



# End-to-end Optimization of Optical Communication Systems using Deep Learning

Laurent Schmalen, Boris Karanov, Vincent Lauinger



- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Recurrent Neural Networks
- Application 2: In-Situ Optimization of Transmitters
- Application 3: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

## ■ Introduction

- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Recurrent Neural Networks
- Application 2: In-Situ Optimization of Transmitters
- Application 3: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

# A Few Words About Myself



- Grew up in Luxembourg



- M.Sc. (Dipl.-Ing.) and PhD in electrical engineering and information technology from RWTH Aachen University in multimedia communications over heterogeneous networks



- 8 years of industry work at Bell Labs conducting fundamental research on error correction coding, fiber optical communication and machine learning for communications



- Contributed to transceiver ASICs with throughputs of 100 GBit/s to 600 GBit/s



- As of April 1st 2019, full professor at KIT and head of communications engineering lab

# Machine Learning Success: Handwritten Digit Recognition

## ■ The MNIST database<sup>1 2</sup>



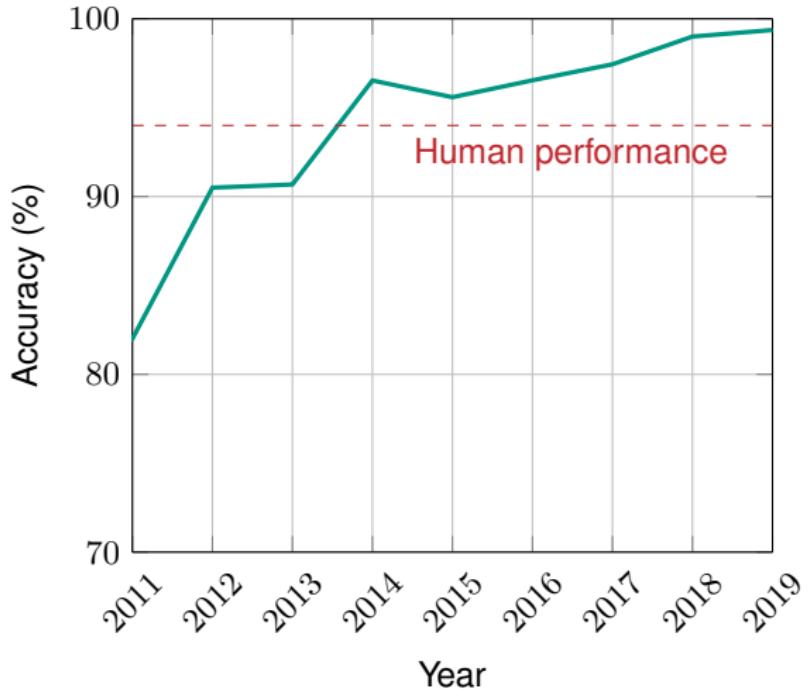
0  
1  
2  
3  
4  
5  
6  
7  
8  
9 9

## ■ Error rates of 0.23% can be achieved today

<sup>1</sup> Available online at: <http://yann.lecun.com/exdb/mnist/>

<sup>2</sup> Image source: <https://upload.wikimedia.org/wikipedia/commons/2/27/MnistExamples.png>

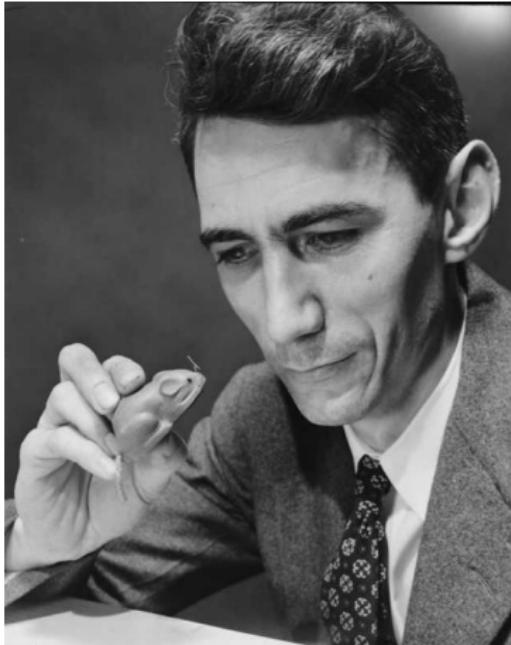
# Machine Learning Success: Image Recognition



- Image recognition accuracy on CIFAR-10 dataset<sup>3</sup>
- In less than 10 years, advances in machine learning and in particular **neural networks** and **deep learning** yield systems that surpass human capabilities
- Almost perfect recognition accuracy

<sup>3</sup>Accuracy data from <https://benchmarks.ai/cifar-10>

# Machine Learning for Communications

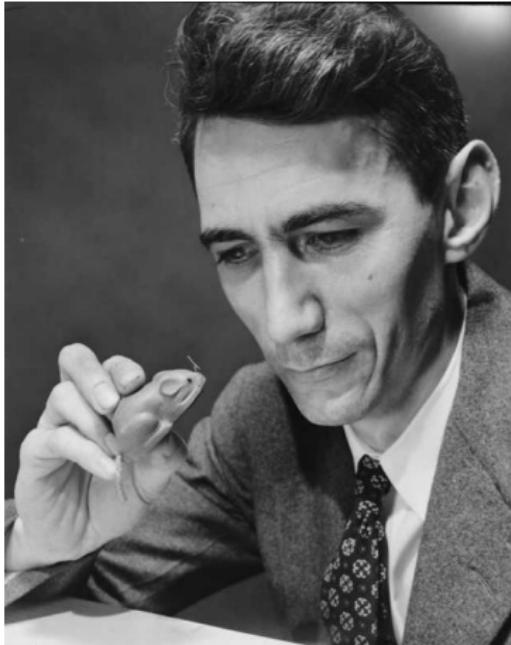


- 70 years of communications engineering have brought us close to the Shannon limit<sup>4</sup>(Shannon capacity)
- **But:** this applies only to a sub-class of all communication problems
- We don't know neither good nor optimal transceivers for many communication systems

---

<sup>4</sup>Image source: [https://commons.wikimedia.org/wiki/File:Claude\\_Shannon\\_1776.jpg](https://commons.wikimedia.org/wiki/File:Claude_Shannon_1776.jpg)

# Machine Learning for Communications



- 70 years of communications engineering have brought us close to the Shannon limit<sup>4</sup> (Shannon capacity)
- **But:** this applies only to a sub-class of all communication problems
- We don't know neither good nor optimal transceivers for many communication systems
- **Machine learning** can help us identify novel transceivers
- **This talk:**
  - Concept of using neural networks for optimizing of optical communication systems
  - Use of neural networks to learn parameters of the communication system

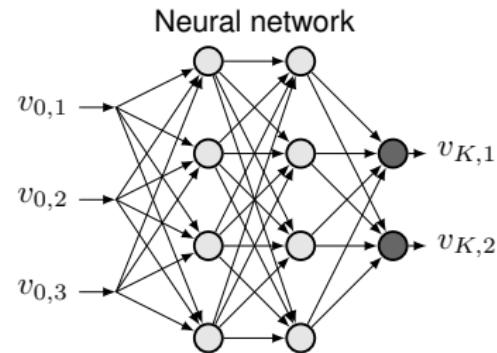
<sup>4</sup>Image source: [https://commons.wikimedia.org/wiki/File:Claude\\_Shannon\\_1776.jpg](https://commons.wikimedia.org/wiki/File:Claude_Shannon_1776.jpg)

# Neural Networks: A Whirlwind Tour

## Feed-Forward Neural Networks (FFNN)

- Maps an input vector  $\mathbf{v}_0 = (v_{0,1} \dots v_{0,M})$  to an output vector  $\mathbf{v}_K = (v_{K,1} \dots v_{K,n}) = f_{\text{NN}}(\mathbf{v}_0)$
- The neural network (NN) is a composed function consisting of layers, where each layer computes

$$\mathbf{v}_k = g_{\text{NL},k}(\mathbf{W}_k \mathbf{v}_{k-1} + \mathbf{b}_k), \quad k = 1, \dots, K$$



# Neural Networks: A Whirlwind Tour

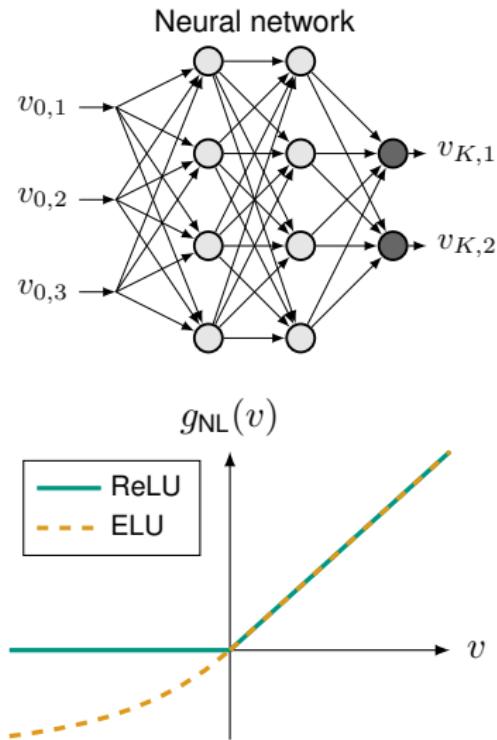
## Feed-Forward Neural Networks (FFNN)

- Maps an input vector  $\mathbf{v}_0 = (v_{0,1} \dots v_{0,M})$  to an output vector  $\mathbf{v}_K = (v_{K,1} \dots v_{K,n}) = f_{\text{NN}}(\mathbf{v}_0)$
- The neural network (NN) is a composed function consisting of layers, where each layer computes

$$\mathbf{v}_k = g_{\text{NL},k}(\mathbf{W}_k \mathbf{v}_{k-1} + \mathbf{b}_k), \quad k = 1, \dots, K$$

- Activation function  $g_{\text{NL}}$  introduces **nonlinear** relation between layers
- A popular choice for  $g_{\text{NL}}$  is the ReLU activation function (or one of its variants, e.g. the ELU function)

$$\mathbf{x} = g_{\text{NL}}(\mathbf{v}) = g_{\text{ReLU}}(\mathbf{v}) \quad \text{with } x_i = \max(0, v_i)$$



# Neural Networks: A Whirlwind Tour

## Feed-Forward Neural Networks (FFNN)

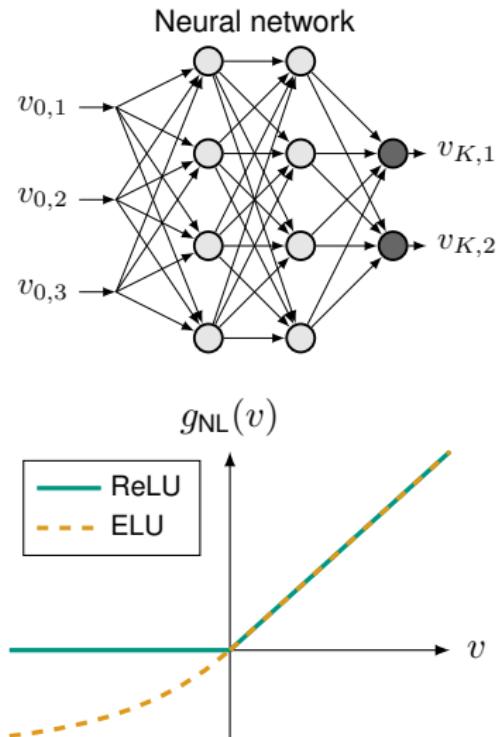
- Maps an input vector  $\mathbf{v}_0 = (v_{0,1} \dots v_{0,M})$  to an output vector  $\mathbf{v}_K = (v_{K,1} \dots v_{K,n}) = f_{\text{NN}}(\mathbf{v}_0)$
- The neural network (NN) is a composed function consisting of layers, where each layer computes

$$\mathbf{v}_k = g_{\text{NL},k}(\mathbf{W}_k \mathbf{v}_{k-1} + \mathbf{b}_k), \quad k = 1, \dots, K$$

- Activation function  $g_{\text{NL}}$  introduces **nonlinear** relation between layers
- A popular choice for  $g_{\text{NL}}$  is the ReLU activation function (or one of its variants, e.g. the ELU function)

$$\mathbf{x} = g_{\text{NL}}(\mathbf{v}) = g_{\text{ReLU}}(\mathbf{v}) \quad \text{with } x_i = \max(0, v_i)$$

- Objective of learning:** Find parameters  $\mathbf{W}_k$  and  $\mathbf{b}_k$  such that  $f_{\text{NN}}(\mathbf{v}_0)$  approximates an (unknown) function whose inputs and outputs can be observed and fed to a training algorithm



# Neural Networks: Computational Graph

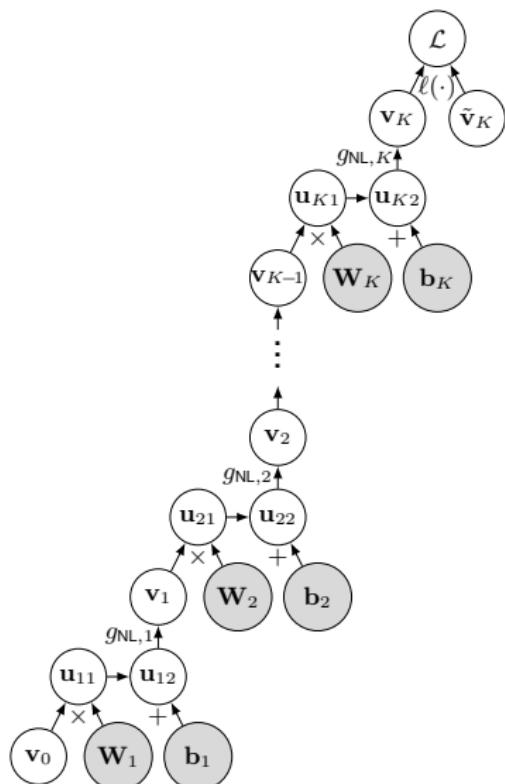
## Deep Learning

- Objective is to find NN parameters  $\{\mathbf{W}_k, \mathbf{b}_k\}$  that minimize a **loss function**

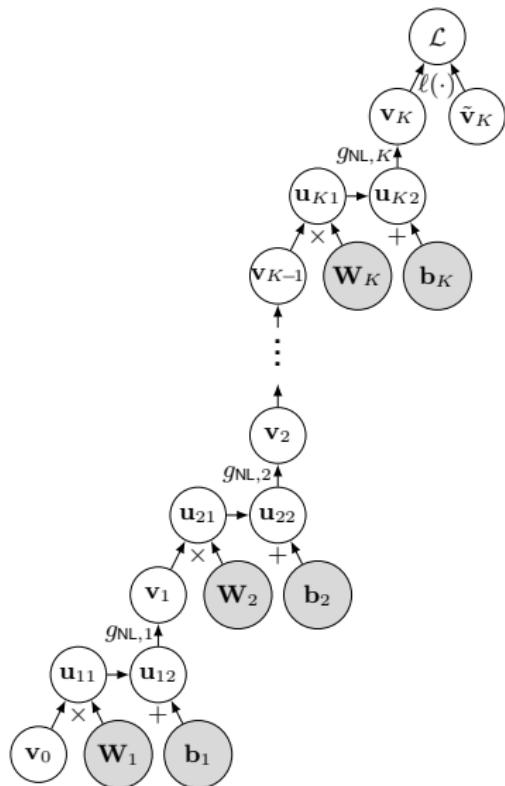
$$\mathcal{L}(\{\mathbf{W}_k, \mathbf{b}_k\}_K) = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{v}_{0,i}, \tilde{\mathbf{v}}_{K,i}) \in \mathcal{S}} L(f_{\text{NN}}(\mathbf{v}_{0,i}), \tilde{\mathbf{v}}_{K,i})$$

where  $(\mathbf{v}_{0,i}, \tilde{\mathbf{v}}_{K,i})$  are **examples** of inputs  $(\mathbf{v}_{0,i})$  and observed outputs  $(\tilde{\mathbf{v}}_{K,i})$  of the system we like to approximate

