Research Article | Vol. 32, No. 22 / 21 Oct 2024 / Optics Express 38642
Optics EXPRESS

Fast neural network inverse model to maximize throughput in ultra-wideband WDM systems

ZELIN GAN,[1,*] MYKYTA SHEVCHENKO,[2,3] SAM NALLAPERUMA HERZBERG,[4] AND SEB J. SAVORY[1]

[1]Electrical Engineering Division, Department of Engineering, University of Cambridge, 9 JJ Thomson Ave, Cambridge CB3 0FA, UK
[2]Department of Electronic and Electrical Engineering, University College London (UCL), Roberts Building, Torrington Place, London WC1E 7JE, UK
[3]National Physical Laboratory (NPL), Hampton Road, Teddington, Middlesex TW11 0LW, UK
[4]Department of Computer Science and Technology, University of Cambridge, William Gates Building, 15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
*zg276@cam.ac.uk

Abstract: Ultra-wideband systems expand the optical bandwidth in wavelength-division multiplexed (WDM) systems to provide increased capacity using the existing fiber infrastructure. In ultra-wideband transmission, power is transferred from shorter-wavelength WDM channels to longer-wavelength WDM channels due to inelastic inter-channel stimulated Raman scattering. Thus, managing launch power is necessary to improve the overall data throughput. While the launch power optimization problem can be solved by the particle swarm optimization method it is sensitive to the objective value and requires intensive objective calculations. Hence, we first propose a fast and accurate data-driven deep neural network-based physical layer in this paper which can achieve 99% – 100% throughput compared to the semi-analytical approach with more than 2 orders of magnitude improvement in computational time. To further reduce the computational time, we propose an iterative greedy algorithm enabled by the inverse model to well approximate a sub-optimal solution with less than 6% performance degradation but almost 3 orders of magnitude reduction in computational time.

1. Introduction

The relentless growth in data traffic in optical fiber communication systems over the past decades is driven by the increase of the internet services such as high-definition video streaming and the Internet of things (IoT). To satisfy this increasing demand, developing new strategies has become imperative. Generally, there are two main approaches: space-division multiplexing (SDM) [1] and ultra-wideband (UWB) wavelength-division multiplexing (WDM) [2]. SDM involves using multicore fiber (MCF) and multimode fiber (MMF), which spatially separate channels by using independent cores and modes respectively. Although the SDM has been demonstrated to achieve high throughput and long reach in optical fiber communication system, it still have much more work to do to be competitive with the existing single-mode fiber (SMF) [1]. On the other hand, the exploitation of the fiber transmission bandwidth beyond the conventional $C-$band combined with erbium-doped fiber amplifiers (EDFA) and other newly developed optical amplifiers is a more achievable solution [3].

The quality of transmission (QoT) of the optical fiber communication system can be calculated by using the split-step Fourier method (SSFM) to numerically solve the Manakov equation. The signal-to-noise ratio (SNR) is constrained due to a combination of the amplified spontaneous emission (ASE) noise and the fiber nonlinearity associated with optical power. In UWB systems,

#536632 https://doi.org/10.1364/OE.536632
Journal © 2024 Received 15 Jul 2024; revised 11 Sep 2024; accepted 25 Sep 2024; published 9 Oct 2024

Optics EXPRESS

interchannel stimulated Raman scattering (ISRS) has a significant impact on optical power, causing power to shift from shorter to longer wavelengths, which must be taken into account. It will further increase the time to obtain the numerical solution. The Gaussian noise (GN) model derived from perturbation analysis can provide an efficient analytic approximation of fiber nonlinear interference (NLI) [4]. The GN model with the presence of the ISRS can be found in [5], which can account for arbitrary power profiles and wavelength-dependent fiber parameters. The closed-form approximation of the ISRS GN model is proposed in [6], which provides a rapid estimate without any time-consuming integral evaluations.

The launch power control strategies have recently attracted a lot of attention which is enabled by the recent developments of accurate physical models. In [7], a simulated annealing (SA) algorithm is proposed to search for the optimal power slopes and offsets of $(C{+}L{+}S)-$band. It has been shown that managing power profiles by introducing a pre-tilt and an offset from dynamically configuring amplifiers can improve the performance in $(C+L)-$band links and optical network [8,9]. The key advantages of implementing the per wavelength launch power optimization enhanced with the optimal allocation of modulated signal bandwidth in the UWB point-to-point transmission systems have been recently demonstrated in [10,11]. Machine learning (ML) is recently exploited to be a scalable and efficient approach in optical fiber communication systems [12]. ML-based approaches are proposed to predict the characteristics of fiber-optic systems including the stimulated Raman scattering effect in [13,14]. A launch power controlled, digital twin assisted flat SNR optimization approach is proposed for $(C{+}L)-$band in [15].

This paper is an extension of the work we presented at the 2024 Optical Fiber Communications Conference and Exhibition (OFC) [16]. We investigate the way to reduce the computational complexity by using a surrogate model to replace the time-consuming physical layer calculations. We propose a novel more effective deep neural network (DNN) based approach in conjunction with particle swarm optimization (PSO) for launch power optimization in point-to-point optical transmission to increase the overall throughput. We also investigate using the inverse model to maximize the system throughput. Inverse models in modern photonics and optical fibre communication systems attract much attention in optimizing system performance, designing new devices, and compensating for various transmission impairments [17,18]. These models essentially work by predicting the necessary input conditions or system configurations to achieve a desired output, thereby reversing the conventional forward model. We implemented the inverse physical layer model by gradient descent and applied an iterative greedy algorithm to solve the throughput maximization problem. The performance is evaluated by bench-marking with the numerical approach. The proposed approach is comparable with the numerical approach without performance degradation and outperforms it in terms of computational complexity.

The paper is organized as follows: Section 2 discusses the detailed physical layer calculation and the proposed DNN structure of the launch power optimization problem. Then, the launch power optimization approaches considered in this paper are presented in Section 3. Section 4 contains the simulation setup and the performance evaluations of the proposed approach. Key results are concluded in Section 5.

