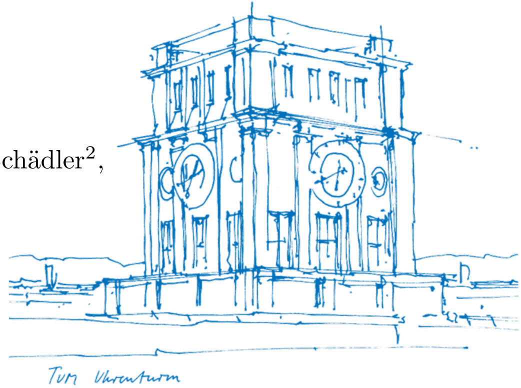
# Nonlinear Equalization for Optical Communications

## Based on Entropy-Regularized Mean Square Error




Francesca Diedolo[1], Georg Böcherer[2], Maximilian Schädler[2], Stefano Calabrò[2]

[1]Technical University of Munich
Institute for Communications Engineering

[2]Huawei Munich Research Center

June 3, 2022

# Outline

- **Motivation:**
  - Nonlinear impairments in optical communications.
  - $\Rightarrow$ Cost functions for nonlinear equalization are required.
- **Contribution:**
  - New cost function for nonlinear equalization.

F. Diedolo, G. Böcherer, M. Schädler, *et al.*, *Nonlinear equalization for optical communications based on entropy-regularized mean square error*, submitted to ECOC, 2022. [Online]. Available: https://arxiv.org/abs/2206.01004

# Outline

# Cost Functions for Nonlinear Equalizers

- Compensation of component non-linearities critical in short-reach optics.
- Two cost functions popular in ML and applied in optics:

  **MSE**[1,2] Mean Squared Error

  **CE**[3,4] Cross Entropy

[1] S. Zhang, F. Yaman, K. Nakamura, *et al.*, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nat. Commun.*, 2019.

[2] P. J. Freire, V. Neskornuik, A. Napoli, *et al.*, "Complex-valued neural network design for mitigation of signal distortions in optical links," *J. Light. Technol.*, 2021.

[3] M. Schädler, G. Böcherer, and S. Pachnicke, "Soft-demapping for short reach optical communication: A comparison of deep neural networks and volterra series," *J. Light. Technol.*, 2021.

[4] S. Deligiannidis, A. Bogris, C. Mesaritakis, *et al.*, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks," *J. Light. Technol.*, 2020.

# Mean Square Error: Suitable for Nonlinear Equalizers?

label observation estimate

$$X \rightarrow \boxed{\phantom{f}} \rightarrow R \rightarrow \boxed{f} \rightarrow Y \quad = X + Z, \quad Z \sim \mathcal{N}(0, \sigma^2)$$

equalizer

- Mean square error: $|X - f(R)|^2$.
- Cross-entropy: $-\log Q_{X|Y}(X|f(R))$.
- **Their Claim:**[5] if $Y = X + Z$ with Gaussian $Z$:

$$\boxed{\text{MSE} \equiv \text{CE}}$$

- **Our Claim:** this statement is wrong!
- How did the authors arrive at this claim?

Advanced convolutional **neural networks** for nonlinearity mitigation in long-haul WDM transmission systems
…, S Sygletos, M Fedoruk, S Turitsyn - Journal of Lightwave …, 2021 - osapublishing.org
… of convolutional **neural networks** for compensating nonlinear … a single-layer convolutional **neural network**. Furthermore, we … in the large multi-layer **network**. We examine application of …
Save Cite Cited by 24 Related articles All 11 versions

Performance versus complexity study of **neural network** equalizers in coherent optical systems
…, N Costa, JE Prilepsky, SK Turitsyn - Journal of Lightwave …, 2021 - osapublishing.org
We present the results of the comparative performance-versus-complexity analysis for the several types of artificial **neural networks** (NNs) used for nonlinear channel equalization in …
Save Cite Cited by 27 Related articles All 10 versions

