

# Multi-Task Learning to Enhance Generalizability of Neural Network Equalizers in Coherent Optical Systems

Sasipim Srivallapanondh<sup>(1)</sup>, Pedro J. Freire<sup>(1)</sup>, Ashraful Alam<sup>(1)</sup>, Nelson Costa<sup>(2)</sup>, Bernhard Spinnler<sup>(3)</sup>, Antonio Napoli<sup>(3)</sup>, Egor Sedov<sup>(1)</sup>, Sergei K. Turitsyn<sup>(1)</sup>, Jaroslav E. Prilepsky<sup>(1)</sup>

<sup>(1)</sup> Aston University, United Kingdom [s.srivallapanondh@aston.ac.uk](mailto:s.srivallapanondh@aston.ac.uk) <sup>(2)</sup> Infinera, Carnaxide, Portugal  
<sup>(3)</sup> Infinera, Munich, Germany

**Abstract** *For the first time, multi-task learning is proposed to improve the flexibility of NN-based equalizers in coherent systems. A “single” NN-based equalizer improves Q-factor by up to 4 dB compared to CDC, without re-training, even with variations in launch power, symbol rate, or transmission distance.*

## Introduction

The demand for high-speed data transmission keeps increasing due to upcoming technologies (6G<sup>[1]</sup>, etc.). Coherent optical systems have emerged as a key solution to meet this demand. Nonetheless, the presence of linear and especially nonlinear distortions in fiber-optic systems limits the achievable information rates<sup>[2]–[4]</sup>. Various digital signal processing (DSP) techniques have been proposed for nonlinear effects mitigation in long-haul systems<sup>[3]</sup>. Neural networks (NNs) have recently emerged as an effective alternative for channel equalization: the NNs have demonstrated excellent capability to approximate the inverse of the optical channel transfer function, potentially outperforming conventional DSP approaches<sup>[5]–[7]</sup>. However, generalizability remains one of the main challenges of NN-based equalizers and attracts more attention<sup>[8]–[10]</sup>. Due to different values of accumulated chromatic dispersion (CD)<sup>[11]</sup>, or the presence of channel distortion, the equalizers in the receiver or transmitter require reconfiguration and must be adjustable to compensate for the variation of impairments as the channel characteristics change.

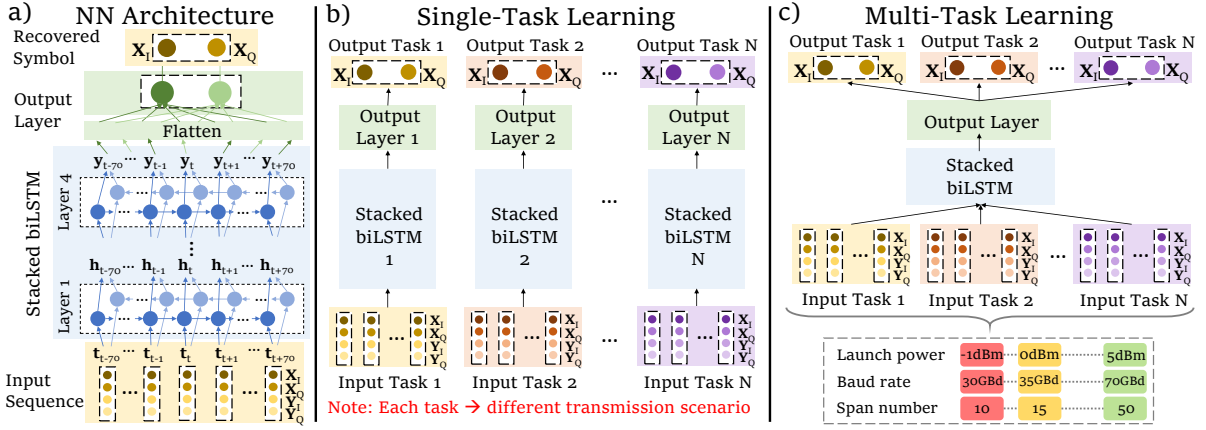
In this work, multi-task learning (MTL)<sup>[12]</sup> is proposed to calibrate the NN-based equalizer used for different transmission conditions in coherent systems. MTL leverages shared representations to enhance the adaptability of NN-based equalizers across different system configurations and optical impairments. This approach does not require re-training or additional data when the channel conditions change. Our results demonstrate the effectiveness of an MTL-based NN equalizer, which not only improves the equalization performance but also works efficiently in different transmission regimes and scenarios, leading to more generalizable and flexible solutions for NN-based nonlinear transmission effect mitigation.

