Model-Based Machine Learning for Joint Digital Backpropagation and PMD Compensation

Christian Häger(1), Henry D. Pfister(2), Rick M. Bütler(3), Gabriele Liga(3), Alex Alvarado(3)

(1) Department of Electrical Engineering, Chalmers University of Technology, Sweden
(2) Department of Electrical and Computer Engineering, Duke University, USA
(3) Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands

Optical Fiber Communications Conference (OFC)
San Diego, USA, March 11, 2020

FORCE
FIBER-OPTIC COMMUNICATIONS RESEARCH CENTER

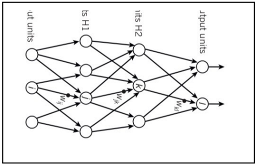
Duke UNIVERSITY

TU/e

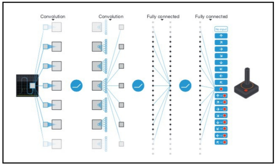
AVANCEZ 1829

CHALMERS

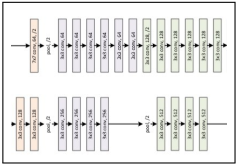
Deep Learning [LeCun et al., 2015]



Deep Q-Learning [Mnih et al., 2015]



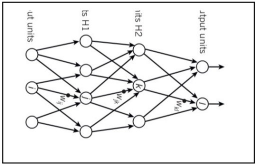
ResNet [He et al., 2015]



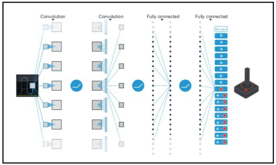
. . .

Multi-layer neural networks: impressive performance, countless applications

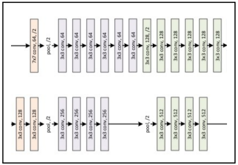
Deep Learning [LeCun et al., 2015]



Deep Q-Learning [Mnih et al., 2015]



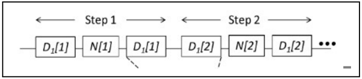
ResNet [He et al., 2015]



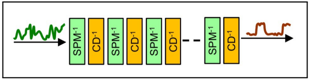
. . .

Multi-layer neural networks: impressive performance, countless applications

---



[Du and Lowery, 2010]



[Nakashima et al., 2017]

Split-step methods for solving the propagation equation in fiber-optics

Agenda

In this talk, we . . .

Agenda

In this talk, we . . .

1. show that multi-layer neural networks and the split-step method have the same functional form: both alternate linear and pointwise nonlinear steps

Agenda

In this talk, we . . .

1. show that multi-layer neural networks and the split-step method have the same functional form: both alternate linear and pointwise nonlinear steps
2. propose a model-based machine-learning approach based on parameterizing the split-step method (no black-box neural networks)

Agenda

In this talk, we . . .

1. show that multi-layer neural networks and the split-step method have the same functional form: both alternate linear and pointwise nonlinear steps
2. propose a model-based machine-learning approach based on parameterizing the split-step method (no black-box neural networks)
3. revisit hardware-efficient digital backpropagation combined with distributed compensation of polarization mode dispersion

Outline

1. Machine Learning and Neural Networks for Communications

2. Model-Based Machine Learning for Fiber-Optic Communications

3. Learned Digital Backpropagation

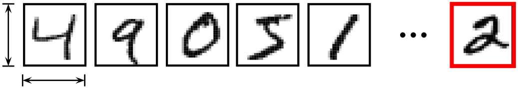
4. Polarization-Dependent Effects

5. Conclusions

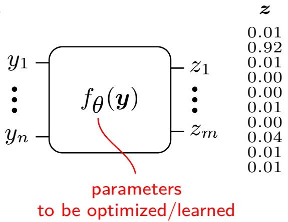
Outline

Supervised Learning

handwritten digit recognition (MNIST: 70,000 images)

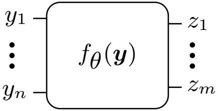


28×28 pixels $\implies n = 784$



Supervised Learning

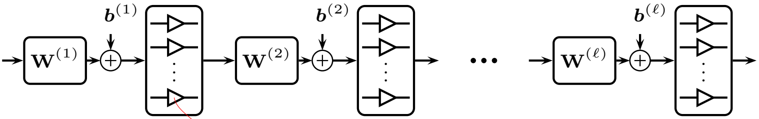
handwritten digit recognition (MNIST: 70,000 images)



$\boldsymbol{z}$

0.01
0.92
0.01
0.00
0.00
0.01
0.00
0.04
0.01
0.01

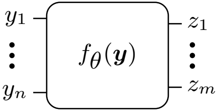
How to choose $f_\theta(\boldsymbol{y})$? Deep feed-forward neural networks



activation function

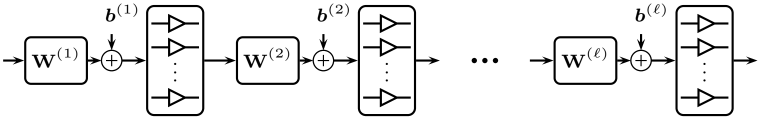
Supervised Learning

handwritten digit recognition (MNIST: 70,000 images)



$\boldsymbol{z}$	$\boldsymbol{x}$
0.01	0
0.92	1
0.01	0
0.00	0
0.00	0
0.01	0
0.00	0
0.04	0
0.01	0
0.01	0

How to choose $f_\theta(\boldsymbol{y})$? Deep feed-forward neural networks



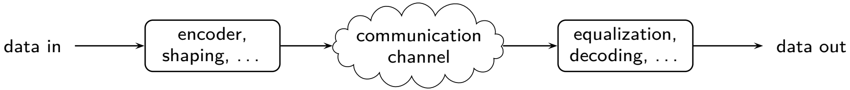
How to optimize $\theta = \{\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(\ell)}, \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(\ell)}\}$? Deep learning

$$\min_\theta \sum_{i=1}^{N} \text{Loss}(f_\theta(\boldsymbol{y}^{(i)}), \boldsymbol{x}^{(i)}) \triangleq g(\theta) \quad \text{using} \quad \theta_{k+1} = \theta_k - \lambda \nabla_\theta g(\theta_k) \qquad (1)$$

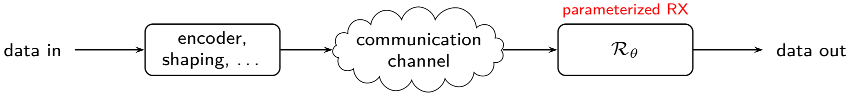
mean squared error
cross-entropy, ...

stochastic gradient descent,
RMSProp, Adam, ...

Machine Learning for Physical-Layer Communications

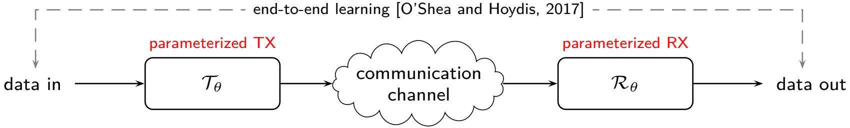


Machine Learning for Physical-Layer Communications



[Shen and Lau, 2011], Fiber nonlinearity compensation using extreme learning machine for DSP-based . . . , (*OECC*)
[Giacoumidis et al., 2015], Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based . . . , (*Opt. Lett.*)
[Zibar et al., 2016], Machine learning techniques in optical communication, (*J. Lightw. Technol.*)
[Kamalov et al., 2018], Evolution from 8qam live traffic to ps 64-qam with neural-network based nonlinearity compensation . . . , (*OFC*)
. . .

