Revisiting Multi-Step Nonlinearity Compensation with Machine Learning

Christian Häger[1]

Joint work with: Henry D. Pfister[2], Rick M. Bütler[3], Gabriele Liga[3], Alex Alvarado[3], Christoffer Fougstedt[4], Lars Svensson[4], and Per Larsson-Edefors[4]

[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
[4]Department of Computer Science and Engineering, Chalmers University of Technology, Sweden

September 2, 2018

FORCE
FIBER-OPTIC COMMUNICATIONS RESEARCH CENTER

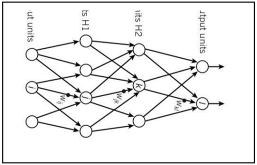
Duke UNIVERSITY

TU/e

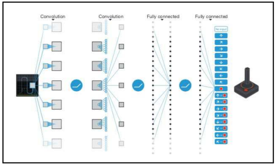
CHALMERS

"Multi-layer" vs. "Multi-step"

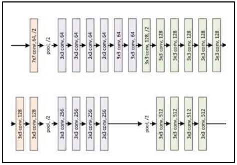
Deep Learning [LeCun et al., 2015]



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



ResNet [He et al., 2015]

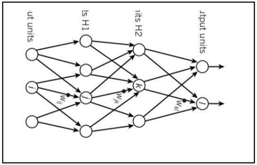


...

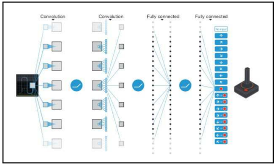
Multi-layer neural networks: impressive performance, countless applications

"Multi-layer" vs. "Multi-step"

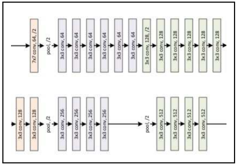
Deep Learning [LeCun et al., 2015]



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



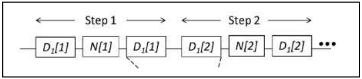
ResNet [He et al., 2015]



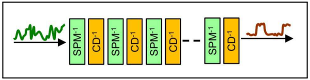
. . .

Multi-layer neural networks: impressive performance, countless applications

---



[Du and Lowery, 2010]



[Nakashima et al., 2017]

Conventional wisdom: Steps are inefficient $\Longrightarrow$ reduce as much as possible

- "with only four steps for the entire link . . . " [Du and Lowery, 2010]
- "up to 80% reduction in required [. . .] steps" [Rafique et al., 2011]
- "it reduces 85% back-propagation stages [. . .]" [Yan et al., 2011]
- "considerably reduces the number of spans needed " [Napoli et al., 2014]
- "single-step digital backpropagation" [Secondini et al., 2016]

"Multi-layer" vs. "Multi-step"

In this talk, we . . .

"Multi-layer" vs. "Multi-step"

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

"Multi-layer" vs. "Multi-step"

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)

“Multi-layer” vs. “Multi-step”

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 efficient multi-step nonlinearity compensation with deep learning: jointly optimize, prune, and quantize all linear substeps

"Multi-layer" vs. "Multi-step"

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 efficient multi-step nonlinearity compensation with deep learning: jointly optimize, prune, and quantize all linear substeps

Complexity $\stackrel{?}{\approx}$ Number of Steps

“Multi-layer” vs. “Multi-step”

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 efficient multi-step nonlinearity compensation with deep learning: jointly optimize, prune, and quantize all linear substeps

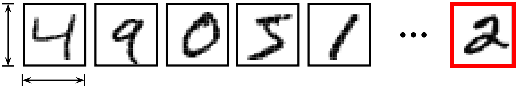
Complexity = Number of Steps × Complexity per Step

Outline

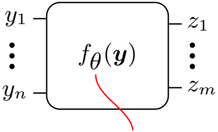
Outline

Supervised Learning

handwritten digit recognition (MNIST: 70,000 images)



28×28 pixels $\implies n = 784$

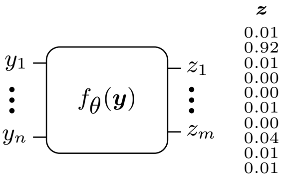


$\boldsymbol{z}$

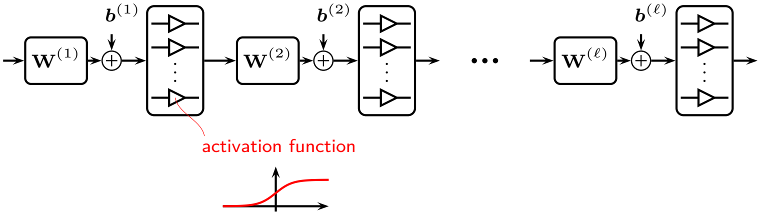
0.01
0.92
0.01
0.00
0.00
0.01
0.00
0.04
0.01
0.01

Supervised Learning

handwritten digit recognition (MNIST: 70,000 images)

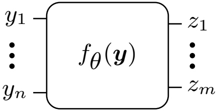


How to choose $f_\theta(\boldsymbol{y})$? Deep feed-forward neural networks



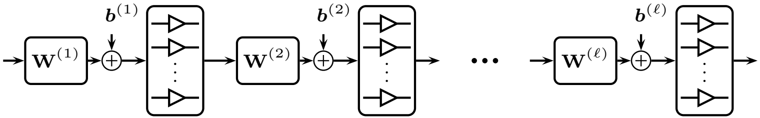
Supervised Learning

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



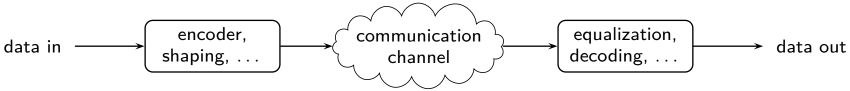
How to optimize $\theta = \{\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(\ell)}, \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(\ell)}\}$? Deep learning

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

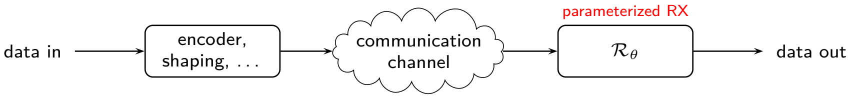
mean squared error
cross-entropy, ...

stochastic gradient descent,
RMSProp, Adam, ...