2. Physical layer

2.1. Analytical model

It is customary to introduce the effective SNR vector $\mathbf{SNR} \triangleq \mathrm{SNR}\,(f_k)\ \forall\, k = \{1, \ldots, N_{\mathrm{ch}}\}$ under the locally-white noise assumption can be decomposed as follows:

$$\mathrm{SNR}\,(f_k) \approx \frac{P\,(f_k)}{\sigma^2_{\mathrm{ASE}}\,(f_k) + \sigma^2_{\mathrm{NLI}}\,(f_k)}, \tag{1}$$

Optics EXPRESS

where f_k is the relative frequency of k-th channel. The overall spectrum of ASE noise power for all spans can be expressed as follows:

$$\sigma^2_{\mathrm{ASE}}(f_k) = N_{\mathrm{span}} 10^{\mathrm{NF}(f_k)/10}\, h(f_0+f_k)\left[G(L_{\mathrm{s}},f_k)-1\right](1+r)R_s, \tag{2}$$

where NF stands for the noise figure , h is the Planck constant, f_0 is the carrier signal frequency, L_{s} is the fiber span length, and r is the roll-off factor, R_s is the symbol rate, and N_{span} is the number of identical fiber spans in a link. The ideal optical amplifier gain at the end of each span can be defined as follows $G(z,f_k) \triangleq P(0,f_k)\, P^{-1}(z,f_k)$ (cf., [11, Eq. (7)]). Without loss of generality, this assumption, i.e., identical spans and identical ideal lumped optical amplifiers, was made to simplify computations at the physical layer. The NLI noise power can be represented as

$$\sigma^2_{\mathrm{NLI}}(f_k) = \eta^{(\mathrm{SPM})}_{kk} P^3(f_k) + P(f_k) \sum_{l\neq k} \eta^{(\mathrm{XPM})}_{kl} P^2(f_l), \tag{3}$$

where the matrix $\boldsymbol{\eta} \triangleq \eta_{kl} \forall\ \{k,l\} = \{1,\ldots,N_{\mathrm{ch}}\}$ evaluates the amount of noise due to NLI indicated in Fig. 2 and N_{ch} is the total number of WDM channels. The approximate closed-form expressions for self-phase modulation (SPM) $\eta^{(\mathrm{SPM})}$, cross-phase modulation (XPM) $\eta^{(\mathrm{XPM})}$ NLI noise coefficients spectral profiles in Eq. (3) can be readily found in [6]. The impact of four-wave mixing (FWM) is neglected owing to the presence of a large amount of chromatic dispersion in a link.

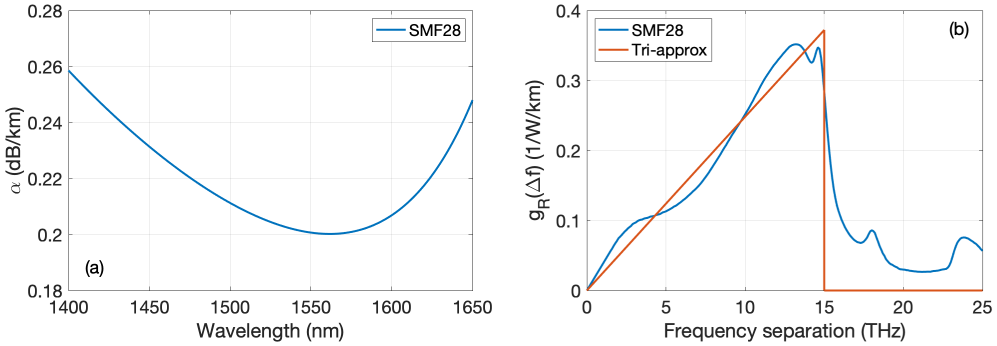

Fig. 1. Wavelength-dependent (a) fiber attenuation and (b) Raman gain spectrum of the standard single mode fiber (SMF)

To obtain the amplifier gain, which entirely compensates both the fibre loss and the power spectral tilt due to and ISRS effect, one needs to solve the Raman coupled ordinary differential equations (ODEs) [19]:

$$\begin{aligned}\frac{dP(z,f_k)}{dz} = &-\alpha(f_k)\,P(z,f_k) - \sum_{l=k+1}^{N_{\mathrm{ch}}} \frac{f_l}{f_k} g_R(|f_l-f_k|)\,P(z,f_l)\,P(z,f_k)\\ &+ \sum_{l=1}^{k-1} g_R(|f_l-f_k|)\,P(z,f_l)\,P(z,f_k),\end{aligned} \tag{4}$$

where $g_R(|f_l-f_k|)$ denotes the polarization-averaged Raman gain spectrum normalized by the fiber effective mode area $A_{\mathrm{eff}}(f_k)$ as a function of frequency separation and $\alpha(f_k)$ is the fiber loss, as shown in Fig. 1.

Optics EXPRESS

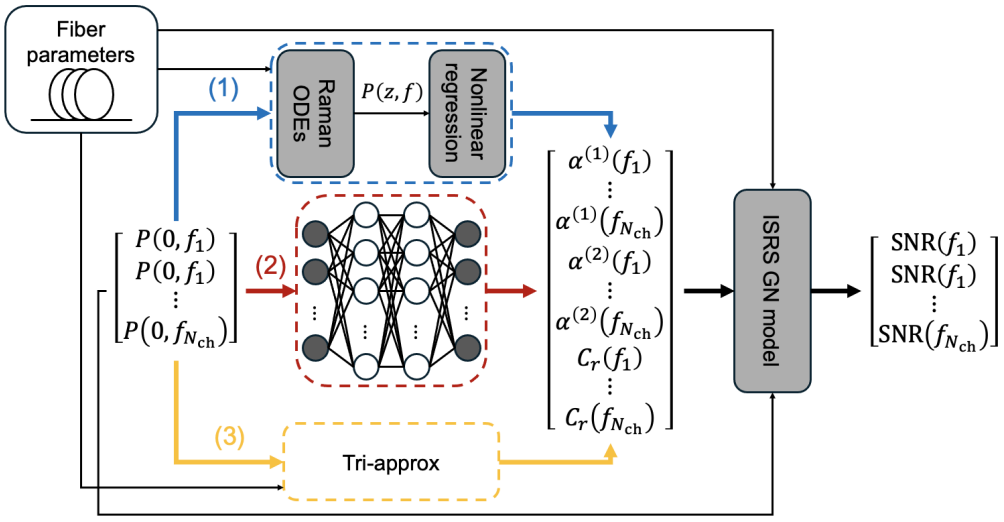

Fig. 2. Schematics for SNR prediction using ISRS GN model with f-dependent parameters $C_r(f_k)$, $\alpha^{(1)}(f_k)$ and $\alpha^{(2)}(f_k)$ in Eq. (6) by (1) numerically solving Raman ODEs followed a nonlinear regression, (2) a fully-connected DNN, and (3) triangle-approximation.