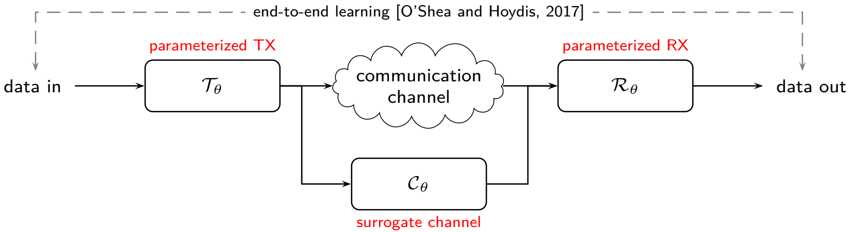
Machine Learning for Physical-Layer Communications



[Shen and Lau, 2011], Fiber nonlinearity compensation using extreme learning machine for DSP-based . . . , (*OECC*)
[Giacoumidis et al., 2015], Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based . . . , (*Opt. Lett.*)
[Zibar et al., 2016], Machine learning techniques in optical communication, (*J. Lightw. Technol.*)
[Kamalov et al., 2018], Evolution from 8qam live traffic to ps 64-qam with neural-network based nonlinearity compensation . . . , (*OFC*)
. . .

[O'Shea and Hoydis, 2017], An introduction to deep learning for the physical layer, (*IEEE Trans. Cogn. Commun. Netw.*)
[Karanov et al., 2018], End-to-end deep learning of optical fiber communications (*J. Lightw. Technol.*)
[Jones et al., 2018], Deep learning of geometric constellation shaping including fiber nonlinearities, (*ECOC*)
[Li et al., 2018], Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning, (*ECOC*)
. . .

Machine Learning for Physical-Layer Communications

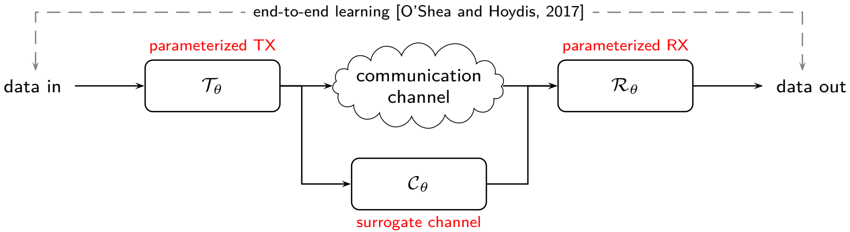


[Shen and Lau, 2011], Fiber nonlinearity compensation using extreme learning machine for DSP-based . . . , (*OECC*)
[Giacoumidis et al., 2015], Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based . . . , (*Opt. Lett.*)
[Zibar et al., 2016], Machine learning techniques in optical communication, (*J. Lightw. Technol.*)
[Kamalov et al., 2018], Evolution from 8qam live traffic to ps 64-qam with neural-network based nonlinearity compensation . . . , (*OFC*)
. . .

[O'Shea and Hoydis, 2017], An introduction to deep learning for the physical layer, (*IEEE Trans. Cogn. Commun. Netw.*)
[Karanov et al., 2018], End-to-end deep learning of optical fiber communications (*J. Lightw. Technol.*)
[Jones et al., 2018], Deep learning of geometric constellation shaping including fiber nonlinearities, (*ECOC*)
[Li et al., 2018], Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning, (*ECOC*)
. . .

[O'Shea et al., 2018], Approximating the void: Learning stochastic channel models from observation with variational GANs, (*arXiv*)
[Ye et al., 2018], Channel agnostic end-to-end learning based communication systems with conditional GAN, (*arXiv*)
. . .

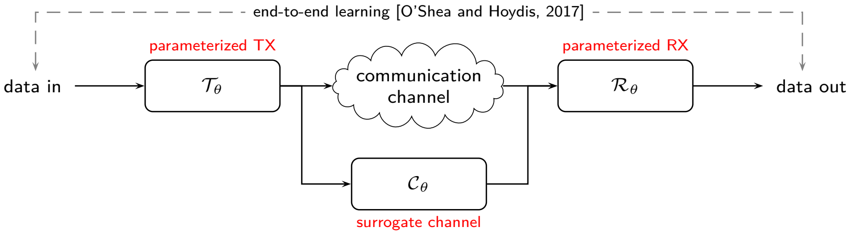
Machine Learning for Physical-Layer Communications



Using neural networks for $\mathcal{T}_\theta, \mathcal{R}_\theta, \mathcal{C}_\theta$

- How to choose network architecture (#layers, activation function)?
- How to initialize parameters?
- How to interpret solutions? Any insight gained?
- ...

Machine Learning for Physical-Layer Communications



Using neural networks for $\mathcal{T}_\theta, \mathcal{R}_\theta, \mathcal{C}_\theta$

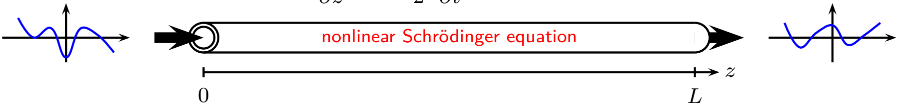
- How to choose network architecture (#layers, activation function)? ✗
- How to initialize parameters? ✗
- How to interpret solutions? Any insight gained? ✗
- ...

Model-based learning: sparse signal recovery [Gregor and Lecun, 2010], [Borgerding and Schniter, 2016], neural belief propagation [Nachmani et al., 2016], radio transformer networks [O'Shea and Hoydis, 2017], ...

Outline

The Split-Step Method

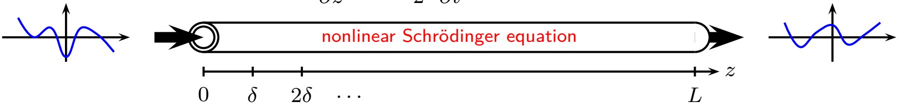
$$\frac{\partial u}{\partial z} = -\jmath \frac{\beta_2}{2} \frac{\partial^2 u}{\partial t^2} + \jmath \gamma u |u|^2$$



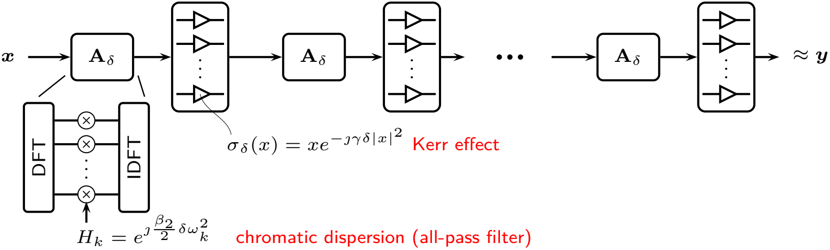
- Deterministic channel model: partial differential equation

The Split-Step Method

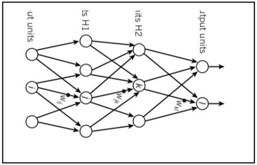
$$\frac{\partial u}{\partial z} = -\jmath \frac{\beta_2}{2} \frac{\partial^2 u}{\partial t^2} + \jmath \gamma u |u|^2$$



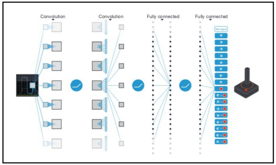
- Deterministic channel model: partial differential equation
- Split-step method with M steps ($\delta = L/M$):



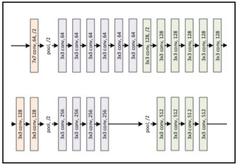
Deep Learning [LeCun et al., 2015]



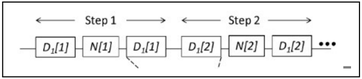
Deep Q-Learning [Mnih et al., 2015]



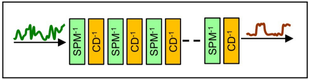
ResNet [He et al., 2015]



. . .