Machine Learning for Physical-Layer Communications

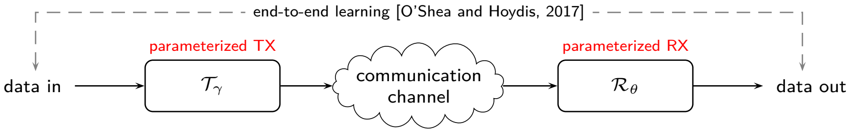


Machine Learning for Physical-Layer Communications



[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*)
. . .

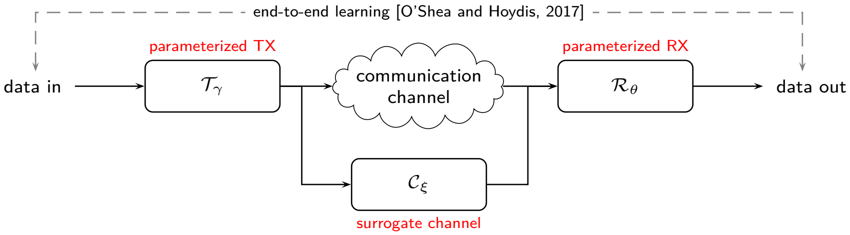
Machine Learning for Physical-Layer Communications



[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

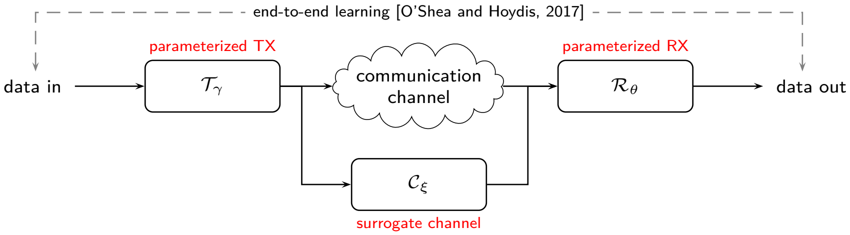


[Shen and Lau, 2011], Fiber nonlinearity compensation using extreme learning machine for DSP-based . . . , (*OECC*)
[Giacoumidis et al., 2015], Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based . . . , (*Opt. Lett.*)
[Zibar et al., 2016], Machine learning techniques in optical communication, (*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*)
. . .

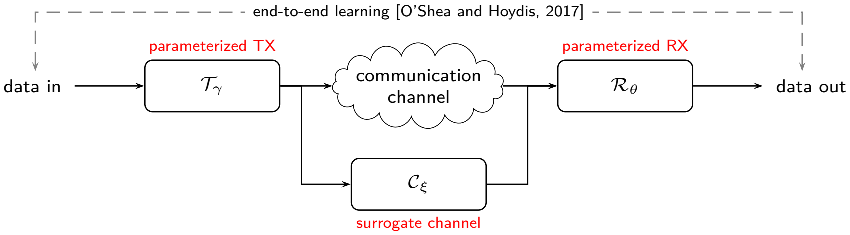
Machine Learning for Physical-Layer Communications



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

- 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 $\mathcal{T}_\gamma, \mathcal{R}_\theta, \mathcal{C}_\xi$

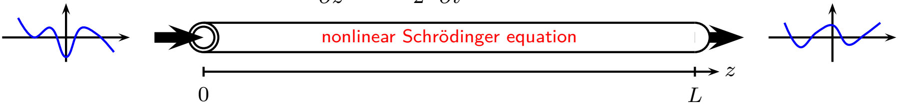
- How to choose network architecture (#layers, activation function)? ✗
- How to initialize parameters? ✗
- How to interpret solutions? Any insight gained? ✗
- ...

Model-based alternatives: sparse signal recovery [Gregor and Lecun, 2010], [Borgerding and Schniter, 2016], channel coding [Nachmani et al., 2016], ...

Outline

The Split-Step Method

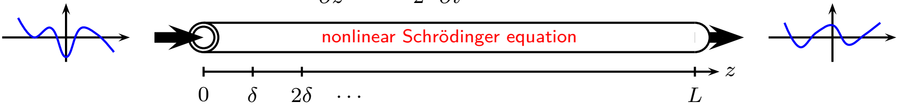
$$\frac{\partial u}{\partial z} = -\jmath \frac{\beta_2}{2} \frac{\partial^2 u}{\partial t^2} + \jmath \gamma u |u|^2$$



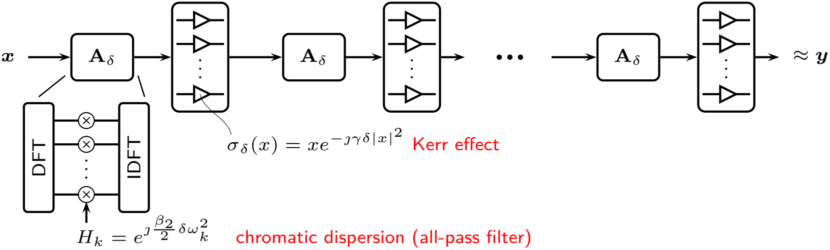
- Deterministic channel model: partial differential equation

The Split-Step Method

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

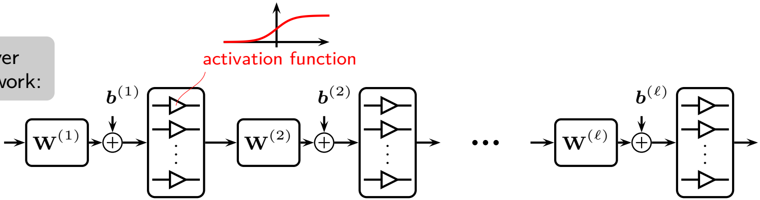


- Deterministic channel model: partial differential equation
- Split-step method with M steps ($\delta = L/M$):