- Training data set  $\mathcal{S}$  contains these examples



# Neural Networks: Computational Graph



## Deep Learning

- Objective is to find NN parameters  $\{\mathbf{W}_k, \mathbf{b}_k\}$  that minimize a **loss function**

$$\mathcal{L}(\{\mathbf{W}_k, \mathbf{b}_k\}_K) = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{v}_{0,i}, \tilde{\mathbf{v}}_{K,i}) \in \mathcal{S}} L(f_{\text{NN}}(\mathbf{v}_{0,i}), \tilde{\mathbf{v}}_{K,i})$$

where  $(\mathbf{v}_{0,i}, \tilde{\mathbf{v}}_{K,i})$  are **examples** of inputs  $(\mathbf{v}_{0,i})$  and observed outputs  $(\tilde{\mathbf{v}}_{K,i})$  of the system we like to approximate

- Training data set  $\mathcal{S}$  contains these examples
- Parameters  $\{\mathbf{W}_k, \mathbf{b}_k\}$  are optimized using **gradient descent**

$$\mathbf{W}_k \leftarrow \mathbf{W}_k - \epsilon \nabla_{\mathbf{W}_k} \mathcal{L}(\{\mathbf{W}_k, \mathbf{b}_k\}_K)$$

$$\mathbf{b}_k \leftarrow \mathbf{b}_k - \epsilon \nabla_{\mathbf{b}_k} \mathcal{L}(\{\mathbf{W}_k, \mathbf{b}_k\}_K)$$

- Efficient computation of gradient using a **computational graph**

# Overview



Introduction

## ■ Application 1: Waveform Optimization for Short-Reach Optical Communications

- Simple Transmitter Using Feed-Forward Neural Networks
- Improved Transceivers Using Recurrent Neural Networks



Application 2: In-Situ Optimization of Transmitters

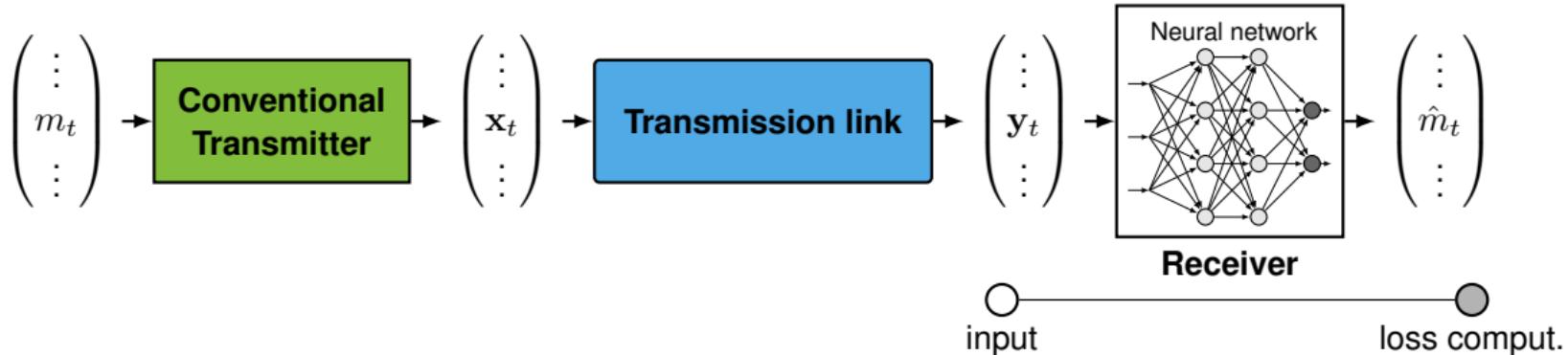


Application 3: Blind Equalization Using Variational Autoencoders



Conclusions and Outlook

# Receiver Processing Using Neural Networks

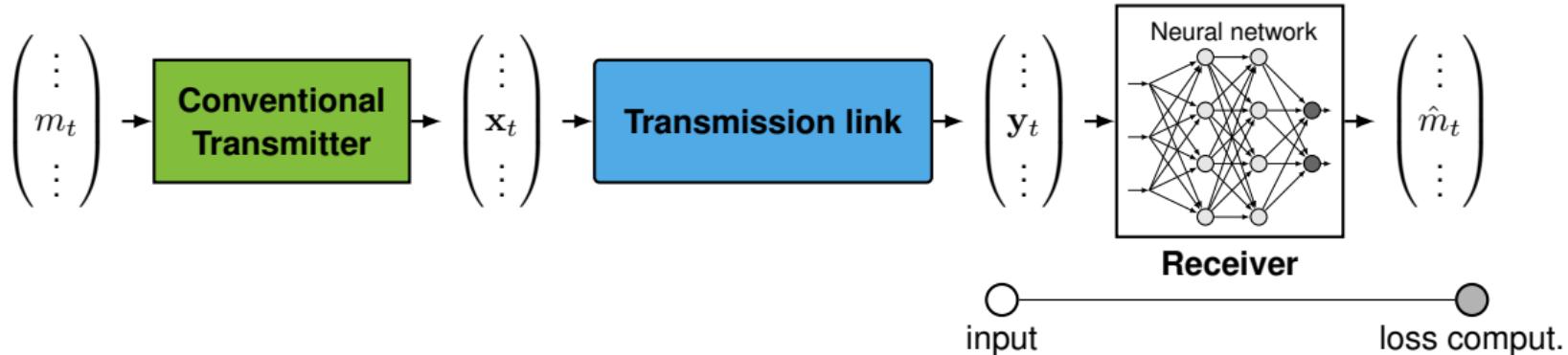


- Transmission of known sequence of messages (e.g., pilot)  $m_t$
- Based on received data ( $y_t$ ), try to recover  $m_t$  as closely as possible
- Training dataset is database of received signals and associated symbols  $\mathcal{S} = \{(y_t, m_t) : t = 1, \dots, N\}$
- Proposed for optical communications [Lyu15] and used for example as receivers in PON systems [HCvV19]

[Lyu15] I. Lyubomirsky, "Machine learning equalization techniques for high speed PAM4 fiber optic communication systems," *CS229 Final Project Report*, Stanford University, 2015

[HCvV19] V. Houtsma, E. Chou, and D. van Veen, "92 and 50 Gbps TDM-PON using neural network enabled receiver equalization specialized for PON," in *Proc. Optical Fiber Communications Conference (OFC)*, 2019

# Receiver Processing Using Neural Networks



- Transmission of known sequence of messages (e.g., pilot)  $m_t$
- Based on received data ( $y_t$ ), try to recover  $m_t$  as closely as possible
- Training dataset is database of received signals and associated symbols  $\mathcal{S} = \{(y_t, m_t) : t = 1, \dots, N\}$
- Proposed for optical communications [Lyu15] and used for example as receivers in PON systems [HCvV19]
- **Example:** AWGN channel

[Lyu15] I. Lyubomirsky, "Machine learning equalization techniques for high speed PAM4 fiber optic communication systems," *CS229 Final Project Report*, Stanford University, 2015

[HCvV19] V. Houtsma, E. Chou, and D. van Veen, "92 and 50 Gbps TDM-PON using neural network enabled receiver equalization specialized for PON," in *Proc. Optical Fiber Communications Conference (OFC)*, 2019

# Example: Deep NN Detection in AWGN Channel

- Implementation using PyTorch<sup>a</sup>
- Source code available online<sup>b</sup>

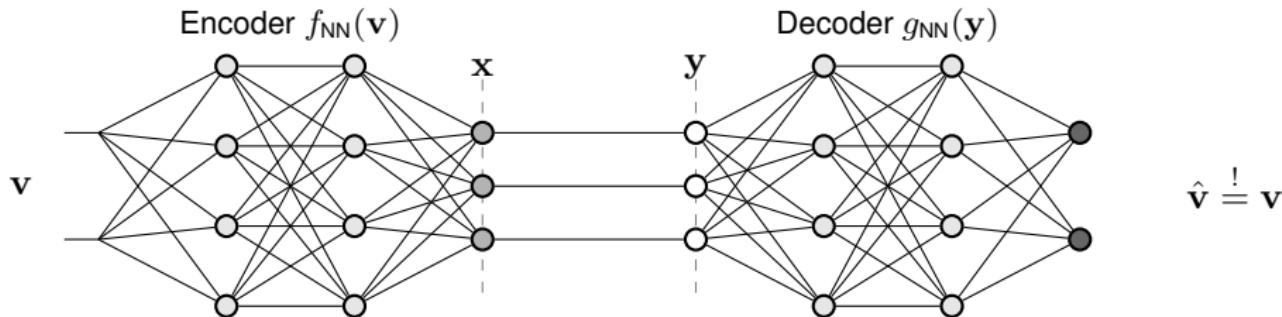


---

<sup>a</sup><http://pytorch.org>

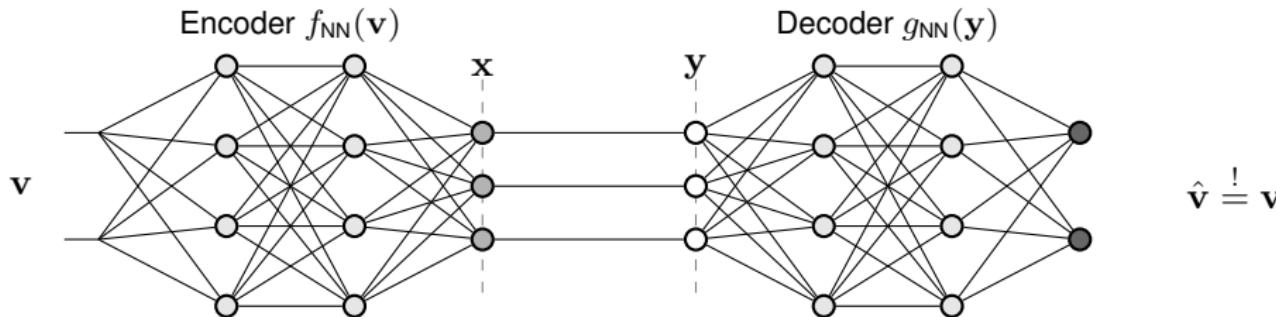
<sup>b</sup>[https://github.com/kit-cel/TUe\\_Guest\\_Lecture\\_Jan2022](https://github.com/kit-cel/TUe_Guest_Lecture_Jan2022)

# Auto-encoders – Basic Concept



- Consists of an **encoder**  $f_{NN}(v)$  and a **decoder**  $g_{NN}(y)$
- **Goal:** Try to reproduce  $v$  by  $\hat{v}$  as close as possible

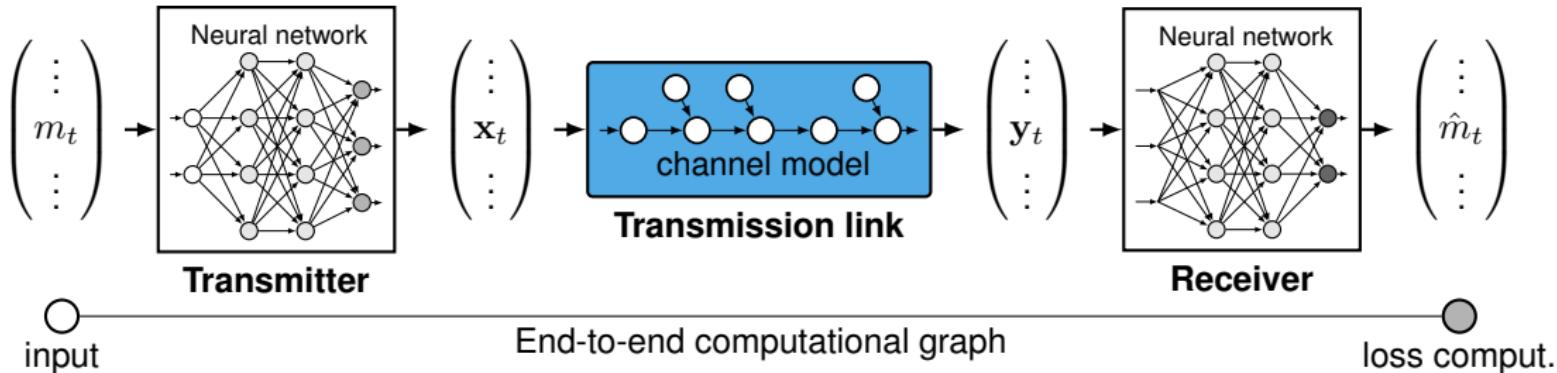
# Auto-encoders – Basic Concept



- Consists of an **encoder**  $f_{NN}(v)$  and a **decoder**  $g_{NN}(y)$
- **Goal:** Try to reproduce  $v$  by  $\hat{v}$  as close as possible
- Resembles a communication system (Goal: reproduce information as close as possible)
- Can we use auto-encoders to design communication systems? [OKC16], [OH17]

- [OKC16] T. O'Shea, K. Karra and T. C. Clancy, "Learning to Communicate: Channel Auto-encoders, Domain Specific Regularizers, and Attention," *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2016
- [OH17] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, 2017.
- [CAD<sup>+</sup>20] S. Cammerer, F. Ait Aoudia, S. Dörner, M. Stark, J. Hoydis, S. ten Brink, "Trainable communication systems: Concepts and prototype," *IEEE Trans. on Commun.*, 2020

# Communication System as Auto-Encoder



- Replacing (parts of) transmitter by neural network
- Transmission link needs to be modeled by an adequate, **differentiable** channel model that can be part of the computational graph
- **Example:** AWGN channel  $y_t = x_t + n_t$  with  $\nabla_W y_t = \nabla_W x_t$

# Example: Auto-Encoder in AWGN Channel

- Implementation using PyTorch
- Source code available online<sup>5</sup>



<sup>5</sup>[https://github.com/kit-cel/TUe\\_Guest\\_Lecture\\_Jan2022](https://github.com/kit-cel/TUe_Guest_Lecture_Jan2022)

# Example: Auto-Encoder in AWGN Channel

- Implementation using PyTorch
- Joint optimization of bit mapping and constellation minimizing bit error rate
- $M = 2^4 = 16$  points,  $E_b/N_0 = 5$  dB
- Source code available online<sup>6</sup>



[Sch20] L. S., "Geometric Constellation Shaping for Optical Data Transport," US Patent 10,834,485, filed Oct. 2018, granted Nov. 2020

<sup>6</sup>[https://github.com/kit-cel/TUE\\_Guest\\_Lecture\\_Jan2022](https://github.com/kit-cel/TUE_Guest_Lecture_Jan2022)

# Example: Auto-Encoder in AWGN Channel

- Implementation using PyTorch
- Joint optimization of bit mapping and constellation minimizing bit error rate
- $M = 2^5 = 32$  points,  $E_b/N_0 = 10$  dB
- Source code available online<sup>6</sup>

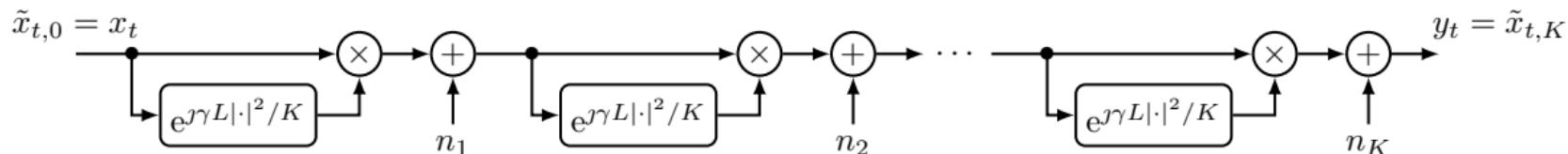


[Sch20] L. S., "Geometric Constellation Shaping for Optical Data Transport," US Patent 10,834,485, filed Oct. 2018, granted Nov. 2020