Given the triangle approximation of the Raman gain spectrum, Eq. (4) has an approximated analytical solution [20]:

$$G\left(z,f_k\right) \approx \frac{\exp\left(\alpha(f_k)\, z\right)}{P_{\text{tot}} \exp\left(-f_k C_r L_{\text{eff}} P_{\text{tot}}\right)} \sum_l P\left(f_k\right) \exp\left[-\left(f_k - f_l\right) C_r L_{\text{eff}} P_{\text{tot}}\right], \tag{5}$$

where $P_{\text{tot}} = \sum_{k=1}^{N_{\text{ch}}} P\left(f_k\right)$ defines the total optical launch power, C_r is the Raman gain slope, and L_{eff} is the fiber effective length. Note that the approximation is not accurate sufficiently if the overall modulated signal bandwidth exceeds 15 THz. In order to get the exact solution for Eq. (4), one needs to numerically solve the coupled Raman ODEs via, e.g., the Runge-Kutta (RK) method. It is customary to fit the obtained numerical solution as follows [6]:

$$\begin{aligned} G\left(z,f_k\right) \approx \Bigg\{ & \left(1 - \frac{P_{\text{tot}} f_k C_r\left(f_k\right)}{\alpha^{(2)}(f_k)}\right) \exp\left(-\alpha^{(1)}(f_k)\, z\right) \\ & + \frac{P_{\text{tot}} f_k C_r\left(f_k\right)}{\alpha^{(2)}(f_k)} \exp\left[-\left(\alpha^{(1)}(f_k) + \alpha^{(2)}(f_k)\right) z\right] \Bigg\}^{-1}, \end{aligned} \tag{6}$$

where $C_r(f_k)$, $\alpha^{(1)}(f_k)$ and $\alpha^{(2)}(f_k)$ are the $f-$dependent nonlinear regression fitting parameters, which account for the actual Raman gain spectrum.

The ultimate throughput is upper-bounded by the Shannon capacity of the additive white Gaussian noise (AWGN) channel, it then reads

$$\mathrm{C}_{\text{tot}} = 2R_s \sum_{k=1}^{N_{\text{ch}}} \log_2\left(1 + \mathrm{SNR}\left(f_k\right)\right). \tag{7}$$

2.2. Deep neural network

In this paper, we propose to use a DNN to significantly reduce the computational time at the physical layer as an efficient alternative method of numerically solving coupled Raman ODEs via the Runge-Kutta (RK) method in conjunction with the nonlinear regression (NLR) to find the

Optics EXPRESS

$f-$dependent parameters $C_r(f_k)$, $\alpha^{(1)}(f_k)$ and $\alpha^{(2)}(f_k)$ for closed-form ISRS GN model which is conventionally a fairly time-consuming process. We also compare the closed-form ISRS GN model with triangle-approximation. The process of above three cases can be described by the schematic shown in Fig. 2. The DNN model can be potentially experimental data-driven if the training data is the actual power evolution in the fiber to be used in a more practical scenario. In this paper, we used only simulation data to train the DNN model, with details of data generation and model training discussed in Section 4.

3. Launch power optimization

3.1. Particle swarm optimization

The impact of launch power on the system performance is nonlinear due to the presence of fiber nonlinearity, such as the nonlinear inter-channel interference (the optical Kerr effect) and ISRS. To maximize the system performance, an unconstrained $N_{\text{ch}}-$dimensional non-convex optimization problem needs to be solved [11,21], which can be expressed as follows:

$$P_{\text{opt}}(f_k) = \underset{P(f_k)}{\arg\max}\ \mathrm{C}_{\text{tot}}, \tag{8}$$

where C_{tot} is the total throughput of the nonlinear channel with launch power profile $P(f_k)$ calculated by Eq. (7) as the objective function. Our goal is to find the optimal launch power allocation $P_{\text{opt}}(f_k)$ to maximize the total throughput. To tackle this problem, the global search numerical optimization algorithms are vital. In this paper, we consider PSO [22], which can effectively search for a good candidate with a near-optimal solution for a given objective. We use the PSO provided by the global optimization toolbox of MATLAB. The PSO parameters we set are as follows: the swarm size of $10 \cdot N_{\text{ch}}$ and the maximum iteration of 10^4 with a function tolerance of 10^{-8}.

3.2. Inverse model

The inverse physical layer model for fiber transmission, finding the desired power profile given a target SNR, can be used to enhance the system performance. The schematic is shown in Fig. 3. The inverse model can be implemented by the gradient descent as the closed-form ISRS GN model is differentiable under triangle approximation. For the UWB system with a bandwidth over 15 THz, the derivative can be approximated by replacing the fitted parameters. For a given target SNR profile, the possible power profile that can achieve it can be found by the gradient descent algorithm. The iteration update is given by:

$$\Delta \boldsymbol{P} = \left(\mathbf{SNR}_{\text{target}} - \overline{\mathbf{SNR}}\right) \cdot \mathbf{J}^{-1}, \tag{9}$$

where $\mathbf{J}$ denotes the Jacobian matrix, which (i,j)-th entry is $J_{ij} = \frac{\partial \overline{\mathrm{SNR}}(f_i)}{\partial P(f_j)}$ representing the derivative of the i-th channel SNR with respect to the j-th channel launch power. The launch power profile $\boldsymbol{P}$ keeps updating as follows: $\boldsymbol{P} := \boldsymbol{P} + \mu \Delta \boldsymbol{P}$, whilst the iteration is proceeding until the convergence is entirely achieved. The learning rate μ can be adapted as follows:

$$10^{-2} \le \mu \triangleq \left(\mathbf{SNR}_{\text{target}} - \overline{\mathbf{SNR}}\right)^{\mathsf{T}} \left(\mathbf{SNR}_{\text{target}} - \overline{\mathbf{SNR}}\right) \cdot 10^{-3} \le 10^{0}\,. \tag{10}$$

3.3. Greedy algorithm

Unlike the PSO, the greedy algorithm is a simple, iterative heuristic to find the local optimum [23]. It maximizes the data rate of each channel. For k-th channel, its target data rate $\mathbf{C}_{\text{target}}(f_k)$

