JOURNAL OF
Optical Communications
and Networking

SkipNet: an adaptive neural network equalization algorithm for future passive optical networking

Stephen L. Murphy,[1,2,*] Paul D. Townsend,[1,2] and Cleitus Antony[1,2]

[1]Photonic Systems Group, Irish Photonic Integration Centre, Tyndall National Institute, Cork, Ireland
[2]School of Physics, University College Cork, Cork, Ireland
*stephen.murphy@tyndall.ie

Received 1 May 2024; revised 18 July 2024; accepted 25 July 2024; published 8 October 2024

In this paper, we propose an original adaptive neural network equalizer (NNE) algorithm named SkipNet, which is suitable for rapid training on a packet-by-packet basis for burst-mode non-linear equalization in upstream PON transmission. SkipNet uses the simple LMS algorithm and avoids complex neural network training algorithms such as backpropagation and mini-batch training. We demonstrate SkipNet on captured continuous mode 100 Gbit/s PAM4 signals using an SOA preamplifier to achieve the challenging 29 dB PON optical loss budget. The adaptive SkipNet equalizer is shown to overcome combinations of severe SOA patterning effects and fiber dispersion impairments to achieve >29 dB dynamic range back-to-back and >22.9 dB dynamic range for up to 81.6 ps/nm accumulated dispersion. It can adapt in as little as 250 training symbols to each impairment scenario, which is equivalent to existing FFE/DFE solutions, while matching the non-linear performance of previously proposed static NNE solutions. To the best of our knowledge, SkipNet is the first ever adaptive NNE framework that can realistically be trained and adapted on a packet-by-packet basis and within strict PON packet preamble lengths.

https://doi.org/10.1364/JOCN.528490

1. INTRODUCTION

Neural network equalizers (NNEs) have attracted widespread interest within the optical communications community due to their ability to overcome non-linear signal impairments [1]. In particular, NNEs have been proposed for future 100 Gbit/s/λ passive optical networks (PONs) employing advanced modulation formats that are susceptible to device and fiber non-linearities [2]. Indeed, recent work has shown NNEs have improved performance over conventional equalizers when mitigating fiber impairments and supporting low-bandwidth directly modulated transceivers in four-level pulse amplitude modulation (PAM4) systems [3–5], a preferred solution for achieving 100G PON. NNEs have also demonstrated remarkable abilities to overcome semiconductor optical amplifier (SOA) non-linearities in the context of SOA preamplification at the optical line terminal (OLT) [6–9] and help exceed the challenging 19.5 dB PON dynamic range (DR) requirement. However, while challenges related to NNE complexity and hardware implementation have received attention [10], questions remain concerning NNE training in the PON setting.

The most appropriate deployment of an NNE in PON is at the OLT, where its benefits and costs can be shared among all optical network units (ONUs) and where it can best influence overall network design. However, burst-mode signaling at the OLT receiver poses a unique challenge for any NNE scheme as the channel characteristics change rapidly between packets. Although the types of linear and non-linear impairments in PON are limited and well-understood, there are many possible permutations of these among the 64 (or potentially 128) ONUs subscribed to a single network. The fiber dispersion accumulated by a received packet can vary and be up to 77 ps/nm depending on the distance to the OLT, while individual ONUs can have distinct bandwidth and chirp induced impairments due to transceiver vendor variability [11]. Further, if an SOA preamplifier is used at the OLT, non-linear gain saturation induced patterning effects will distort high-power "loud" bursts, while "soft" bursts are unaffected. Considering that channel variability and degradation due to environmental factors cannot be disregarded either, it is clear that any NNE to be deployed at the OLT must be robust and flexible to a wide range of channel impairments.

Adaptive equalization is a possible solution to ensure such equalizer flexibility at the OLT and has been investigated for conventional DSP in both IM/DD [12] and coherent [13] PON systems. Here, the equalizer parameters are adapted on a packet-by-packet basis using training symbols embedded

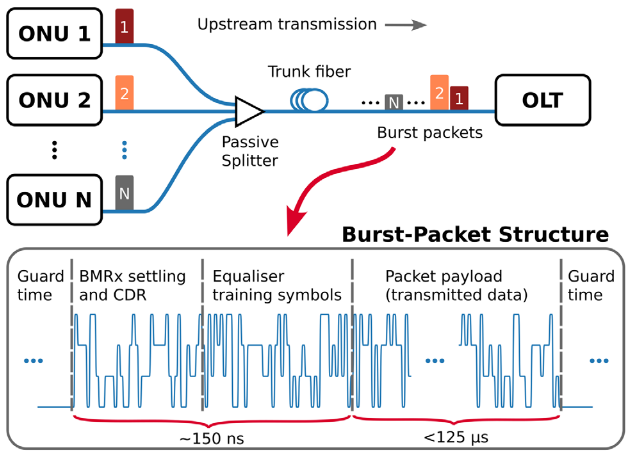

Fig. 1. The burst-packet structure of upstream PON transmission according to the latest 50 Gibt/s HS-PON standards is shown. Only a 150 ns preamble is allocated for 25 Gbit/s upstream transmission, and 610 ns in a worst-case scenario [14].