<sup>6</sup>[https://github.com/kit-cel/TUe\\_Guest\\_Lecture\\_Jan2022](https://github.com/kit-cel/TUe_Guest_Lecture_Jan2022)

# Example: Zero-Dispersion Optical Fiber

- A more complicated model is the zero-dispersion optical fiber model used in [LHG<sup>+</sup>18]
- Transmission over an optical fiber of length  $L$  having nonlinearity parameter  $\gamma$  and zero dispersion ( $\beta_i = 0$ )



- Let  $\tilde{x}_{x,0} = x_t$  and recursively compute

$$\tilde{x}_{t,i+1} = \tilde{x}_{t,i} \cdot e^{jL\gamma|\tilde{x}_{t,i}|^2/K} + n_{i+1}, \quad 0 \leq i < K$$

where  $n_{i+1} \sim \mathcal{CN}(0, P_N/K)$  is complex, circularly symmetric Gaussian noise (assuming ideal distributed amplification) and  $K$  the number of steps (ideally,  $K \rightarrow \infty$ )

- Computation graph is fully **differentiable**

[LHG<sup>+</sup>18] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning," *Proc. ECOC*, Rome, Sep. 2018

# Example: Zero-Dispersion Optical Fiber

- Implementation using PyTorch
- Source code available online<sup>7</sup>



---

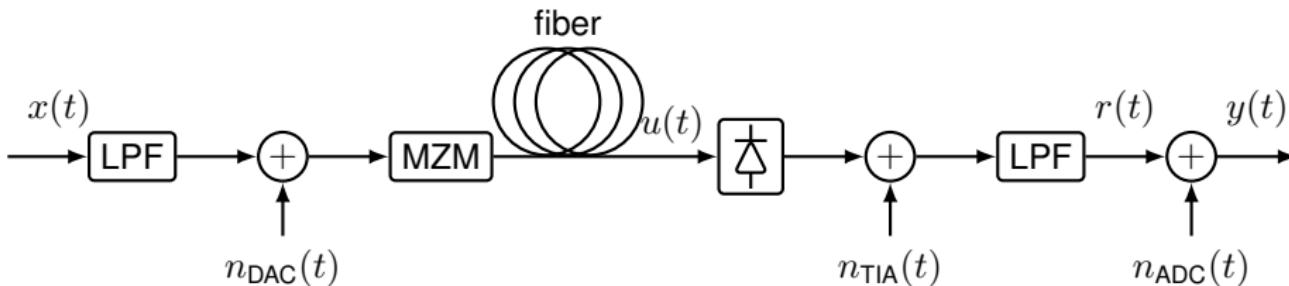
<sup>7</sup>[https://github.com/kit-cel/TUE\\_Guest\\_Lecture\\_Jan2022](https://github.com/kit-cel/TUE_Guest_Lecture_Jan2022)

# Auto-Encoders in Optical Communications

- End-to-end learning applied to optimize multidimensional constellations for IM/DD links [KCT<sup>+</sup>18], [CBS18]
- End-to-end learning of transceivers for the nonlinear Fourier transform [GJD<sup>+</sup>20]
- Constellation optimization for the zero-dispersion channel [LHG<sup>+</sup>18]
- Constellation optimization for geometric shaping [JEY<sup>+</sup>18], [JYZ19], [GAC<sup>+</sup>20]
- Differentiable channel models and nonlinearity compensation for coherent systems [HP20]

- [KCT<sup>+</sup>18] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018
- [CBS18] M. Chagnon, B. Karanov, and L. Schmalen, "Experimental demonstration of a dispersion tolerant end-to-end deep learning-based IM-DD transmission system," *Proc. European Conference on Optical Communication (ECOC)*, 2018
- [GJD<sup>+</sup>20] S. Gaiarin, R. T. Jones, F. Da Ros, and D. Zibar, "End-to-end optimized nonlinear Fourier transform-based coherent communications," *Proc. CLEO: Science and Innovations*, May 2020
- [LHG<sup>+</sup>18] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning," *Proc. European Conference on Optical Communication (ECOC)*, Sep. 2018
- [JEY<sup>+</sup>18] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," *Proc. European Conference on Optical Communication (ECOC)*, 2018
- [JYZ19] R. T. Jones, M. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," arXiv:1907.08535, 2019
- [GAC<sup>+</sup>20] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-End Learning of Geometrical Shaping Maximizing Generalized Mutual Information," *Proc. Optical Fiber Communications Conference (OFC)*, Mar. 2020
- [HP20] C. Häger and H. D. Pfister, "Physics-Based Deep Learning for Fiber-Optic Communication Systems," *Journal of Selected Areas in Communications*, 2021

# End-to-end Learning for IM/DD

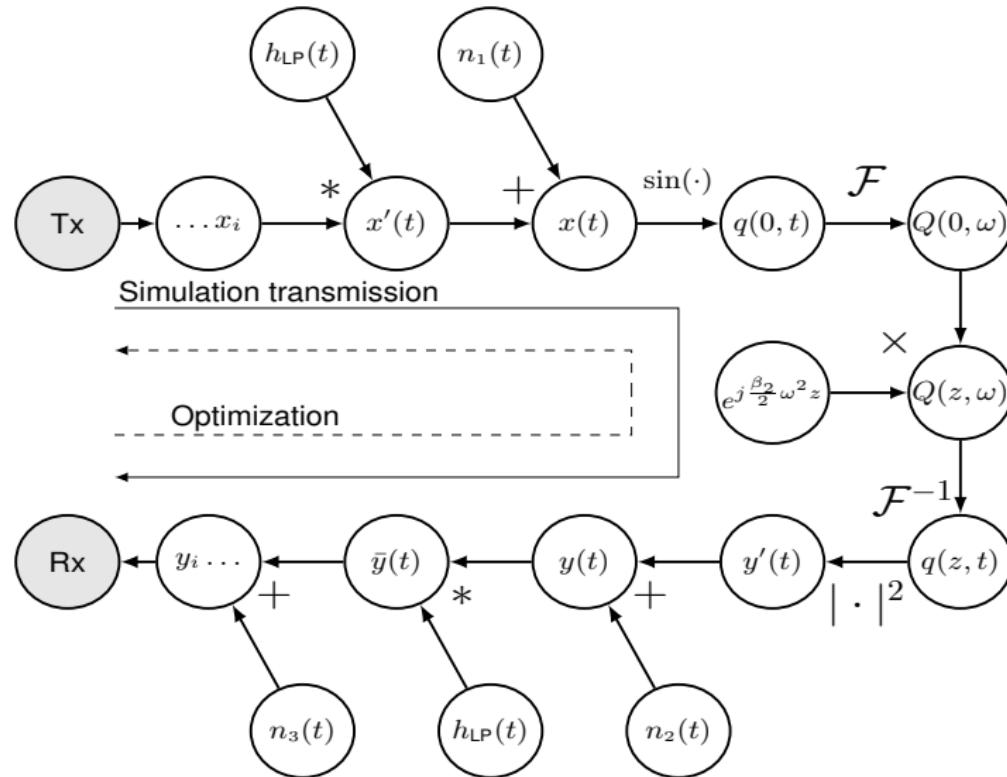


- IM/DD (intensity modul., direct detect.) ubiquitous in access networks and data center interconnects

$$y(t) = \left| \hat{h}_{\text{Fiber}} \left( \hat{h}_{\text{MZM}} (x(t) + n_{\text{DAC}}(t)) \right) \right|^2 + n_{\text{TIA}}(t) + n_{\text{ADC}}(t)$$

- Fiber adds only chromatic dispersion (no nonlinearities) with  $H(\omega, z) = \exp \left( j \frac{\beta_2}{2} z \omega^2 \right)$
- Nonlinear channel with memory, *however*, optimal **computationally feasible algorithms absent**
- Model is fully **differentiable**
- Contrary to previous examples, channel input is **waveform**, not modulation symbols
- We need to **learn optimal waveforms**

# Computational Graph of Channel Model



- Chromatic dispersion is added in frequency domain
- Fourier transform is linear and differentiable
- Channel model can be directly implemented in machine learning software

# Overview



Introduction

## ■ Application 1: Waveform Optimization for Short-Reach Optical Communications

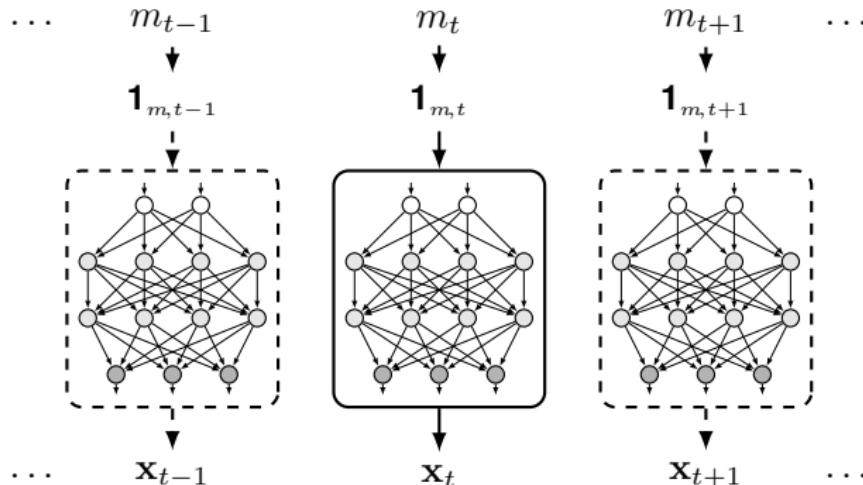
- Simple Transmitter Using Feed-Forward Neural Networks
- Improved Transceivers Using Recurrent Neural Networks

## ■ Application 2: In-Situ Optimization of Transmitters

## ■ Application 3: Blind Equalization Using Variational Autoencoders

## ■ Conclusions and Outlook

# First Transmitter Attempt Using FFNN

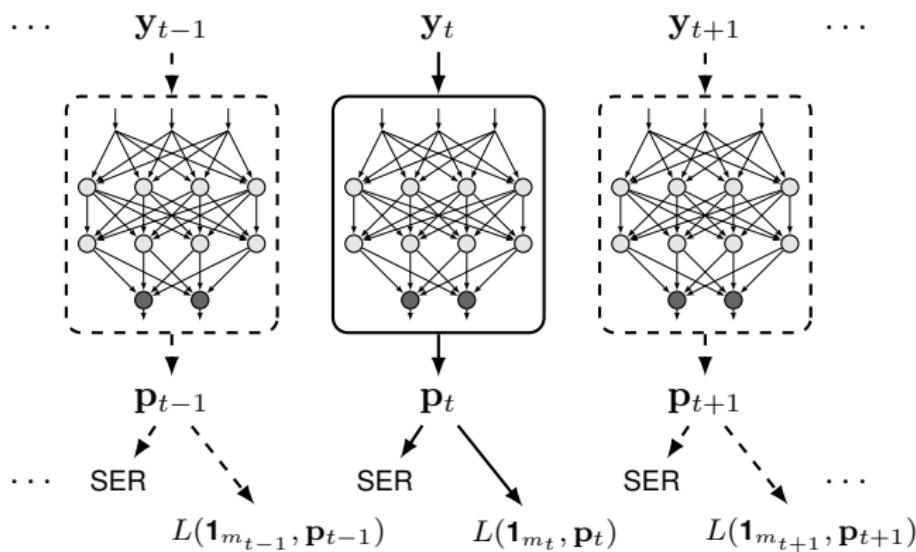


- Transform message  $m_t$  (from an alphabet containing  $M$  messages (e.g., a modulation alphabet)) into **one-hot<sup>8</sup>**vector  $1_{m,t}$
- FNN encodes  $1_{m,t}$  into  $n$  **oversampled waveform samples**  $\mathbf{x}_t$
- Apply low-pass-filter (LPF, part of the channel model) to **smear** the waveforms

[KCT<sup>+</sup>18] B. Karanov, M. Chagnon, F. Thouin, T. Eriksson, H. Bülow, D. Lavery, P. Bayvel, L. S., "End-to-end Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, Oct. 2018

<sup>8</sup>A one-hot vector  $1_{m,t}$  of length  $M$  contains only zeros except a single "1" at position  $m_t$

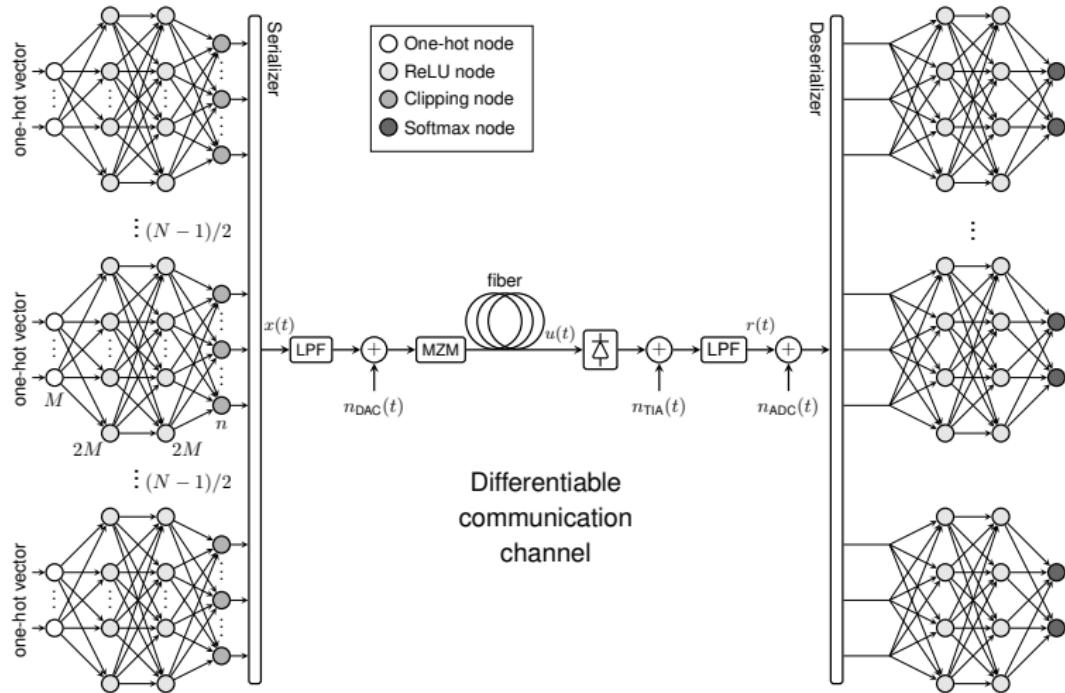
# First Receiver Attempt Using FFNN



- Receiver input is chopped into blocks  $y_t$  of  $n$  received **samples**
- Samples are processed by an FFNN with a **softmax** output activation function (“●”)
- Softmax output function generates a probability vector  $\mathbf{p}_t$
- Decision according to most probable symbol
- Training using cross-entropy loss function

$$L(\mathbf{1}_{m_t}, \mathbf{p}_t) = -\log(p_{t,m_t})$$

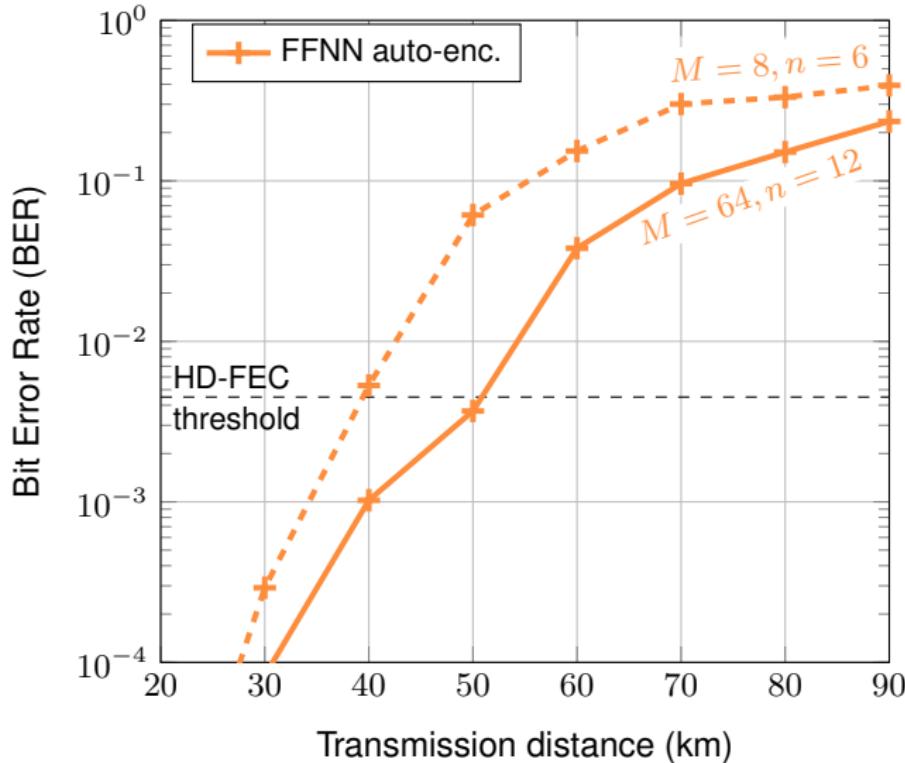
# Putting Everything Together



[KCT<sup>+</sup>18]

