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

Laurent Schmalen, Boris Karanov, Vincent Lauinger

Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

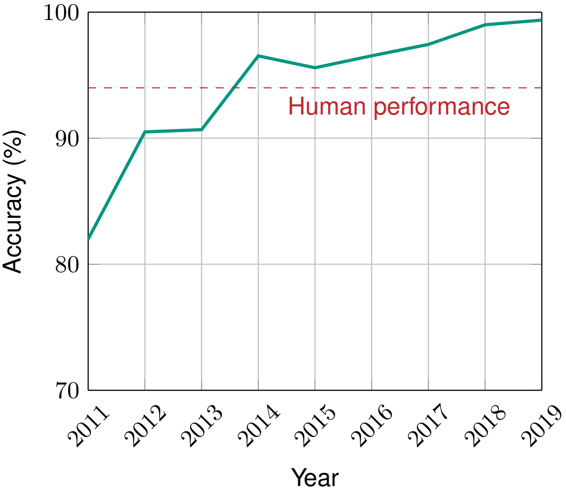
Machine Learning Success: Handwritten Digit Recognition

- **The MNIST database**[1] [2]

- **Error rates of 0.23%** can be achieved today

[1] Available online at: `http://yann.lecun.com/exdb/mnist/`
[2] Image source: `https://upload.wikimedia.org/wikipedia/commons/2/27/MnistExamples.png`

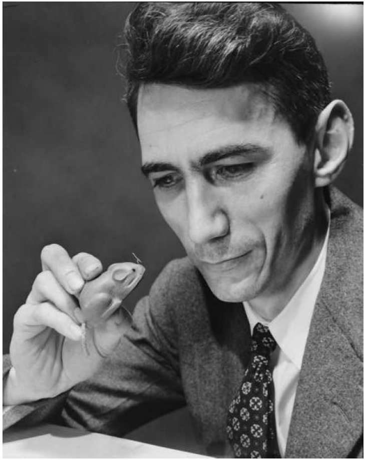
Machine Learning Success: Image Recognition



- Image recognition accuracy on CIFAR-10 dataset[3]
- 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

[3] Accuracy data from `https://benchmarks.ai/cifar-10`

Machine Learning for Communications



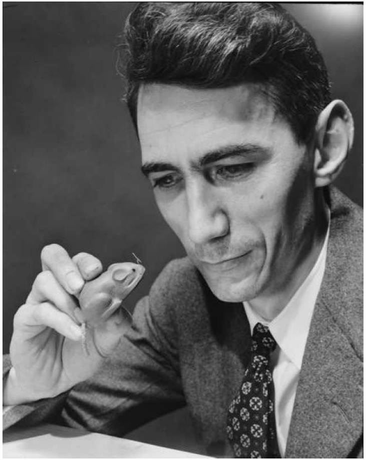
- 70 years of communications engineering have brought us close to the Shannon limit[4] (Shannon capacity)

- **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

[4] Image source: `https://commons.wikimedia.org/wiki/File:Claude_Shannon_1776.jpg`

Machine Learning for Communications

Karlsruher Institut für Technologie



- 70 years of communications engineering have brought us close to the Shannon limit[4] (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

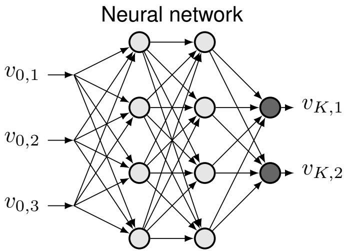
[4] Image source: `https://commons.wikimedia.org/wiki/File:Claude_Shannon_1776.jpg`

Neural Networks: A Whirlwind Tour

Feed-Forward Neural Networks (FFNN)

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

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



Neural Networks: A Whirlwind Tour

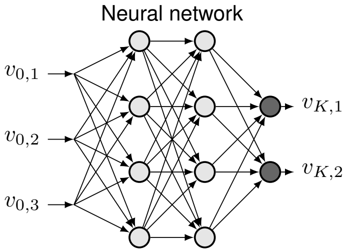
Feed-Forward Neural Networks (FFNN)

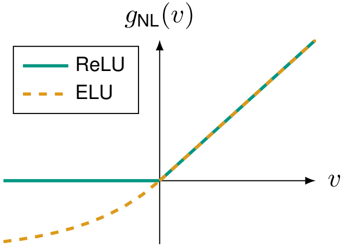
- Maps an input vector $\mathbf{v}_0 = \begin{pmatrix} v_{0,1} & \ldots & v_{0,M} \end{pmatrix}$ to an output vector $\mathbf{v}_K = \begin{pmatrix} v_{K,1} & \ldots & v_{K,n} \end{pmatrix} = f_{\text{NN}}(\mathbf{v}_0)$
- The neural network (NN) is a composed function consisting of layers, where each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

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

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

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





Neural Networks: A Whirlwind Tour

Feed-Forward Neural Networks (FFNN)

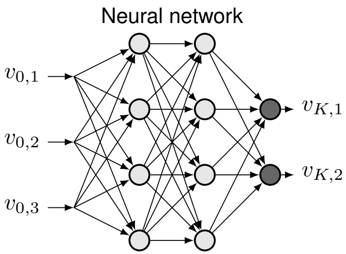
- Maps an input vector $\mathbf{v}_0 = \begin{pmatrix} v_{0,1} & \dots & v_{0,M} \end{pmatrix}$ to an output vector $\mathbf{v}_K = \begin{pmatrix} v_{K,1} & \dots & v_{K,n} \end{pmatrix} = f_{\text{NN}}(\mathbf{v}_0)$
- The neural network (NN) is a composed function consisting of layers, where each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

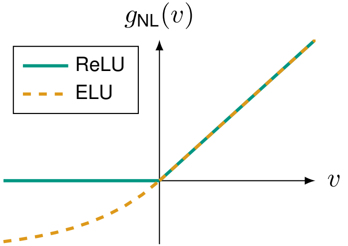
$$\mathbf{v}_k = g_{\text{NL},k}(\mathbf{W}_k \mathbf{v}_{k-1} + \mathbf{b}_k), \qquad k = 1, \dots, K$$

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

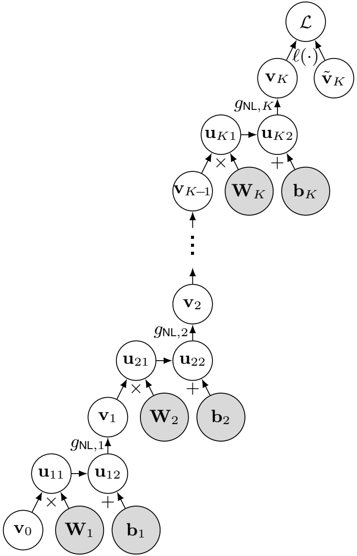
$$\mathbf{x} = g_{\text{NL}}(\mathbf{v}) = g_{\text{ReLU}}(\mathbf{v}) \qquad \text{with } x_i = \max(0, v_i)$$

- **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



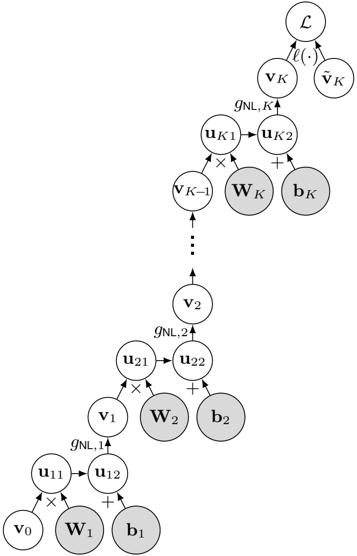
Deep Learning

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

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

  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**

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

  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**

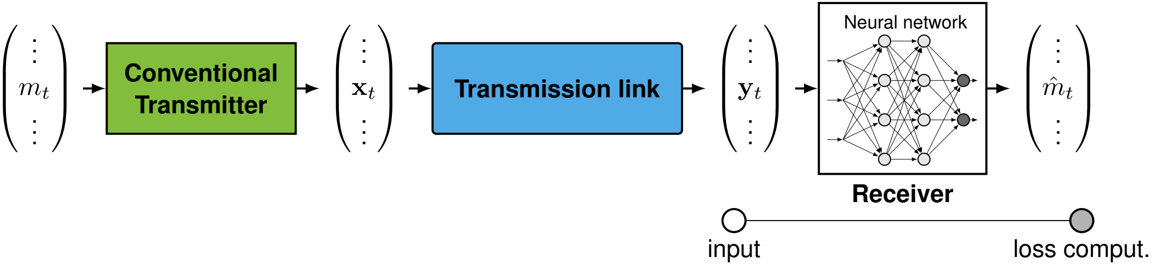
$$\mathbf{W}_k \leftarrow \mathbf{W}_k - \epsilon \nabla_{\mathbf{W}_k} \mathcal{L}\left(\{\mathbf{W}_k, \mathbf{b}_k\}_K\right)$$
$$\mathbf{b}_k \leftarrow \mathbf{b}_k - \epsilon \nabla_{\mathbf{b}_k} \mathcal{L}\left(\{\mathbf{W}_k, \mathbf{b}_k\}_K\right)$$

- 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 Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

Receiver Processing Using Neural Networks

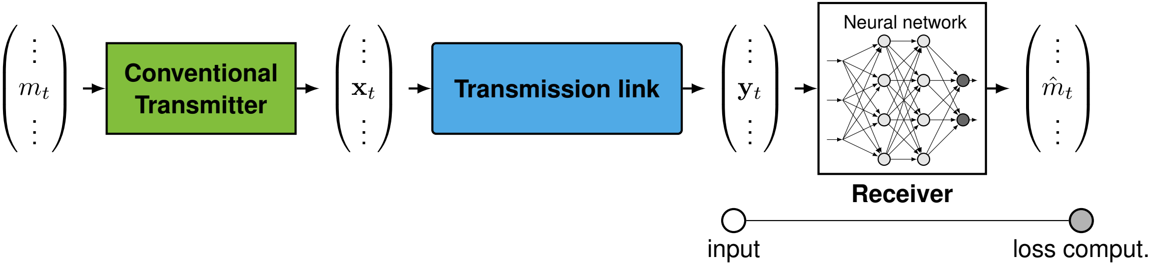


- Transmission of known sequence of messages (e.g., pilot) m_t
- Based on received data ($\mathbf{y}_t$), try to recover m_t as closely as possible
- Training dataset is database of received signals and associated symbols $\mathcal{S} = \{(\mathbf{y}_t, m_t) : t = 1, \dots N\}$
- Proposed for optical 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