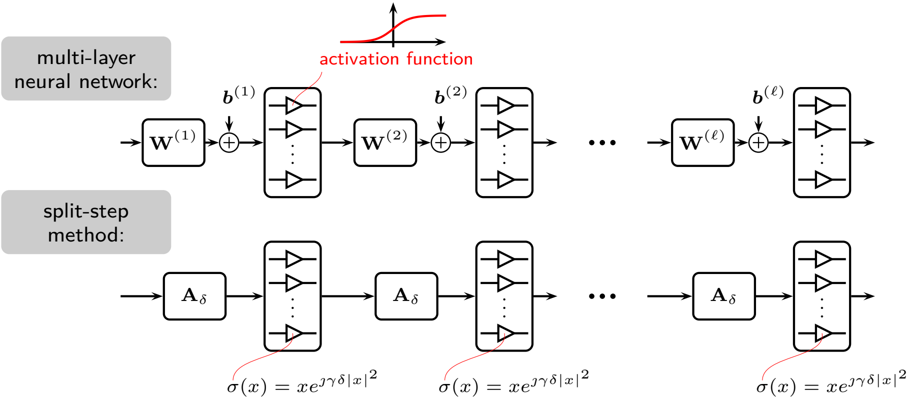


Parameterizing the Split-Step Method

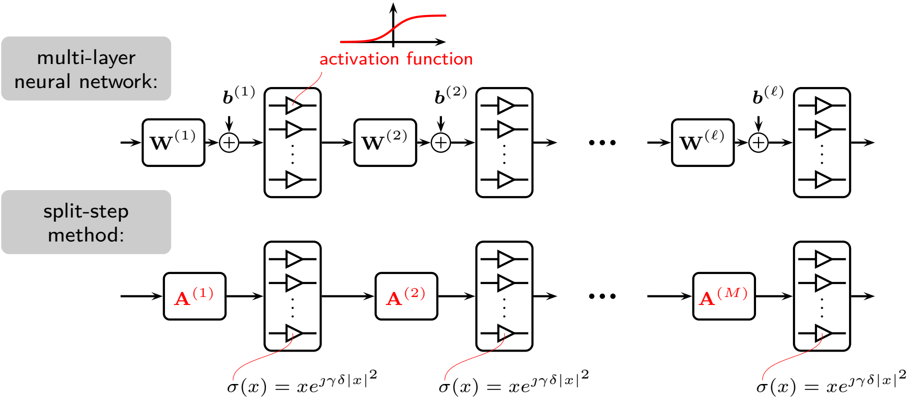
multi-layer neural network:



Parameterizing the Split-Step Method

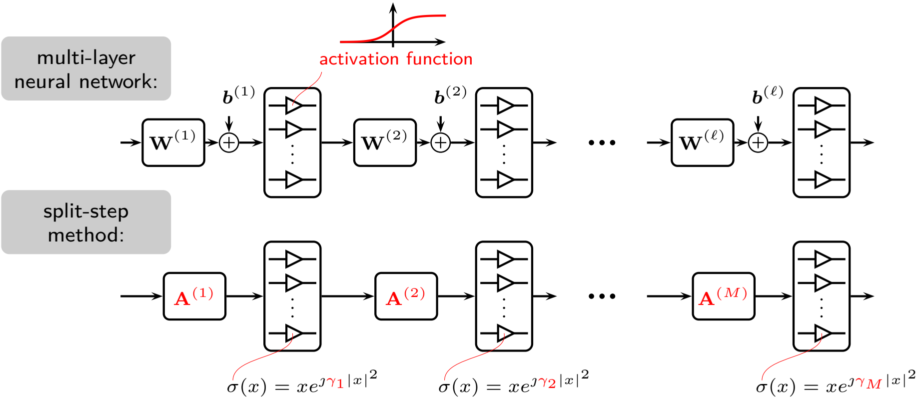


Parameterizing the Split-Step Method



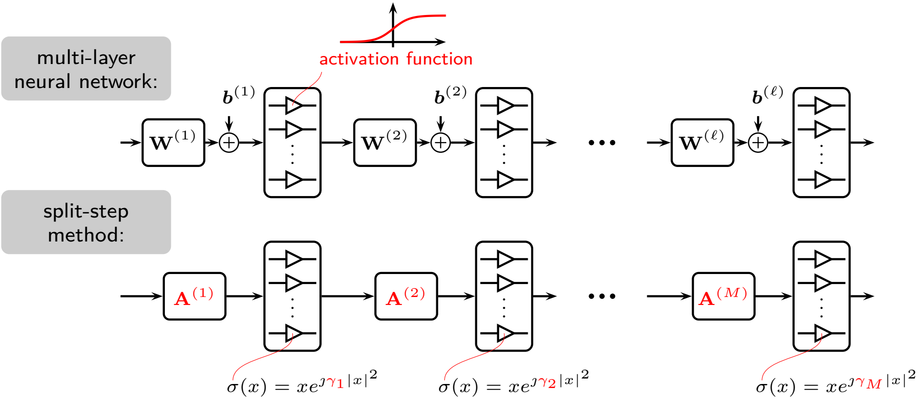
[Häger & Pfister, 2018], Nonlinear Interference Mitigation via Deep Neural Networks, *(OFC)*
[Häger & Pfister, 2018], Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications, *(ISIT)*