Convolutional long short-term memory **neural network** equalizer for nonlinear Fourier transform-based optical transmission systems
…, A Vasylchenkova, JE Prilepsky, SK Turitsyn - Optics …, 2021 - osapublishing.org
… of such equalisers utilising artificial **neural networks** (NN) of two types… methods such as feed-forward **neural networks** (FFNN), k-… Nonetheless, when using the recurrent **neural networks** (…
Save Cite Cited by 15 Related articles All 12 versions

Transfer learning for **neural networks**-based equalizers in coherent optical systems
…, B Spinnler, A Napoli, SK Turitsyn - Journal of Lightwave …, 2021 - osapublishing.org
In this work, we address the question of the adaptability of artificial **neural networks** (NNs) used for impairments mitigation in optical transmission systems. We demonstrate that by using …
Save Cite Cited by 9 Related articles All 11 versions

[HTML] **Neural networks** for computing and denoising the continuous nonlinear Fourier spectrum in focusing nonlinear Schrödinger equation
…, M Kamalian-Kopae, IS Chekhovskoy, SK Turitsyn… - Scientific Reports, 2021 - nature.com
… **Net** as compared to the conventional numerical NFT methods becomes evident when we deal with noise-corrupted signals, where the **neural networks**… we train the **neural network** on the …
Save Cite Cited by 1 Related articles All 14 versions

Realization of **neural network**-based optical channel equalizer in restricted hardware
…, A Napoli, M Kamalian-Kopae, SK Turitsyn - arXiv preprint arXiv …, 2021 - arxiv.org
We quantify the achievable reduction of the processing complexity of artificial **neural network**-based equalizers in a coherent optical channel using the pruning and quantization …
Save Cite Cited by 2 Related articles All 2 versions

**Neural networks** based post-equalization in coherent optical systems: regression versus classification
…, JE Prilepsky, Y Osadchuk, SK Turitsyn… - arXiv preprint arXiv …, 2021 - arxiv.org
In this paper, we address the question of which type of predictive modeling, classification, or regression, fits better the task of equalization using **neural networks** (NN) based post-…
Save Cite Cited by 1 Related articles All 2 versions

[5] Section II.A, P. J. Freire, J. E. Prilepsky, Y. Osadchuk, *et al.*, "Neural networks based post-equalization in coherent optical systems: Regression versus classification," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2109.13843v3

# Likelihood Versus Posterior Probability

label observation estimate

$$X \longrightarrow \boxed{\phantom{x}} \longrightarrow R \longrightarrow \boxed{f} \longrightarrow Y$$

## Conditional Log-Likelihood and Mean Squared Error[6]

- Assume

$$X = Y + Z, \quad Z \sim \mathcal{N}(0, \sigma^2)$$

- Cross-entropy is then

$$-\log Q_{X|Y}(X|f(R)) = \log \sqrt{2\pi\sigma^2} + \frac{|X - f(R)|^2}{2\sigma^2} \log e$$

- Obviously,

$$\boxed{\mathrm{MSE} \equiv \mathrm{CE}}.$$

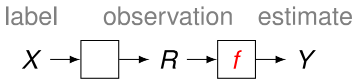
**Clash of terminology:**

| | Communications | Goodfellow (2016) Section 5.5.1 |
|---|---|---|
| $Q_{X\|Y}$ | posterior | likelihood |
| $Q_{Y\|X}$ | likelihood | |

$\Rightarrow$ Supposedly, the author's[5] claim is based on confusing the posterior $Q_{X|Y}(\cdot|y)$ being Gaussian with the likelihood $Q_{Y|X}(\cdot|x)$ being Gaussian.

[6] Section 5.5.1 of I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016

# When MSE $\equiv$ CE




- Suppose

$$Y = X + Z$$

with $Z$ **and** $X$ independent Gaussian.

$\Rightarrow$ The posterior $Q_{X|Y}(\cdot|y)$ is Gaussian[7].