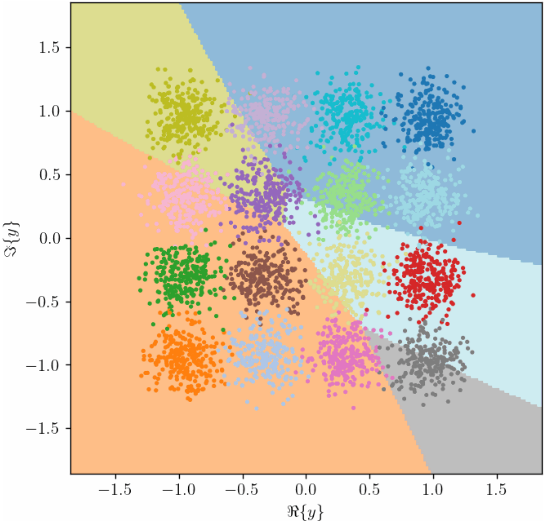


- Transmission of known sequence of messages (e.g., pilot) m_t
- Based on received data ($\mathbf{y}_t$), try to recover m_t as closely as possible
- Training dataset is database of received signals and associated symbols $\mathcal{S} = \{(\mathbf{y}_t, m_t) : t = 1, \dots N\}$
- Proposed for optical 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

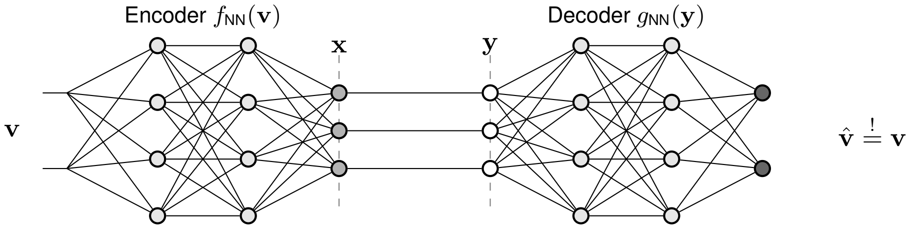
Example: Deep NN Detection in AWGN Channel



- Implementation using PyTorch[a]
- Source code available online[b]

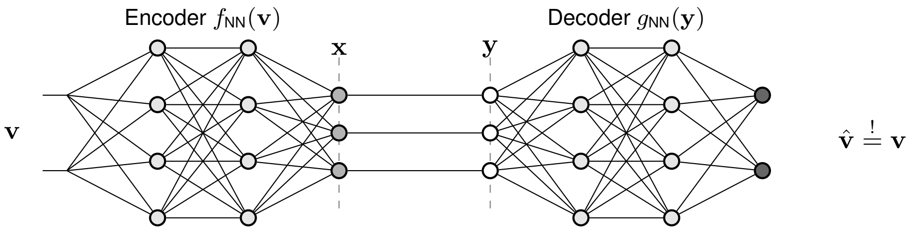
[a] http://pytorch.org
[b] https://github.com/kit-cel/HHI_SummerSchool_2021

Auto-encoders – Basic Concept



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

Auto-encoders – Basic Concept



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

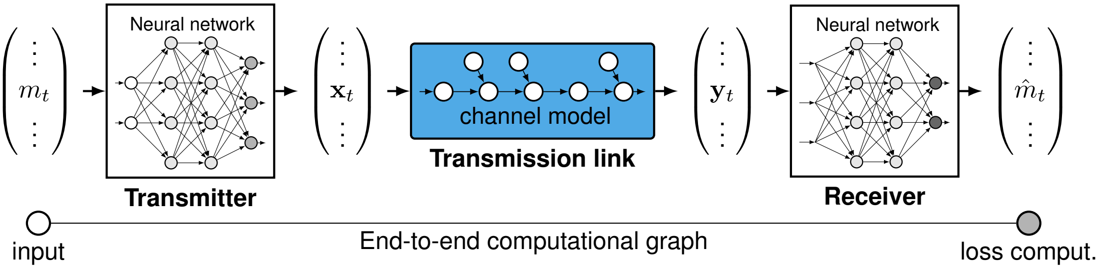
- Resembles a communication system (Goal: reproduce information as close as possible)
- Can we use auto-encoders to design communication systems? **[OKC16]**, **[OH17]**

[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+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

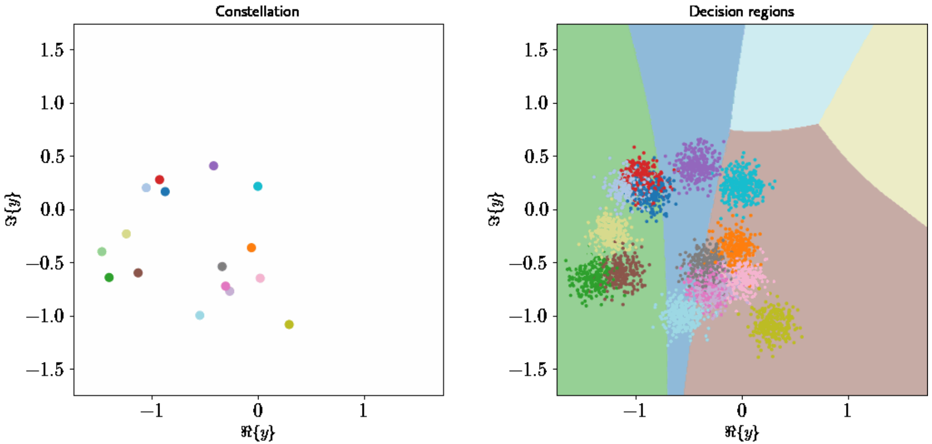
Communication System as Auto-Encoder



- 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 $\mathbf{y}_t = \mathbf{x}_t + \mathbf{n}_t$ with $\nabla_{\mathbf{W}} \mathbf{y}_t = \nabla_{\mathbf{W}} \mathbf{x}_t$

Example: Auto-Encoder in AWGN Channel

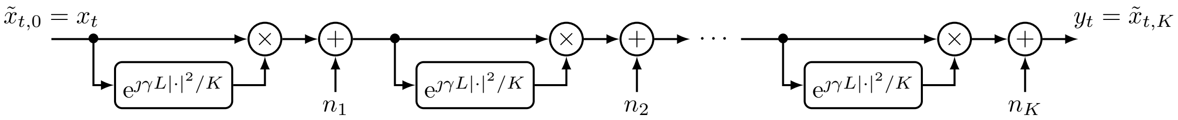


- Implementation using PyTorch
- Source code available online[5]

[5] https://github.com/kit-cel/HHI_SummerSchool_2021

Example: Zero-Dispersion Optical Fiber

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



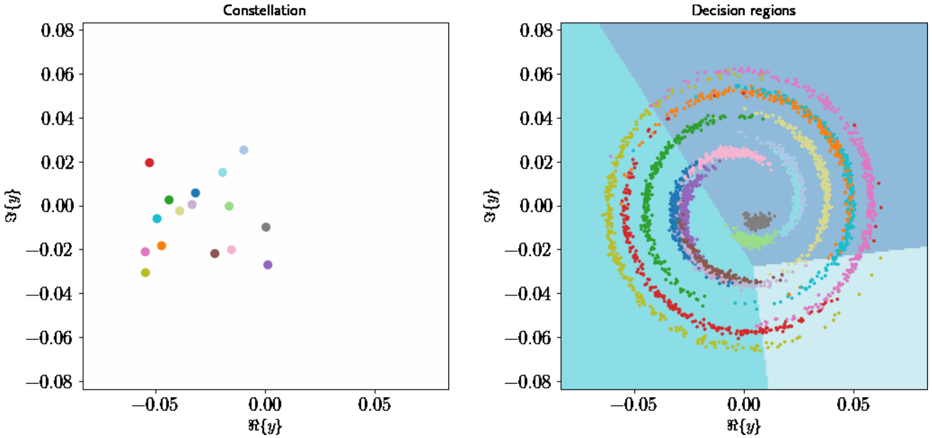
- Let $\tilde{x}_{x,0} = x_t$ and recursively compute

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

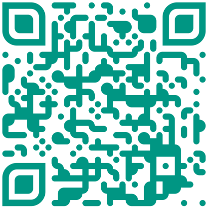
  where $n_{i+1} \sim \mathcal{CN}(0, P_N/K)$ is complex, circularly symmetric Gaussian noise (assuming ideal distributed amplification) and K the number of steps (ideally, $K \to \infty$)
- Computation graph is fully **differentiable**

[LHG+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[6]



[6] https://github.com/kit-cel/HHI_SummerSchool_2021

Auto-Encoders in Optical Communications

Karlsruher Institut für Technologie

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

[KCT+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+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+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

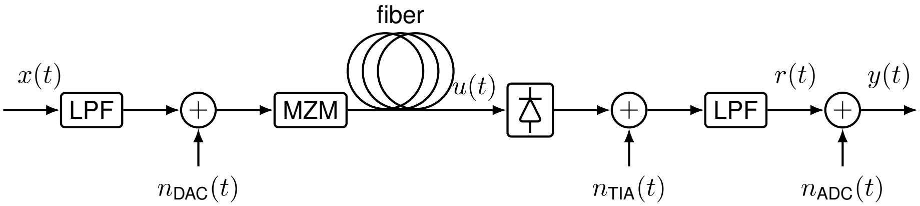
[JEY+18] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities,", *Proc. European Conference on Optical Communication (ECOC)*, 2018

[JYZ19] R. T. Jones, M. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," arXiv:1907.08535, 2019