Optics EXPRESS

Algorithm 1. Greedy algorithm pseudocode

Input: Initial target data rates $\mathbf{C}_0$ and initial power profile $\boldsymbol{P}_0$
Output: Desired power profile $\boldsymbol{P}$
  $\mathbf{C}_{\text{target}} \leftarrow \mathbf{C}_0$
  $\mathbf{SNR}_{\text{target}} \leftarrow \mathbf{C}_{\text{target}}$
  $\boldsymbol{P} \leftarrow \boldsymbol{P}_0$
  for channel k in N_{ch} **do**
    while $\mathbf{C}_{\text{target}}$ is achieved **do**
      $\mathbf{C}_{\text{target}}(f_k) \leftarrow \mathbf{C}_{\text{target}}(f_k) + \Delta\text{C}$
      $\mathbf{SNR}_{\text{target}} \leftarrow \mathbf{C}_{\text{target}}$
      $\boldsymbol{P} \leftarrow$ `InverseModel` $(\boldsymbol{P}, \mathbf{SNR}_{\text{target}})$
    end while
  end for

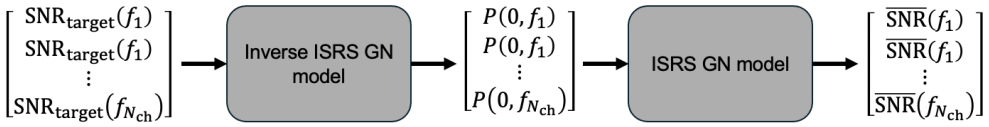

Fig. 3. Schematics for the inverse ISRS GN model.

will be increased 100 Gbps in every iteration until the overall target data rates $\mathbf{C}_{\text{target}}$ is not achievable. The condition is determined by the convergence of the gradient descent algorithm in Eq. (9). For the gradient descent algorithm, $\mathbf{C}_{\text{target}}$ and $\mathbf{SNR}_{\text{target}}$ are convertible through Eq. (7). In this paper, we set $\mathbf{C}_{\text{target}}$ to be an integer multiple of 100 Gbps which ΔC is equal to. The impact of ΔC on the algorithm can be found in Section 4. Other than maximizing a single channel per iteration, the complexity of the algorithm can be reduced by assigning multiple channels, in which k will be a set of integers. The pseudocode is shown in Algorithm 1.

The order and amount of the selected channels in each iteration will affect the performance of the greedy algorithm. Due to the SRS, the amplifier gain and ASE noise power vary across the bandwidth and are larger in the $S-$band, so typically higher power needs to be assigned to the $S-$band to account for it. Then, if maximizing $S-$band data rate first, the higher P_{tot} will stop the algorithm from finding the optimal data rate for $C-$ and $L-$band. Therefore, the greedy algorithm starts from $L-$band. Maximizing a single channel by assigning more power can achieve a higher SNR and data rate for itself but introduce higher NLI to its adjacent channels which will draw the greedy algorithm away from the potential optimal solution. To avoid this, we can maximize multiple adjacent channels or a sub-band of channels i.e., channel index k becomes a set of integers.

4. Results

4.1. Simulation setup

In the simulation, we consider a 10×100 km optical long-haul transmission system within $(C + L + S)-$band. The total modulated bandwidth is 20.2 THz from 185.95 THz to 206.15 THz following the ITU-T G.694.1 [24]. The frequency ranges of each band are set to 185.95 THz~190.75 THz for $L-$band, 191.25 THz~196.05 THz for $C-$band, and 196.55 THz~206.15 THz for $S-$band, including a practically relevant 500 GHz S/C and C/L guard band. Each channel is set to be modulated at a symbol rate of 96 GBd with a 100 GHz spacing in between, given 48 channels in $L-$band, 48 channels in $C-$band, and 96 channels in $S-$band which has a

Optics EXPRESS

total $N_{\text{ch}} = 192$ WDM channels. We consider a flat noise figure spectrum for each band, which are 7 dB for $S-$band, 4 dB for $C-$band, and 5 dB for $L-$band.

4.2. Performance analysis

The overall data contains $N_{\text{data}} = 1.2 \cdot 10^4$ launch power profiles from -12 to 12 dBm generated by smoothed random Gaussian walk as input and the corresponding fitted parameters as labels. The random Gaussian walk process can be described as, $P_{\text{in}}(n) = P_{\text{in}}(n-1) + w_n$, where $P_{\text{in}}(n)$ is the launch power at the n-th channel and $w_n \sim \mathcal{N}\left(0, \sigma_W^2\right)$ is a zero-mean Gaussian distributed random variable with a variance of σ_W^2 [25]. P_{in} is initiated from a uniform distribution $[-12, 12]$ dBm. The generated launch power profile is then smoothened by a box filter with a variable length for adjustable smoothness. The excursion of the profile can be controlled by the variance σ_W^2. There is a total power constraint for the small ISRS assumption from the closed-form ISRS GN model. Only the profiles that satisfy the total and boundary conditions will be used. The Levenberg-Marquard algorithm is used to solve the NLR problem. The input power profiles and fitted parameters were normalized with respect to their lower and upper bounds. We used 70% data in the training process and the rest in the validation process, given that $N_{\text{train}} = 8400$ and $N_{\text{validation}} = 3600$. We optimize the hyper-parameters of the DNN by means of the tree-structured Parzen estimator (TPE) algorithm implemented in Optuna [26]. It is followed by a multi-objective optimization of validation MAE and number of floating-point operations per second (FLOPs) to balance the trade-off between the accuracy and inference time. The search space and optimized values for the hyper-parameters are provided in the Table 1. The trials searched by multi-objective Optuna are shown in Fig. 4(a) and the Pareto front is labeled by a red dashed line. The parameters in the yellow marked trail are selected. The optimized DNN has an architecture of an input layer with 192 neurons corresponding to the launch power profile for each channel, 4 hidden layers with 400 neurons per layer following the *sigmoid* activation function, and an output layer with 576 neurons to predict the 3 parameters of each channel. The *Adam* optimizer with a learning rate of $2.47 \cdot 10^{-4}$ is used. The training process contains 1300 epochs and a batch size of 112. Given the mean absolute error (MAE) and mean squared error (MSE), we consider the MAE between the predicted and labeled fitted parameters as the loss. The training and validation curve of the DNN with selected parameters are shown in Fig. 4(b).