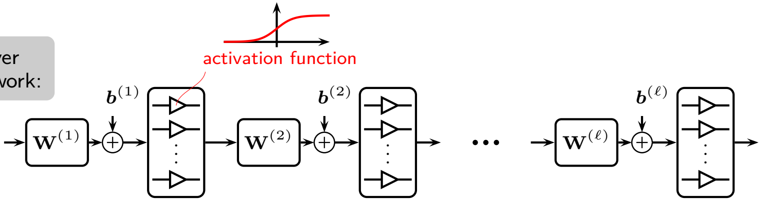
[Du and Lowery, 2010]



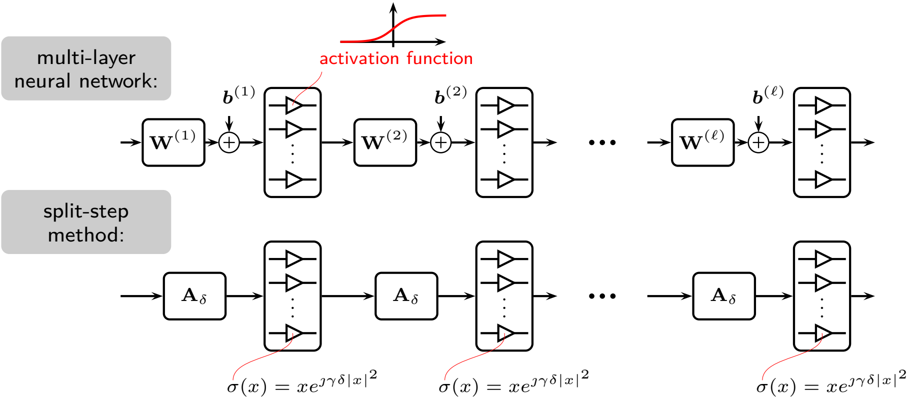
[Nakashima et al., 2017]

Parameterizing the Split-Step Method

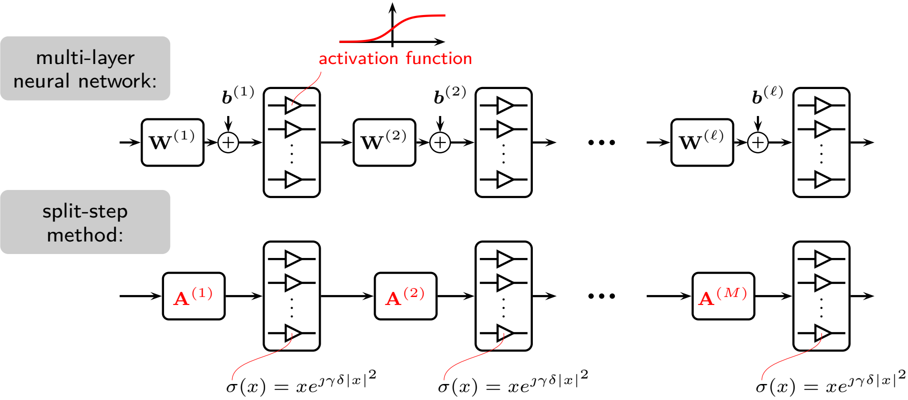
multi-layer neural network:



Parameterizing the Split-Step Method

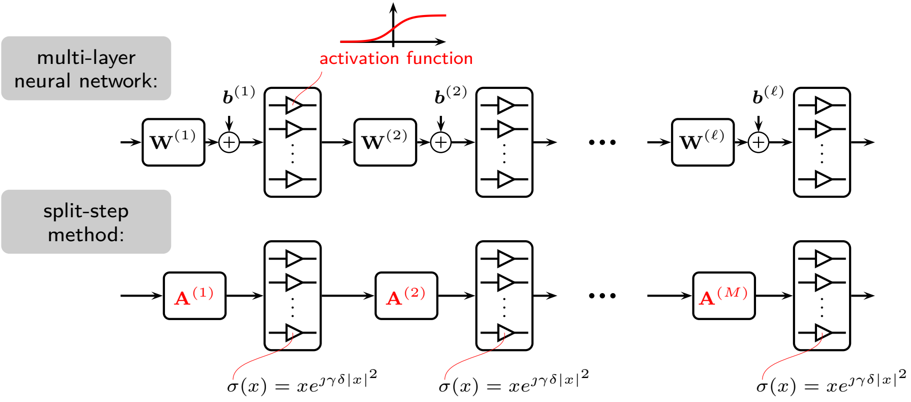


Parameterizing the Split-Step Method



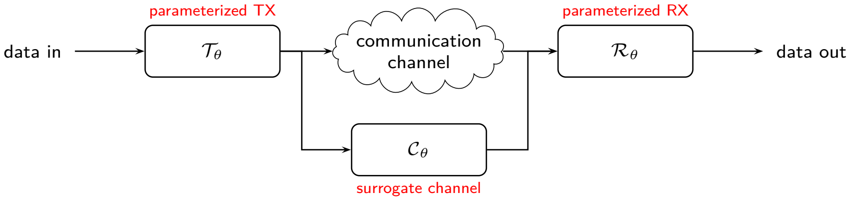
[Häger & Pfister, 2018], Nonlinear Interference Mitigation via Deep Neural Networks, *(OFC)*
[Häger & Pfister, 2018], Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications, *(ISIT)*

Parameterizing the Split-Step Method

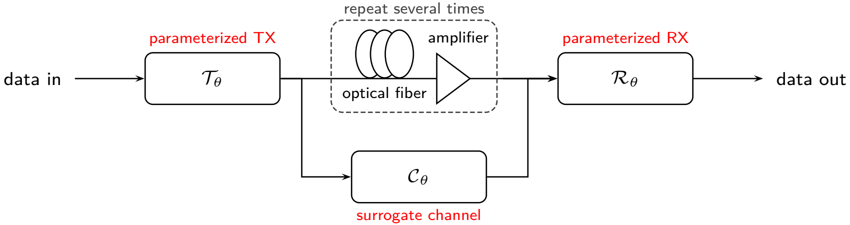


- Parameterized model f_θ with $\theta = \{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(M)}\}$

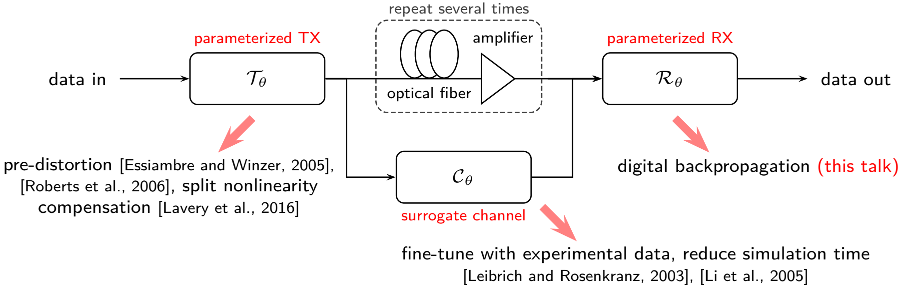
Possible Applications



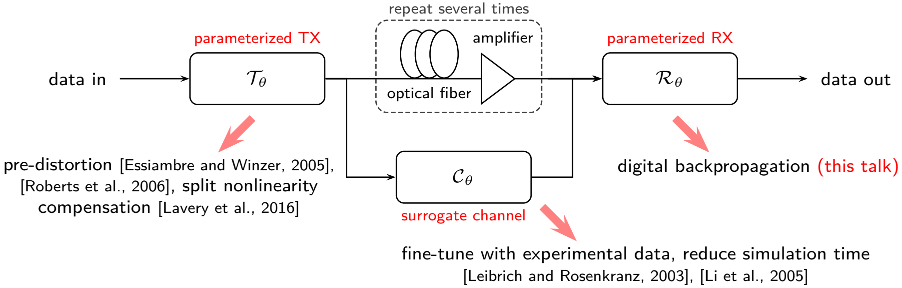
Possible Applications



Possible Applications



Possible Applications

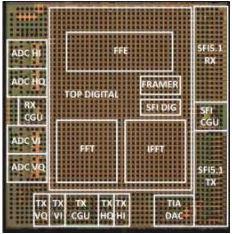


Model-based learning approaches

- How to choose network architecture (#layers, activation function)? ✓
- How to initialize parameters? ✓
- How to interpret solutions? Any insight gained? ✓