$\Rightarrow$ In this case[8]

$$\boxed{\text{MSE} \equiv \text{CE}}$$

## Consequences

- Information theory often assumes Gaussian inputs, and in this case MSE $\equiv$ CE.
- When $X$ and $Y$ are oversampled, the assumption $Q_{X|Y}$ being Gaussian may not be that bad, which explains why MSE works well in this case.
- When $X$ and $Y$ are multiplexed digital subcarrier signals, $X$ is pretty much Gaussian and MSE works well on the multiplexed signals.

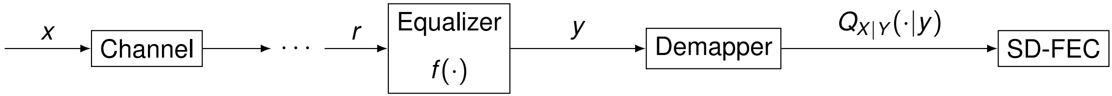
> In the following, we consider practically relevant 1 sample per symbol (SPS) signals. As we will see, MSE $\not\equiv$ CE.

[4][8, Sec. 3.5] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013

[5][8, Sec. 10.7]

# Outline

# Channel Model




The transmit symbols $x$ are drawn from a finite alphabet $\mathcal{X}$, e.g., QAM, ASK.

Most systems use binary FEC and the demapper output is a log-likelihood ratio (LLR)

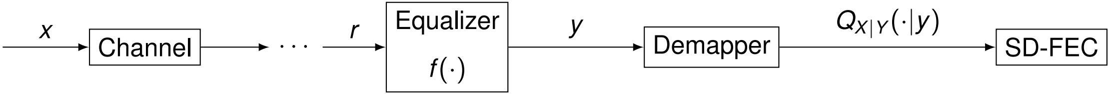
$$\ell_i = \log \frac{Q_{B_i|Y}(0|y)}{Q_{B_i|Y}(1|y)},$$

which can be calculated from $Q_{X|Y}$ through

$$Q_{B_i|Y}(b|y) = \sum_{x \in \mathcal{X}_i^b} Q_{X|Y}(x|y),$$

where $\mathcal{X}_i^b$ is the set of constellation points with the $i$-th label bit equal to $b$, $b \in \{0, 1\}$.

# Design Criterium for SD-FEC




An achievable information rate (AIR) of a system with demapper $Q_{X|Y}$ is[9,10]

$$\left[\mathrm{H}(X) - \mathbb{E}[-\log Q_{X|Y}(X|Y)]\right]^{+}$$

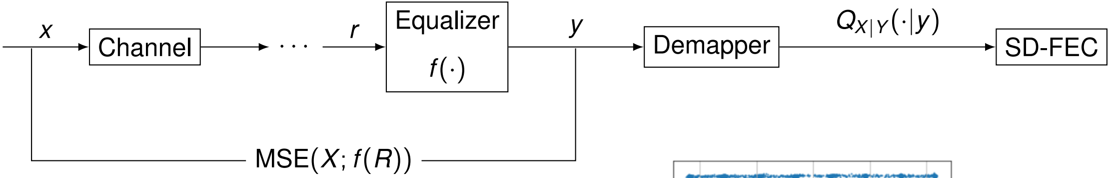
As the input entropy $\mathrm{H}(X)$ does not depend on the receiver, the design problem can be rephrased as

$$\underset{\text{Equalizer, Demapper}}{\text{minimize}} \quad \mathbb{E}[-\log Q_{X|Y}(X|Y)]$$

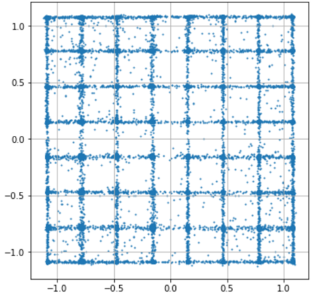
[9] N. Merhav, G. Kaplan, A. Lapidoth, *et al.*, “On information rates for mismatched decoders,” *IEEE Trans. Inf. Theory*, 1994.
[10] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic shaping and forward error correction for fiber-optic communication systems,” *J. Light. Technol.*, 2019.