Parameterizing the Split-Step Method



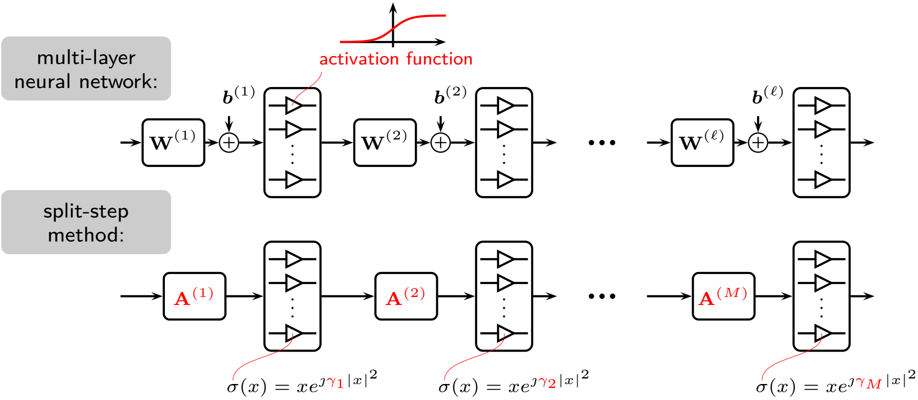
[Häger & Pfister, 2018], Nonlinear Interference Mitigation via Deep Neural Networks, *(OFC)*
[Häger & Pfister, 2018], Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications, *(ISIT)*

Parameterizing the Split-Step Method



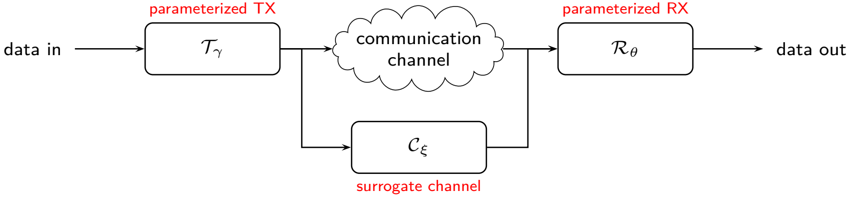
- Parameterized model f_θ with $\theta = \{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(M)}, \gamma_1, \ldots, \gamma_M\}$

Parameterizing the Split-Step Method

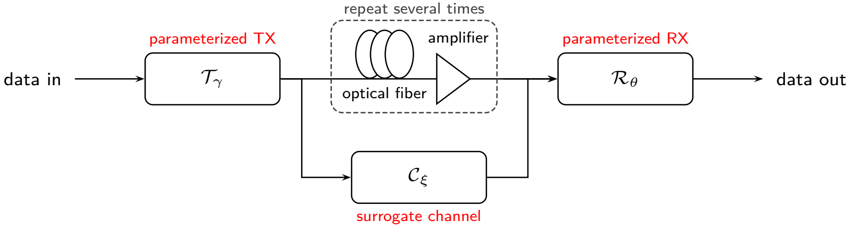


- Parameterized model f_θ with $\theta = \{\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(M)}, \gamma_1, \dots, \gamma_M\}$
- Includes as special cases: step-size optimization, "placement" of nonlinear operator, higher-order dispersion, matched filtering ...

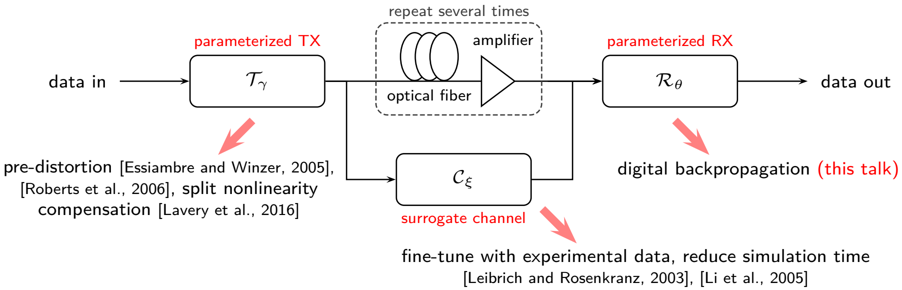
Possible Applications



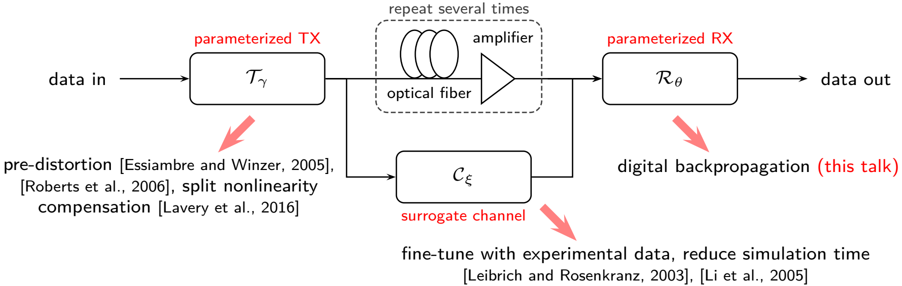
Possible Applications



Possible Applications



Possible Applications

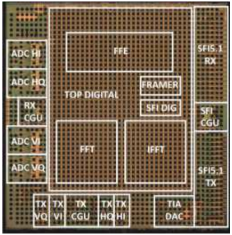


Model-based learning approaches

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

Outline

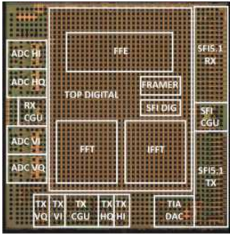
Real-Time Digital Backpropagation



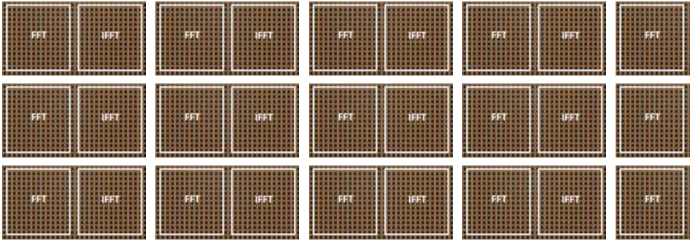
[Crivelli et al., 2014]

- 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

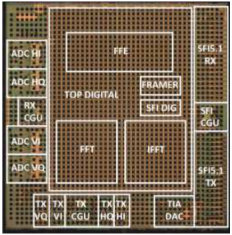


[Crivelli et al., 2014]