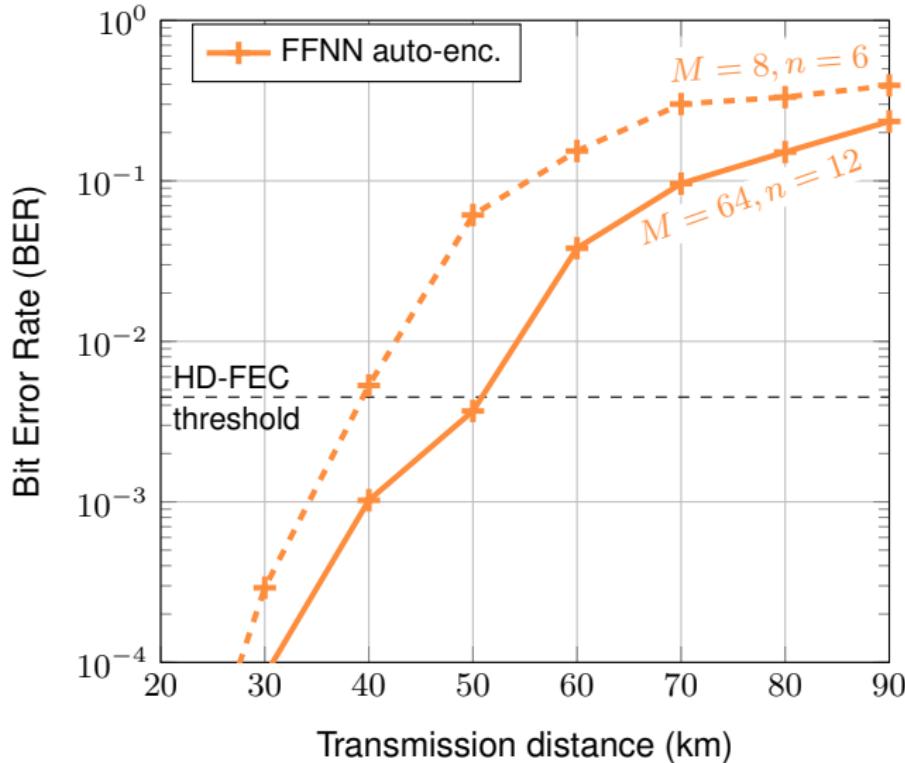
B. Karanov, M. Chagnon, F. Thouin, T. Eriksson, H. Bülow, D. Lavery, P. Bayvel, L. S., "End-to-end Deep Learning of Optical Communications," *Journal of Lightwave Technology*, Oct. 2018

# Simulation Results



- State-of-the-art IM/DD system with 42 Gb/s net rate
- For small alphabets ( $M$  small) and neural networks, reach is limited
- Increasing the alphabet and the neural networks enables extra reach (multidimensional constellations) as more dispersion can be compensated for
- Increasing the NN size (i.e.,  $M$  and thus  $n$ ) leads to unfeasibly large networks

# Simulation Results



- State-of-the-art IM/DD system with 42 Gb/s net rate
- For small alphabets ( $M$  small) and neural networks, reach is limited
- Increasing the alphabet and the neural networks enables extra reach (multidimensional constellations) as more dispersion can be compensated for
- Increasing the NN size (i.e.,  $M$  and thus  $n$ ) leads to unfeasibly large networks
- **New approach tailored to dispersive nature of channel needed**



Introduction

## ■ Application 1: Waveform Optimization for Short-Reach Optical Communications

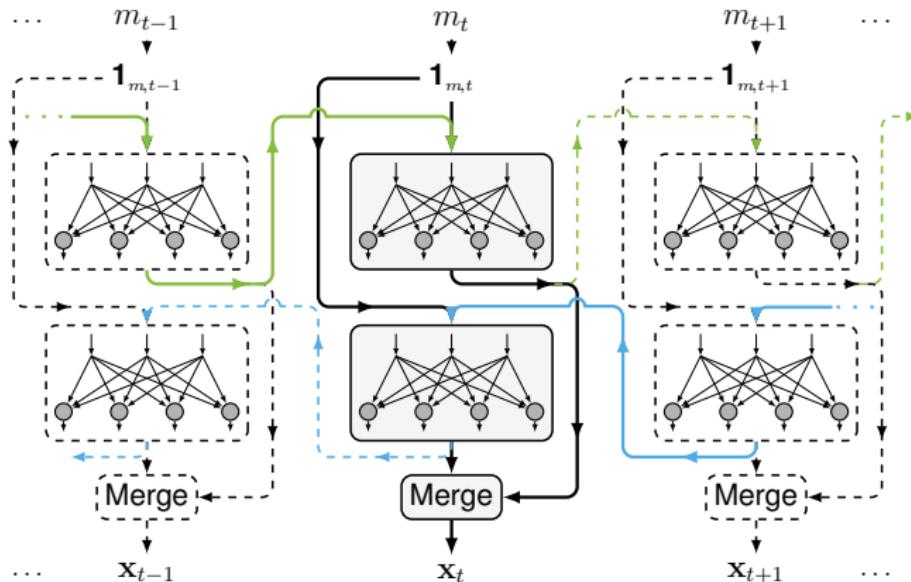
- Simple Transmitter Using Feed-Forward Neural Networks
- Improved Transceivers Using Recurrent Neural Networks

## ■ Application 2: In-Situ Optimization of Transmitters

## ■ Application 3: Blind Equalization Using Variational Autoencoders

## ■ Conclusions and Outlook

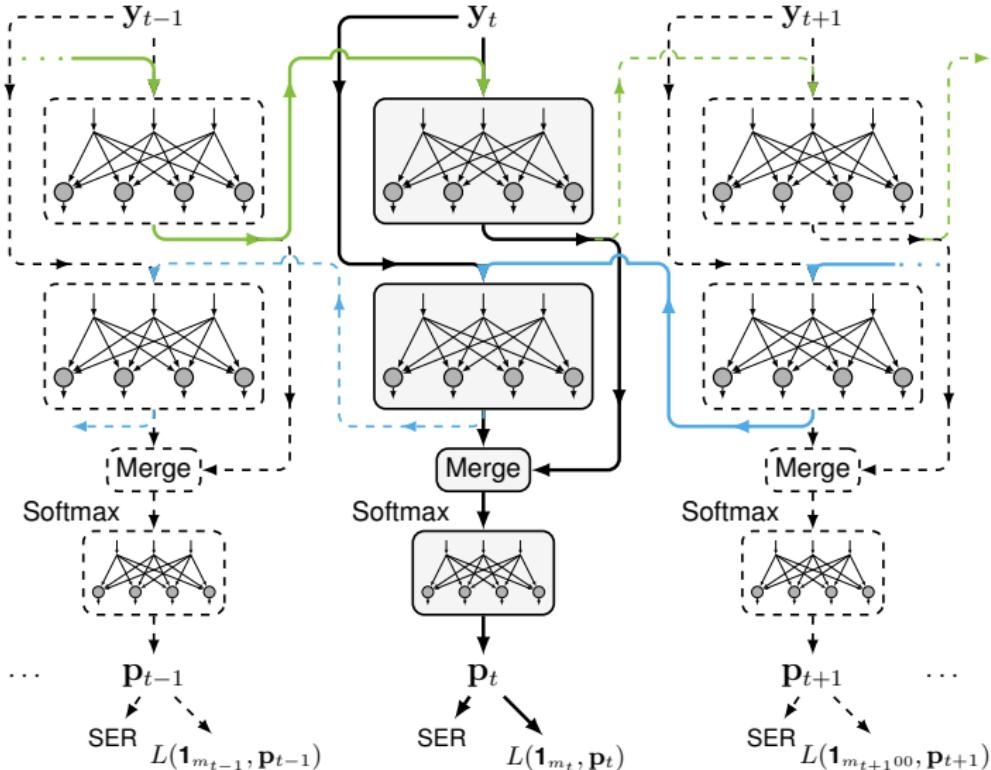
# Bidirectional Recurrent NN Based Transmitter



- Transmitter based on **bidirectional recurrent neural networks (BRNN)** [FG18] to account for memory due to chromatic dispersion
- Current message  $m_t$  jointly processed with previous and future messages via **bidirectional recurrent neural network (BRNN)**
- Complexity gain using **small** networks inside RNN cells

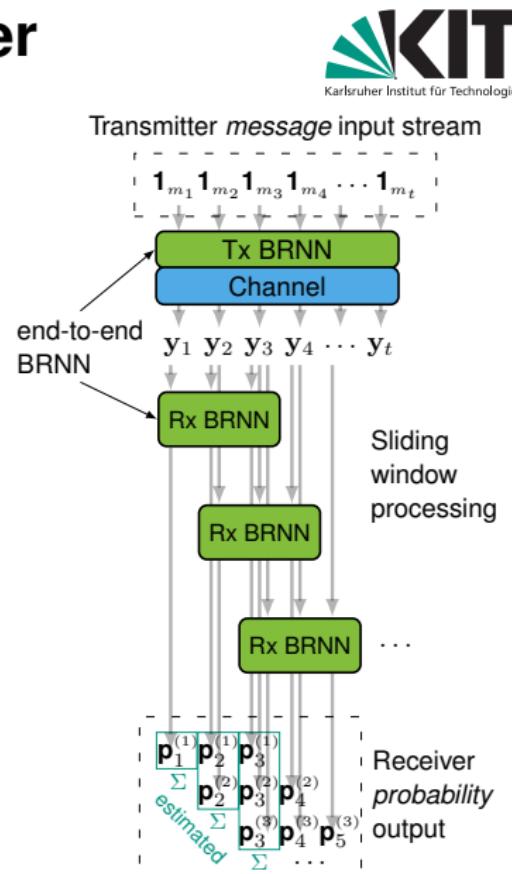
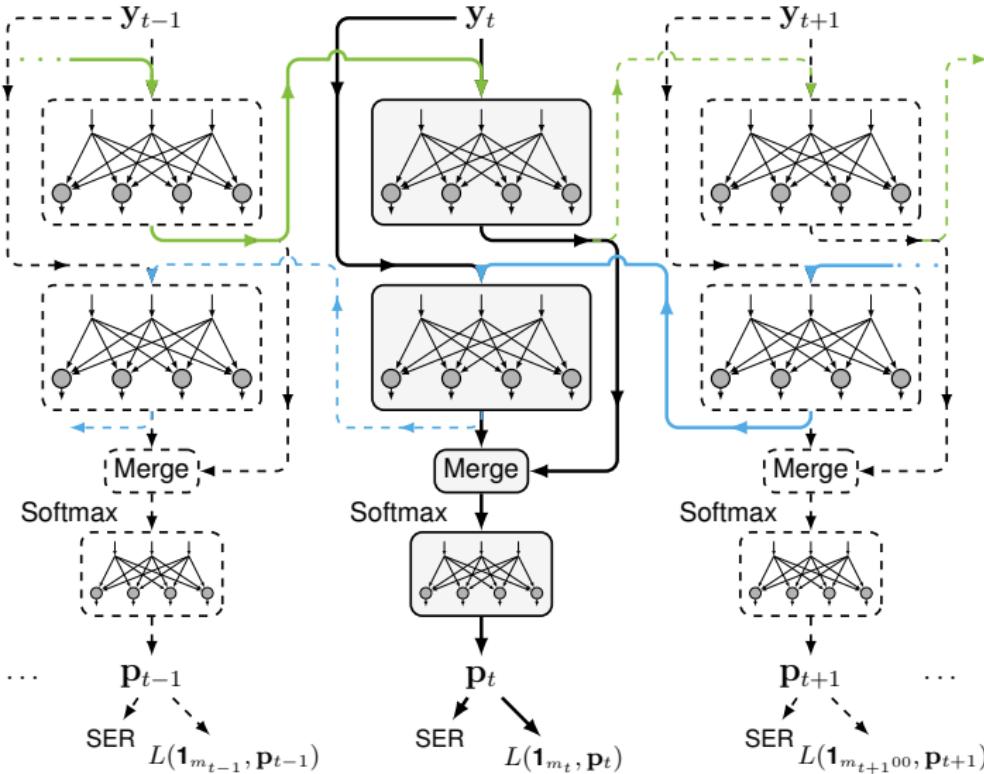
- [KLB<sup>+</sup>19] B. Karanov, D. Lavery, P. Bayvel, L. Schmalen, "End-to-end optimized transmission over dispersive intensity modulated channels using bidirectional recurrent neural networks," *Optics Express*, Jul. 2019
- [FG18] N. Farsad, A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Trans. Signal Process*, 2018

# Bidirectional Recurrent NN Based Receiver

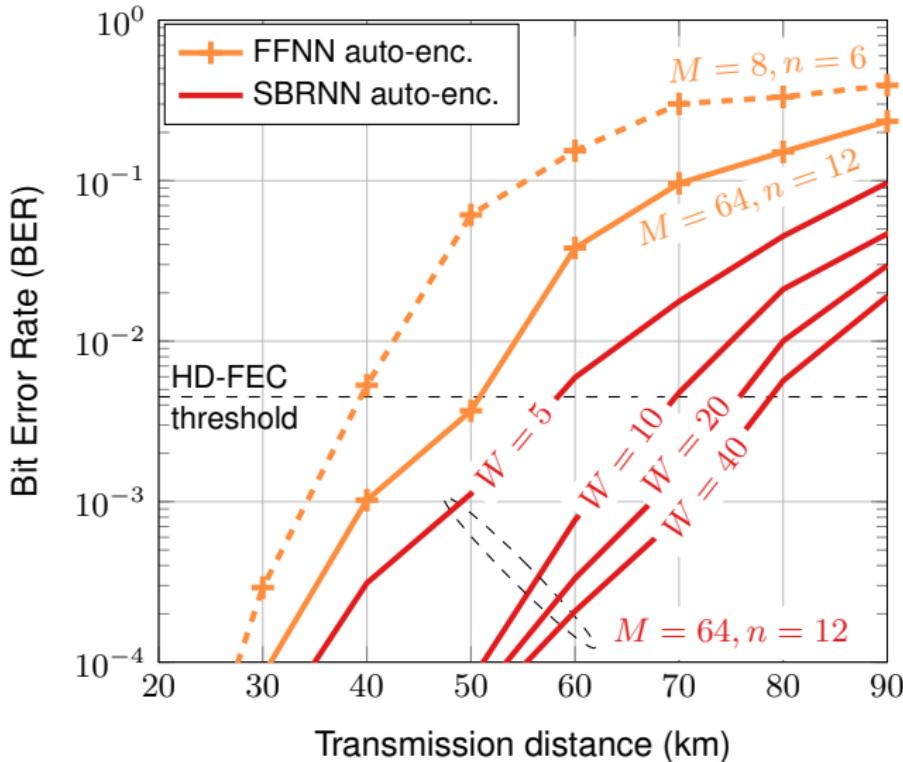


- Receiver structure similar to transmitter
- Additional **softmax** layer to generate probability vectors
- We use **sliding window** processing at the receiver to trade complexity and latency with performance
- A **window** of  $W$  stages is jointly processed at the receiver [FG18]