Outline

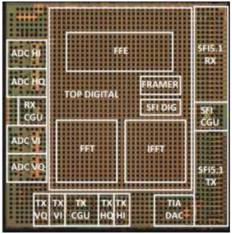
Real-Time Digital Backpropagation



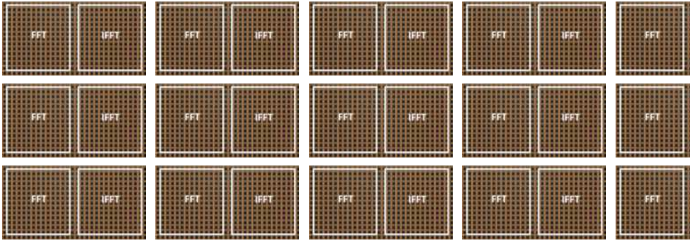
[Crivelli et al., 2014]

- Invert a partial differential equation in real time ([Paré et al., 1996], [Essiambre and Winzer, 2005], [Roberts et al., 2006], [Li et al., 2008], [Ip and Kahn, 2008])

Real-Time Digital Backpropagation

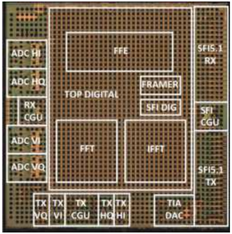


[Crivelli et al., 2014]

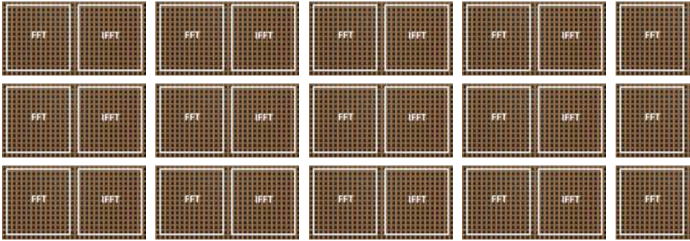


- Invert a partial differential equation in real time ([Paré et al., 1996], [Essiambre and Winzer, 2005], [Roberts et al., 2006], [Li et al., 2008], [Ip and Kahn, 2008])
- Widely considered to be impractical (too complex): linear equalization is already one of the most power-hungry DSP blocks in coherent receivers

Real-Time Digital Backpropagation



[Crivelli et al., 2014]



- Invert a partial differential equation in real time ([Paré et al., 1996], [Essiambre and Winzer, 2005], [Roberts et al., 2006], [Li et al., 2008], [Ip and Kahn, 2008])
- Widely considered to be impractical (too complex): linear equalization is already one of the most power-hungry DSP blocks in coherent receivers

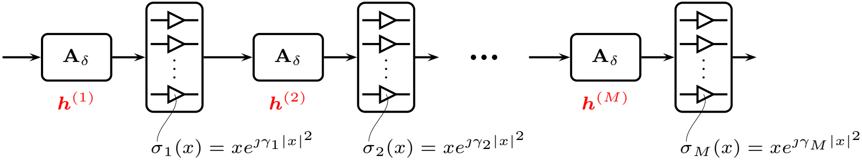
Our approach

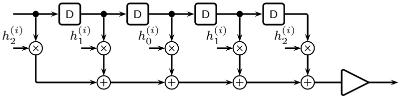
Joint optimization, pruning, and quantization of all chromatic-dispersion filters leads to efficient digital backpropagation, even with many steps.

Learned Digital Backpropagation

Learned Digital Backpropagation

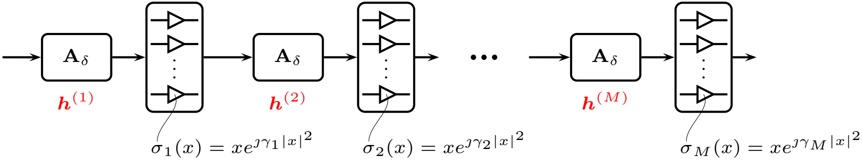
TensorFlow implementation of the computation graph $f_\theta(\boldsymbol{y})$:





Learned Digital Backpropagation

TensorFlow implementation of the computation graph $f_\theta(\boldsymbol{y})$:



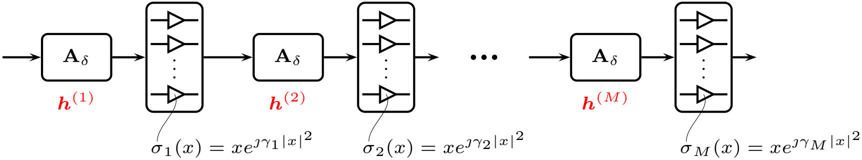
Deep learning of parameters $\theta = \{\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(M)}\}$:

$$\min_\theta \sum_{i=1}^{N} \text{Loss}(f_\theta(\boldsymbol{y}^{(i)}), \boldsymbol{x}^{(i)}) \triangleq g(\theta) \qquad \text{using} \quad \theta_{k+1} = \theta_k - \lambda \nabla_\theta g(\theta_k)$$

mean squared error — Adam optimizer, fixed learning rate

Learned Digital Backpropagation

TensorFlow implementation of the computation graph $f_\theta(\boldsymbol{y})$:



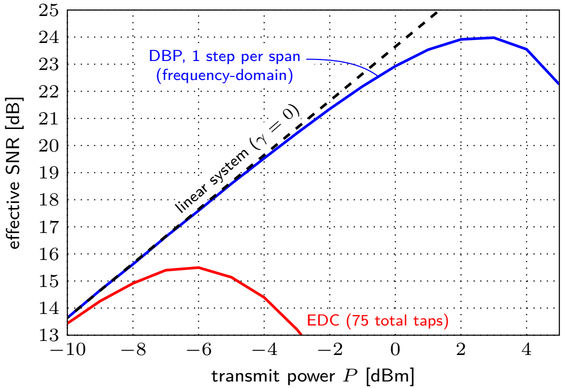
Deep learning of parameters $\theta = \{\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(M)}\}$:

$$\min_\theta \sum_{i=1}^{N} \text{Loss}(f_\theta(\boldsymbol{y}^{(i)}), \boldsymbol{x}^{(i)}) \triangleq g(\theta) \qquad \text{using} \quad \theta_{k+1} = \theta_k - \lambda \nabla_\theta g(\theta_k)$$

mean squared error — Adam optimizer, fixed learning rate

Iteratively prune (set to 0) outermost filter taps during gradient descent

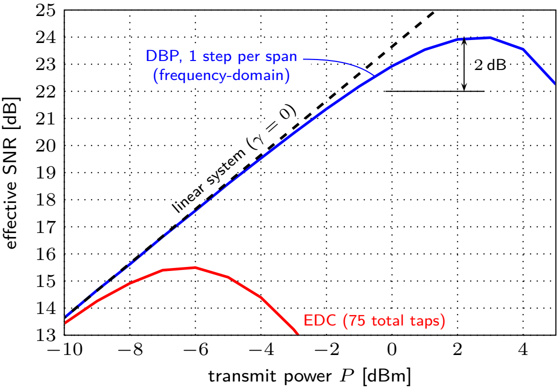
Revisiting Ip and Kahn (2008)



Parameters similar to [Ip and Kahn, 2008]:

- 25×80 km SSFM
- Gaussian modulation
- RRC pulses (0.1 roll-off)
- 10.7 Gbaud
- 2 samples/symbol processing
- single channel, single pol.