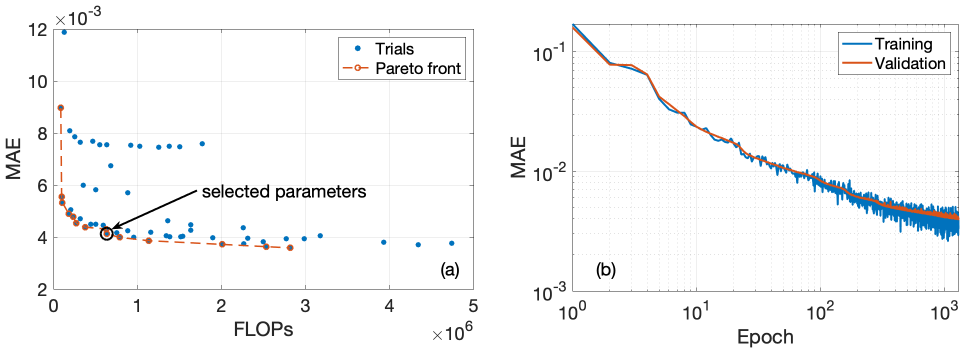

Fig. 4. (a) MAE and FLOPs of trails searched by Optuna; (b) training and validation curve of the DNN with selected parameters

We compare the mean SNR achieved by the launch power profiles from the validation dataset in one span calculated by using labels, predictions, and triangle approximation in Fig. 5(a) and the corresponding MAE of the SNR with respect to the value calculated from labels in Fig. 5(b). The maximum MAE of SNR along all channels is always less than 0.034 dB which provides an accurate prediction for the proposed DNN. If we apply the triangle-approximation beyond 15 THz,

Optics EXPRESS

Table 1. Hyper-parameters search space and its optimal values (where "$\in$" means a set membership, e.g., N_n is an integer between 100 and 1000 with an increment of 50)

Hyper-parameters	Notation	Search space	Optimal value
# of hidden layers	N_ℓ	$\{N_\ell \in [1, 6] \text{ and } N_\ell \in \mathbb{Z}\}$	4
# of neurons per layer	N_n	$\left\{N_n \in [100, 1000] \text{ and } \frac{N_n}{50} \in \mathbb{Z}\right\}$	400
Activation function	-	{Relu, Sigmoid, Tanh}	Sigmoid
Optimizer	-	{Adam, SGD}	Adam
Learning rate	l	$\left\{l \in \left[10^{-5}, 10^{-1}\right]\right\}$	$2.47 \cdot 10^{-4}$
# of epochs	N_E	$\left\{N_E \in [100, 1500] \text{ and } \frac{N_E}{100} \in \mathbb{Z}\right\}$	1300
Batch size	B	$\{B \in \mathbb{Z} \mid B \neq 0 : N_{\text{train}} \bmod B = 0\}$	112
Loss function	-	{MAE, MSE}	MAE

the mean SNR will have a maximum of 1.20 dB and an average of 0.36 dB deviation in $S-$band, a maximum of 0.45 dB and an average of 0.31 dB deviation in $C-$band, and a maximum of 1.43 dB and an average of 0.97 dB deviation in $L-$band. For the launch power optimization problem, we first compare the performance of PSO with different physical layer models, RK+NLR, DNN, and Triangle-approx. One of the initial swarm for the PSO is given as the optimal flat launch power found by sweeping from −12 to 12 dBm and others are generated uniformly at random. PSO is converged as the relative change in the objective function value is less than the function tolerance. The optimized launch power profiles are shown in Fig. 6(a). The optimized launch power profile by DNN is well aligned with the one by RK+NLR. The relatively scattered launch powers at $S-$band can be explained by the large validation MAE of the DNN at the corresponding wavelengths. However, the PSO with triangle-approx is far from the optimal value misled by the incorrect approximation of the ASE and NLI noise power. The SNR of the optimized launch power profiles are shown in Fig. 6(a) which supports that DNN can be considered a feasible replacement model for the RK+NLR.

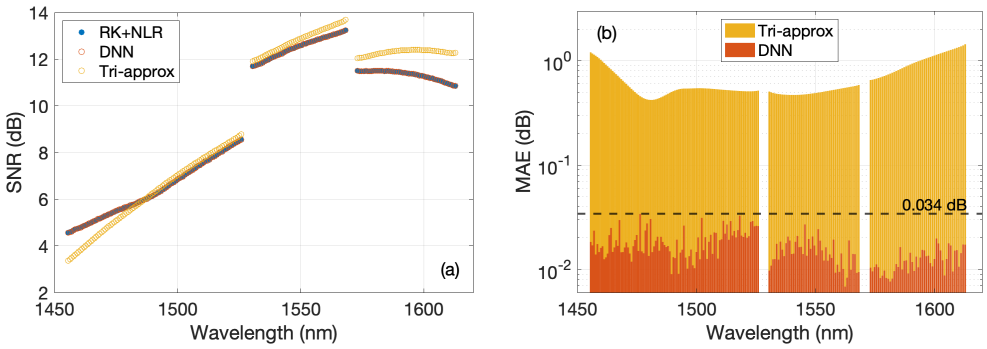

Fig. 5. Performance evaluation of RK+NLR, proposed DNN, and triangle-approximation on: (a) mean SNR per channel of the validation dataset; (b) MAE of the mean SNR with respect to the labeled dataset

The inverse physical layer model using DNN for the given target SNR is demonstrated in Fig. 7. In Fig. 7(a) and 7(b), we consider the scenario with a flat 4 dB NF and flat target SNR across the entire bandwidth. The target SNR profiles are shown as dashed lines and can be increased from 5.1 to 12.3 dB to achieve 400 to 800 Gbps data rate per channel. The inverse physical layer model will fail to converge if the target SNR is not achievable or the desired power profile exceeds the bound. We then re-evaluate the launch power profiles found by the inverse physical