[GAC+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



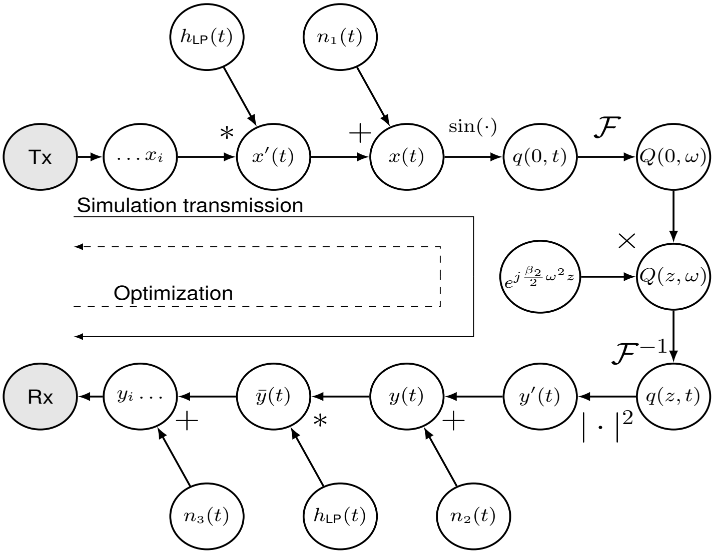
- IM/DD (intensity modul., direct detect.) ubiquitous in access networks and data center interconnects

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

- Fiber adds only chromatic dispersion (no nonlinearities) with $H(\omega, z) = \exp\left(\jmath\frac{\beta_2}{2}z\omega^2\right)$
- 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

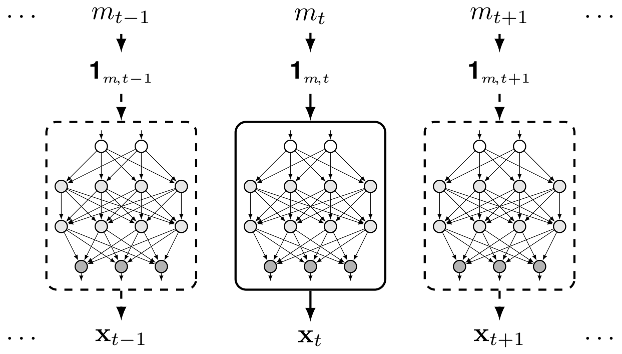


- Chromatic dispersion is added in frequency domain
- Fourier transform is linear and differentiable
- Channel model can be directly implemented in machine learning software

Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

First Transmitter Attempt Using FFNN

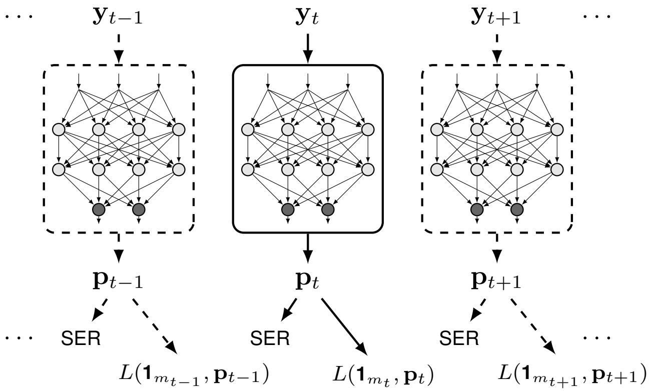


- Transform message m_t (from an alphabet containing M messages (e.g., a modulation alphabet)) into **one-hot**[7] vector $1_{m,t}$
- FNN encodes $1_{m,t}$ into n **oversampled waveform samples** $\mathbf{x}_t$
- Apply low-pass-filter (LPF, part of the channel model) to **smear** the waveforms

[KCT+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

[7] A one-hot vector $\mathbf{1}_{m,t}$ of length M contains only zeros except a single "1" at position m_t

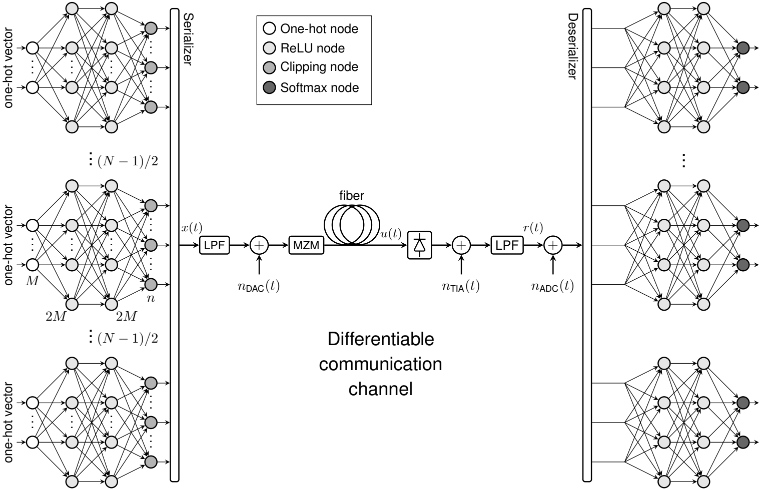
First Receiver Attempt Using FFNN



- Receiver input is chopped into blocks $\mathbf{y}_t$ 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

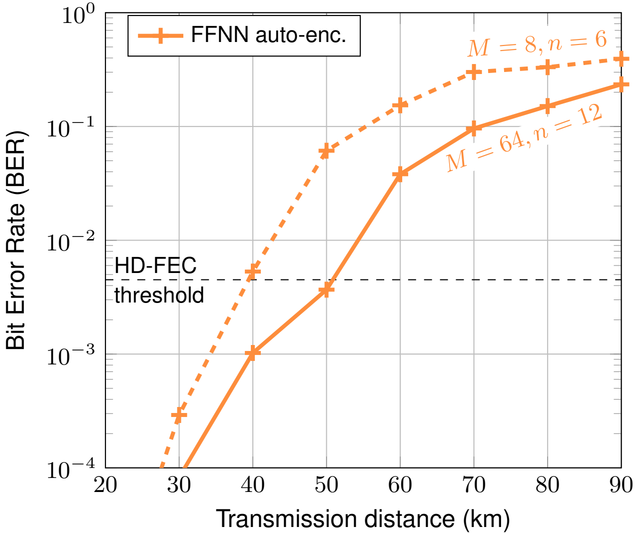
$$L(\mathbf{1}_{m_t}, \mathbf{p}_t) = -\log(p_{t,m_t})$$

Putting Everything Together



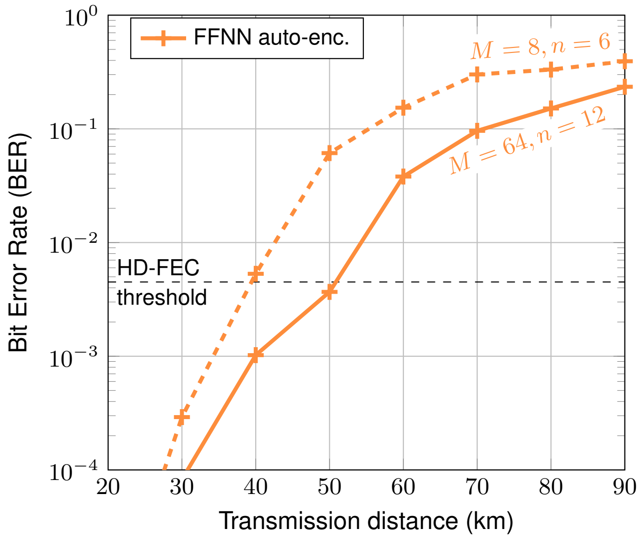
[KCT+18] B. Karanov, M. Chagnon, F. Thouin, T. Eriksson, H. Bülow, D. Lavery, P. Bayvel, L. S., "End-to-end Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, Oct. 2018

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 the NN size (i.e., M and thus n) leads to unfeasibly large networks

Simulation Results

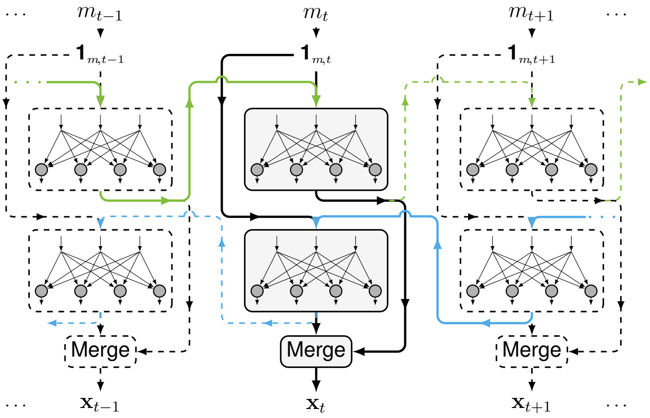


- State-of-the-art IM/DD system with 42 Gb/s net rate
- For small alphabets (M small) and neural networks, reach is limited
- Increasing the alphabet and the neural networks enables extra reach (multidimensional constellations) as more dispersion can be compensated for
- Increasing the the NN size (i.e.,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

Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

Bidirectional Recurrent NN Based Transmitter

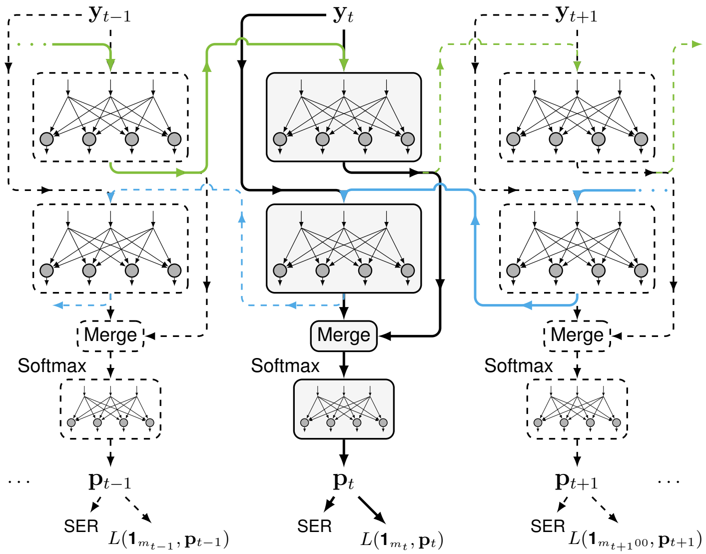


- Transmitter based on **bidirectional recurrent neural networks (BRNN)** [FG18] to account for memory due to chromatic dispersion
- Current message m_t jointly processed with previous and future messages vie **bidirectional recurrent neural network (BRNN)**
- Complexity gain using **small** networks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