Revisiting Ip and Kahn (2008)

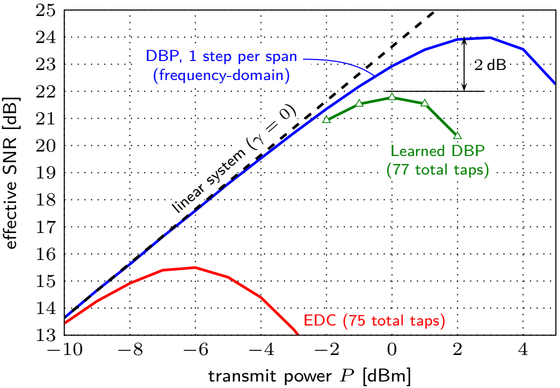


Parameters similar to [Ip and Kahn, 2008]:

- 25×80 km SSFM
- Gaussian modulation
- RRC pulses (0.1 roll-off)
- 10.7 Gbaud
- 2 samples/symbol processing
- single channel, single pol.

- $\gg 1000$ total taps (70 taps/step) $\implies$ $> 100\times$ complexity of EDC

Revisiting Ip and Kahn (2008)

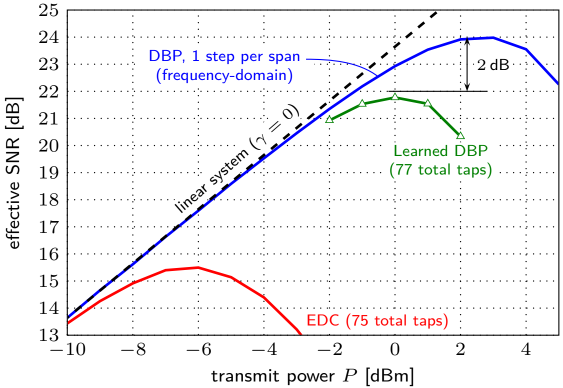


Parameters similar to [Ip and Kahn, 2008]:

- 25×80 km SSFM
- Gaussian modulation
- RRC pulses (0.1 roll-off)
- 10.7 Gbaud
- 2 samples/symbol processing
- single channel, single pol.

- $\gg 1000$ total taps (70 taps/step) $\Longrightarrow$ $> 100\times$ complexity of EDC
- Learned approach uses only 77 total taps: alternate 5 and 3 taps/step and use different filter coefficients in all steps [Häger and Pfister, 2018a]

Revisiting Ip and Kahn (2008)

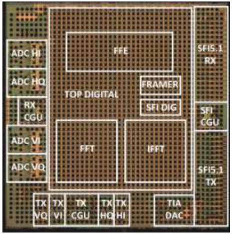


Parameters similar to [Ip and Kahn, 2008]:

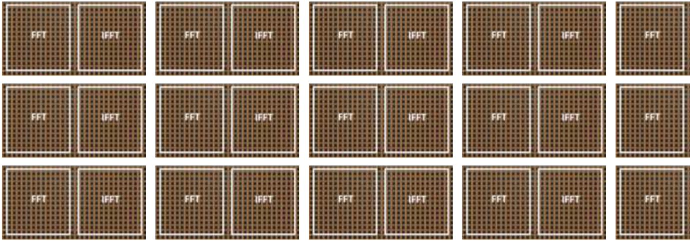
- 25×80 km SSFM
- Gaussian modulation
- RRC pulses (0.1 roll-off)
- 10.7 Gbaud
- 2 samples/symbol processing
- single channel, single pol.

- $\gg 1000$ total taps (70 taps/step) $\implies$ $> 100\times$ complexity of EDC
- Learned approach uses only 77 total taps: alternate 5 and 3 taps/step and use different filter coefficients in all steps [Häger and Pfister, 2018a]
- Can outperform "ideal DBP" in the nonlinear regime [Häger and Pfister, 2018b]

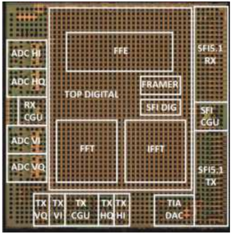
Real-Time ASIC Implementation



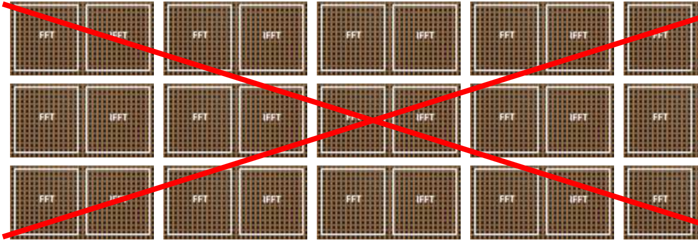
[Crivelli et al., 2014]



Real-Time ASIC Implementation

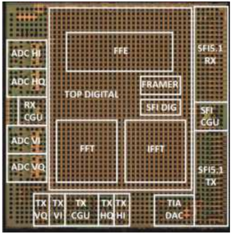


[Crivelli et al., 2014]

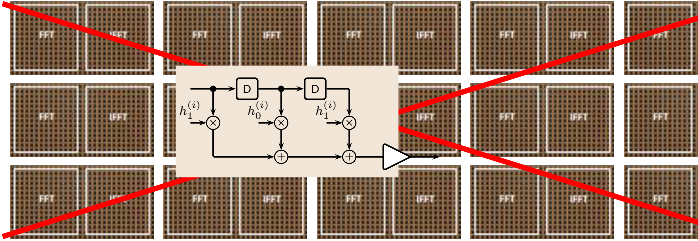


[Fougstedt et al., 2017], Time-domain digital back propagation: Algorithm and finite-precision implementation aspects, (*OFC*)
[Fougstedt et al., 2018], ASIC implementation of time-domain digital back propagation for coherent receivers, *(PTL)*
[Sherborne et al., 2018], On the impact of fixed point hardware for optical fiber nonlinearity compensation algorithms, (*JLT*)

Real-Time ASIC Implementation

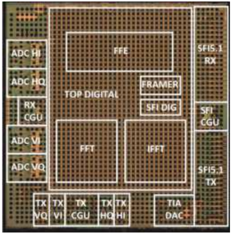


[Crivelli et al., 2014]

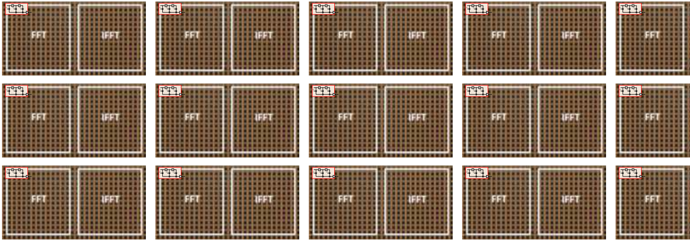


- Our linear steps are very short symmetric FIR filters (as few as 3 taps)

Real-Time ASIC Implementation



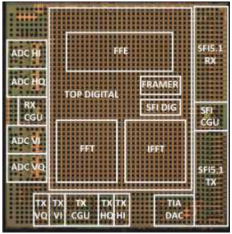
[Crivelli et al., 2014]



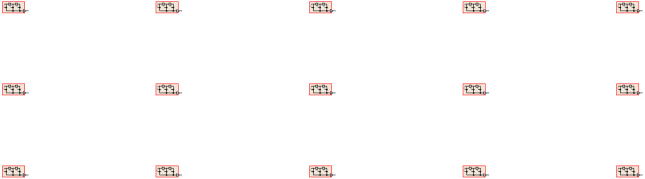
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
  - Only 5-6 bit filter coefficients via learned quantization
  - Hardware-friendly nonlinear steps (Taylor expansion)
  - All FIR filters are fully reconfigurable