Optics EXPRESS

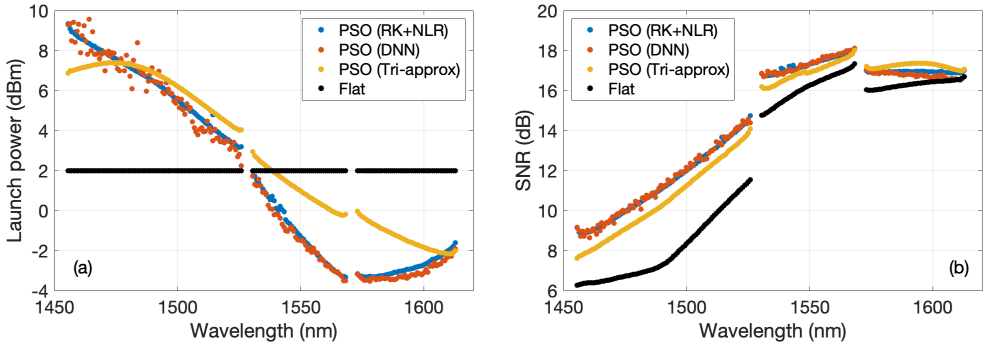

Fig. 6. Performance evaluation of RK+NLR, proposed DNN, and triangle-approximation on: (a) flat and optimized launch power profiles by PSO and (b) the corresponding SNR.

layer model through RK+NLR method as shown by the scattered points in Fig. 7(a). The case that has different NF per band, 7 dB for $S-$band, 4 dB for $C-$band, and 5 dB for $L-$band, is investigated in Fig. 7(c) and 7(d). The maximum achievable flat data rate per channel is 700 Gbps. The inverse physical layer model can effectively find the desired power profile to achieve a target SNR with the ripple always less than 0.1 dB.

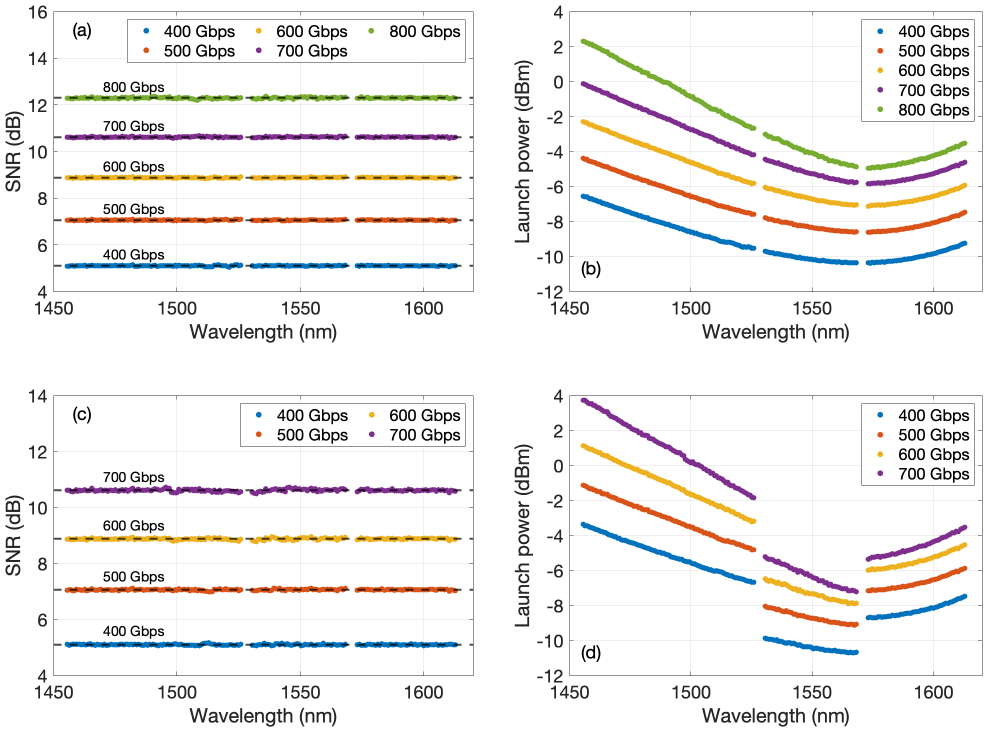

Fig. 7. (a) $\mathbf{SNR}_{\text{target}}$ and $\overline{\mathbf{SNR}}$ with flat noise figure; (b) launch power profile found by the inverse model; different noise figure for each band in (c) and (d).

For the greedy algorithm, we compare two cases: applying the DNN and triangle-approximation in the inverse physical layer model. We first find the optimal number of channels per sub-band to balance the achievable throughput and computational time of the greedy algorithm as shown in

Optics EXPRESS

Fig. 8(a) and 8(b). Given 6 channels per sub-band for both cases, the greedy algorithm can be completed with fewer iterations and avoid converging to a bad local optimum. The throughput gap between the greedy algorithm with the DNN and the PSO with RK+NLR is caused by the ΔC of 100 Gbps. This difference can be reduced by decreasing ΔC, however, it will require an extra computational time as shown in Fig. 8(c). The bad performance of the greedy algorithm with triangle-approximation is caused by the difference between $\mathbf{SNR}_{\text{target}}$ evaluated under triangle-approximation and $\overline{\mathbf{SNR}}$ re-evaluated using RK+NLR approach. Given the optimal number of channels per sub-band, the target and achieved SNR of the greedy algorithm with the DNN are shown in Fig. 9(a) and the desired launch power profile is shown in Fig. 9(b). The results of the greedy algorithm with triangle-approximation are shown in Fig. 9(c) and 9(d).

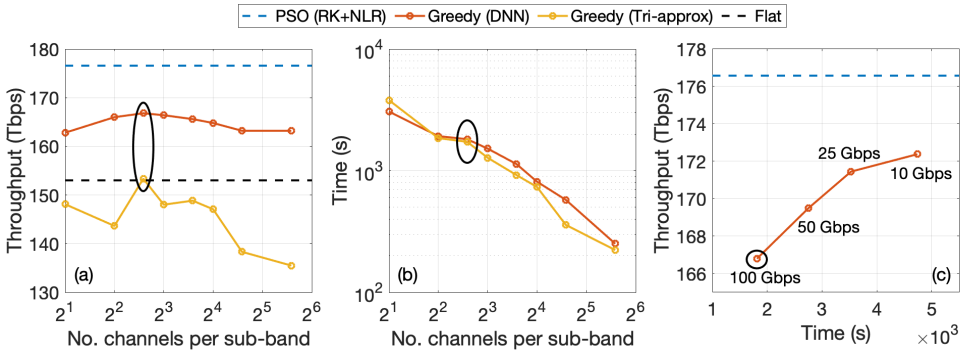

Fig. 8. (a) Achievable throughput and (b) computational time of the greedy algorithm with different number of channels per sub-band and (c) achievable throughput with different ΔC at 6 channels per sub-band.