# Bidirectional Recurrent NN Based Receiver



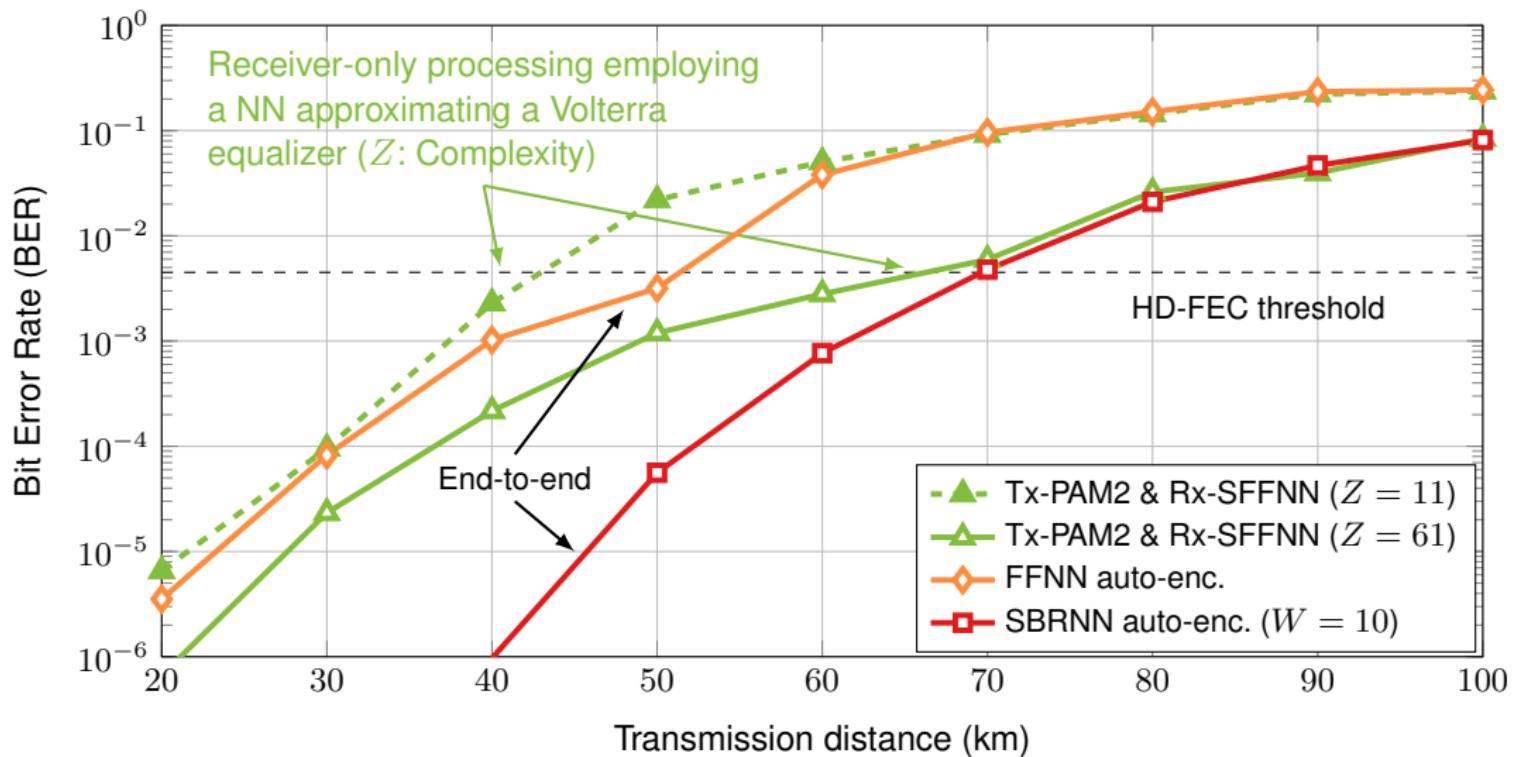
# Simulation Results



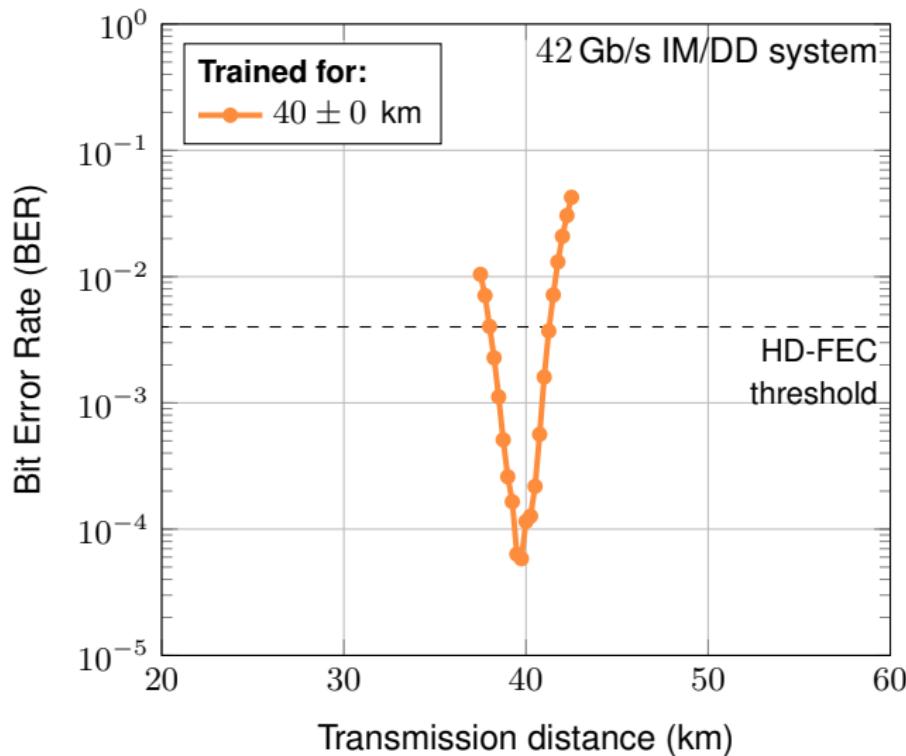
- SBRNN outperforms simple FFNN implementation at significantly **lower complexity**
- 20 km distance gain for  $W = 10$
- SBRNN can **outperform** MLSD if complexity is constrained [KLA<sup>+</sup>19]
- Now: Comparison with pure receiver NN processing approximating a Volterra equalizer [Lyu15]

- [KLA<sup>+</sup>19] B. Karanov, G. Liga, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "Deep learning for communication over dispersive nonlinear channels: Performance and comparison with classical digital signal processing", *Proc. Allerton Conf. on Commun., Control, and Computing*, 2019
- [Lyu15] I. Lyubomirsky, "Machine learning equalization techniques for high speed PAM4 fiber optic communication systems," *CS229 Final Project Report*, Stanford University, 2015

# Simulation Results (2)

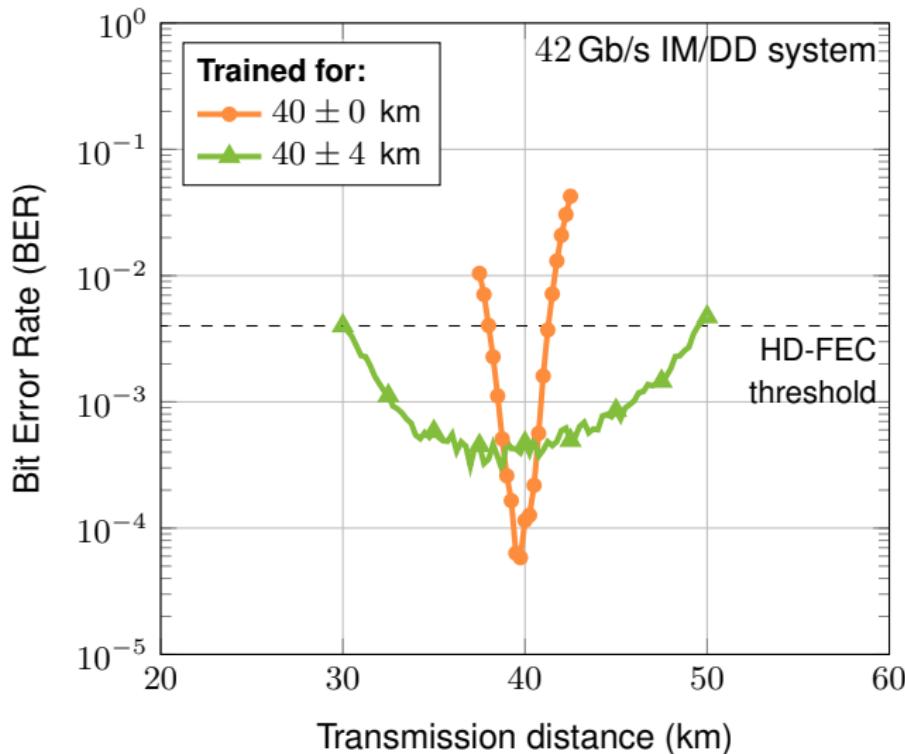


# Training Generalization: Simulation Results



- How well does the training generalize?
- If no precautions are taken, not too well

# Training Generalization: Simulation Results



- How well does the training generalize?
- If no precautions are taken, not too well
- Multi-task learning enables transceivers that operate on a range of distances **without reconfiguration** [KCT<sup>+</sup>18]
- Big step towards **distance agnostic** transceivers

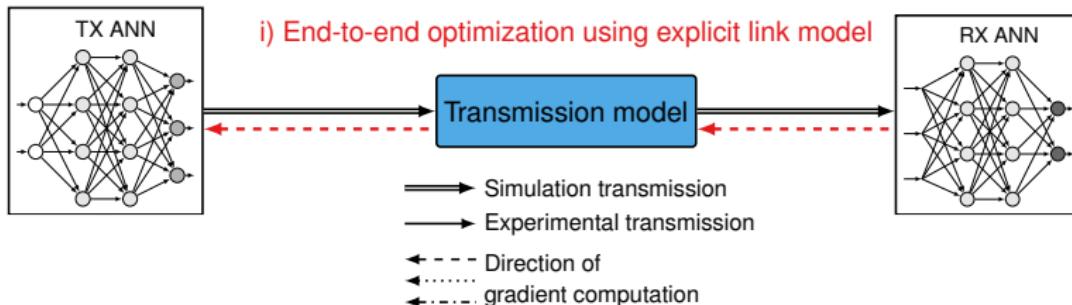
[KCT<sup>+</sup>18]

B. Karanov, M. Chagnon, F. Thouin, T. Eriksson, H. Bülow, D. Lavery, P. Bayvel, L. S., "End-to-end Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, Oct. 2018

# Overview

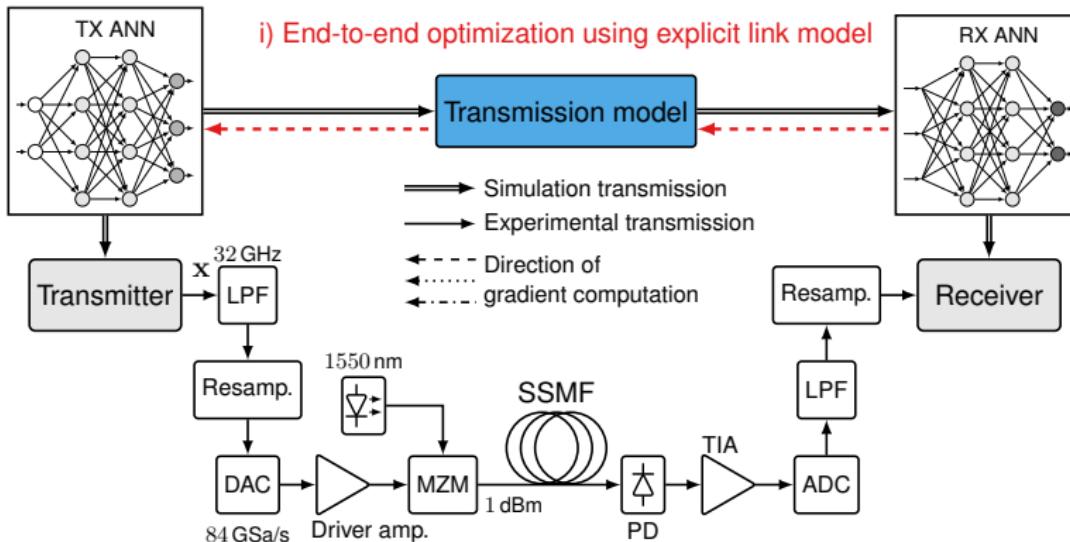
- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Recurrent Neural Networks
- Application 2: In-Situ Optimization of Transmitters
- Application 3: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

# Experimental Demonstration



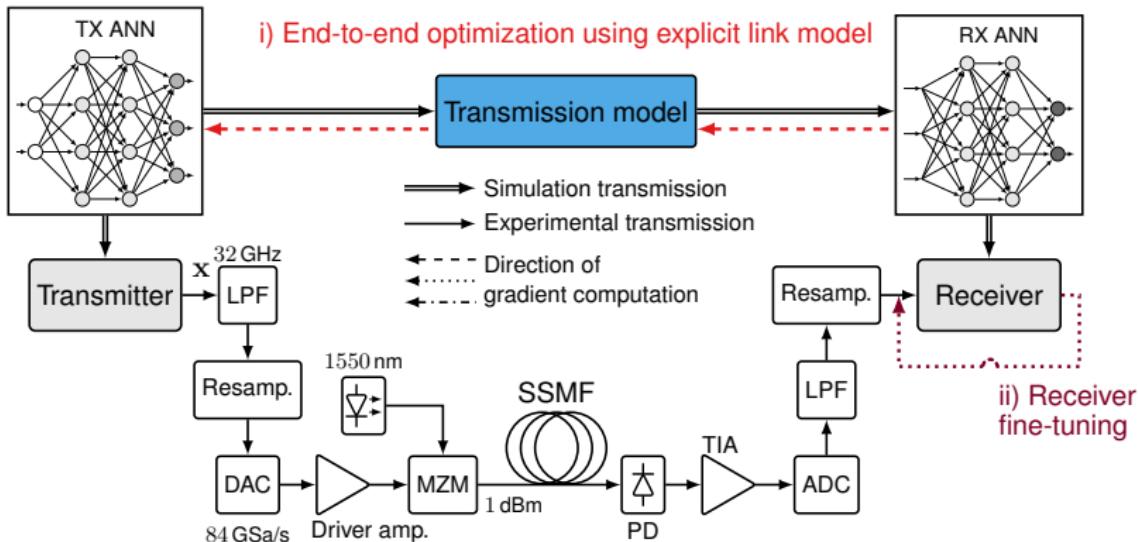
- Optimization of transmitter and receiver using differentiable model