[KLB+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

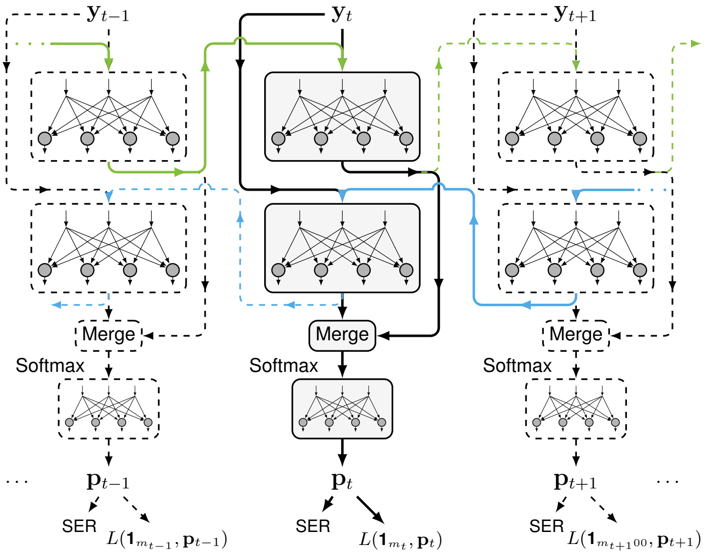
[FG18] N. Farsad, A. Goldsmith, "Neural network detection of data sequences in communication systems," *IEEE Trans. Signal Process*, 2018

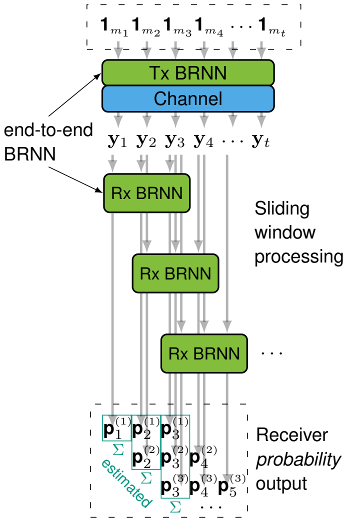
Bidirectional Recurrent NN Based Receiver



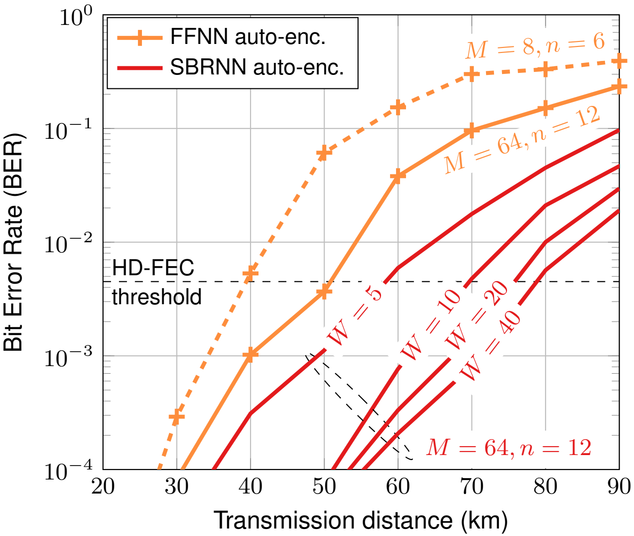
- Receiver structure similar to transmitter
- Additional **softmax** layer to generate probability vectors
- We use **sliding window** processing at the receiver to trade complexity and latency with performance
- A **window** of W stages is jointly processed at the receiver [FG18]

Bidirectional Recurrent NN Based Receiver





Simulation Results



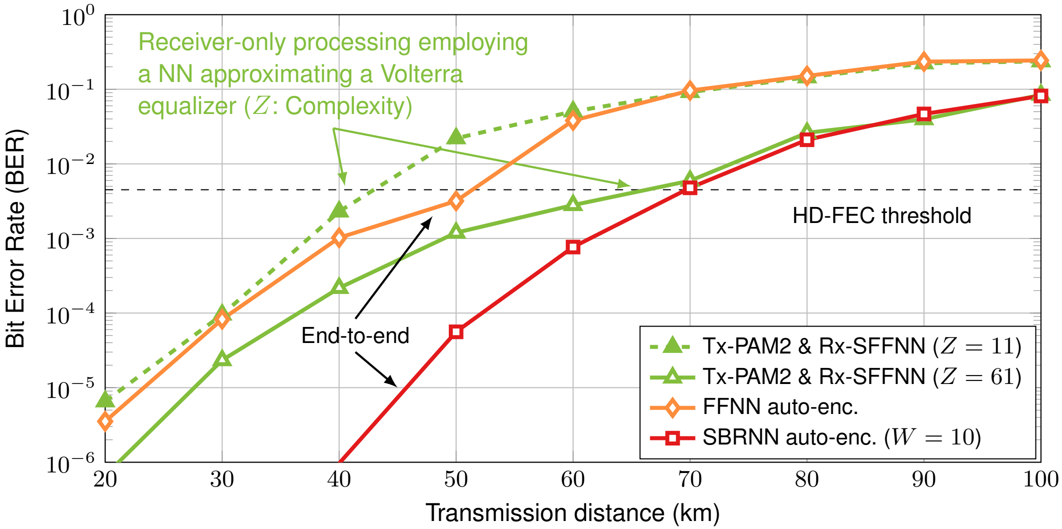
- SBRNN outperforms simple FFNN implementation at significantly **lower complexity**
- 20 km distance gain for $W = 10$
- SBRNN can **outperform** MLSD if complexity is constrained **[KLA+19]**

- Now: Comparison with pure receiver NN processing approximating a Volterra equalizer **[Lyu15]**

[KLA+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

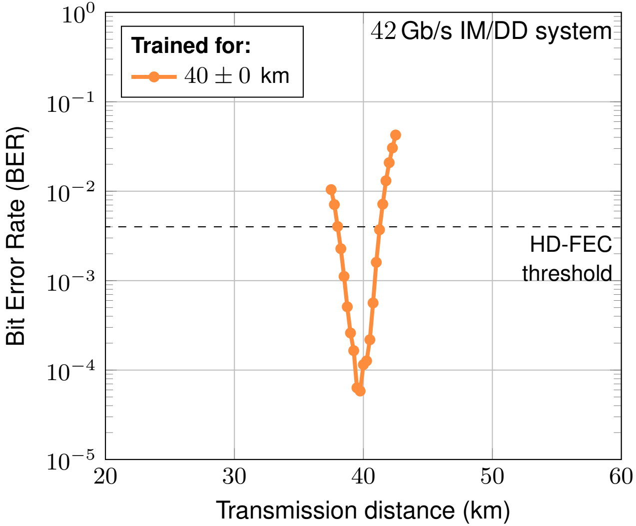
Simulation Results (2)



Overview

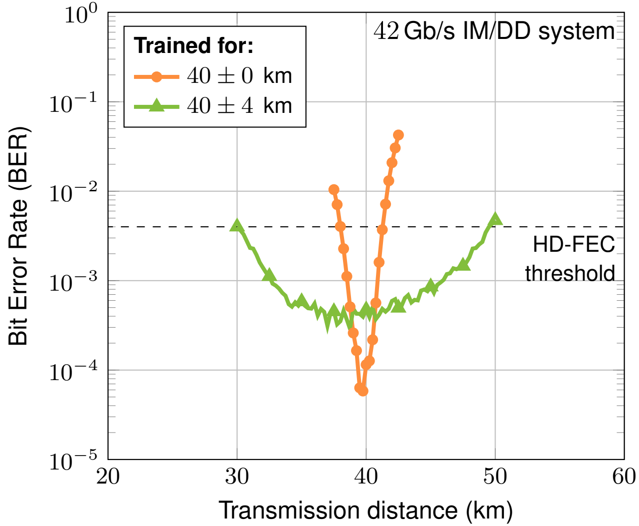
- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

Training Generalization: Simulation Results



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

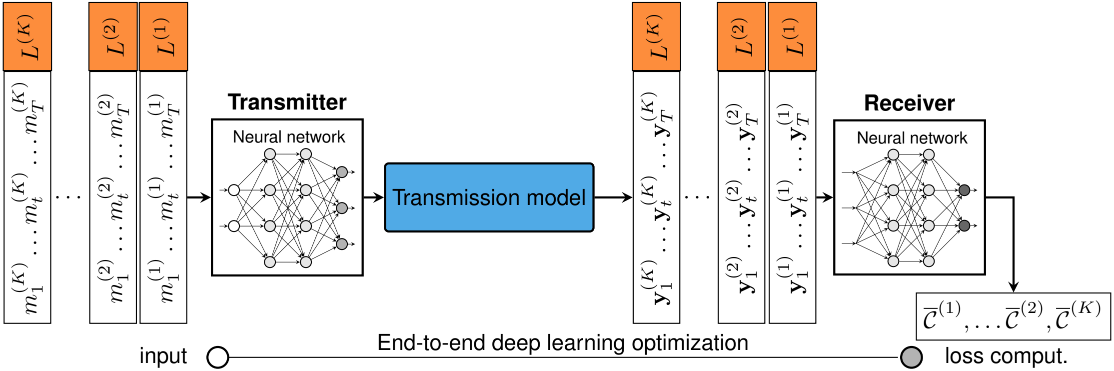
Training Generalization: Simulation Results



- **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+18]
- Big step towards **distance agnostic** transceivers