[Fougstedt et al., 2018], ASIC implementation of time-domain digital backpropagation with deep-learned chromatic dispersion filters, *(ECOC)*

Real-Time ASIC Implementation



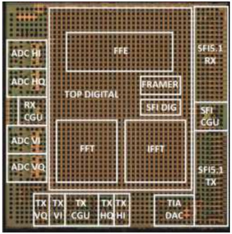
[Crivelli et al., 2014]



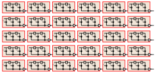
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
  - Only 5-6 bit filter coefficients via learned quantization
  - Hardware-friendly nonlinear steps (Taylor expansion)
  - All FIR filters are fully reconfigurable

[Fougstedt et al., 2018], ASIC implementation of time-domain digital backpropagation with deep-learned chromatic dispersion filters, *(ECOC)*

Real-Time ASIC Implementation



[Crivelli et al., 2014]



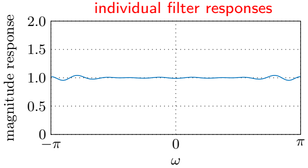
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
  - Only 5-6 bit filter coefficients via learned quantization
  - Hardware-friendly nonlinear steps (Taylor expansion)
  - All FIR filters are fully reconfigurable
- $< 2\times$ power compared to EDC [Crivelli et al., 2014, Pillai et al., 2014]

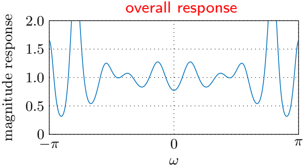
[Fougstedt et al., 2018], ASIC implementation of time-domain digital backpropagation with deep-learned chromatic dispersion filters, *(ECOC)*

Why Does The Learning Approach Work?

Previous work: design a single filter or filter pair and use it repeatedly.

$\Longrightarrow$ Good overall response only possible with very long filters.





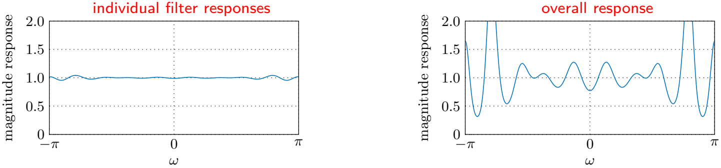
From [Ip and Kahn, 2009]:

- "We also note that [...] 70 taps, is much larger than expected"
- "This is due to amplitude ringing in the frequency domain"
- "Since backpropagation requires multiple iterations of the linear filter, amplitude distortion due to ringing accumulates (Goldfarb & Li, 2009)"

Why Does The Learning Approach Work?

Previous work: design a single filter or filter pair and use it repeatedly.

$\Longrightarrow$ Good overall response only possible with very long filters.



From [Ip and Kahn, 2009]:

- "We also note that [. . .] 70 taps, is much larger than expected"
- "This is due to amplitude ringing in the frequency domain"
- "Since backpropagation requires multiple iterations of the linear filter, amplitude distortion due to ringing accumulates (Goldfarb & Li, 2009)"

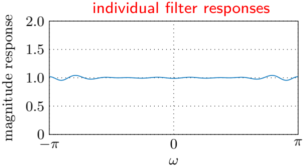
The learning approach uncovered that there is no such requirement!

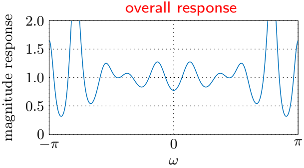
[Lian, Häger, Pfister, 2018], What can machine learning teach us about communications? (*ITW*)

Why Does The Learning Approach Work?

Previous work: design a single filter or filter pair and use it repeatedly.

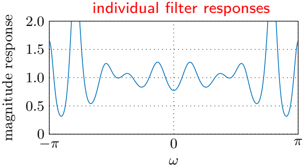
$\Longrightarrow$ Good overall response only possible with very long filters.

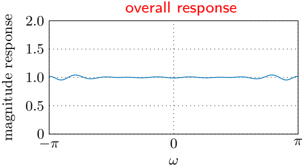




Sacrifice individual filter accuracy, but different response per step.

$\Longrightarrow$ Good overall response even with very short filters by joint optimization.

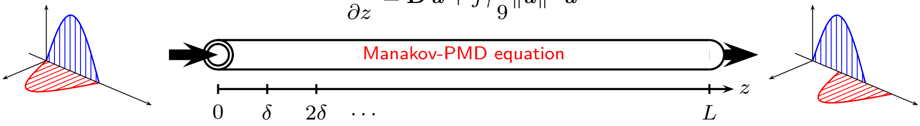




Outline

Evolution of Polarization-Multiplexed Signals

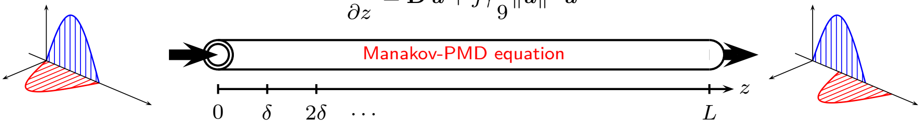
$$\frac{\partial \boldsymbol{u}}{\partial z} = \hat{\mathbf{D}}\,\boldsymbol{u} + \jmath\gamma\frac{8}{9}\|\boldsymbol{u}\|^2\,\boldsymbol{u}$$



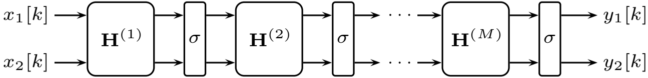
- Jones vector $\boldsymbol{u} \triangleq (u_1(t,z), u_2(t,z))^\top$ with complex baseband signals
- linear operator $\hat{\mathbf{D}}$: attentuation, chromatic & polarization mode dispersion

Evolution of Polarization-Multiplexed Signals

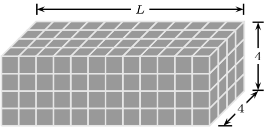
$$\frac{\partial \boldsymbol{u}}{\partial z} = \hat{\mathbf{D}}\,\boldsymbol{u} + \jmath\gamma\frac{8}{9}\|\boldsymbol{u}\|^2\,\boldsymbol{u}$$



- Jones vector $\boldsymbol{u} \triangleq (u_1(t,z), u_2(t,z))^\top$ with complex baseband signals
- linear operator $\hat{\mathbf{D}}$: attentuation, chromatic & polarization mode dispersion
- Split-step method: alternate linear and nonlinear steps $\sigma(\boldsymbol{x}) = \boldsymbol{x}e^{\jmath\gamma\frac{8}{9}\delta\|\boldsymbol{x}\|^2}$

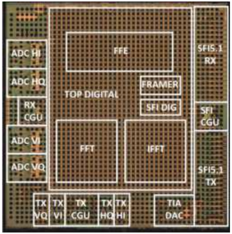


$e^{\delta\hat{\mathbf{D}}} \approx$

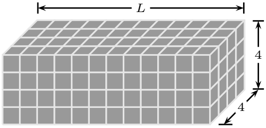


2×2 complex or
4×4 real MIMO filters $\Longrightarrow$ complexity!