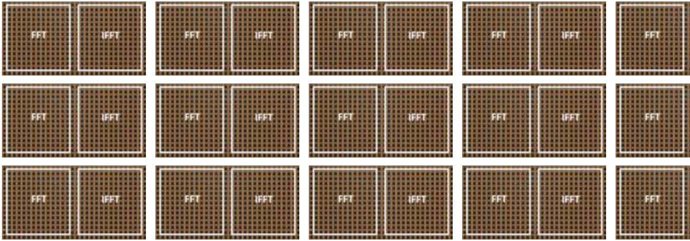


- Invert a partial differential equation in real time ([Paré et al., 1996], [Essiambre and Winzer, 2005], [Roberts et al., 2006], [Li et al., 2008], [Ip and Kahn, 2008])
- Widely considered to be impractical (too complex): linear equalization is already one of the most power-hungry DSP blocks in coherent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

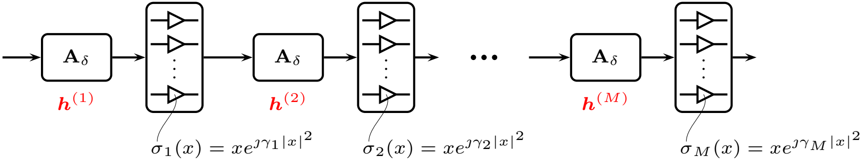
Our approach

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

Learned Digital Backpropagation

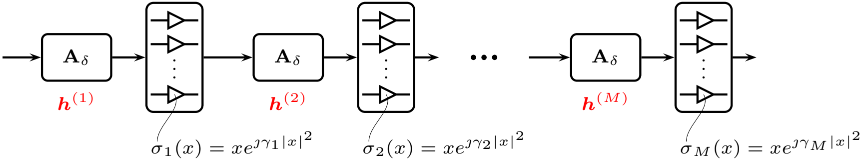
Learned Digital Backpropagation

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



Learned Digital Backpropagation

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



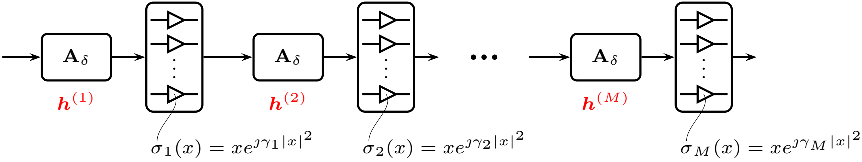
Deep learning of parameters $\theta = \{\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(M)}\}$:

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

mean squared error — Adam optimizer, fixed learning rate

Learned Digital Backpropagation

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



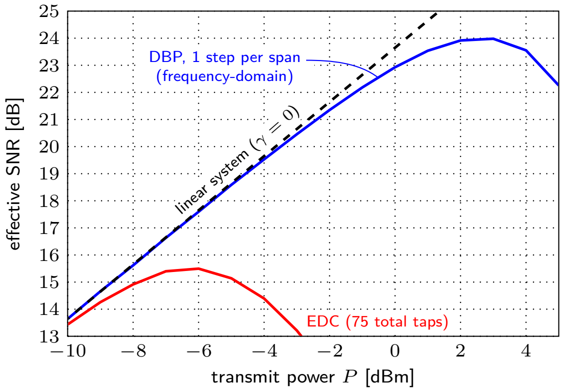
Deep learning of parameters $\theta = \{\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(M)}\}$:

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

mean squared error — Adam optimizer, fixed learning rate

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

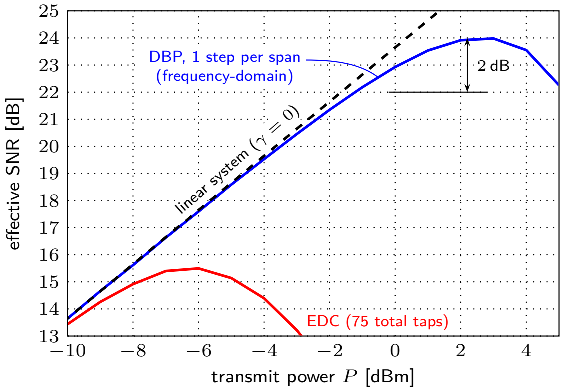
Revisiting Ip and Kahn (2008)



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)

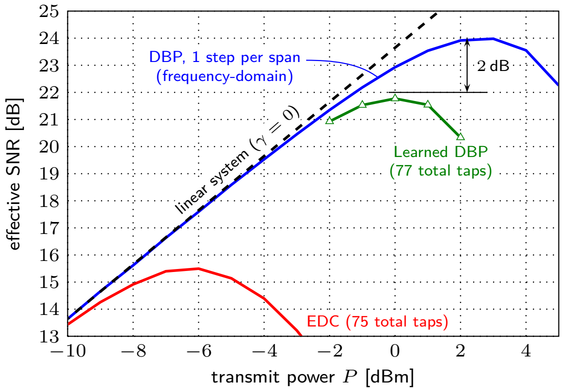


Parameters similar to [Ip and Kahn, 2008]:

- 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)

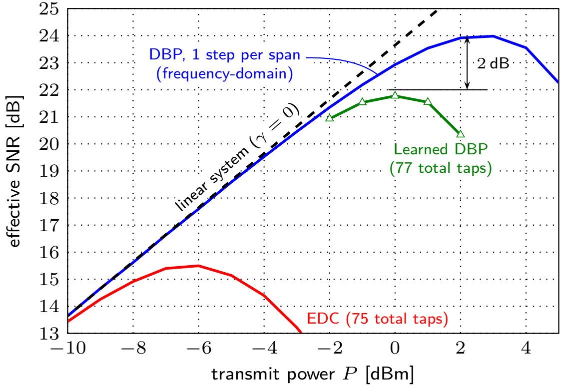


Parameters similar to [Ip and Kahn, 2008]:

- 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)

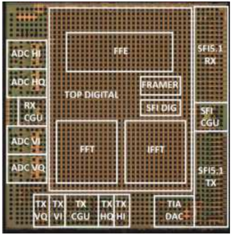


Parameters similar to [Ip and Kahn, 2008]:

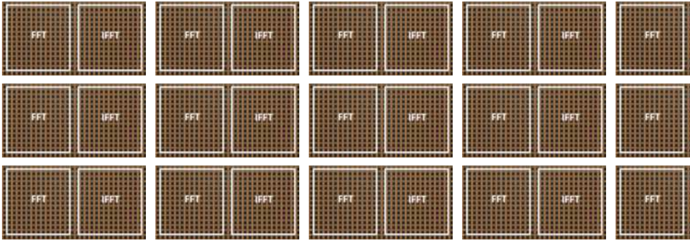
- 25×80 km SSFM
- Gaussian modulation
- RRC pulses (0.1 roll-off)
- 10.7 Gbaud
- 2 samples/symbol processing
- single channel, single pol.