# MSE Training

$$\boxed{\text{minimize } \mathrm{MSE}(X, f(R)) = \underset{f}{\text{minimize}}\, \mathbb{E}[|f(R) - X|^2]}$$

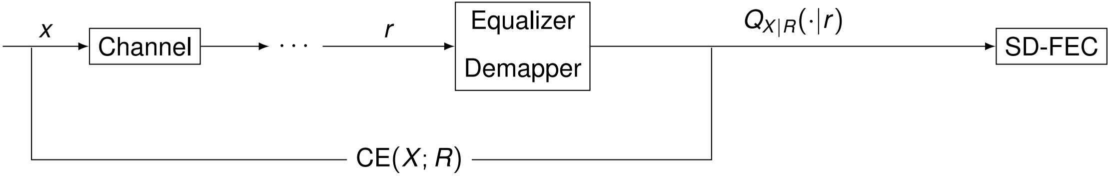



- It minimizes the pre-FEC bit error rate (BER)
- Does not maximizes the AIR
- Hence suboptimal for SD-FEC systems



# CE Training

$$\text{minimize } \mathrm{CE}(X,R) = \underset{Q_{X|R}}{\text{minimize}}\, \mathbb{E}[-\log Q_{X|R}(X|R)]$$




- It maximizes the AIR
- The trained device acts as an equalizer and demapper jointly
- Some algorithms, e.g. carrier and timing recovery, need access to the equalized signal which is lost

# A New Cost Function

We wish to find a cost function which

- Maximizes the AIR for SD-FEC systems
- Achieves same pre-FEC BER as MSE cost function
- Preserves the block structure and the access to an equalized signal

# Entropy-Regularized MSE (I)

Our approach consists in optimizing the equalizer based on the demapper output

$$\underset{f}{\text{minimize}} \qquad \mathbb{E}[-\log Q_{X|Y}(X|f(R))]$$

where we choose the **AWGN demapper**

$$Q_{X|Y}(x|y) = \frac{P_X(x)Q_{Y|X}(y|x)}{Q_Y(y)}$$

where $P_X$ is the input distribution and

$$Q_{Y|X}(y|x) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right]$$

is a Gaussian channel and

$$Q_Y(y) = \sum_{x' \in \mathcal{X}} P_X(x')Q_{Y|X}(y|x').$$

Note that $x \in \mathcal{X}$ and the demapper parameters are $\sigma^2$ and $\mathcal{X}$.

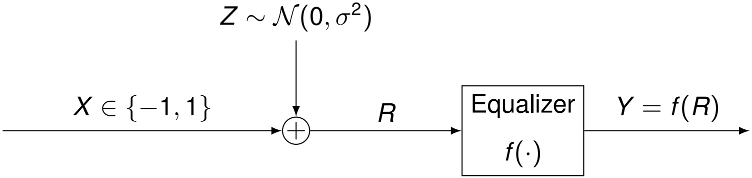
# Entropy-Regularized MSE (II)

Now, we solve the optimization problem

$$\begin{aligned}
\arg\min_f \mathbb{E}[-\log Q_{X|Y}(X|f(R))] &= \arg\min_f \mathbb{E}\left[-\log \frac{P_X(X)Q_{Y|X}(f(R)|X)}{Q_Y(f(R))}\right] \\
&= \arg\min_f \mathbb{E}\Big[-\log \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-|f(R)-X|^2}{2\sigma^2}}\Big] + \mathbb{E}\Big[-\log P_X(X)\Big] - \mathbb{E}\Big[-\log Q_Y(Y)\Big] \\
&= \arg\min_f \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2}\mathbb{E}[|f(R)-X|^2] + \mathrm{h}(X) - \mathbb{E}\Big[-\log Q_Y(f(R))\Big]
\end{aligned}$$

After dropping all the terms which do not depend on $f$ the optimization problem is reduced to