Then, we compared the achieved throughput and the computation time of all the approaches discussed in this paper: the PSO with the physical layer model using (1) RK+NLR, (2) the DNN, (3) triangle-approximation, and the greedy algorithm using with the inverse physical layer model using (4) the DNN and (5) triangle-approximation. The throughput values achieved by the optimized launch power profiles through all approaches are 176.6, 176.3, 171.6, 166.8, and 153.3 Tbps for cases 1, 2, 3, 4, and 5 respectively re-evaluated by RK+NLR approach as shown in Fig. 10(a). The throughput achieved by the PSO with the DNN and triangle-approximation has less than 1% and around 3% difference, compared to PSO with RK+NLR respectively. The greedy algorithm with the DNN can achieve a difference lower than 6% throughput of the PSO with RK+NLR, but with triangle-approximation can only achieve around 13%. The computational time taken by each approach is shown in Fig. 10(b). Every approach runs on a machine with a 3.0 GHZ 8-core CPU under the same conditions. Using a DNN in the physical layer can reduce the time in seconds by more than 2 orders of magnitudes without performance degradation. A DNN also enables fast calculation of the inverse model and the greedy algorithm for a fast approximation of optimal solution with a time reduction of almost 3 orders of magnitudes in seconds.

Optics EXPRESS

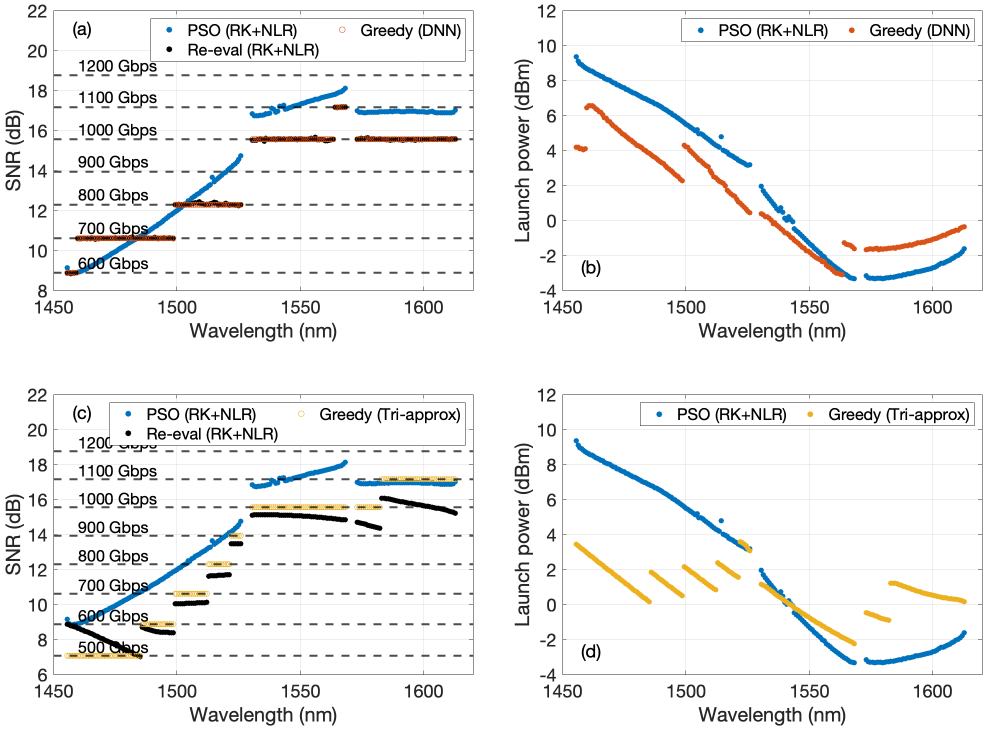

Fig. 9. (a, c) Optimal SNR and (b, d) power profile found by the greedy algorithm using DNN and Tri-approx respectively with considering 6 channels per sub-band.

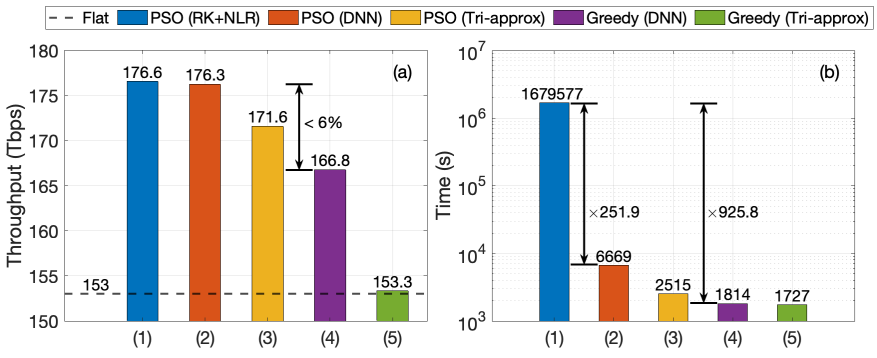

Fig. 10. (a) Achieved throughput and (b) computational time of PSO with physical layer models using (1), (2), and (3) and greedy algorithm with inverse physical layer models using (4) and (5).

Optics EXPRESS

5. Discussions and conclusion

In this work, we made assumptions to simplify the physical layer of our simulations, such as identical spans and ideal lumped amplification. Within these assumptions, the launch power profile at the beginning of each span as well as the lumped amplifier gain profile at the end of each fibre span remain the same throughout the transmission. Hence, the overall ASE noise power linearly scales with the number of fibre span N_{span} (see, Eq. (2)). The power-dependent NLI coefficients $\eta^{(\text{SPM})}$ and $\eta^{(\text{XPM})}$ are assumed to be the same for each span with only one evaluation needed for any arbitrary launch power profile. The total NLI coefficients can be modelled as a coherent accumulation process over multiple spans with a given coherence factor ϵ [4, Eq. (22)]. Since the DNN model is trained with launch power profiles generated via a smooth Gaussian random-walk process, the accuracy of the DNN model mentioned in this paper can still be achieved for any arbitrary amplifier gain profiles and any different values of launch power per span. A non-identical multi-span transmission can be modelled as a multiplication of unique DNN-based models, which must be specifically trained accounting for the fibre span parameters.