- $\gg 1000$ 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]

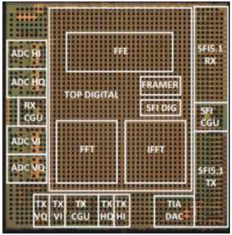
Real-Time ASIC Implementation



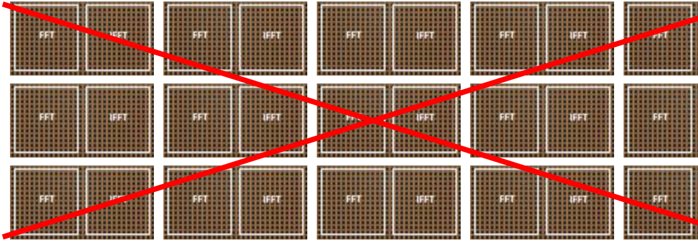
[Crivelli et al., 2014]



Real-Time ASIC Implementation

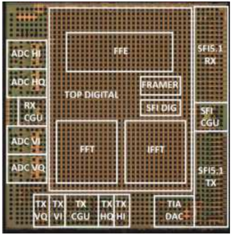


[Crivelli et al., 2014]

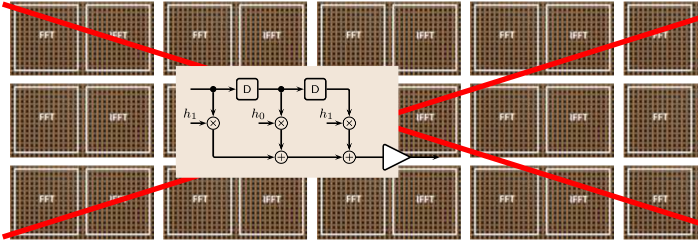


[Fougstedt et al., 2017], Time-domain digital back propagation: Algorithm and finite-precision implementation aspects, (*OFC*)
[Fougstedt et al., 2018], ASIC implementation of time-domain digital back propagation for coherent receivers, *(PTL)*
[Sherborne et al., 2018], On the impact of fixed point hardware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

Real-Time ASIC Implementation

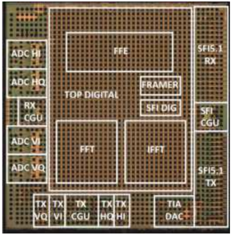


[Crivelli et al., 2014]

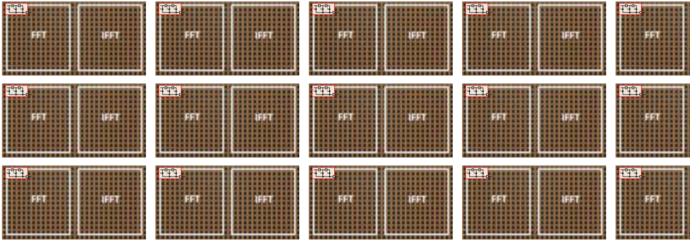


- Our linear steps are very short symmetric FIR filters (as few as 3 taps)

Real-Time ASIC Implementation



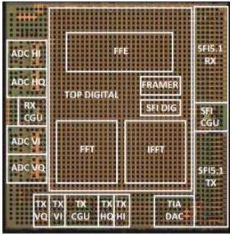
[Crivelli et al., 2014]



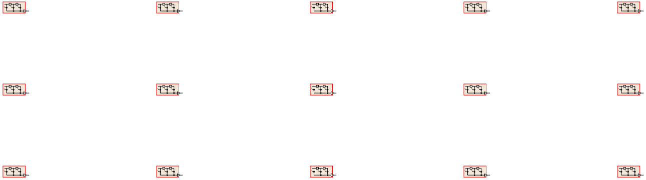
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
    - Only 5-6 bit filter coefficients via learned quantization
    - Hardware-friendly nonlinear steps (Taylor expansion)
    - All FIR filters are fully reconfigurable

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

Real-Time ASIC Implementation



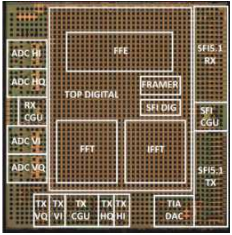
[Crivelli et al., 2014]



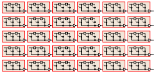
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
  - Only 5-6 bit filter coefficients via learned quantization
  - Hardware-friendly nonlinear steps (Taylor expansion)
  - All FIR filters are fully reconfigurable

[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]



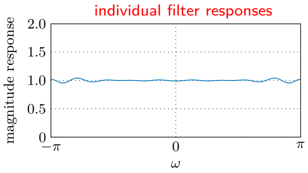
- Our linear steps are very short symmetric FIR filters (as few as 3 taps)
- 28-nm ASIC at 416.7 MHz clock speed (40 GHz signal)
  - Only 5-6 bit filter coefficients via learned quantization
  - Hardware-friendly nonlinear steps (Taylor expansion)
  - All FIR filters are fully reconfigurable
- $< 2\times$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



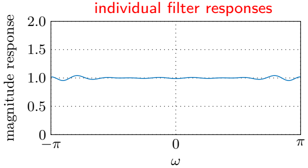
From [Ip and Kahn, 2009]:

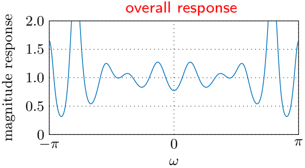
- "We also note that [. . .] 70 taps, is much larger than expected"
- "This is due to amplitude ringing in the frequency domain"
- "Since backpropagation requires multiple iterations of the linear filter, amplitude distortion due to ringing accumulates (Goldfarb & Li, 2009)"

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)"

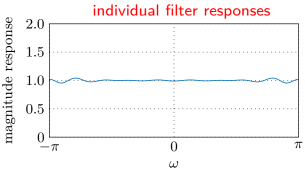
The learning approach uncovered that there is no such requirement!