# Experimental Demonstration



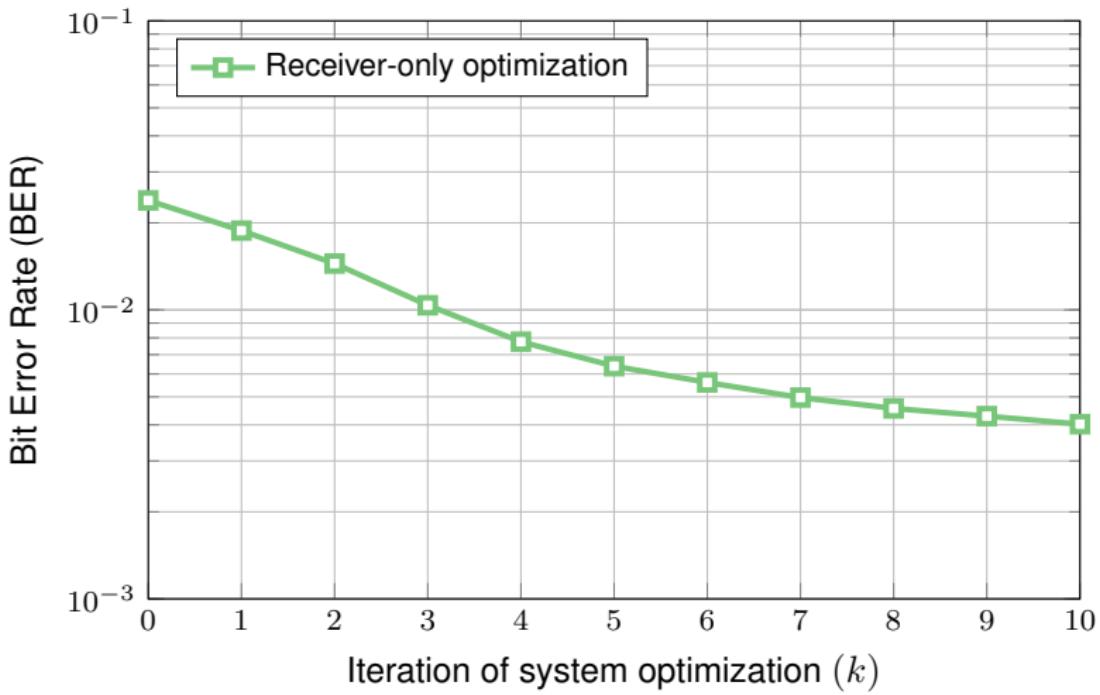
- Optimization of transmitter and receiver using differentiable model
- Mismatch between model and experiment

# Experimental Demonstration



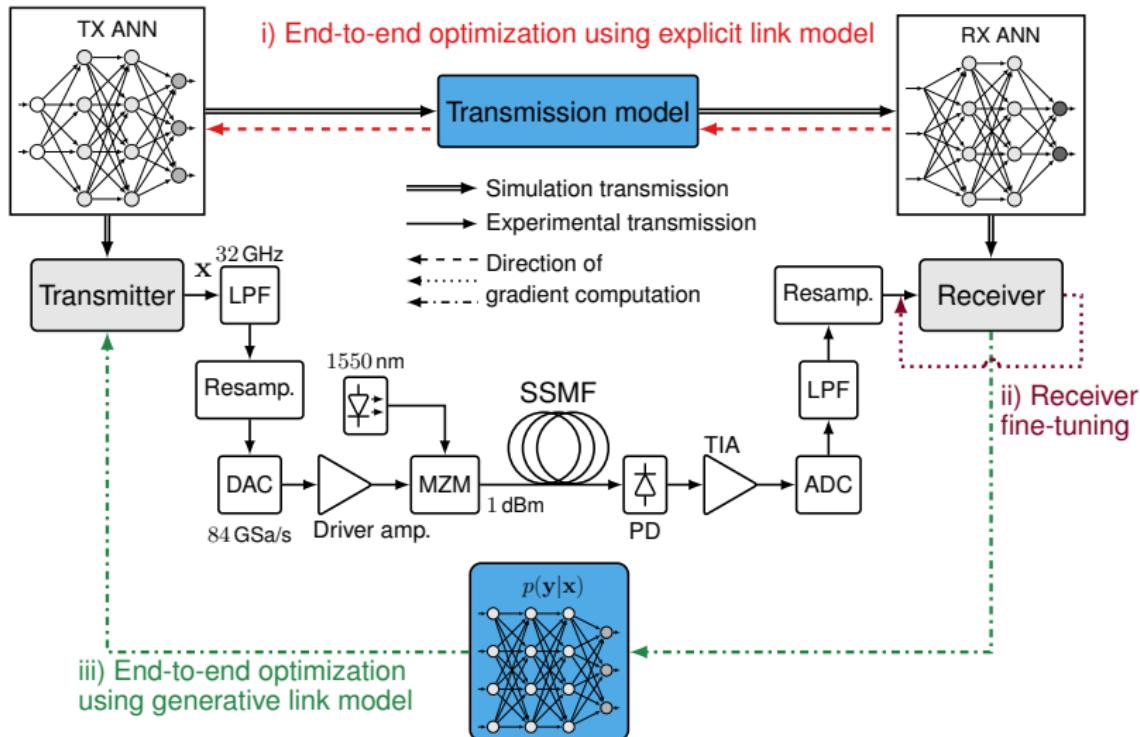
- Optimization of transmitter and receiver using differentiable model
- Mismatch between model and experiment
- Receiver fine-tuning via supervised learning

# Experimental Results: Receiver Fine-tuning



- Improvement of symbol and bit error probabilities
- Saturation after a few iterations
- (Had to stop iterating due to unavailability of experimental setup)

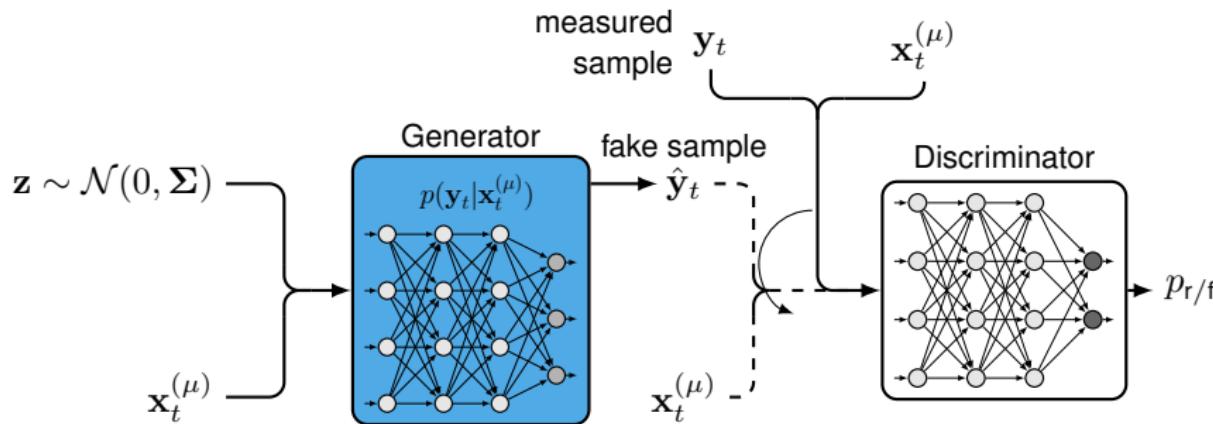
# Experimental Demonstration



- Optimazion of transmitter and receiver using differentiable model
- Mismatch between model and experiment
- Receiver fine-tuning via supervised learning
- Transmitter fine-tuning via model of experimental testbed

# Channel Modelling Using a Conditional GAN

- Neural network that behaves like a random number generator with the same properties as the channel
- Combining  $\mathbf{z} \sim \mathcal{N}(0, \Sigma)$  and the channel input  $\mathbf{x}_t$ , the generator generates a channel output distributed according to  $p(\mathbf{y}|\vec{\mathbf{x}}_t)$

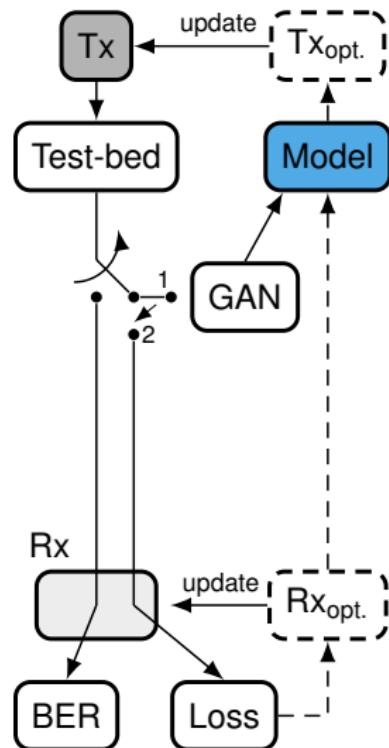


[GPM<sup>+</sup>14]  
[KCA<sup>+</sup>20]

I. Goodfellow *et al.*, “Generative adversarial nets,” in *Proc. NeurIPS*, 2014

B. Karanov, M. Chagnon, V. Aref, D. Lavery, P. Bayvel, L. S., “Concept and Experimental Demonstration of Optical IM/DD End-to-End System Optimization using a Generative Model,” *Proc. OFC*, 2020

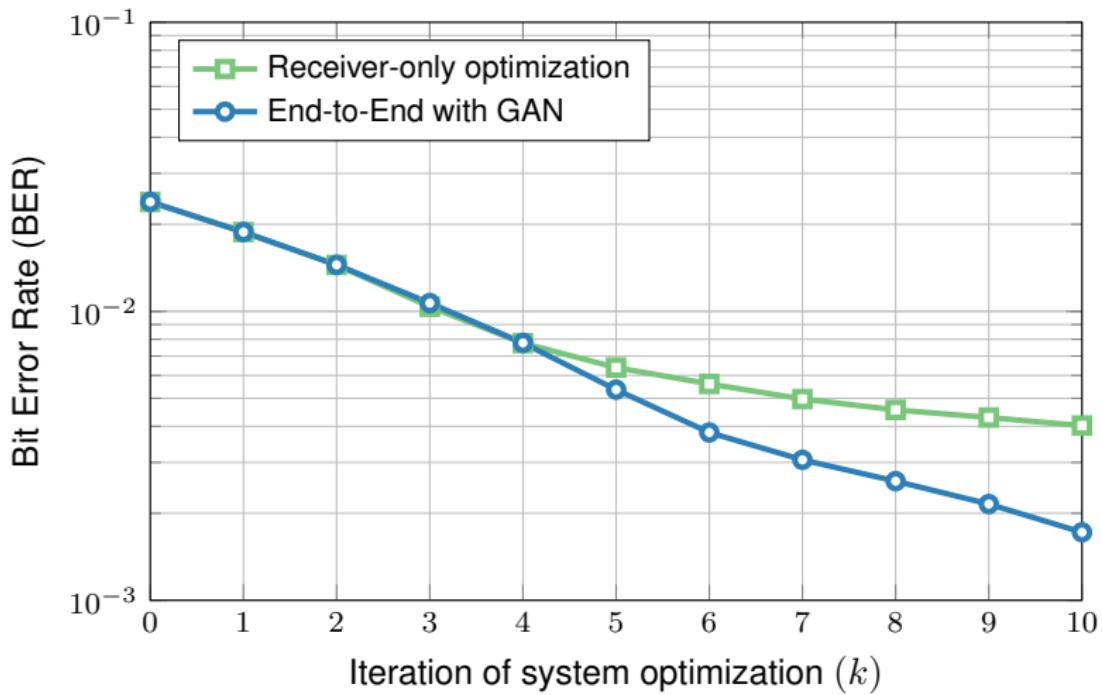
# Channel Modelling Using a Conditional GAN



## Training procedure:

1. Use testbed to generate input-output samples
2. Apply GAN training to get a channel model
3. Use this model to update and optimize the transmitter
4. With the new transmitter, use the test-bed to get input-output samples and optimize the receiver using supervised learning
5. Evaluate the system and re-iterate if not converged

# Experimental Results

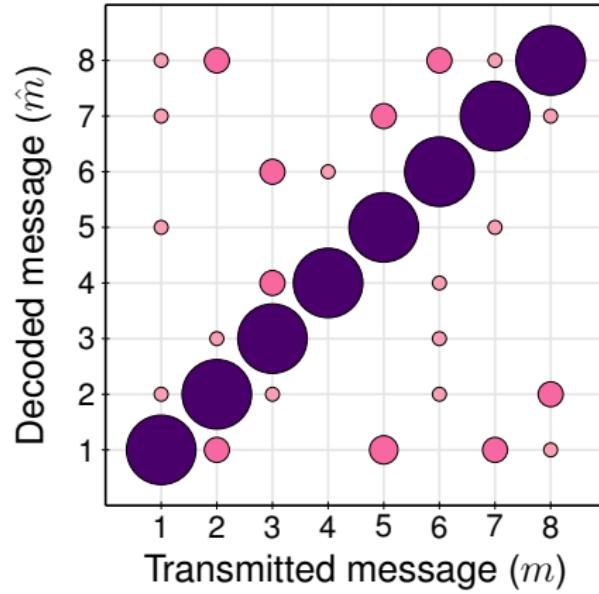
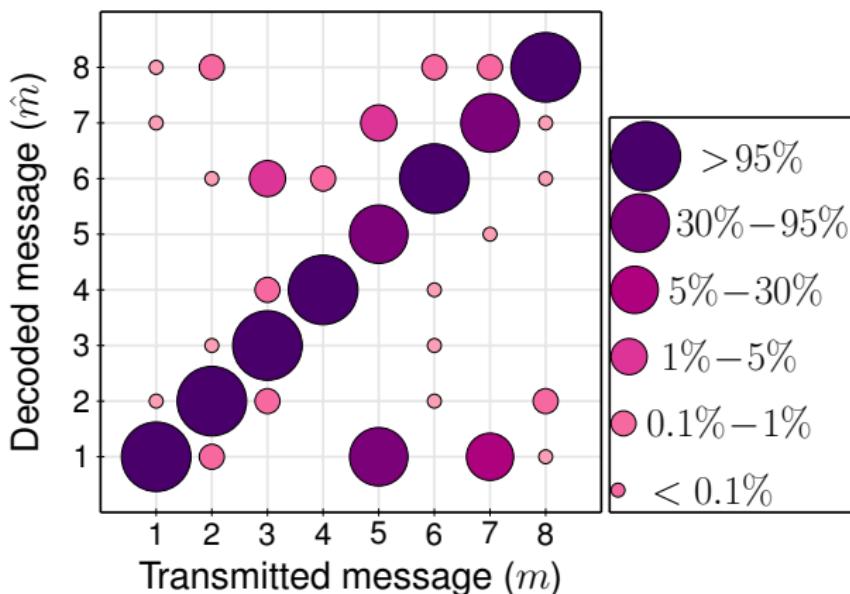


- Improvement of symbol and bit error probabilities
- More iterations will even be more beneficial
- (Had to stop iterating due to unavailability of experimental setup)

# Experimental Results (2)

- End-to-end cross-over probabilities
- Initial cross-over probabilities

- After  $k = 10$  iterations



# In-Situ Transmitter Optimization Approaches

## Approach 1: Generative adversarial nets (GANs)

- Demonstrated often ([YLJ+18], [ORW19], [SD19], [KCA<sup>+</sup>20], [DHC<sup>+</sup>20]), however, GAN training tends to be difficult, requiring careful hyperparameter tuning

## Approach 2: Policy gradient methods

- Approach from reinforcement learning to estimate gradient [AH19], [GAH19]. Potentially slow convergence and requiring careful hyperparameter tuning