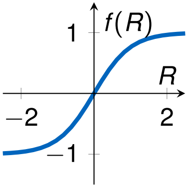
$$\begin{aligned}
&\arg\min_f \underbrace{\mathbb{E}[|f(R)-X|^2]}_{\text{MSE}(X,f(R))} - \underbrace{2\sigma^2\mathbb{E}[-\log Q_Y(f(R))]}_{\text{Entropy regularization}} \\
&= \arg\min_f \text{MSE-X}(X, f(R))
\end{aligned}$$




## MSE

$$\arg\min_f \quad \mathbb{E}[|f(R) - X|^2]$$

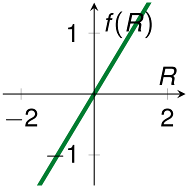
$$f^*(R) = \tanh(R)$$



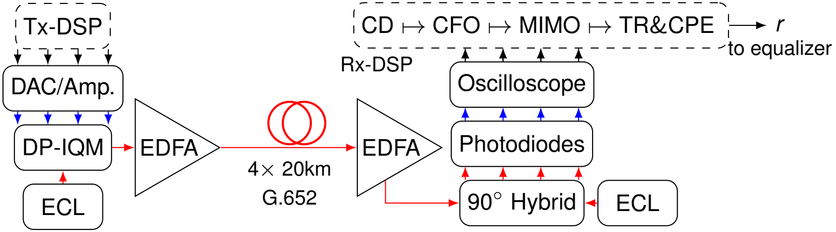

## MSE-X

$$\arg\min_f \quad \mathbb{E}[|f(R) - X|^2] - 2\sigma^2 \mathbb{E}[-\log Q_Y(f(R))]$$

$$f^*(R) = R$$



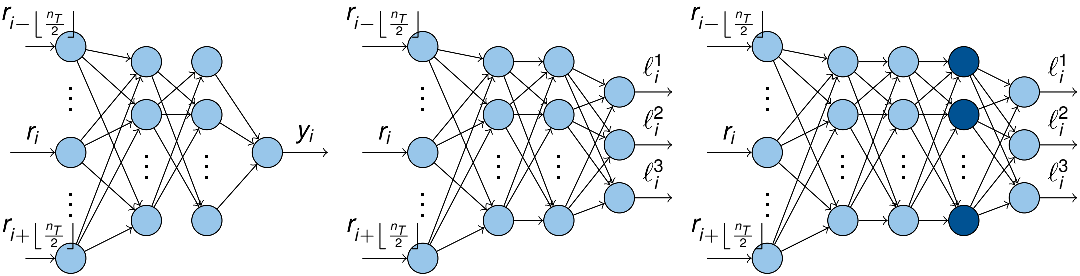
# Experimental Setup




- 80 GBd DP-64QAM transmitted signal with gross data rate of 960 Gb/s and net rate 800 Gb/s
- CAZAC training sequence for frame and carrier frequency synchronization, and channel estimation.
- Four 120GSa/s digital-to-analog converters (DACs) generate an electrical signal amplified by four 60GHz 3dB-bandwidth amplifiers.
- A tunable 100kHz external cavity laser (ECL) generates a continuous wave that is modulated by a 32GHz 3dB-bandwidth DP-I/Q modulator.
- The receiver has an optical 90°-hybrid and four 100GHz balanced photodiodes
- E/O conversion by an oscilloscope with 256GSa/s and 110GHz 3dB-bandwidth.

# Neural Networks Structure

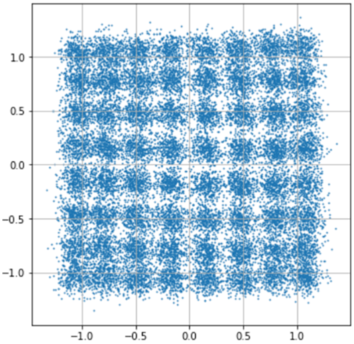
| Structure | Cost function | Name | Activation |
|---|---|---|---|
| 17\|32\|26\|1 | MSE, MSE-X | $\mathrm{NN}_{\mathrm{eq}}$ | ReLU |
| 17\|32\|26\|3 | BCE | $\mathrm{NN}^{1}_{\mathrm{joint}}$ | ReLU |
| 17\|32\|26\|16\|3 | BCE | $\mathrm{NN}^{2}_{\mathrm{joint}}$ | ReLU |



