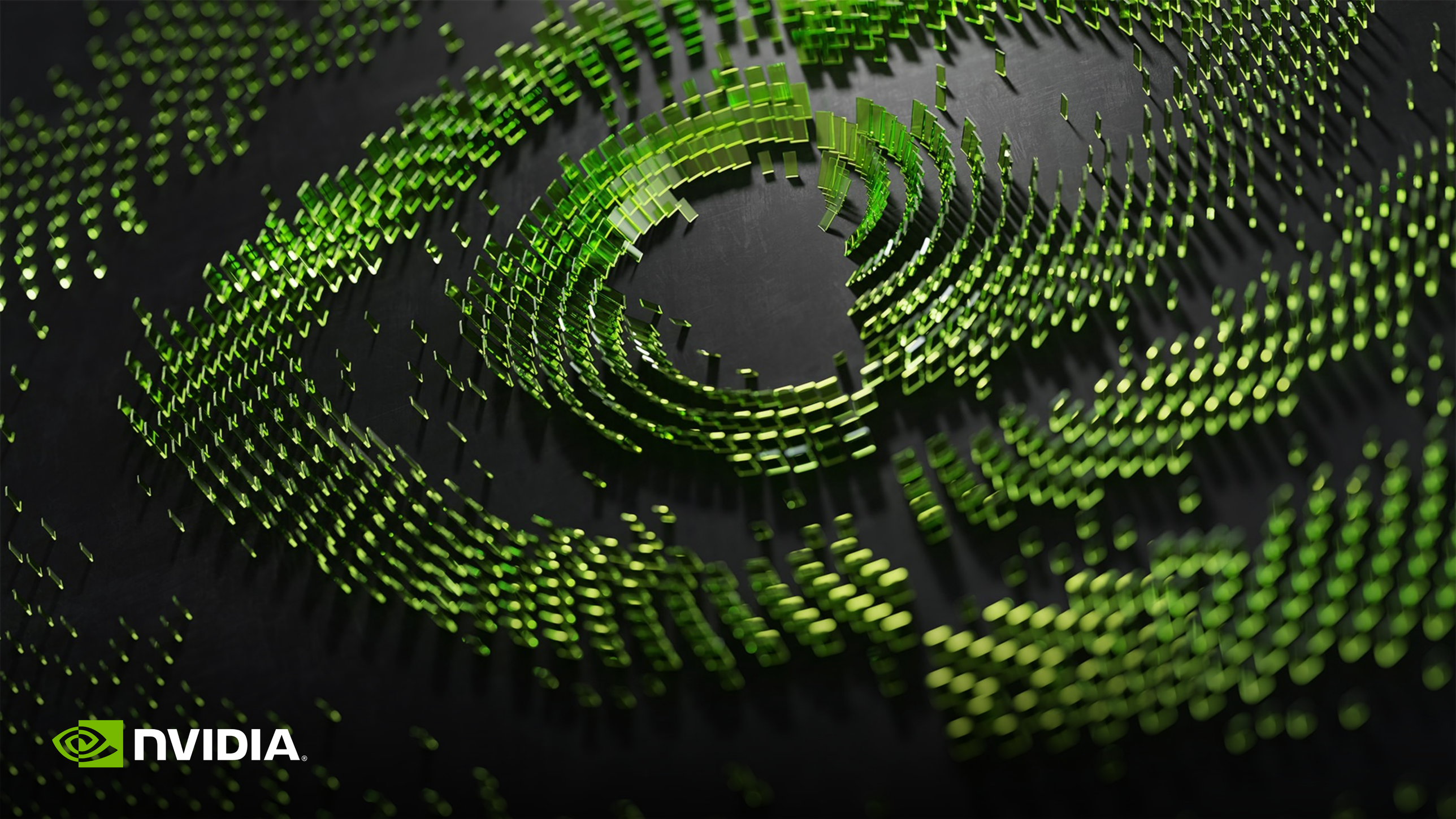


Dilated separable convolutions

- Finding the right neural network architectures is key for:
  - Optimal performance
  - Low complexity and energy consumption
- Unprecedented throughput & latency requirements

# FUTURE CHALLENGES AND DIRECTIONS

- Paradigm change in communication systems design
  - Is research mature enough & industry ready?
- Energy efficiency
  - Are fully learned solutions competitive?
- Standardization
  - Learning outcome or procedures & signaling?
- Trust & Reliability
  - How do we test & trouble-shoot fully learned systems?
- New applications and use-cases:
  - THz frequencies
  - Joint communication & sensing


NVIDIA