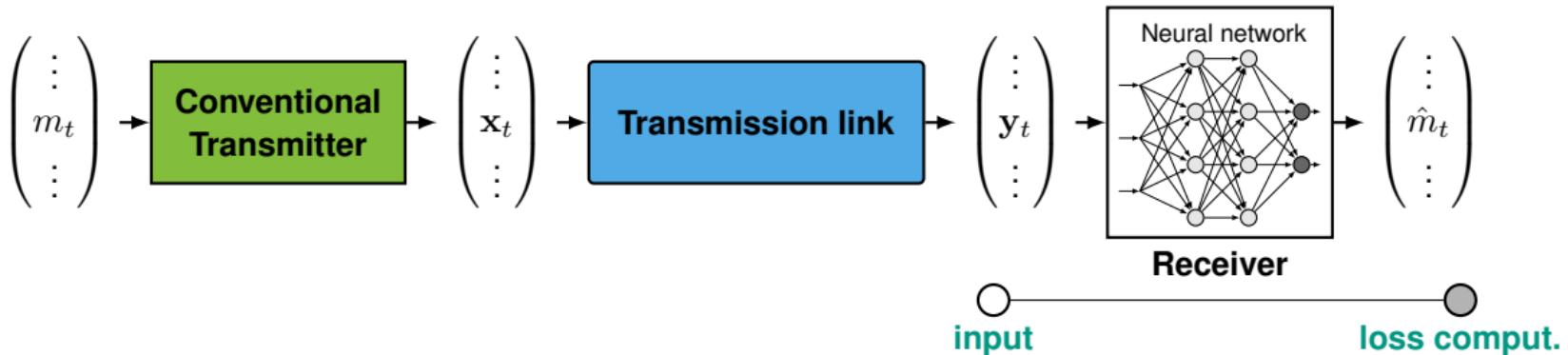
## Approach 3: Bayesian filtering

- Recent approach [JYdR<sup>+</sup>], still many open questions

[YLJ+18]	H. Ye, G. Y. Li, B. F. Juang and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in <i>Proc. GLOBECOM</i> , 2018
[ORW19]	T. O'Shea, T. Roy and N. West, "Approximating the void: Learning stochastic channel models from observation with variational generative adversarial networks," in <i>Proc. of 2019 International Conference on Computing, Networking and Communications</i> , 2019
[SD19]	A. Smith and J. Downey, "A communication channel density estimating generative adversarial network," <i>NASA Technical Reports</i> 2019
[KCA <sup>+</sup> 20]	B. Karanov, M. Chagnon, V. Aref, D. Lavery, P. Bayvel, L. S., "Concept and Experimental Demonstration of Optical IM/DD End-to-End System Optimization using a Generative Model," <i>Proc. OFC</i> , 2020
[DHC <sup>+</sup> 20]	S. Dörner, M. Henninger, S. Cammerer, S. ten Brink, "WGAN-based Autoencoder Training Over-the-air," <i>Proc. SPAWC</i> , 2020
[AH19]	F. Ait Aoudia, J. Hoydis, "Model-free training of end-to-end communication systems," <i>IEEE J. Selected Areas in Communications</i> , 2019
[GAH19]	M. Goutay, F. Ait Aoudia, J. Hoydis, "Deep reinforcement learning autoencoder with noisy feedback," <i>Proc. WiOPT</i> , 2019
[JYdR <sup>+</sup> ]	O. Jovanovic, M. Yankov, F. da Ros, D. Zibar, "Gradient-free training of autoencoders for non-differentiable communication channels," <i>Journal of Lightwave Technology</i> , 2021

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Recurrent Neural Networks
- Application 2: In-Situ Optimization of Transmitters
- Application 3: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

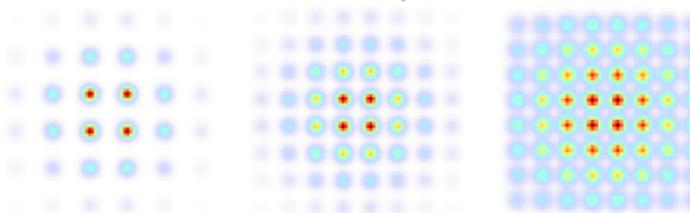
# Receiver Processing Using Neural Networks



- So far, we have considered **supervised learning**
- To compute the loss function and update the network parameters, we need to know the transmit sequence, e.g., a **pilot sequence**
- If the transmission link changes rapidly, **large overhead** for transmitting pilots
- Can we find a **blind** scheme that works purely on the received data?

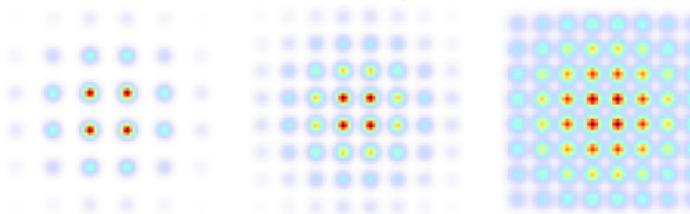
# Motivation

- Great demand for **blind** equalizers
  - + saved data rate can increase FEC overhead or data throughput
  - ⚡ have to be flexible and adaptive
  - ⚡ must support higher-order modulation formats and probabilistic constellation shaping (PCS)



# Motivation

- Great demand for **blind** equalizers
  - + saved data rate can increase FEC overhead or data throughput
  - ⚡ have to be flexible and adaptive
  - ⚡ must support higher-order modulation formats and probabilistic constellation shaping (PCS)

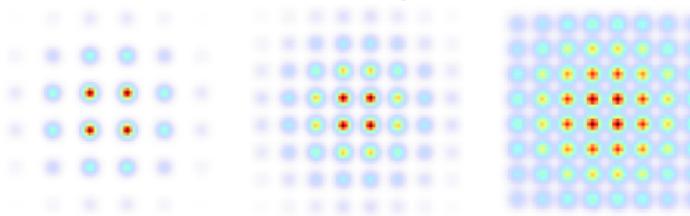


- **State-of-the-art:** Constant modulus algorithm (CMA) / multi modulus algorithm (MMA) [Yan02]
  - CMA: ill-matched for higher-order modulation formats and probabilistic shaping
  - MMA: high implementation complexity and low convergence rate

[Yan02] J. Yang *et al.*, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas. Commun.* **20**(5), pp. 997ff (2002).  
[GW92] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Optical Engineering* **31**(6), pp. 1224–1229 (1992).

# Motivation

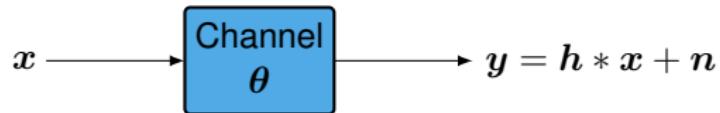
- Great demand for **blind** equalizers
  - + saved data rate can increase FEC overhead or data throughput
  - ⚡ have to be flexible and adaptive
  - ⚡ must support higher-order modulation formats and probabilistic constellation shaping (PCS)



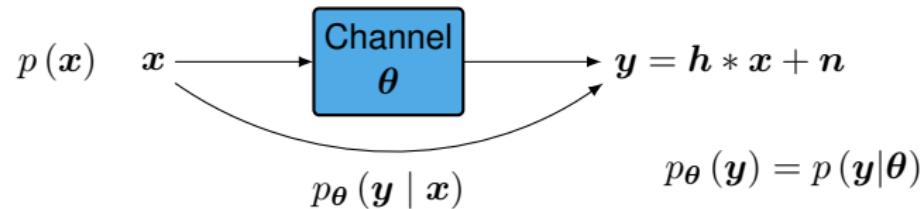
- **State-of-the-art:** Constant modulus algorithm (CMA) / multi modulus algorithm (MMA) [Yan02]
  - CMA: ill-matched for higher-order modulation formats and probabilistic shaping
  - MMA: high implementation complexity and low convergence rate
- **Ideally:** Maximum likelihood detector, e.g. [GW92]

[Yan02] J. Yang *et al.*, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas. Commun.* **20**(5), pp. 997ff (2002).  
[GW92] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Optical Engineering* **31**(6), pp. 1224–1229 (1992).

# System and Problem Description

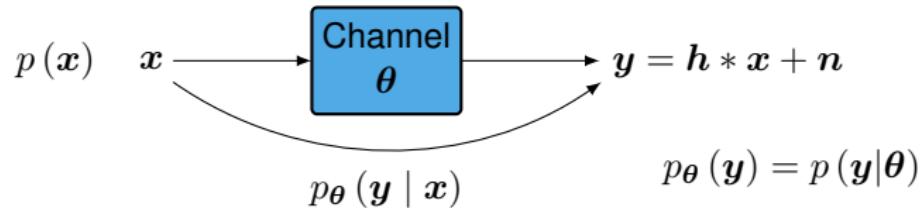


# System and Problem Description



**Goal:**  $\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p_{\theta}(y)$

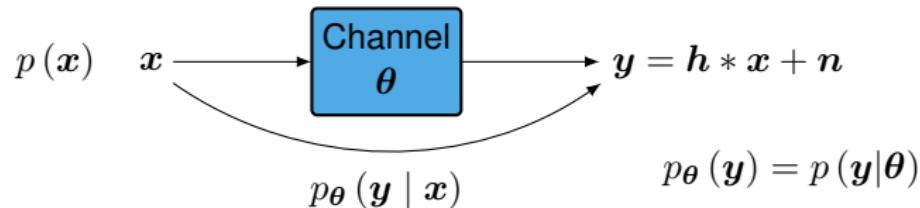
# System and Problem Description



**Goal:**  $\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p_{\theta}(y)$

Given:  $p(x)$  ,  $p_{\theta}(y | x) = \mathcal{CN}(\mathbf{h} * \mathbf{x}, \sigma_w^2 \mathbf{I}_N)$

# System and Problem Description

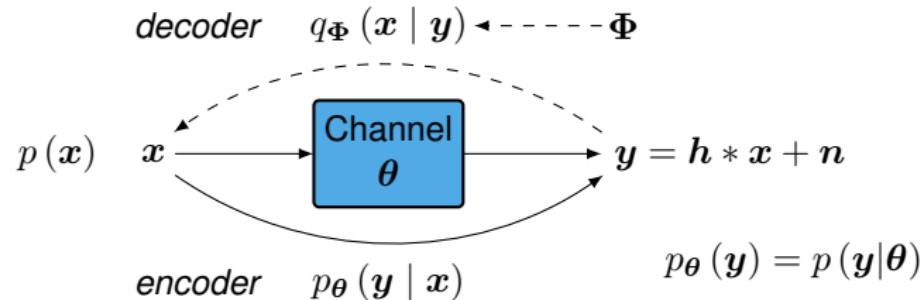


**Goal:**  $\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p_{\theta}(y)$

Given:  $p(x)$  ,  $p_{\theta}(y | x) = \mathcal{CN}(\mathbf{h} * x, \sigma_w^2 \mathbf{I}_N)$

Problem:  $p_{\theta}(y) = \sum_x p(x) p_{\theta}(y | x) \rightarrow \text{intractable}$

# System and Problem Description



**Goal:**  $\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p_{\theta}(y)$

Given:  $p(x)$ ,  $p_{\theta}(y | x) = \mathcal{CN}(h * x, \sigma_w^2 I_N)$

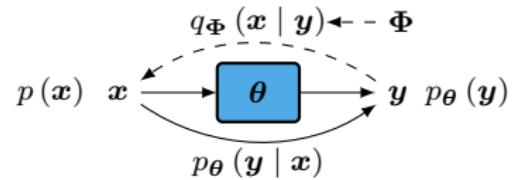
Problem:  $p_{\theta}(y) = \sum_x p(x) p_{\theta}(y | x) \rightarrow \text{intractable}$

■ **Idea:** use variational inference to approximate (intractable) ML problem [CB18]

[CB18] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders," *Proc. IEEE ICC*, Kansas City, USA, 2018.

# Variational Inference via VAE Paradigm (1)

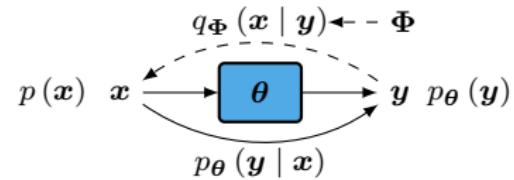
■ ML criterion:  $\hat{\theta} = \arg \max_{\theta} p_{\theta}(y) = \arg \max_{\theta} \ln p_{\theta}(y)$



$$\ln p_{\theta}(y) = \dots = \underbrace{\sum_x q_{\Phi}(x | y) \cdot \ln \frac{p_{\theta}(x, y)}{q_{\Phi}(x | y)}}_{=: \mathcal{L}(\theta, \Phi, y)} + \underbrace{\sum_x q_{\Phi}(x | y) \cdot \ln \frac{q_{\Phi}(x | y)}{p_{\theta}(x | y)}}_{=: D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y))}$$

- [KW14] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. ICLR*, 2014.  
[BKM17] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Statist. Assoc.* **112**(518) (2017).

# Variational Inference via VAE Paradigm (1)



■ ML criterion:  $\hat{\theta} = \arg \max_{\theta} p_{\theta}(y) = \arg \max_{\theta} \ln p_{\theta}(y)$

$$\ln p_{\theta}(y) = \dots = \underbrace{\sum_x q_{\Phi}(x|y) \cdot \ln \frac{p_{\theta}(x,y)}{q_{\Phi}(x|y)}}_{=: \mathcal{L}(\theta, \Phi, y)} + \underbrace{\sum_x q_{\Phi}(x|y) \cdot \ln \frac{q_{\Phi}(x|y)}{p_{\theta}(x|y)}}_{=: D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y))}$$

■ Kullback-Leibler divergence  $D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y)) \geq 0$

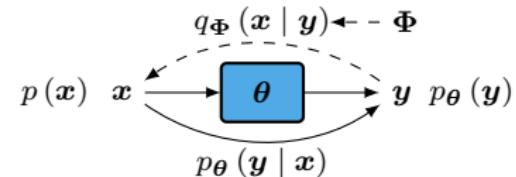
⇒ Evidence lower bound (ELBO):  $\mathcal{L}(\theta, \Phi, y)$

[KW14]  
[BKM17]

D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. ICLR*, 2014.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Statist. Assoc.* **112**(518) (2017).

# Variational Inference via VAE Paradigm (1)



■ **ML criterion:**  $\hat{\theta} = \arg \max_{\theta} p_{\theta}(y) = \arg \max_{\theta} \ln p_{\theta}(y)$

$$\ln p_{\theta}(y) = \dots = \underbrace{\sum_x q_{\Phi}(x|y) \cdot \ln \frac{p_{\theta}(x,y)}{q_{\Phi}(x|y)}}_{=: \mathcal{L}(\theta, \Phi, y)} + \underbrace{\sum_x q_{\Phi}(x|y) \cdot \ln \frac{q_{\Phi}(x|y)}{p_{\theta}(x|y)}}_{=: D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y))}$$

■ Kullback-Leibler divergence  $D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y)) \geq 0$

⇒ Evidence lower bound (**ELBO**):  $\mathcal{L}(\theta, \Phi, y)$

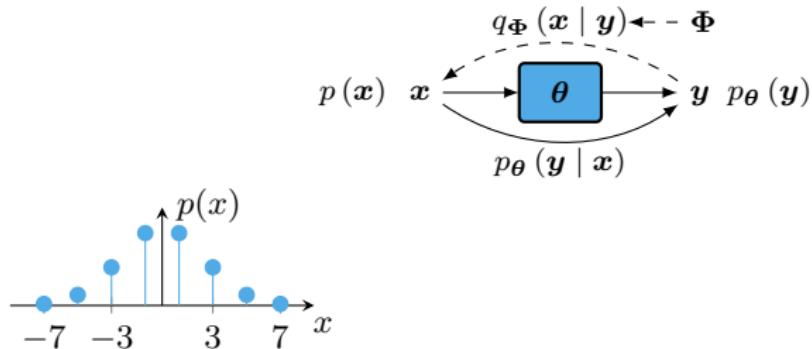
■ **Optimization:**  $\mathcal{L}(\theta, \Phi, y) \uparrow \Rightarrow D_{KL}(q_{\Phi}(x|y) \| p_{\theta}(x|y)) \rightarrow 0$

[KW14] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. ICLR*, 2014.

[BKM17] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Statist. Assoc.* **112**(518) (2017).