## Multi-Task Learning for NN-based Equalizers

Single Task Learning (STL) is a commonly used approach to train NNs. STL refers to the training in which the NN learns the representation of the function to provide the output of a “specific” task<sup>[12]</sup>. One advantage of STL is that it allows the NN to focus solely on a specific task, usually leading to very good performance in that task. However, the NN may behave poorly when applied to different tasks (e.g., when the transmission scenario of interest is not included in the initial training dataset). As shown in Fig. 1b, if STL is used for channel equalization in different transmission scenarios, multiple NN models are usually required to provide acceptable performance. In MTL, the NN is trained with multiple datasets from multiple related tasks. In this case, the common representations learned from different but related tasks are shared<sup>[12],[13]</sup>. As depicted in Fig. 1c, MTL enables a single NN to equalize the signal in different ranges of launch power, symbol rate, and transmission distance by the joint training on the datasets from different transmission scenarios. MTL allows the NN to generalize better by using the domain-specific information contained in the different related tasks<sup>[12]</sup>.

Besides the generalization feature enabled by the MTL, it reduces hardware costs. In fact, the shared weights are fixed, which results in the simplification of the multipliers<sup>[11]</sup>. However, MTL can also lead to some disadvantages compared to the STL. Firstly, there is a trade-off between the performance of individual tasks and the overall performance of the equalizer. Secondly, the degree of information sharing between tasks has to be carefully controlled. Too much sharing can cause a negative information transfer, resulting in performance degradation for each task<sup>[13]</sup>.

In this work, we investigate the performance of NN-based equalizers using MTL where a sin-



**Fig. 1:** a) Equalizer architecture with 4-layer biLSTM and a dense layer; b) STL: multiple models are required for multiple transmission scenarios; c) MTL: only one model is required for multiple transmission scenarios.

gle NN, without re-training, is potentially capable of recovering the transmitted symbol independently of the specific parameters of the transmission systems. The considered transmission setup is altered by changing the symbol rate ( $R_S$ ) and launch power ( $P$ ) of data channels and the transmission distance (number of spans,  $N_{Span}$ ). For the MTL, the NN is trained with different datasets resulting from the combination of different transmission setups (to share the weights and biases).

### Numerical Setup

The dataset was obtained by numerical simulation assuming the transmission of a single 16-QAM dual-polarization channel along the standard single-mode fiber (SSMF). The signal propagation through the fiber was represented by a generalized Manakov equation using the GPU-accelerated split-step Fourier method<sup>[14]</sup>. The SSMF is characterized by the effective nonlinearity coefficient  $\gamma = 1.2 \text{ (W} \cdot \text{km)}^{-1}$ , chromatic dispersion coefficient  $D = 16.8 \text{ ps/(nm} \cdot \text{km)}$ , and attenuation parameter  $\alpha = 0.21 \text{ dB/km}$ . At the end of each fiber span, the optical fiber losses were compensated by an erbium-doped fiber amplifier with a noise figure of 4.5 dB. Downsampling and CD compensation (CDC) were performed on the receiver end. Afterwards, the received symbols were normalized and used as inputs of the NN.

### Methodology

The NN architecture, depicted in Fig. 1a, contains a stack of four bidirectional-Long Short-Term Memory (biLSTM) layers with 100 hidden units in each layer coupled with a dense output layer of 2 neurons to deliver the real and imaginary values for the  $X$ -polarization. The biLSTM was selected because it outperformed other types of NNs when used for nonlinear compensation<sup>[7],[15]</sup>. The model took four input features resulting from the in-phase and quadrature components of the complex signal ( $X_I, X_Q, Y_I$ , and  $Y_Q$ ) where  $X_I +$

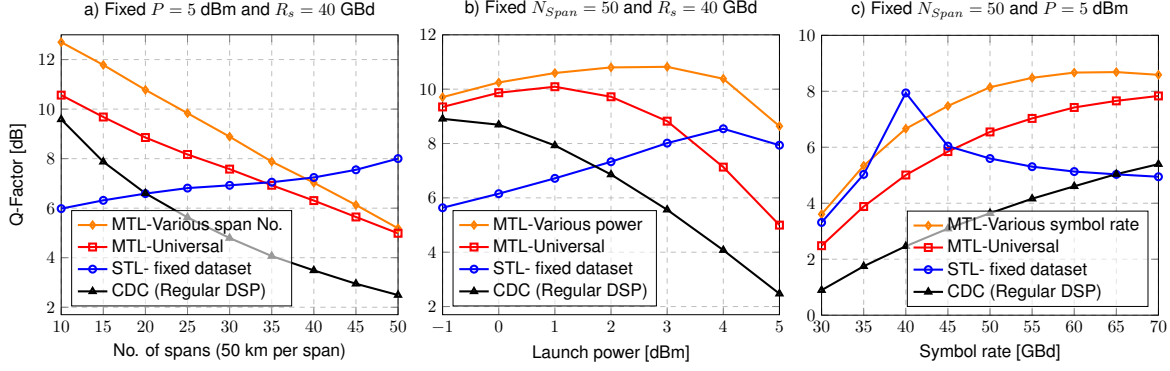
$jX_Q$  and  $Y_I + jY_Q$  were the signals in the  $X$  and  $Y$  polarizations, respectively. A set of 141 input symbols was fed to the NN to recover one symbol at the output. A new set of synthetic data of size  $2^{18}$  was randomly created with different system parameters and used in each training epoch to allow the model to learn different transmission scenarios. The entire training was carried out with a mini-batch size of 2000, and a learning rate of 0.001. The mean square error (MSE) loss estimator and the classical Adam algorithm<sup>[16]</sup> were applied when training the weights and biases.