We presented an effective launch power assignment approach for UWB systems for combining effective learning capabilities from surrogate machine learning models with realistic optical transmission physical layer modeling. The proposed DNN model with PSO has achieved a near-optimal launch power profile exhibiting the potential of a DNN to learn from the power-dependent and $f-$dependent physical layer. Moreover, the proposed approach has significantly reduced the computational time by a magnitude of 2 compared to conventional physical layer estimation using numerical methods while the accuracy is preserved within the range of 99% – 100%. The proposed iterative greedy algorithm enabled by the inverse physical layer model using a DNN can effectively find the sub-optimal solutions with less than 6% performance degradation and substantial time reduction close to 3 orders of magnitude. In real optical networks, multiple fiber links in the topology increase the complexity of the launch power optimization problem significantly. The proposed greedy algorithm with the DNN approach might be deemed as a power-management tool, which can be potentially applied for dynamic UWB optical networks to attain higher throughput and lower margin values.

Funding. EPSRC programme grant TRANSNET (EP/R035342/1).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in [27].

References

1. D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," Nat. Photonics **7**(5), 354–362 (2013).
2. H. J. Thiele and M. Nebeling, *Coarse wavelength division multiplexing: Technologies and Applications* (CRC Press, 2007).
3. P. J. Winzer, "Would scaling to extreme ultraviolet or soft X-ray communications resolve the capacity crunch?" J. Lightwave Technol. **36**(24), 5786–5793 (2018).
4. P. Poggiolini, "The gn model of non-linear propagation in uncompensated coherent optical systems," J. Lightwave Technol. **30**(24), 3857–3879 (2012).
5. D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian noise model in the presence of inter-channel stimulated Raman scattering," J. Lightwave Technol. **36**(14), 3046–3055 (2018).
6. D. Semrau, R. I. Killey, and P. Bayvel, "A closed-form approximation of the Gaussian noise model in the presence of inter-channel stimulated Raman scattering," J. Lightwave Technol. **37**(9), 1924–1936 (2019).
7. H. Luo, J. Lu, and Z. Huang, "Optimization strategy of power control for C+L+S band transmission using a simulated annealing algorithm," Opt. Express **30**(1), 664–675 (2022).
8. Y. Song, Q. Fan, and C. Lu, "Efficient three-step amplifier configuration algorithm for dynamic C+L-band links in presence of stimulated Raman scattering," J. Lightwave Technol. **41**(5), 1445–1453 (2023).
9. Z. Huang, L. Dou, and J. Cheng, "Performance improvements by dynamic amplifier reconfigurations for C+L-band optical networks in the presence of stimulated Raman scattering," J. Opt. Commun. Netw. **15**(6), 344–356 (2023).
10. N. A. Shevchenko, S. Nallaperuma, and S. J. Savory, "Ultra-wideband information throughput attained via launch power allocation," in *2021 International Conference on Optical Network Design and Modeling (ONDM)*, (IEEE, 2021), pp. 1–3.

Optics EXPRESS

11. N. A. Shevchenko, S. Nallaperuma, and S. J. Savory, "Maximizing the information throughput of ultra-wideband fiber-optic communication systems," Opt. Express **30**(11), 19320–19331 (2022).
12. J. W. Nevin, S. Nallaperuma, and N. A. Shevchenko, "Machine learning for optical fiber communication systems: An introduction and overview," APL Photonics **6**(12), 121101 (2021).
13. X. Ye, A. Arnould, A. Ghazisaeidi, *et al.*, "Experimental prediction and design of ultra-wideband Raman amplifiers using neural networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, (2020), pp. 1–3.
14. A. M. R. Brusin, A. Nespola, M. R. Zefreh, *et al.*, "Ml-based spectral power profiles prediction in presence of isrs for ultra-wideband transmission," J. Lightwave Technol. **42**(1), 37–47 (2024).
15. Y. Zhang, X. Pang, and Y. Song, "Optical power control for GSNR optimization based on C+L-band digital twin systems," J. Lightwave Technol. **1**, 1–7 (2023).
16. Z. Gan, M. Shevchenko, S. N. Herzberg, *et al.*, "Fast and accurate DNN-based approach in maximizing ultra-wideband fiber-optic systems throughput," in *2024 Optical Fiber Communication Conference (OFC)*, (2024), p. M4K.6.
17. D. Zibar, A. M. Rosa Brusin, and U. C. de Moura, "Inverse system design using machine learning: The Raman amplifier case," J. Lightwave Technol. **38**(4), 736–753 (2020).
18. S. Molesky, Z. Lin, and A. Y. Piggott, "Inverse design in nanophotonics," Nat. Photonics **12**(11), 659–670 (2018).
19. S. Tariq and J. C. Palais, "A computer model of non-dispersion-limited stimulated Raman scattering in optical fiber multiple-channel communications," J. Lightwave Technol. **11**(12), 1914–1924 (1993).
20. D. Christodoulides and R. Jander, "Evolution of stimulated raman crosstalk in wavelength division multiplexed systems," IEEE Photonics Technol. Lett. **8**(12), 1722–1724 (1996).
21. I. Roberts, J. M. Kahn, J. Harley, *et al.*, "Channel power optimization of WDM systems following Gaussian noise nonlinearity model in presence of stimulated Raman scattering," J. Lightwave Technol. **35**(23), 5237–5249 (2017).
22. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4 (1995), pp. 1942–1948 vol.4.
23. T. H. Cormen, C. E. Leiserson, R. L. Rivest, *et al.*, *Introduction to Algorithms*, 3rd ed (MIT Press, 2009).
24. Telecommunication Union International, "G.694.1: Spectral grids for WDM applications: DWDM frequency grid," (2020).
25. M. P. Yankov, U. C. de Moura, and F. D. Ros, "Power evolution modeling and optimization of fiber optic communication systems with EDFA repeaters," J. Lightwave Technol. **39**(10), 3154–3161 (2021).
26. T. Akiba, S. Sano, T. Yanase, *et al.*, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2019).
27. Z. Gan, M. Shevchenko, S. N. Herzberg, *et al.*, "Data for: Fast neural network inverse model to maximize throughput in ultra-wideband WDM systems," University of Cambridge, 2024, https://doi.org/10.17863/CAM.110401.