# Variational Inference via VAE Paradigm (2)

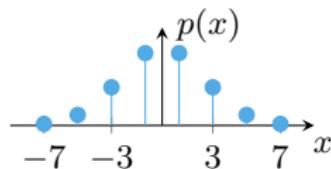
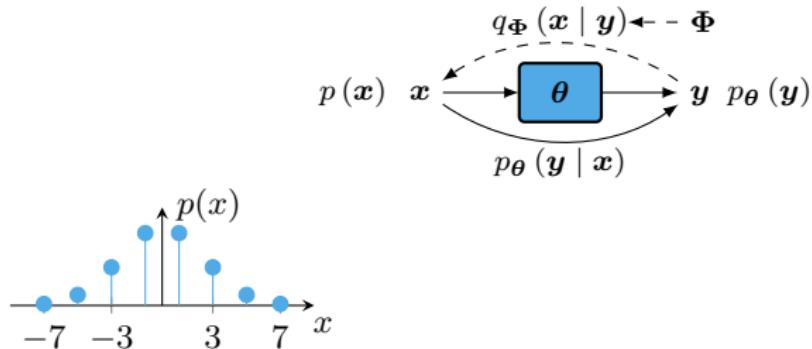
$$\mathcal{L}(\theta, \Phi, y) = B(h, \Phi, y) - A(\Phi, y)$$



- $A(\Phi, y) = D_{KL}(q_\Phi(x | y) \| p(x))$
- Tries to shape  $q_\Phi(x | y)$  similar to  $p(x)$

# Variational Inference via VAE Paradigm (2)

$$\mathcal{L}(\theta, \Phi, y) = B(h, \Phi, y) - A(\Phi, y)$$

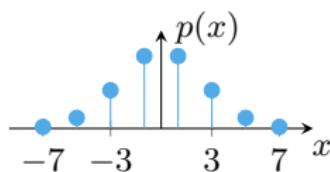
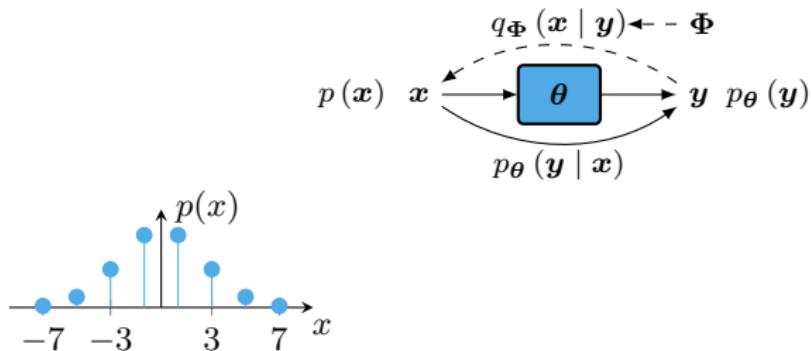


- $A(\Phi, y) = D_{KL}(q_\Phi(x|y) \| p(x))$ 
  - Tries to shape  $q_\Phi(x|y)$  similar to  $p(x)$

- $B(h, \Phi, y)$ 
  - Matches equalizer output to received symbols

# Variational Inference via VAE Paradigm (2)

$$\mathcal{L}(\theta, \Phi, y) = B(h, \Phi, y) - A(\Phi, y)$$



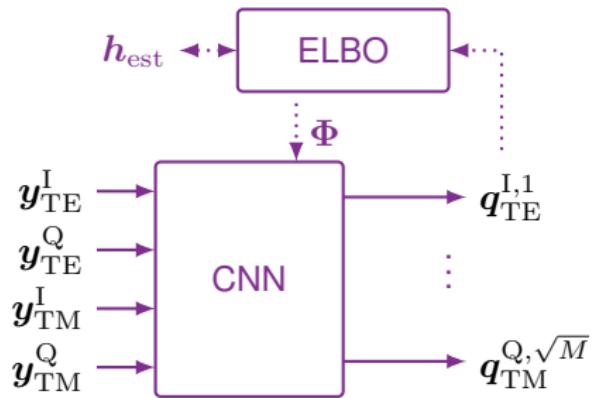
- $A(\Phi, y) = \text{D}_{\text{KL}}(q_{\Phi}(x | y) \| p(x))$ 
  - Tries to shape  $q_{\Phi}(x | y)$  similar to  $p(x)$

- $B(h, \Phi, y)$ 
  - Matches equalizer output to received symbols

⇒ **Blind approach:** requires only  $p(x)$ , not  $x$

⇒  $q_{\Phi}(x | y)$  can be used in conjunction with Soft-FEC

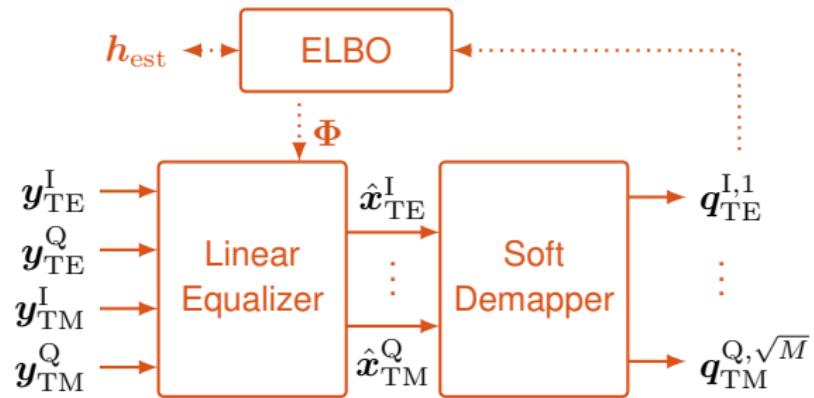
# VAE Equalizer – Neural Network (VAE-NN)



- Convolutional Neural Network (CNN) has two tasks: equalization *and* demapping
- Outputs  $q_{\Phi}(x | y)$
- Potentially capable of compensating non-linearities

[CB18] A. Caciularu and D. Burshtein, "Blind channel equalization using variational autoencoders," Proc. IEEE ICC, Kansas City, USA (2018).

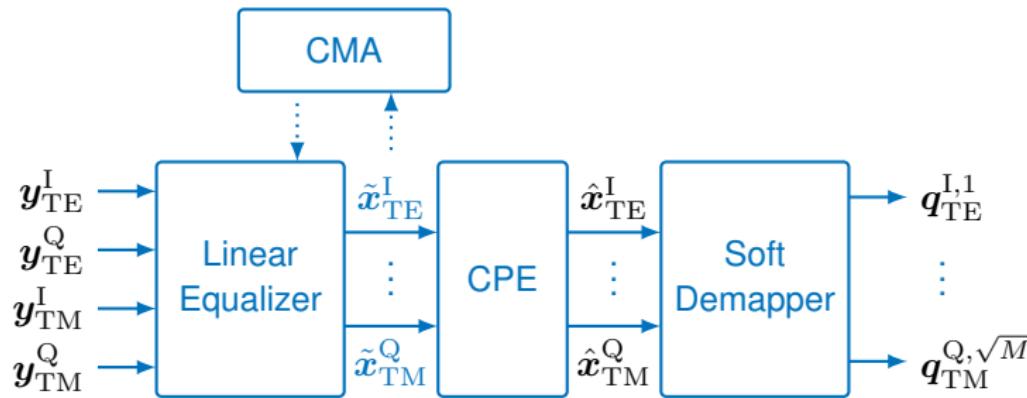
# VAE Equalizer – Linear Equalizer (VAE-LE)



- Equalization and demapping are split
- Classical butterfly structure with FIR filters
- Outputs estimated symbols and soft demapper computes  $q_{\Phi}(x | y)$
- Targets compensation of linear impairments

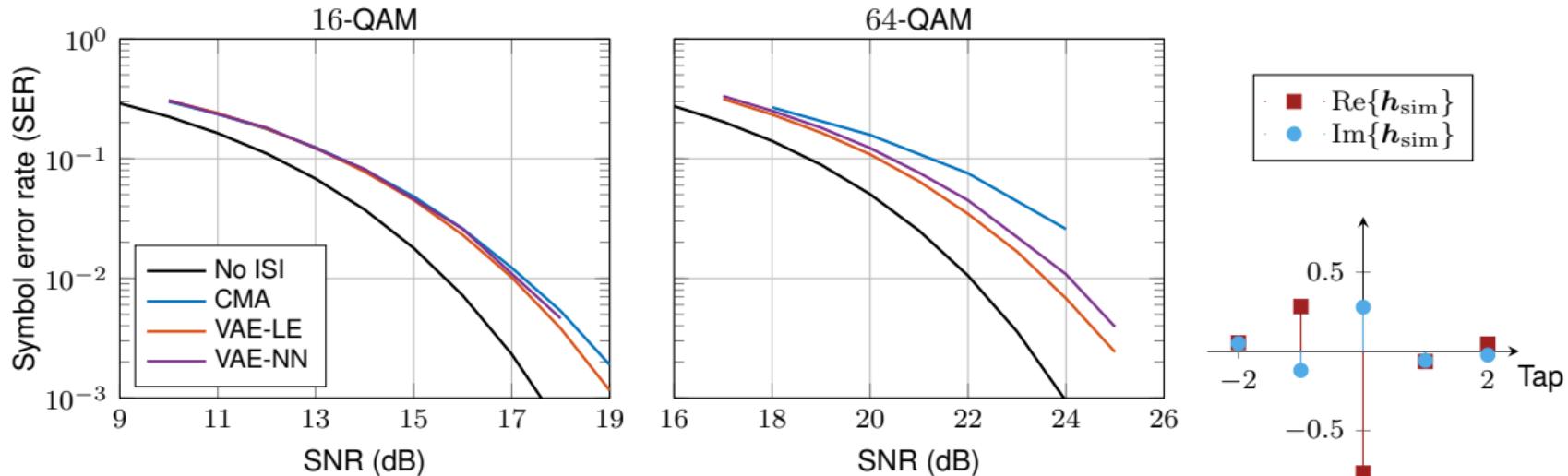
[LBS21] V. Lauinger, F. Buchali, L. S., "Blind equalization for coherent optical communications based on variational inference," *Proc. Adv. Photon. Congr., SPPCOM*, Jul. 2021

# Reference: CMA-based Equalizer



- **State-of-the-art** blind adaptive equalizer in optical comm.
- Carrier Phase Estimation (CPE) required (e.g. Viterbi-Viterbi algorithm)
- Targets compensation of linear impairments

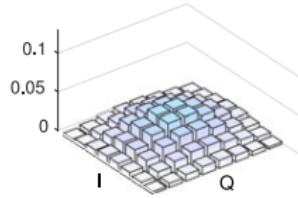
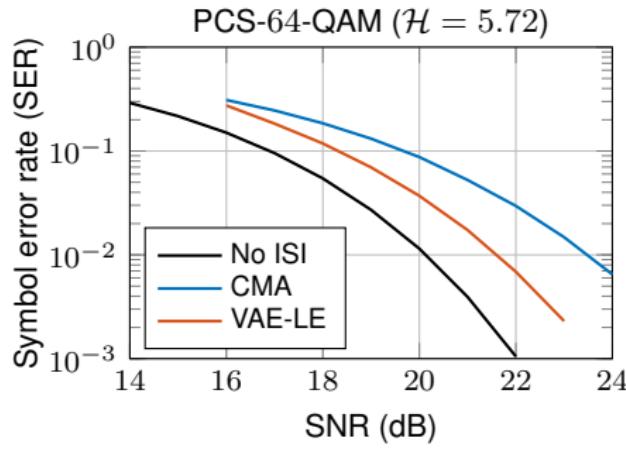
# Simulation Results: Conventional QAM



- Single polarization
- 2 samples per symbol
- Batch-wise update with Adam optimizer

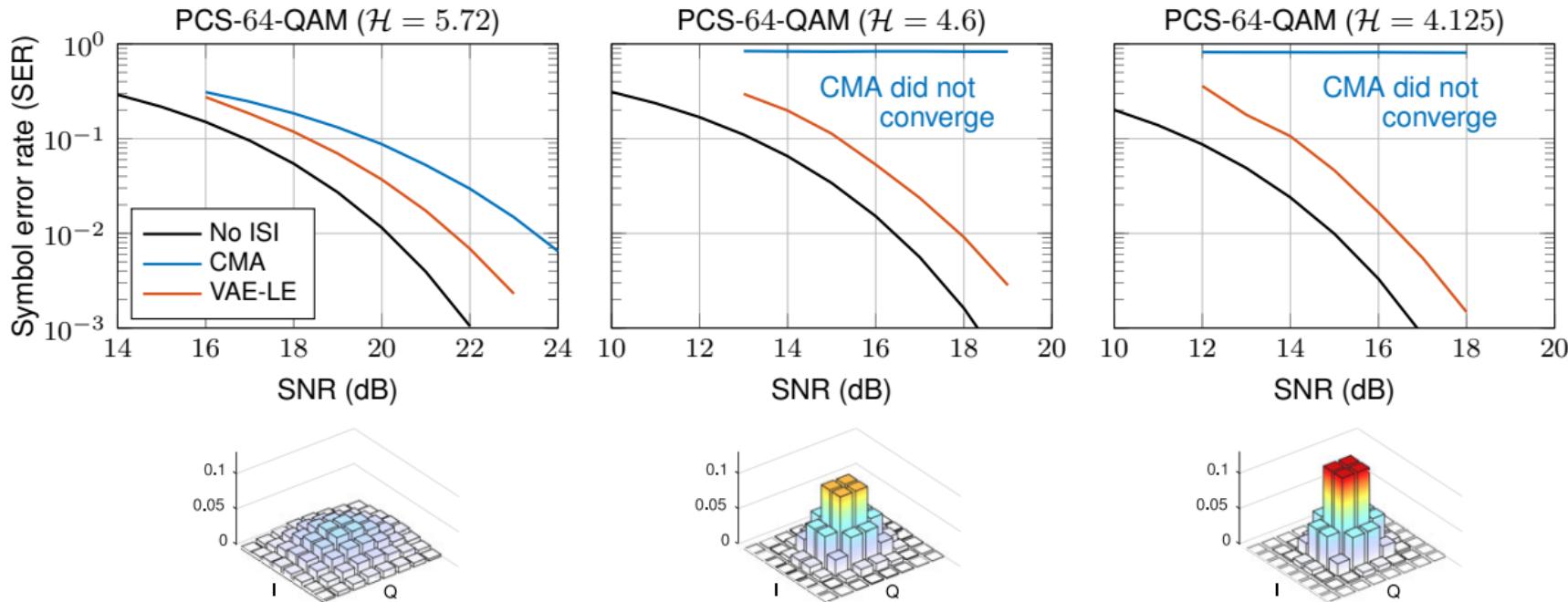
[LBS21] V. Lauinger, F. Buchali, L. S., "Blind equalization for coherent optical communications based on variational inference," *Proc. Adv. Photon. Congr., SPPCOM*, Jul. 2021

# Simulation Results: PCS-QAM



[LBS21] V. Lauinger, F. Buchali, L. S., "Blind equalization for coherent optical communications based on variational inference," *Proc. Adv. Photon. Congr., SPPCOM*, Jul. 2021

# Simulation Results: PCS-QAM



[LBS21] V. Lauinger, F. Buchali, L. S., "Blind equalization for coherent optical communications based on variational inference," *Proc. Adv. Photon. Congr., SPPCOM*, Jul. 2021

# Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Recurrent Neural Networks
- Application 2: In-Situ Optimization of Transmitters
- Application 3: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

# Conclusion and Outlook

- Neural networks are and will be **integral part** of optical communication systems
- Machine learning enables design of **novel waveforms and modulation formats**
- Machine learning enables **novel applications** that are difficult with traditional approaches, e.g. distance agnostic transceivers
- Generative networks can be used to **model** (part of) the communication system for model-based optimization
- Powerful tools from the machine learning community can be used to design **novel receiver algorithms**

# Questions?



European Research Council  
Established by the European Commission



GEFÖRDERT VOM  
Bundesministerium  
für Bildung  
und Forschung

Part of this work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001899) and part of it was carried out in the framework of the CELTIC-NEXT project AI-NET-ANTILLAS (C2019/3-3) and received funding from the German Federal Ministry of Education and Research (BMBF) under grant agreement 16KIS1316.