[KCT+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

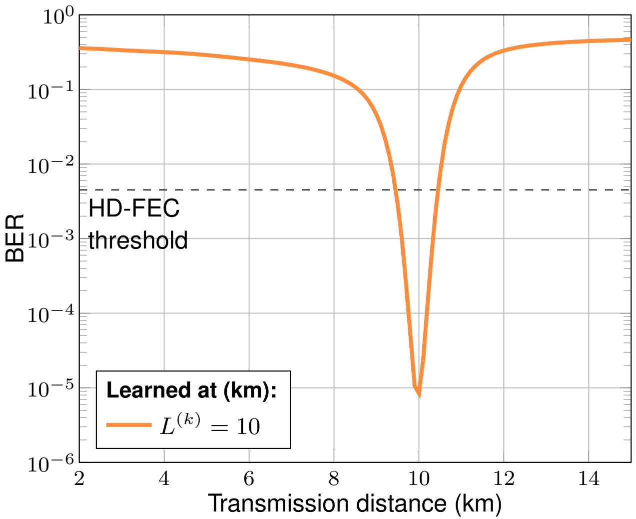
Strategies for Distance-Agnostic Transceivers

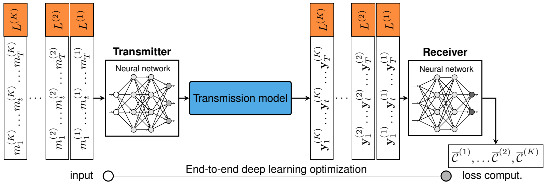


- Loss function

$$\overline{\mathcal{C}}^{(k)}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \ell\left(m_t^{(k)}, f_{\mathrm{Rx}}\left(\mathcal{H}_{\mathrm{ch}}\left\{\ldots, f_{\mathrm{Tx}}\left(m_t^{(k)}\right), \ldots\right\}_{L^{(k)}}\right)\right) = \frac{1}{T}\sum_{t=1}^{T} \ell\left(m_t^{(k)}, f_{\mathrm{Rx}}\left(\mathrm{y}_t^{(k)}\right)\right)$$

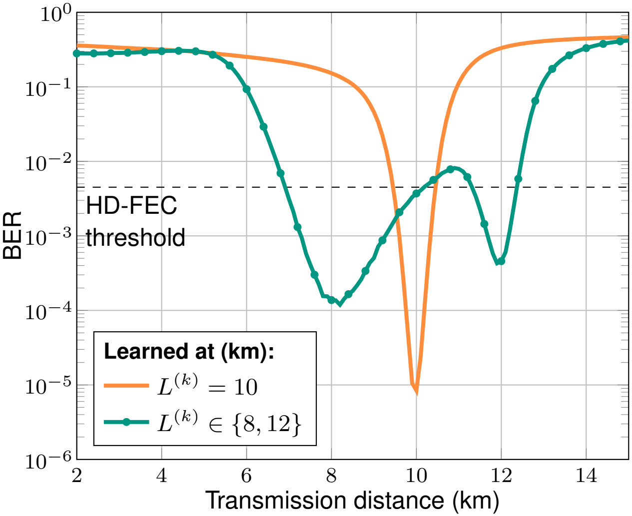
Strategies for Distance-agnostic Transceivers

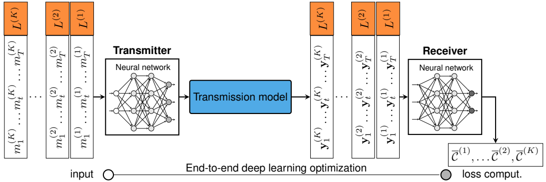




- 100 Gb/s IM/DD system
- Transmitter optimized at fixed distance
- Rapid performance degradation as distance changes

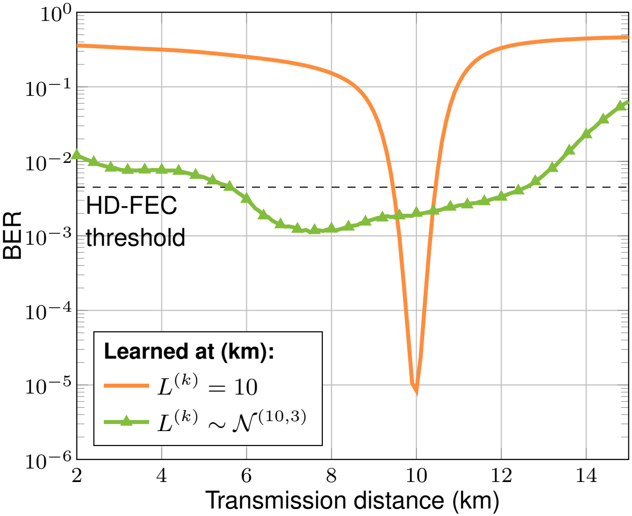
Strategies for Distance-agnostic Transceivers

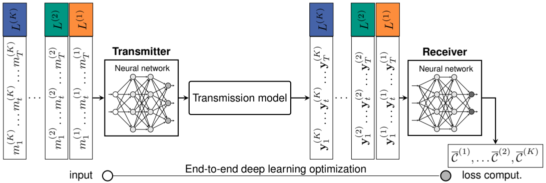




- 100 Gb/s IM/DD system
- Transmitter optimized at two distances $L^{(k)} \in \{8\,\mathrm{km}, 12\,\mathrm{km}\}$
- Strong BER variation but wider allowed range of operation

Strategies for Distance-agnostic Transceivers



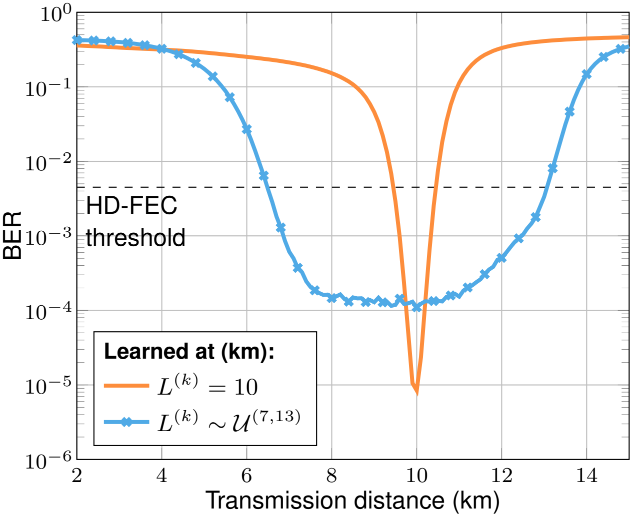


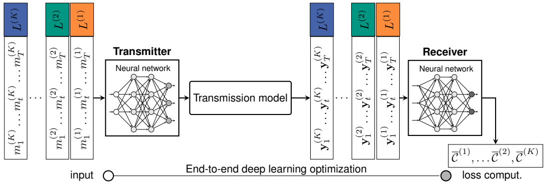
- 100 Gb/s IM/DD system
- Training examples with different accumulated link dispersions
- Gaussian distribution

[CKS18] M. Chagnon, B. Karanov, L. S., "Experimental demonstration of dispersion tolerant end-to-end deep learning-based IM-DD transmission system," *Proc. ECOC*, 2018

[KLA21] B. Karanov, L. S., A. Alvarado, "Distance-Agnostic Auto-Encoders for Short Reach Fiber Communications," *Proc. OFC*, 2021

Strategies for Distance-agnostic Transceivers



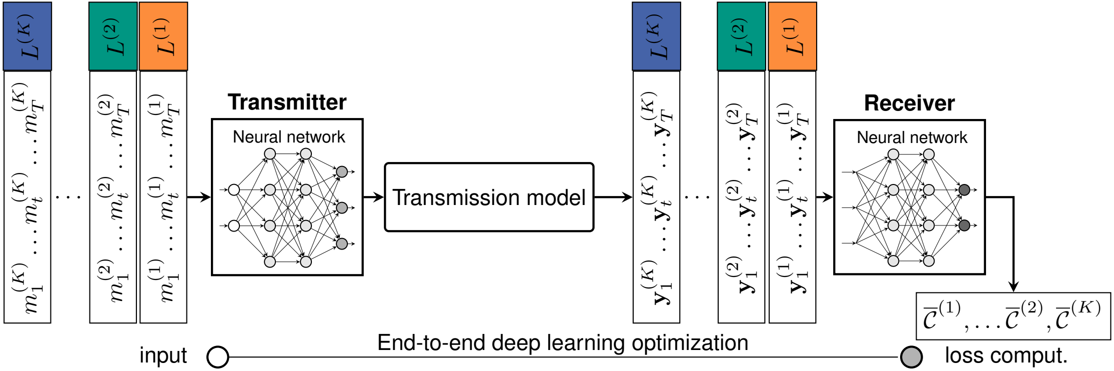


- 100 Gb/s IM/DD system
- Training examples with different accumulated link dispersions
- Uniform distribution

[CKS18] M. Chagnon, B. Karanov, L. S., "Experimental demonstration of dispersion tolerant end-to-end deep learning-based IM-DD transmission system," *Proc. ECOC*, 2018

[KLA21] B. Karanov, L. S., A. Alvarado, "Distance-Agnostic Auto-Encoders for Short Reach Fiber Communications," *Proc. OFC*, 2021