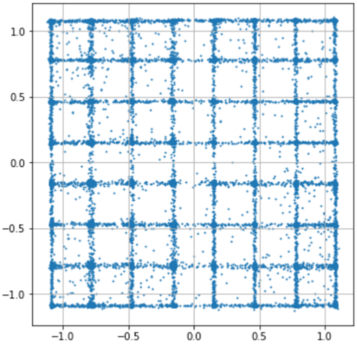

**Note**: $\mathrm{NN}_{\mathrm{eq}}$ is followed by the demapper. The parameters $\sigma^2$ and $\mathcal{X}$ are left constant during training, then learned from the equalized train data, then fixed and applied to the test data.

# Constellation Plot

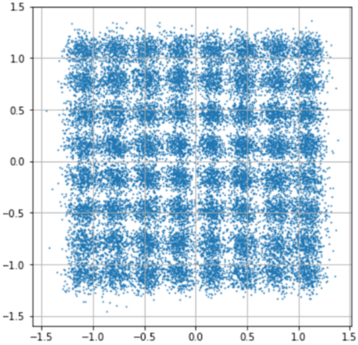
Constellation before equalizer



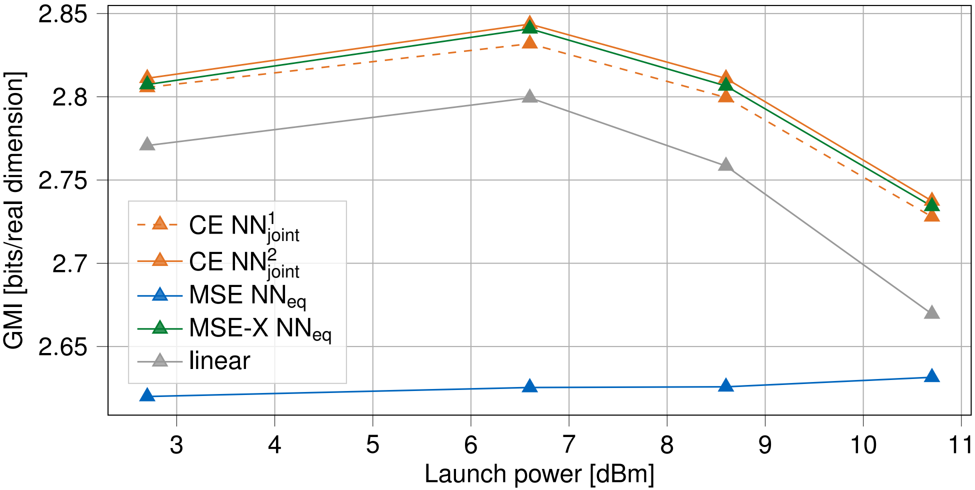

Constellation after MSE equalizer



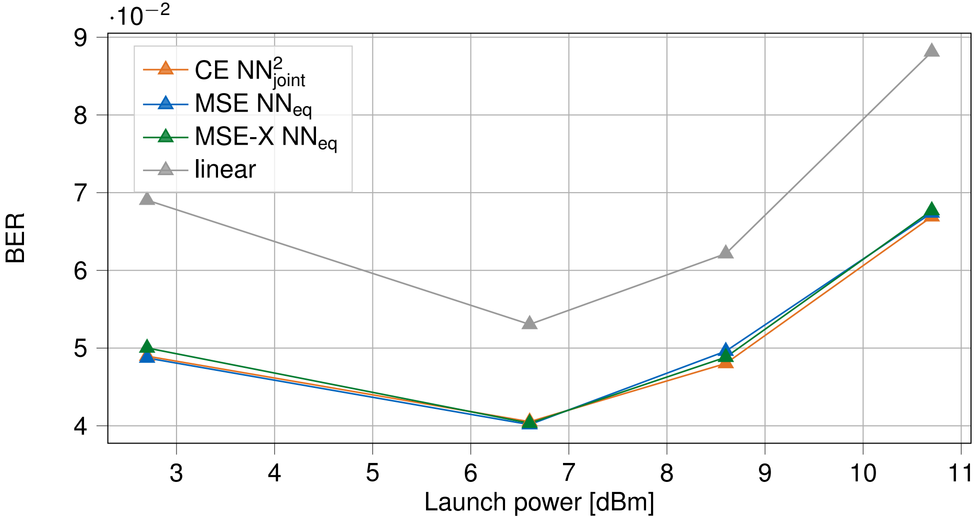

Constellation after MSE-X equalizer



# Experimental Results - GMI



# Experimental Results - BER



# Volterra Nonlinear Equalization

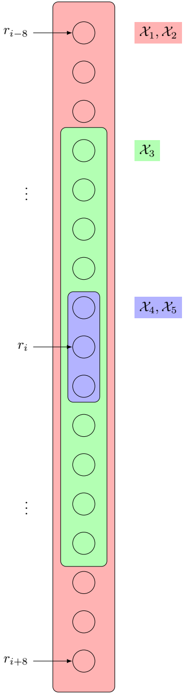