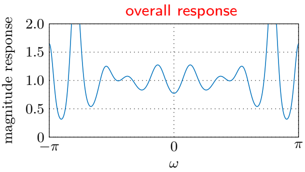
[Lian, Häger, Pfister, 2018], What can machine learning teach us about communications? (*ITW*)

Why Does The Learning Approach Work?

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

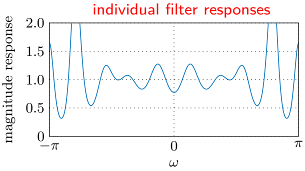
$\Longrightarrow$ Good overall response only possible with very long filters.

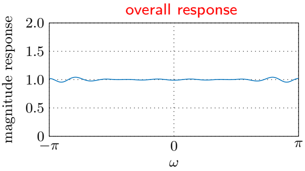




Sacrifice individual filter accuracy, but different response per step.

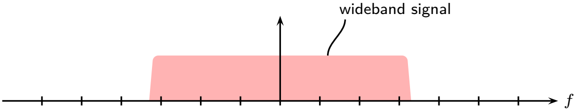
$\Longrightarrow$ Good overall response even with very short filters by joint optimization.



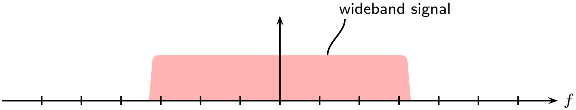


Outline

Wideband Signals and Subband Processing

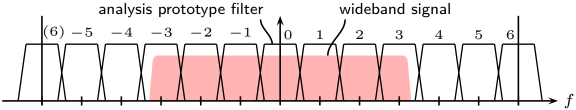


Wideband Signals and Subband Processing



- Quadratic dependence of overall system memory on the backpropagated bandwidth $\implies$ problematic for time-domain DBP (#taps $\propto f^2$)

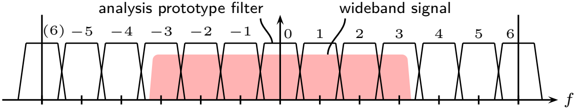
Wideband Signals and Subband Processing



- Quadratic dependence of overall system memory on the backpropagated bandwidth $\implies$ problematic for time-domain DBP (#taps $\propto f^2$)
- Subband processing: split received signal into N parallel signals

[Taylor, 2008], Compact digital dispersion compensation algorithms, *(OFC)*
[Ho, 2009], Subband equaliser for chromatic dispersion of optical fibre, *(Electronics Lett.)*
[Slim et al., 2013], Delayed single-tap frequency-domain chromatic-dispersion compensation, *(PTL)*
[Nazarathy and Tolmachev, 2014], Subbanded DSP architectures based on underdecimated filter banks . . . , *(Signal Proc. Mag.)*
[Mateo et al., 2010], Efficient compensation of inter-channel nonlinear effects via digital backward . . . , *(Opt. Express)*
[Ip et al., 2011], Complexity versus performance tradeoff for fiber nonlinearity compensation . . . *(OFC)*
[Oyama et al., 2015], Complexity reduction of perturbation-based nonlinear compensator by sub-band processing, *(OFC)*
. . .

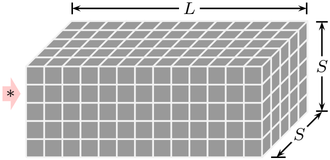
Wideband Signals and Subband Processing



- Quadratic dependence of overall system memory on the backpropagated bandwidth $\implies$ problematic for time-domain DBP (#taps $\propto f^2$)
- Subband processing: split received signal into N parallel signals
- MIMO filter accounts for cross-phase modulation (XPM) between subbands [Leibrich and Rosenkranz, 2003]

[Taylor, 2008], Compact digital dispersion compensation algorithms, *(OFC)*
[Ho, 2009], Subband equaliser for chromatic dispersion of optical fibre, *(Electronics Lett.)*
[Slim et al., 2013], Delayed single-tap frequency-domain chromatic-dispersion compensation, *(PTL)*
[Nazarathy and Tolmachev, 2014], Subbanded DSP architectures based on underdecimated filter banks . . . , *(Signal Proc. Mag.)*
[Mateo et al., 2010], Efficient compensation of inter-channel nonlinear effects via digital backward . . . , *(Opt. Express)*
[Ip et al., 2011], Complexity versus performance tradeoff for fiber nonlinearity compensation . . . *(OFC)*
[Oyama et al., 2015], Complexity reduction of perturbation-based nonlinear compensator by sub-band processing, *(OFC)*
. . .

Wideband Signals and Subband Processing

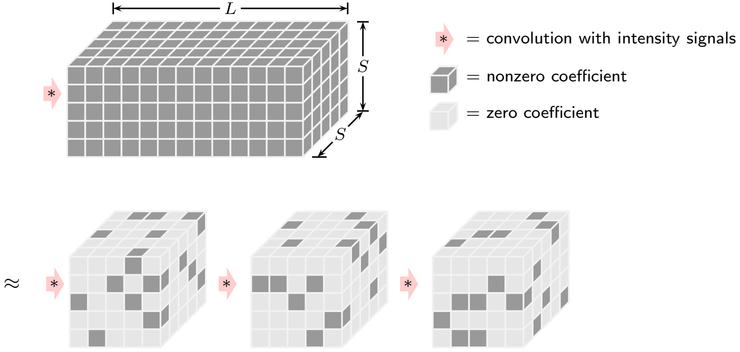


= convolution with intensity signals


= nonzero coefficient

= zero coefficient

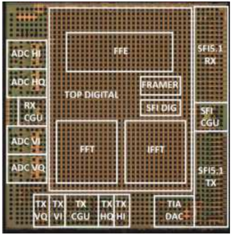
Wideband Signals and Subband Processing



- L_1-norm regularization applied to filter coefficients during gradient descent
- $\implies$ 92% of coefficients are zero with little performance penality