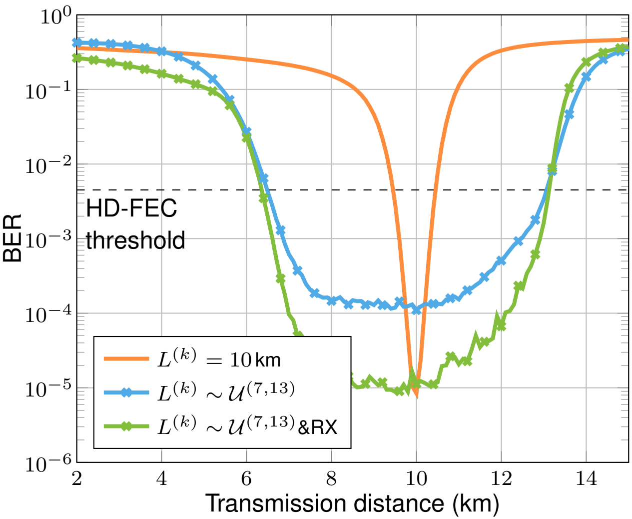
Enhanced Distance-agnostic Transceivers

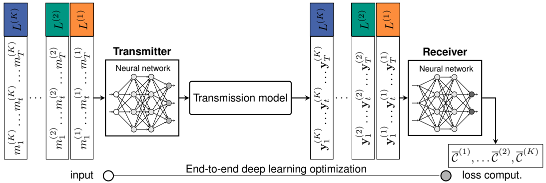


- Receiver has access to length $L^{(k)}$
- Loss function

$$\overline{\mathcal{C}}^{(k)}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \ell\left(m_t^{(k)}, f_{\mathrm{Rx}}\left(\mathcal{H}_{\mathrm{ch}}\left\{\ldots, f_{\mathrm{Tx}}\left(m_t^{(k)}\right), \ldots\right\}_{L^{(k)}}, L^{(k)}\right)\right) = \frac{1}{T}\sum_{t=1}^{T} \ell\left(m_t^{(k)}, f_{\mathrm{Rx}}\left(\mathrm{y}_t^{(k)}, L^{(k)}\right)\right)$$

Enhanced Distance-agnostic Transceivers





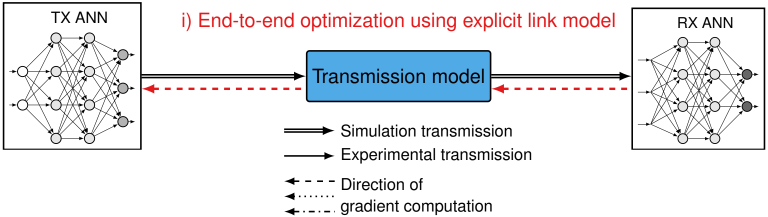
- 100 Gb/s IM/DD system
- Training examples with different accumulated link dispersions
- Receiver has knowledge of link distance (or coarse estimate thereof)
- Significant BER reduction

[KLA21] B. Karanov, L. S., A. Alvarado, "Distance-Agnostic Auto-Encoders for Short Reach Fiber Communications," *Proc. OFC*, 2021

Overview

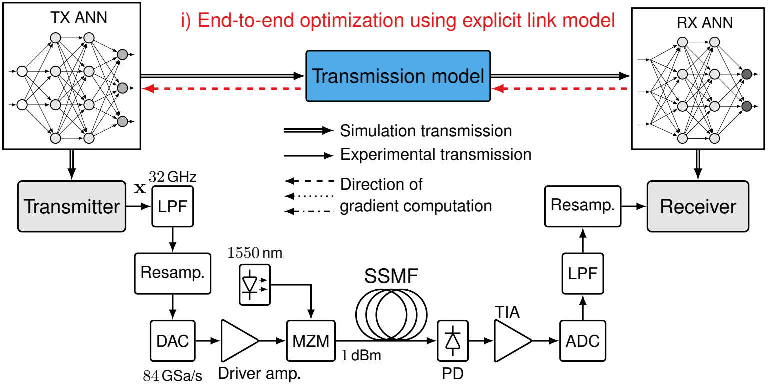
- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: Blind Equalization Using Variational Autoencoders
- Conclusions and Outlook

Experimental Demonstration



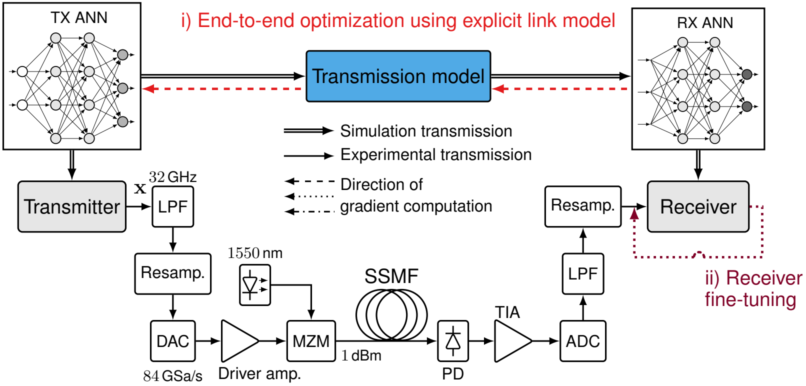
- Optimization of transmitter and receiver using differentiable model

Experimental Demonstration



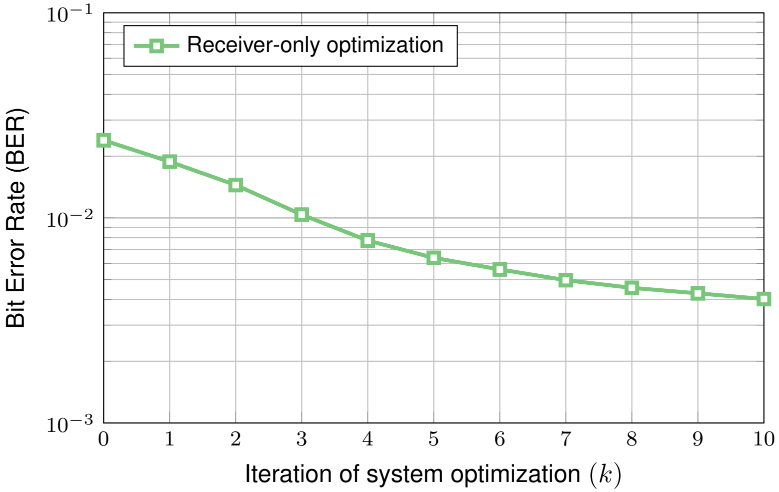
- Optimization of transmitter and receiver using differentiable model
- Mismatch between model and experiment

Experimental Demonstration



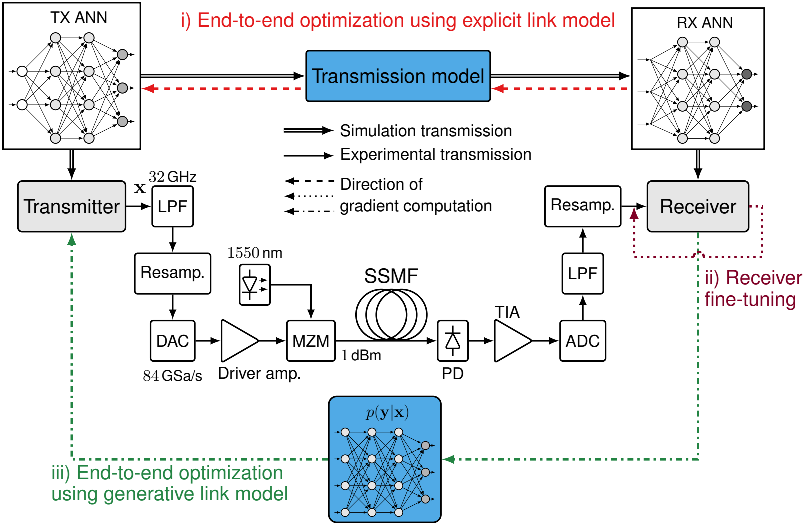
- Optimization of transmitter and receiver using differentiable model
- Mismatch between model and experiment
- Receiver fine-tuning via supervised learning

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)

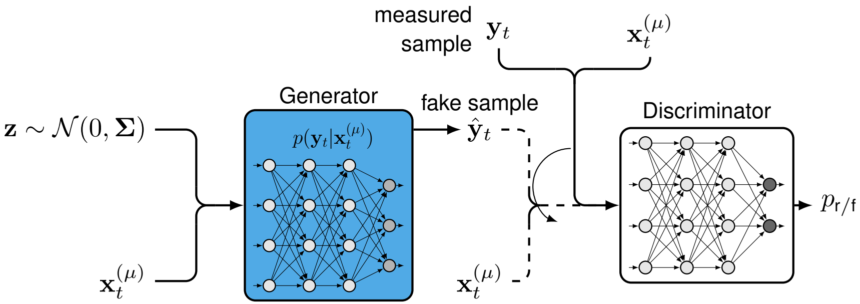
Experimental Demonstration



- 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

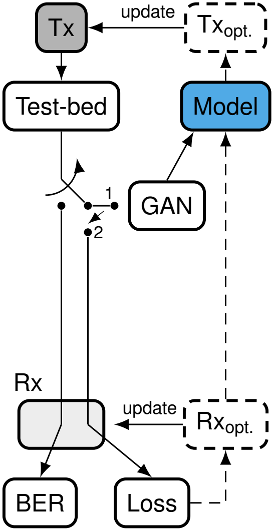
- Neural network that behaves like a random number generator with the same properties as the channel
- Combining $\mathbf{z} \sim \mathcal{N}(0, \mathbf{\Sigma})$ and the channel input $\mathbf{x}_t$, the generator generates a channel output distributed according to $p(\mathbf{y}|\vec{\mathbf{x}}_t)$



[GPM+14] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Proc. NeurIPS*, 2014

[KCA+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,” *Proc. OFC*, 2020

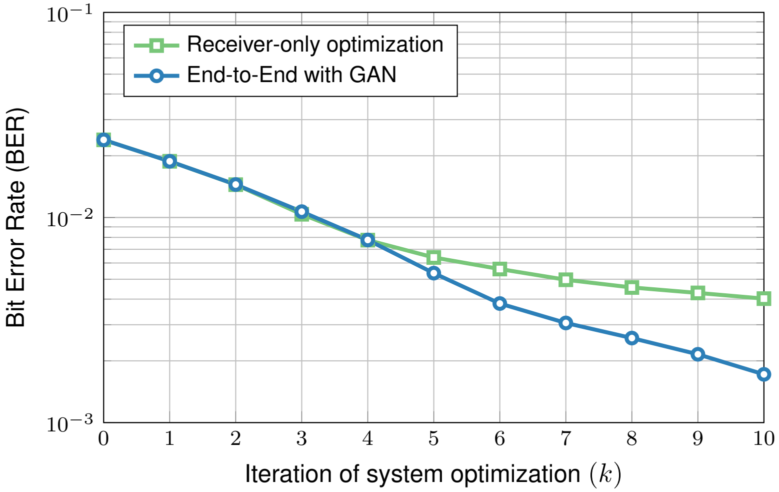
Channel Modelling Using a Conditional GAN