- Linear FIR filter: use $n_T$ taps $\mathcal{X}_1 = \{r_{i-\lfloor \frac{n_T}{2} \rfloor}, \ldots, r_i, \ldots, r_{i+\lfloor \frac{n_T}{2} \rfloor}\}$
- This corresponds to order 1:

$$\mathcal{F}_1 = \{a \colon a \in \mathcal{X}_1\} = \mathcal{X}_1$$

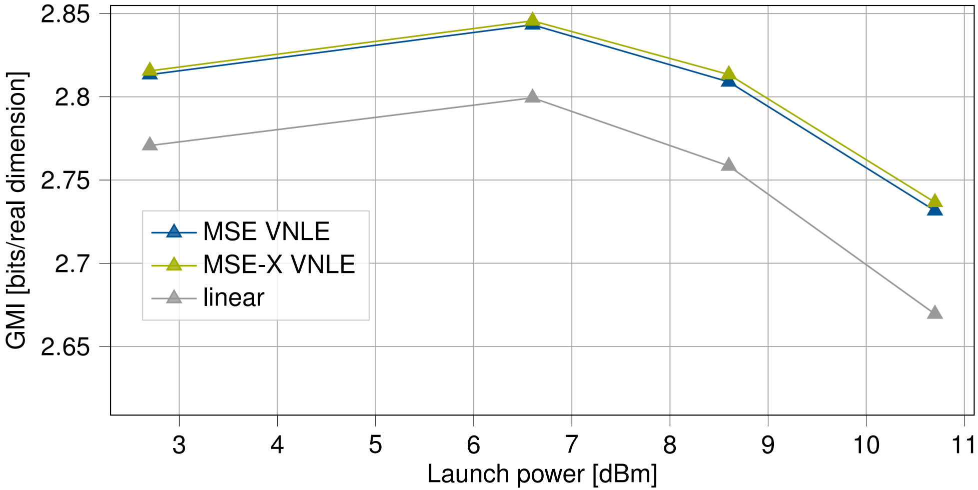
- Volterra equalizer adds higher orders:

$$\begin{aligned}
\mathcal{F}_2 &= \{a \cdot b \colon a, b \in \mathcal{X}_2\} \\
\mathcal{F}_3 &= \{a \cdot b \cdot c \colon a, b, c \in \mathcal{X}_3\} \\
&\vdots \\
\mathcal{F}_k &= \{a_1 \cdot a_2 \cdot \cdots \cdot a_k \colon a_1, \ldots, a_k \in \mathcal{X}_k\}
\end{aligned}$$