The transmission scenarios include  $R_S$  ranging from 30 to 70 GBd, number of spans ranging between 10 and 50 (with fixed 50 km span length), and launch power ranging between -1 and 5 dBm. The NNs were trained with MTL or STL as follows:

1. MTL trained for 1000 epochs with datasets including different  $N_{Span}$ , but fixed  $R_S = 40 \text{ GBd}$  and  $P = 5 \text{ dBm}$ .
2. MTL trained for 1000 epochs with datasets including different  $P$ , but fixed  $N_{Span} = 50$  and  $R_S = 40 \text{ GBd}$ <sup>1</sup>.
3. MTL trained for 1000 epochs with datasets including different  $R_S$  but fixed  $N_{Span} = 50$  and  $P = 5 \text{ dBm}$ .
4. MTL trained for 1200 epochs with datasets including different combinations of  $N_{Span}$ ,  $R_S$ , and  $P$ . This NN is referred to as the “Universal model”<sup>2</sup>.
5. STL (without MTL) trained for 1000 epochs with fixed parameters:  $R_S = 40 \text{ GBd}$ ,  $N_{Span} = 50$  and  $P = 5 \text{ dBm}$ .

<sup>1</sup>This model has one extra input feature, which is the launch power. The model learns the data during the training using a normalized launch power. Therefore, it could not learn to generalize well without knowing the actual launch power.

<sup>2</sup>Here, the values of  $R_S$  and  $N_{Span}$  are randomly selected from the list of possible baud rate values with 5 GBd increment and the list of span number with the increments of 5 spans, respectively, to decrease the possible number of combinations for the NN’s learning.



**Fig. 2:** Q-factor resulting from using MTL (orange and red) and STL model (blue) in the following test cases; a) when the transmission distance changes but the launch power and symbol rate are set to 5 dBm and 40 GBd, respectively; b) when the launch power changes but the number of span and symbol rate are set to 50 and 40 GBd, respectively; c) when the symbol rate changes but the number of spans and launch power are set to 50 and 5 dBm, respectively.

## Results and Discussion

We considered MTL for multiple symbol rates, transmission distances, and launch powers. To evaluate equalization performance and generalizability, the MTL models were compared to CDC and the STL model trained with a fixed dataset.

**Variation of transmission distance:** Fig. 2a shows the optical performance for different reaches considering a fixed launch power of 5 dBm and a signal baud rate of 40 GBd. The STL model performed the best when  $N_{Span}$  was 50 (because it was trained for this specific transmission scenario), significantly outperforming the remaining approaches. However, its performance was significantly impacted in the shorter reaches as it could not generalize. On the other hand, the MTL trained with different  $N_{span}$  showed much better performance than STL for the shorter reaches, achieving a better Q-factor (about 3 dB Q-factor improvement) than CDC only for all considered scenarios. The universal MTL model also showed better performance than the CDC alone, leading to a maximum Q-factor improvement of about 2.5 dB at  $50 \times 50$  km.

**Variation of launch powers:** Fig. 2b depicts the Q-factor as a function of the launch power for a fixed  $R_S$  of 40 GBd and transmission distance of  $50 \times 50$  km. Again, the STL model showed the best gain for launch powers close to the one it was trained with (5 dBm), but revealed quite poor results for the remaining launch powers. In contrast, the universal MTL model enabled a Q-factor improvement exceeding 2 dB for the most relevant launch powers. The MTL, trained with various  $P$  but fixed  $N_{SPAN}$  and  $R_S$ , revealed the best performance, enabling a Q-factor improvement exceeding 4 dB for the most relevant launch powers. Interestingly, we can see that, at 5 dBm, the MTL outperformed STL. The reason for this may be that the STL is overfitting and cannot adapt

to the unseen test data as effectively as the MTL model, which is more generalized. Ref.<sup>[17]</sup> supported the claim that a more generalized model can perform better.