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

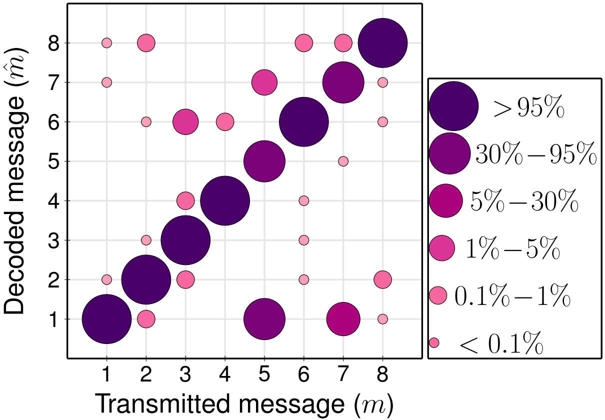
Experimental Results



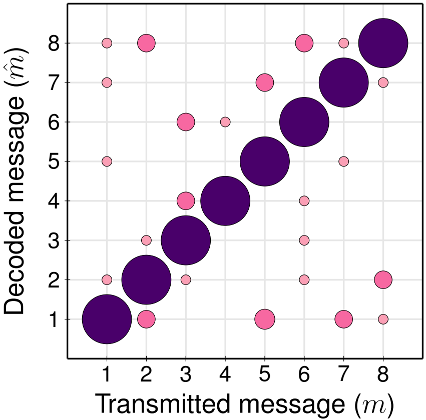
- Improvement of symbol and bit error probabilities
- More iterations will even be more beneficial
- (Had to stop iterating due to unavailability of experimental setup)

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 adverserial nets (GANs)

- Demonstrated often ([YLJ+18], [ORW19], [SD19], [KCA+20], [DHC+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+], 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 *Proc. GLOBECOM*, 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 *Proc. of 2019 International Conference on Computing, Networking and Communications*, 2019

[SD19] A. Smith and J. Downey, "A communication channel density estimating generative adversarial network," *NASA Technical Reports* 2019

[KCA+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," *Proc. OFC*, 2020

[DHC+20] S. Dörner, M. Henninger, S. Cammerer, S. ten Brink, "WGAN-based Autoencoder Training Over-the-air," *Proc. SPAWC*, 2020

[AH19] F. Ait Aoudia, J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE J. Selected Areas in Communications*, 2019

[GAH19] M. Goutay, F. Ait Aoudia, J. Hoydis, "Deep reinforcement learning autoencoder with noisy feedback," *Proc. WiOPT*, 2019

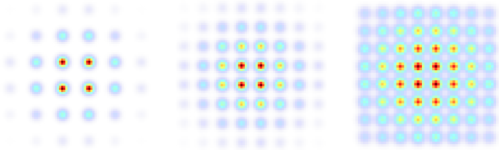
[JYdR+] O. Jovanovic, M. Yankov, F. da Ros, D. Zibar, "Gradient-free training of autoencoders for non-differentiable communication channels," *Journal of Lightwave Technology*, 2021

Overview

- Introduction
- Application 1: Waveform Optimization for Short-Reach Optical Communications
  - Simple Transmitter Using Feed-Forward Neural Networks
  - Improved Transceivers Using Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- **Application 4: Blind Equalization Using Variational Autoencoders**
- Conclusions and Outlook

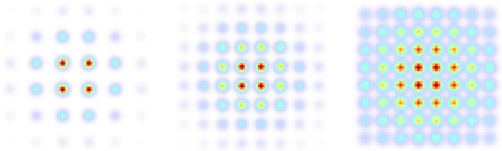
Motivation

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

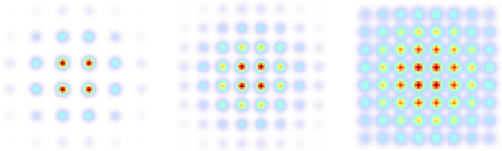


- **State-of-the-art**: Constant modulus algorithm (CMA) / multi modulus algorithm (MMA) **[Yan02]**
  - CMA: ill-matched for higher-order modulation formats and probabilistic shaping
  - MMA: high implementation complexity and low convergence rate

[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

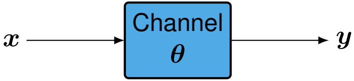
- Great demand for **blind** equalizers
  - + saved data rate can increase FEC overhead or data throughput
  - ϟ have to be flexible and adaptive
  - ϟ must support higher-order modulation formats and probabilistic constellation shaping (PCS)



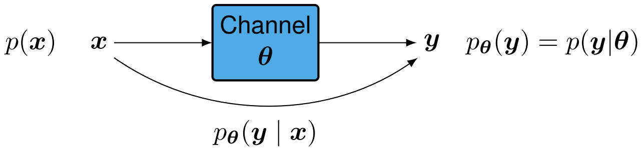
- **State-of-the-art**: Constant modulus algorithm (CMA) / multi modulus algorithm (MMA) **[Yan02]**
  - – CMA: ill-matched for higher-order modulation formats and probabilistic shaping
  - – MMA: high implementation complexity and low convergence rate
- **Ideally**: Maximum likelihood detector, e.g.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

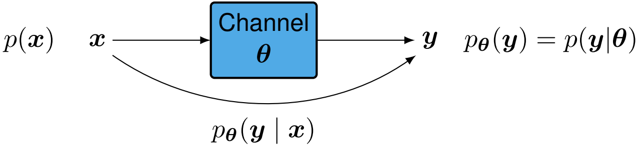


System and Problem Description



Goal: $$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \; p_{\boldsymbol{\theta}}(\boldsymbol{y})$$

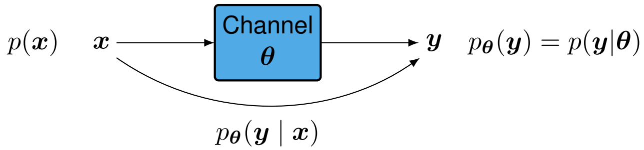
System and Problem Description



Goal: $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \; p_{\boldsymbol{\theta}}(\boldsymbol{y})$

Given: $p(\boldsymbol{x})$, $\quad p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{CN}\left(\boldsymbol{h} * \boldsymbol{x},\, \sigma_{\mathrm{w}}^2 \boldsymbol{I}_N\right)$

System and Problem Description

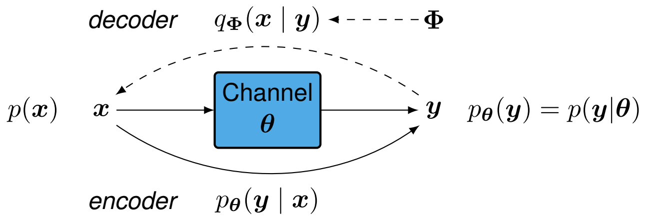


Goal: $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \; p_{\boldsymbol{\theta}}(\boldsymbol{y})$

Given: $p(\boldsymbol{x})\,, \qquad p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{CN}\left(\boldsymbol{h} * \boldsymbol{x},\, \sigma_{\mathrm{w}}^2 \boldsymbol{I}_N\right)$

Problem: $p_{\boldsymbol{\theta}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{x})\, p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x})$ $\rightarrow$ intractable

System and Problem Description



Goal: $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \; p_{\boldsymbol{\theta}}(\boldsymbol{y})$

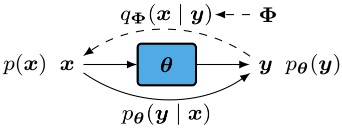
Given: $p(\boldsymbol{x})\,, \qquad p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{CN}\left(\boldsymbol{h} * \boldsymbol{x},\, \sigma_{\mathrm{w}}^2 \boldsymbol{I}_N\right)$

Problem: $p_{\boldsymbol{\theta}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{x})\, p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}) \quad \rightarrow$ 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)

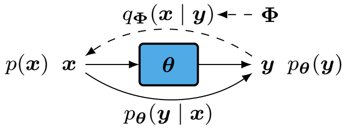


- **ML criterion**: $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\ p_\theta(\boldsymbol{y}) = \arg\max_{\boldsymbol{\theta}}\ \ln p_\theta(\boldsymbol{y})$

$$\ln p_{\boldsymbol{\theta}}(\boldsymbol{y}) = \ldots = \underbrace{\sum_{\boldsymbol{x}} q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})}{q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=:\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\Phi},\boldsymbol{y})} + \underbrace{\sum_{\boldsymbol{x}} q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}{p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=\mathrm{D}_{\mathrm{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{y}))}$$

[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**: $\hat{\theta} = \arg\max_{\boldsymbol{\theta}} \ p_\theta(\boldsymbol{y}) = \arg\max_{\boldsymbol{\theta}} \ \ln p_\theta(\boldsymbol{y})$

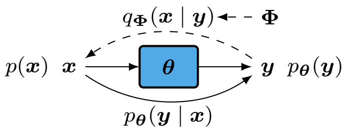
$$\ln p_{\boldsymbol{\theta}}(\boldsymbol{y}) = \ldots = \underbrace{\sum_{\boldsymbol{x}} q_{\mathbf{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})}{q_{\mathbf{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=:\mathcal{L}(\boldsymbol{\theta},\mathbf{\Phi},\boldsymbol{y})} + \underbrace{\sum_{\boldsymbol{x}} q_{\mathbf{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{q_{\mathbf{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}{p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=\mathrm{D}_{\mathsf{KL}}(q_{\mathbf{\Phi}}(\boldsymbol{x}|\boldsymbol{y})\|p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{y}))}$$

- Kullback-Leibler divergence $\mathrm{D}_{\mathsf{KL}}(q_{\mathbf{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})) \ \geq \ 0$