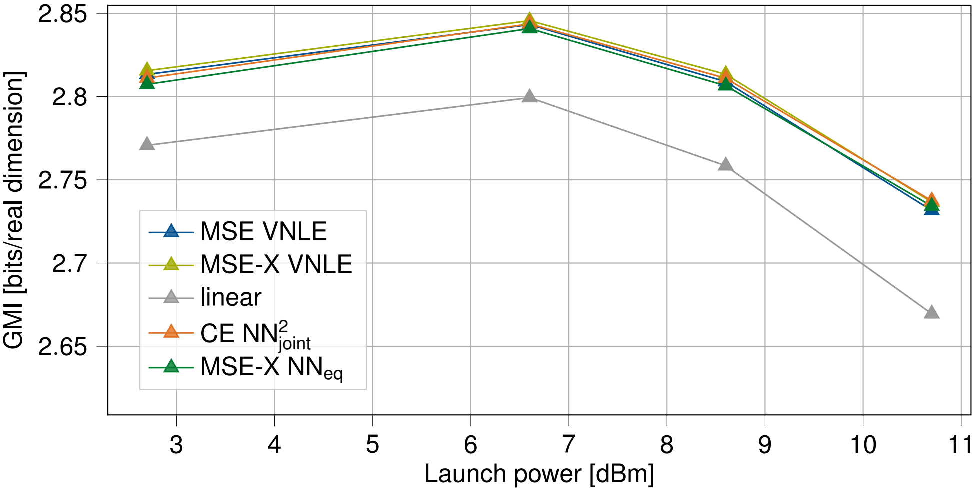
- The features $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are inputs to a linear FIR filter.
- The linear FIR filter has $|\mathcal{F}_1| + \cdots + |\mathcal{F}_k|$ taps.
- Our Volterra equalizer specified to the right has 492 taps.



# Experimental Results - VNLE



# Experimental Results - Comparison



# Outline

# Summary and Future Work

**Summary:**

- MSE cost function is suboptimal for SD-FEC systems
- CE training causes the loss of the equalized signal, useful for DSP algorithms
- We proposed a new objective to train nonlinear equalizers:

$$\boxed{\text{MSE-X}(f(R), X) = \mathbb{E}[|f(R) - X|^2] - 2\sigma^2\mathbb{E}[-\log Q_Y(f(R))]}$$

- We tested the cost function on experimental data and show compatibility with VNLE

**Future work:**

- Test on IM-DD systems
- Optimize the demapper parameters:
  - Noise power parameter $\sigma^2$
  - Target constellation $\mathcal{X}$
- Nonlinear equalization for multicarrier systems

# References I


[1] F. Diedolo, G. Böcherer, M. Schädler, and S. Calabrò, *Nonlinear equalization for optical communications based on entropy-regularized mean square error*, submitted to ECOC, 2022. [Online]. Available: https://arxiv.org/abs/2206.01004.

[2] S. Zhang, F. Yaman, K. Nakamura, *et al.*, “Field and lab experimental demonstration of nonlinear impairment compensation using neural networks,” *Nat. Commun.*, 2019.

[3] P. J. Freire, V. Neskornuik, A. Napoli, *et al.*, “Complex-valued neural network design for mitigation of signal distortions in optical links,” *J. Light. Technol.*, 2021.

[4] M. Schädler, G. Böcherer, and S. Pachnicke, “Soft-demapping for short reach optical communication: A comparison of deep neural networks and volterra series,” *J. Light. Technol.*, 2021.

[5] S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, “Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks,” *J. Light. Technol.*, 2020.

[6] P. J. Freire, J. E. Prilepsky, Y. Osadchuk, S. K. Turitsyn, and V. Aref, “Neural networks based post-equalization in coherent optical systems: Regression versus classification,” *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2109.13843v3.

# References II


[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[8] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[9] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, “On information rates for mismatched decoders,” *IEEE Trans. Inf. Theory*, 1994.

[10] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic shaping and forward error correction for fiber-optic communication systems,” *J. Light. Technol.*, 2019.

[11] G. Böcherer, *Lecture notes on machine learning for communications*, 2022. [Online]. Available: http://georg-boecherer.de/mlcomm.