Real-Time Compensation of Polarization Impairments

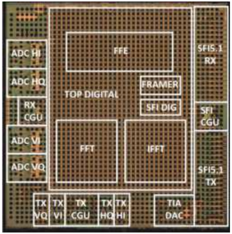


[Crivelli et al., 2014]

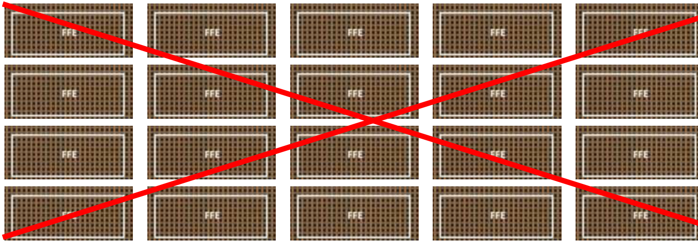


- time-varying effects (e.g., drifts) & apriori unknown realizations
- $\implies$ adaptive filtering (via stochastic gradient descent) required

Real-Time Compensation of Polarization Impairments



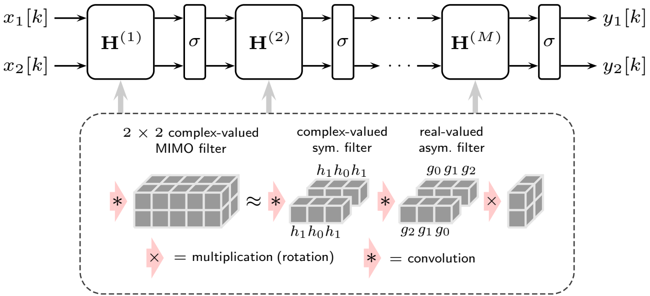
[Crivelli et al., 2014]



- time-varying effects (e.g., drifts) & apriori unknown realizations
- $\implies$ adaptive filtering (via stochastic gradient descent) required

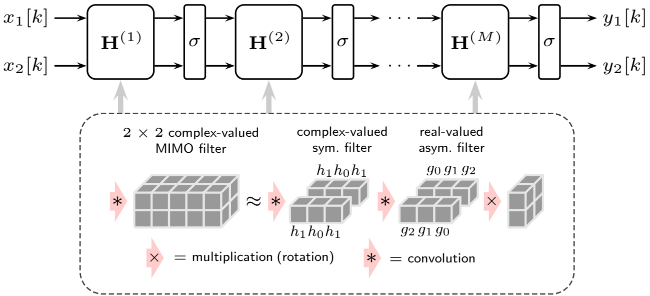
Using (and updating) full MIMO filters in each step is not feasible.

Our approach: Factorize each MIMO Filter



- 5-tap real-valued filters to approximate first-order PMD (DGD)
- Memoryless rotations $\begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix}$, where $a, b \in \mathbb{C}$ (4 real parameters)

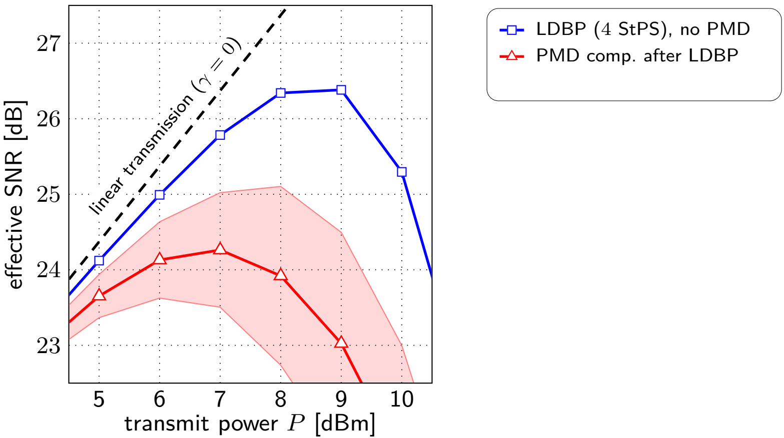
Our approach: Factorize each MIMO Filter



- 5-tap real-valued filters to approximate first-order PMD (DGD)
- Memoryless rotations $\begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix}$, where $a, b \in \mathbb{C}$ (4 real parameters)
- Assumes no knowledge about PMD realizations or accumulated PMD
- FIR-filter based! Avoids frequency-domain (FFT-based) filtering

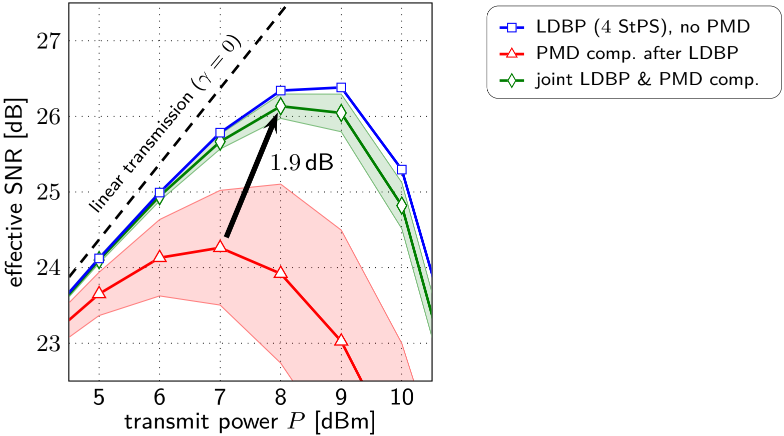
[Goroshko et al., 2016], Overcoming performance limitations of digital back propagation due to polarization mode dispersion, *(CTON)*
[Czegledi et al., 2017], Digital backpropagation accounting for polarization-mode dispersion, *(Opt. Express)*
[Liga et al., 2018], A PMD-adaptive DBP receiver based on SNR optimization, *(OFC)*

Results (32 Gbaud, 10×100 km, 0.2 ps/$\sqrt{\text{km}}$ PMD)



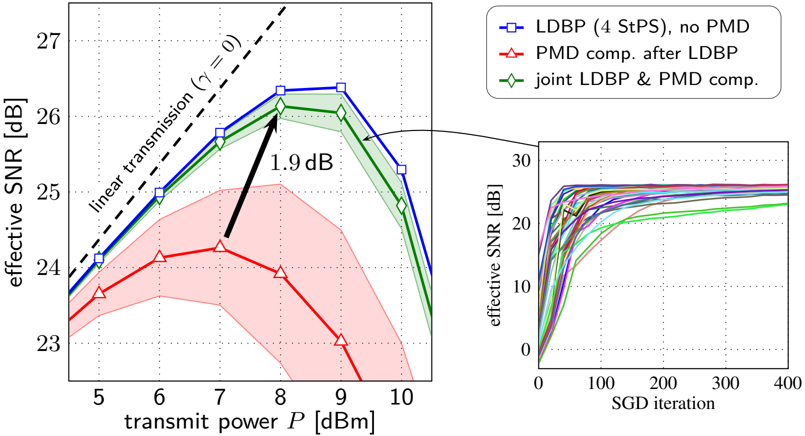
- Similar parameters & simulation setup compared to [Czegledi et al., 2016], results averaged over 40 PMD realizations

Results (32 Gbaud, 10×100 km, 0.2 ps/$\sqrt{\text{km}}$ PMD)



- Similar parameters & simulation setup compared to [Czegledi et al., 2016], results averaged over 40 PMD realizations

Results (32 Gbaud, 10×100 km, 0.2 ps/$\sqrt{\text{km}}$ PMD)



- Similar parameters & simulation setup compared to [Czegledi et al., 2016], results averaged over 40 PMD realizations
- Reliable convergence "from scratch" + only 9 real parameters per step

Related and Recent Works

Learned digital backpropagation

- [Häger & Pfister, 2018], Nonlinear Interference Mitigation via Deep Neural Networks, *(OFC)*
- [Häger & Pfister, 2018], Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications, *(ISIT)*
- [Lian, Häger, Pfister, 2018], What can machine learning teach us about communications? (*ITW*)
- [Häger et al., 2019], Revisiting multi-step nonlinearity compensation with machine learning (*ECOC*)

Wideband & WDM signals (alternating filter banks and nonlinearities)

- [Häger and Pfister, 2018], Wideband time-domain digital backpropagation via subband processing and deep learning, *(ECOC)*