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

Polarization Mode Dispersion



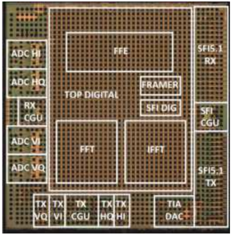
[Crivelli et al., 2014]

- Modeling via PMD sections, $\mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$, in the split-step method:
  - $\mathbf{R}^{(i)}$: complex unitary rotation matrix with determinant one
  - $\mathbf{J}^{(i)}(\omega)$: first-order PMD matrix with differential group delay (DGD) τ_i, i.e.,

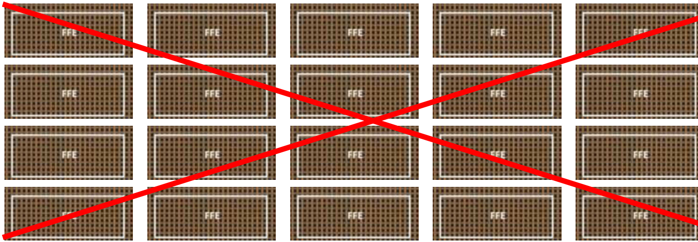
$$\mathbf{J}^{(i)}(\omega) = \begin{pmatrix} e^{-\jmath\omega\frac{\tau_i}{2}} & 0 \\ 0 & e^{\jmath\omega\frac{\tau_i}{2}} \end{pmatrix}$$

- PMD transfer matrix: $\mathbf{J}(\omega) = \prod_{i=1}^{M} \mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$ for large M

Polarization Mode Dispersion



[Crivelli et al., 2014]

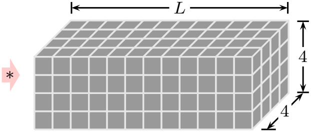


- Modeling via PMD sections, $\mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$, in the split-step method:
  - $\mathbf{R}^{(i)}$: complex unitary rotation matrix with determinant one
  - $\mathbf{J}^{(i)}(\omega)$: first-order PMD matrix with differential group delay (DGD) τ_i, i.e.,

$$\mathbf{J}^{(i)}(\omega) = \begin{pmatrix} e^{-\jmath\omega\frac{\tau_i}{2}} & 0 \\ 0 & e^{\jmath\omega\frac{\tau_i}{2}} \end{pmatrix}$$

- PMD transfer matrix: $\mathbf{J}(\omega) = \prod_{i=1}^{M} \mathbf{R}^{(i)}\mathbf{J}^{(i)}(\omega)$ for large M

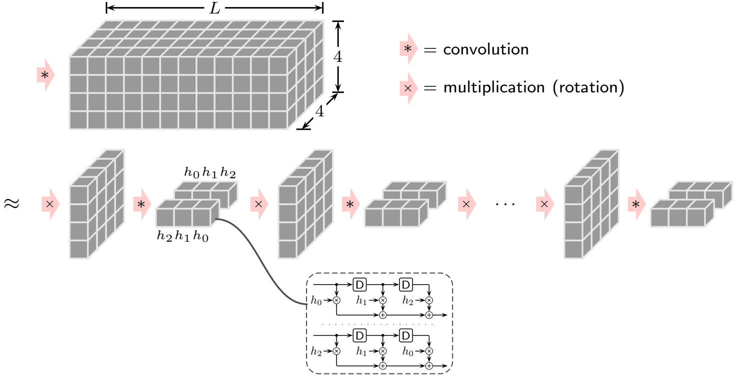
Adaptive PMD Compensation



= convolution

= multiplication (rotation)

Adaptive PMD Compensation



- Ongoing work: characterize optimization behavior (saddle points), integrate into digital backpropagation, . . .

[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)*

Conclusions

Conclusions

- We have proposed a model-based machine-learning approach for fiber-optic communication systems
- We have revisited efficient multi-step digital backpropagation and shown that deep-learning tools can be used to
  - jointly optimize all linear substeps
  - prune filter taps to get very short filters
  - jointly quantize all filter coefficients
- Multi-step enables factorization into simple, elementary steps

Conclusions

- We have proposed a model-based machine-learning approach for fiber-optic communication systems
- We have revisited efficient multi-step digital backpropagation and shown that deep-learning tools can be used to
  - jointly optimize all linear substeps
  - prune filter taps to get very short filters
  - jointly quantize all filter coefficients
- Multi-step enables factorization into simple, elementary steps

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.

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*.

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.

References II

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.

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.

References III

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.

Napoli, A., Maalej, Z., Sleiffer, V. A. J. M., Kuschnerov, M., Rafique, D., Timmers, E., Spinnler, B., Rahman, T., Coelho, L. D., and Hanik, N. (2014).
Reduced complexity digital back-propagation methods for optical communication systems.
J. Lightw. Technol., 32(7).

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.

Rafique, D., Zhao, J., and Ellis, A. D. (2011).
Digital back-propagation for spectrally efficient wdm 112 gbit/s pm m-ary qam transmission.
Opt. Express, 19(6):5219–5224.

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.

Secondini, M., Rommel, S., Meloni, G., Fresi, F., Forestieri, E., and Poti, L. (2016).
Single-step digital backpropagation for nonlinearity mitigation.
Photon. Netw. Commun., 31(3):493–502.

Yan, W., Tao, Z., Dou, L., Li, L., Oda, S., Tanimura, T., Hoshida, T., and Rasmussen, J. C. (2011).
Low complexity digital perturbation back-propagation.
In *Proc. European Conf. Optical Communication (ECOC)*, page Tu.3.A.2, Geneva, Switzerland.