**Variation of symbol rates:** Fig. 2c illustrates the Q-factor as a function of the data signal baud rate for a fixed transmission distance and launch power of  $50 \times 50$  km and 5 dBm, respectively. STL led to very good results for the 40 GBd transmission scenario (training scenario) but showed very poor generalization capability. The MTL, trained with multiple  $R_S$  but fixed  $N_{Span}$  and  $P$ , enabled a Q-factor improvement of up to 4.5 dB with respect to the CDC only, whereas the universal MTL model showed up to 2.5 dB improvement. The MTL provided a good gain in most cases.

The aforementioned results show that, although STL may lead to outstanding performance in specific transmission conditions, it is not suitable for real-world system application because it lacks the adaptability to dynamic optical network parameters. MTL overcomes this limitation, allowing the equalizer to be more flexible, but at the cost of small performance degradation compared to models trained only for a specific task.

## Conclusions

Multi-task learning is proposed to allow a “single” NN-based equalizer, without re-training, to recover received symbols when the transmission scenarios change. The results showed that the MTL can provide up to 4 dB improvement in Q-factor with respect to CDC alone even if the transmission distance, launch power, and symbol rate vary, thus highlighting the adaptability of the MTL NN-based equalizer to the real-world dynamic optical network.

**Acknowledgements:** This work is supported by the EU H2020 Marie Skłodowska-Curie Action project MENTOR (No. 956713), SMARTNET EMJMD program under Grant 586686-EPP-1-2017-1-U.K.-EPPKA1-JMDMOB, EP-SRC project TRANSNET (EP/R035342/1), and the Horizon Europe project ALLEGRO, GA n. 101092766.

## References

- [1] H. Holma, H. Viswanathan, and P. Mogensen, "Extreme massive mimo for macro cell capacity boost in 5g-advanced and 6g", in *White paper*, Nokia Bell Labs, 2021.
- [2] R.-J. Essiambre and R. W. Tkach, "Capacity trends and limits of optical communication networks", *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1035–1055, 2012.
- [3] J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities", *Optics express*, vol. 25, no. 3, pp. 1916–1936, 2017.
- [4] P. Bayvel, R. Maher, T. Xu, *et al.*, "Maximizing the optical network capacity", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2062, p. 20 140 440, 2016.
- [5] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications", *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, 2019.
- [6] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks", in *Optical fiber communication conference*, Optical Society of America, 2018, W3A–4.
- [7] P. J. Freire, Y. Osadchuk, B. Spinnler, *et al.*, "Performance versus complexity study of neural network equalizers in coherent optical systems", *Journal of Lightwave Technology*, vol. 39, no. 19, pp. 6085–6096, 2021.
- [8] P. J. Freire, D. Abode, J. E. Prilepsky, *et al.*, "Transfer learning for neural networks-based equalizers in coherent optical systems", *Journal of Lightwave Technology*, vol. 39, no. 21, pp. 6733–6745, 2021.
- [9] P. J. Freire, B. Spinnler, D. Abode, *et al.*, "Domain adaptation: The key enabler of neural network equalizers in coherent optical systems", in *Optical Fiber Communication Conference*, Optica Publishing Group, 2022, Th2A–35.
- [10] J. Zhang, T. Xu, T. Jin, *et al.*, "Meta-learning assisted source domain optimization for transfer learning based optical fiber nonlinear equalization", *Journal of Lightwave Technology*, vol. 41, no. 5, pp. 1269–1277, 2022.
- [11] B. Liu, C. Bluemm, S. Calabrò, B. Li, and U. Schlichtmann, "Area-efficient neural network CD equalizer for 4x200Gb/s PAM4 CWDM4 systems", in *Optical Fiber Communication Conference (OFC)*, Optica, 2023, pp. 1–3.
- [12] R. Caruana, *Multitask learning*. Springer, 1998.
- [13] M. Crawshaw, "Multi-task learning with deep neural networks: A survey", *arXiv preprint arXiv:2009.09796*, 2020.
- [14] E. Sedov, *Hpcom*, version v0.1.4, Apr. 2023. DOI: 10.5281/zenodo.7880552. [Online]. Available: <https://doi.org/10.5281/zenodo.7880552>.
- [15] S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks", *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5991–5999, 2020.
- [16] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.
- [17] S. Srivallapanondh, P. J. Freire, B. Spinnler, *et al.*, "Knowledge distillation applied to optical channel equalization: Solving the parallelization problem of recurrent connection", in *Optical Fiber Communication Conference*, Optica Publishing Group, 2023.