ASIC implementation & finite-precision aspects

- [Fougstedt et al., 2018], ASIC implementation of time-domain digital backpropagation with deep-learned chromatic dispersion filters, *(ECOC)*

Experimental demonstrations & implementation aspects (e.g., phase noise)

- [Sillekens et al., 2020], Experimental demonstration of learned time-domain digital back-propagation, *(arXiv)*
- [Bitachon et al., 2020], Deep learning based digital back propagation with polarization state rotation & phase noise invariance, *(OFC)*

The Bigger Picture

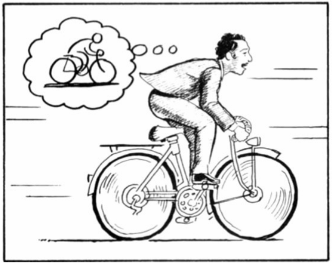
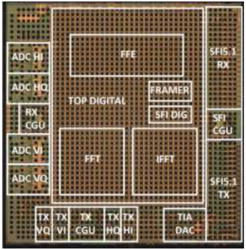


Figure 1. A World Model, from Scott McCloud's *Understanding Comics*. (McCloud, 1993; E, 2012)



[Crivelli et al., 2014]

- Optical receivers build models of their "environment"

Ha and Schmidhuber, 2018, "World Models", arXiv:1803.10122 [cs.LG]

The Bigger Picture

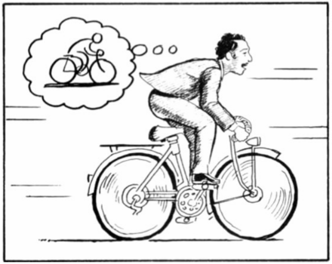
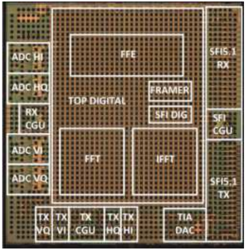


Figure 1. A World Model, from Scott McCloud's *Understanding Comics.* (McCloud, 1993; E, 2012)



[Crivelli et al., 2014]

- Optical receivers build models of their "environment"
- Currently these models are linear and/or rigid (non-adaptive)
- Interpretable physics-based “multi-layer” models for machine learning can be obtained by exploiting our existing domain knowledge

Ha and Schmidhuber, 2018, "World Models", arXiv:1803.10122 [cs.LG]

Conclusions

Conclusions

neural-network-based ML
universal function approximators
good designs require experience and fine-tuning
black boxes, difficult to "open"

Conclusions

neural-network-based ML	model-based ML
universal function approximators	application-tailored
good designs require experience and fine-tuning	relies on domain knowledge (algorithms, physics, . . .)
black boxes, difficult to “open”	familiar building blocks (e.g., FIR filters) can enable interpretability

Conclusions

neural-network-based ML	**model-based ML**
universal function approximators	application-tailored
good designs require experience and fine-tuning	relies on domain knowledge (algorithms, physics, . . .)
black boxes, difficult to “open”	familiar building blocks (e.g., FIR filters) can enable interpretability

Thank you!

FIBER-OPTIC COMMUNICATIONS RESEARCH CENTER

References I

Borgerding, M. and Schniter, P. (2016).
Onsager-corrected deep learning for sparse linear inverse problems.
In *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, Washington, DC.

Crivelli, D. E., Hueda, M. R., Carrer, H. S., Del Barco, M., López, R. R., Gianni, P., Finochietto, J., Swenson, N., Voois, P., and Agazzi, O. E. (2014).
Architecture of a single-chip 50 Gb/s DP-QPSK/BPSK transceiver with electronic dispersion compensation for coherent optical channels.
IEEE Trans. Circuits Syst. I: Reg. Papers, 61(4):1012–1025.

Czegledi, C. B., Liga, G., Lavery, D., Karlsson, M., Agrell, E., Savory, S. J., and Bayvel, P. (2016).
Polarization-mode dispersion aware digital backpropagation.
In *Proc. European Conf. Optical Communication (ECOC)*, Düsseldorf, Germany.

Du, L. B. and Lowery, A. J. (2010).
Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems.
Opt. Express, 18(16):17075–17088.

Essiambre, R.-J. and Winzer, P. J. (2005).
Fibre nonlinearities in electronically pre-distorted transmission.
In *Proc. European Conf. Optical Communication (ECOC)*, Glasgow, UK.

Gregor, K. and Lecun, Y. (2010).
Learning fast approximations of sparse coding.
In *Proc. Int. Conf. Mach. Learning*.

References II

Häger, C. and Pfister, H. D. (2018a).
Deep learning of the nonlinear Schrödinger equation in fiber-optic communications.
In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Vail, CO.

Häger, C. and Pfister, H. D. (2018b).
Nonlinear interference mitigation via deep neural networks.
In *Proc. Optical Fiber Communication Conf. (OFC)*, San Diego, CA.

He, K., Zhang, X., Ren, S., and Sun, J. (2015).
Deep residual learning for image recognition.

Ip, E. and Kahn, J. M. (2008).
Compensation of dispersion and nonlinear impairments using digital backpropagation.
J. Lightw. Technol., 26(20):3416–3425.

Ip, E. and Kahn, J. M. (2009).
Nonlinear impairment compensation using backpropagation.
Optical Fiber New Developments, Chapter 10.

Lavery, D., Ives, D., Liga, G., Alvarado, A., Savory, S. J., and Bayvel, P. (2016).
The benefit of split nonlinearity compensation for single-channel optical fiber communications.
IEEE Photon. Technol. Lett., 28(17):1803–1806.

LeCun, Y., Bengio, Y., and Hinton, G. (2015).
Deep learning.
Nature, 521(7553):436–444.

References III

Leibrich, J. and Rosenkranz, W. (2003).
Efficient numerical simulation of multichannel WDM transmission systems limited by XPM.
IEEE Photon. Technol. Lett., 15(3):395–397.

Li, X., Chen, X., Goldfarb, G., Mateo, E., Kim, I., Yaman, F., and Li, G. (2008).
Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing.
Opt. Express, 16(2):880–888.

Li, Y., Ho, C. K., Wu, Y., and Sun, S. (2005).
Bit-to-symbol mapping in LDPC coded modulation.
In *Proc. Vehicular Technology Conf. (VTC)*, Stockholm, Sweden.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015).
Human-level control through deep reinforcement learning.
Nature, 518(7540):529–533.

Nachmani, E., Be'ery, Y., and Burshtein, D. (2016).
Learning to decode linear codes using deep learning.
In *Proc. Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL.

Nakashima, H., Oyama, T., Ohshima, C., Akiyama, Y., Tao, Z., and Hoshida, T. (2017).
Digital nonlinear compensation technologies in coherent optical communication systems.
In *Proc. Optical Fiber Communication Conf. (OFC)*, page W1G.5, Los Angeles, CA.

References IV

O'Shea, T. and Hoydis, J. (2017).
An introduction to deep learning for the physical layer.
IEEE Trans. Cogn. Commun. Netw., 3(4):563–575.

Paré, C., Villeneuve, A., Bélanger, P.-A. A., and Doran, N. J. (1996).
Compensating for dispersion and the nonlinear Kerr effect without phase conjugation.
Optics Letters, 21(7):459–461.

Pillai, B. S. G., Sedighi, B., Guan, K., Anthapadmanabhan, N. P., Shieh, W., Hinton, K. J., and Tucker, R. S. (2014).
End-to-end energy modeling and analysis of long-haul coherent transmission systems.
J. Lightw. Technol., 32(18):3093–3111.

Roberts, K., Li, C., Strawczynski, L., O'Sullivan, M., and Hardcastle, I. (2006).
Electronic precompensation of optical nonlinearity.
IEEE Photon. Technol. Lett., 18(2):403–405.