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

[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**: $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\ p_\theta(\boldsymbol{y}) = \arg\max_{\boldsymbol{\theta}}\ \ln p_\theta(\boldsymbol{y})$

$$\ln p_{\boldsymbol{\theta}}(\boldsymbol{y}) = \ldots = \underbrace{\sum_{\boldsymbol{x}} q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})}{q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=:\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\Phi},\boldsymbol{y})} + \underbrace{\sum_{\boldsymbol{x}} q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \ln \frac{q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})}{p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})}}_{=\mathrm{D}_{\mathsf{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{y}))}$$

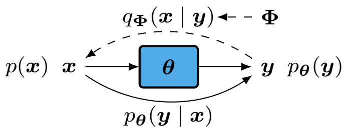
- Kullback-Leibler divergence $\mathrm{D}_{\mathsf{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})) \ \geq\ 0$

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

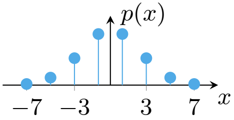
- **Optimization**: $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{y}) \uparrow \ \Rightarrow \ \mathrm{D}_{\mathsf{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x} \mid \boldsymbol{y})) \to 0$

[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)

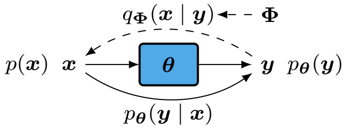


$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{y}) = B(\boldsymbol{h}, \boldsymbol{\Phi}, \boldsymbol{y}) - A(\boldsymbol{\Phi}, \boldsymbol{y})$$

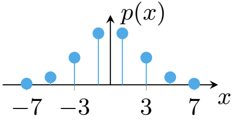


- $A(\boldsymbol{\Phi}, \boldsymbol{y}) = \mathrm{D}_{\mathrm{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p(\boldsymbol{x}))$
  - *Tries to shape* $q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})$ *similar to* $p(\boldsymbol{x})$

Variational Inference via VAE Paradigm (2)

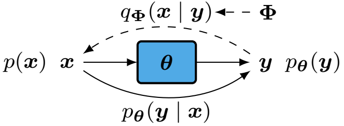


$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{y}) = B(\boldsymbol{h}, \boldsymbol{\Phi}, \boldsymbol{y}) - A(\boldsymbol{\Phi}, \boldsymbol{y})$$

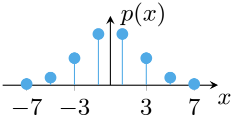


- $A(\boldsymbol{\Phi}, \boldsymbol{y}) = \mathrm{D}_{\mathrm{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p(\boldsymbol{x}))$
  - *Tries to shape* $q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})$ *similar to* $p(\boldsymbol{x})$
- $B(\boldsymbol{h}, \boldsymbol{\Phi}, \boldsymbol{y})$
  - *Matches equalizer output to received symbols*

Variational Inference via VAE Paradigm (2)



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

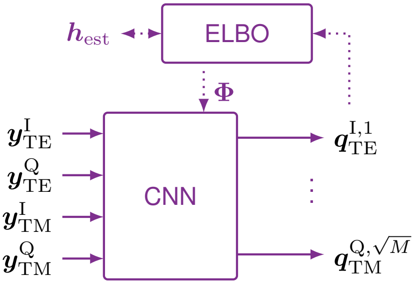


- $A(\boldsymbol{\Phi}, \boldsymbol{y}) = \mathrm{D}_{\mathrm{KL}}(q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y}) \| p(\boldsymbol{x}))$
  - *Tries to shape* $q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})$ *similar to* $p(\boldsymbol{x})$
- $B(\boldsymbol{h}, \boldsymbol{\Phi}, \boldsymbol{y})$
  - *Matches equalizer output to received symbols*

$\Rightarrow$ **Blind approach**: requires only $p(\boldsymbol{x})$, not $\boldsymbol{x}$

$\Rightarrow$ $q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})$ can be used in conjunction with Soft-FEC

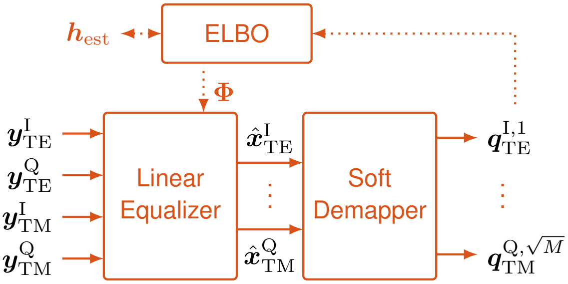
VAE Equalizer – Neural Network (VAE-NN)



- Convolutional Neural Network (CNN) has two tasks: equalization *and* demapping
- Outputs $q_{\boldsymbol{\Phi}}(\boldsymbol{x} \mid \boldsymbol{y})$
- 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).

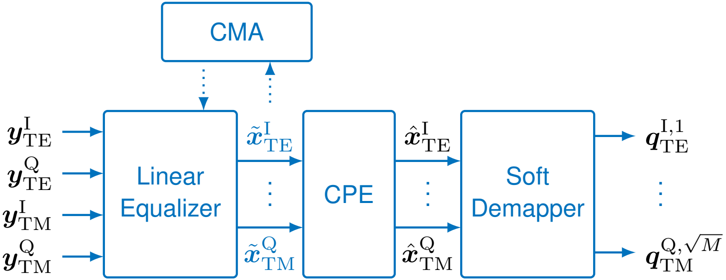
VAE Equalizer – Linear Equalizer (VAE-LE)



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

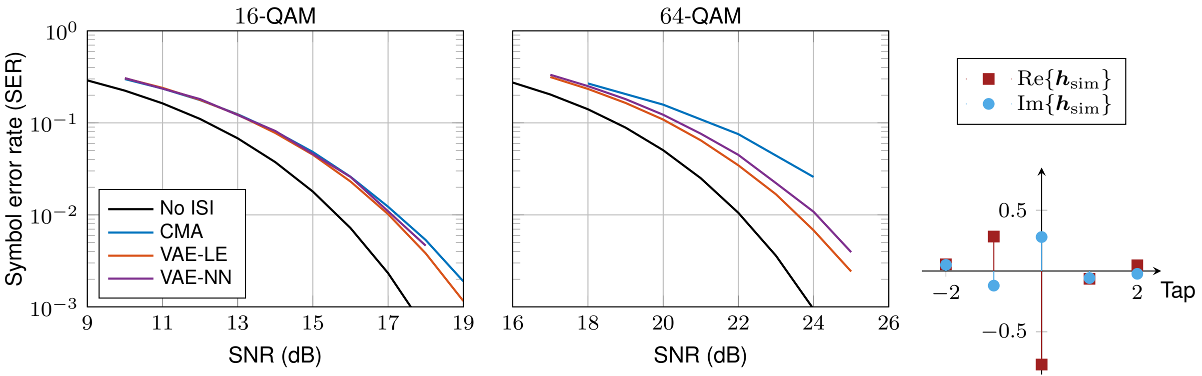
[LBS21] V. Lauinger, F. Buchali, L. S., "Blind equalization for coherent optical communications based on variational inference," *Proc. Adv. Photon. Congr., SPPCOM*, Jul. 2021

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

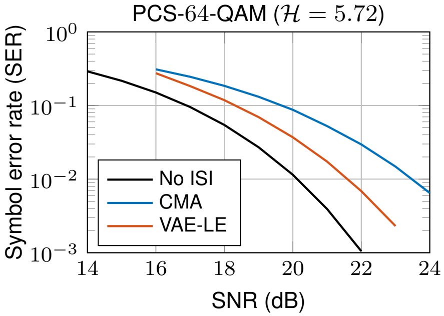
Simulation Results: Conventional QAM

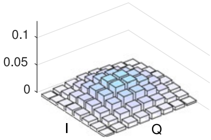


- Single polarization
- 2 samples per symbol
- Batch-wise update with Adam optimizer

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

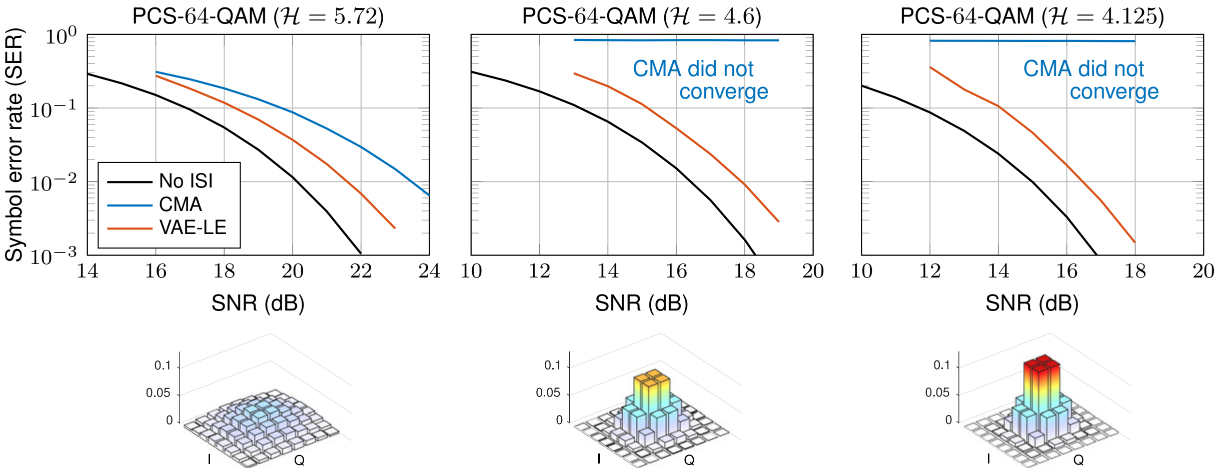
Simulation Results: PCS-QAM





[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 Reccurent Neural Networks
- Application 2: Distance-agnostic Transceivers
- Application 3: In-Situ Optimization of Transmitters
- Application 4: 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**

FIN

Karlsruher Institut für Technologie

Questions?

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

European Research Council
Established by the European Commission

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.