in each packet's preamble. But, as shown in Fig. 1, the latest ITU-T 50 Gbit/s Higher Speed PON standards outline only a 150 ns packet preamble time for OLT reconfiguration [14], which must accommodate burst-mode Rx (BMRx) settling time and clock and data recovery (CDR) alongside any equalizer adaption. Because of this, adaptive equalizers typically use variants of the simple least mean squares algorithm, which can be implemented at high frequencies using hundreds of training symbols. In comparison, NNEs appear incompatible with such rapid adaption since their training algorithms are considerably slower and more complex and involve iterating multiple times (epochs) over large datasets consisting of tens of thousands of samples. Therefore, while impressive results have been reported for OLT-side NNE, such as in [2,3,7,9,10], these rely on retraining and testing static NNEs "offline" for each considered packet scenario and receiver power, which would be impossible to implement in burst-mode.

Fixed or so-called "unified" equalization [15], which avoids adaptive training between packets, also faces challenges. In theory, a sufficiently large NNE could be trained offline to equalize multiple ONU scenarios, allowing it to be deployed in a "fixed" state, such as in [16]. The authors of [17] propose such a "universal" NNE for 50 Gbit/s PAM4 PON, which is pre-trained to compensate for fiber dispersion for varying ONU wavelengths and transmission distances. However, they observe that such a universal NNE's performance degrades when applied to packet conditions it was not exposed to during training. Therefore, costly data collection overheads are a feature of fixed NNE solutions, since training datasets should be exhaustive representations of possible ONU packet conditions, which can differ between network deployments. Further, fixed NNEs require larger neural network structures to effectively equalize multiple ONU scenarios [17] and so consume more hardware resources and power [18]. Consequently, the cost of fixed NNE methods may be prohibitive given the numerous possible ONU impairment permutations.

In this paper we introduce SkipNet: a novel adaptive NNE framework that enables the application of NNEs in PON and that we first proposed in [19]. SkipNet is named for its characteristic skip connection structures, which allows it to incorporate fixed, low-complexity NNE kernels targeting specific non-linear impairments without compromising on equalization performance elsewhere. These NNE kernels are pre-trained and deployed in a fixed state while an online adaptable output layer processes the combined kernel outputs and raw sample inputs from the skip connections, resulting in a rapidly re-configurable and flexible equalization solution that intrinsically promotes transfer learning [20]. This hybrid approach combines the best of the adaptive and fixed/universal methods: being adaptable on a packet-by-packet basis without using conventional NNE training methods and having low complexity by avoiding the model bloat associated with fixed NNE solutions. Here we outline the theory underpinning SkipNet and investigate its performance in a 100 Gbit/s/λ PAM4 PON scenario in continuous mode, with an SOA acting as an OLT preamplifier in order to satisfy the PON 29 dB optical power budget, as in [21]. Packet impairments investigated include permutations of non-linear SOA gain saturation patterning effects and accumulated fiber dispersion impairments of up to 85 ps/nm. The performance results of the scheme are reported in terms of the achievable PON DR and benchmarked against an "ideal" NNE trained using conventional methods that have been shown to extend the SOA DR [22], as well as adaptive feed forward (FFE) and decision feedback equalization (DFE). An analysis of SkipNet's convergence properties is also carried out.

Section 2 details the challenges facing NNE at the OLT, while Section 3 develops the SkipNet framework and its relation to existing NNE training and adaptive equalization schemes. Section 4 details the 100G PON experimental setup used to evaluate SkipNet's performance, and Section 5 reports the achieved results.

2. BURST-MODE EQUALIZATION AT THE OLT

Burst-mode signaling in high-speed PON presents a unique challenge for adaptive equalization due to the extremely short packet preamble available for equalizer reconfiguration. The packet preamble length for 50 Gbit/s upstream transmission outlined in the ITU-T's Higher Speed (HS-) PON recommendations is still under study, while the 25 Gbit/s objective preamble is only 150 ns [14], which corresponds to 7500 symbols for 100 Gbit/s PAM4. But as previously mentioned, this preamble must be shared among several OLT reconfiguration functions, as shown in Fig. 1. Also, the frequency at which equalizer parameter updates can be processed is unlikely to match the raw line-rate due to DSP hardware limitations. This implies that only a fraction of assigned preamble training symbols will be used during equalizer adaption (e.g., $^1/_2$, $^1/_4$, …), and so the actual training symbols available for adaption could be significantly less than that suggested by a 150 ns preamble. Therefore, a key metric when designing adaptive equalization for PON is the training convergence rate, i.e., how many symbols it takes for the equalizer to converge. This must be balanced against the training algorithm's complexity, which is intimately linked to power consumption and clock cycles per parameter update. For example, an algorithm that halves the required training symbols but whose hardware implementation overly increases power consumption or requires more clock cycles per update may not be advantageous.

The restrictive preamble time available for training has meant that existing PON adaptive equalization research has focused on low-cost, low-complexity FFE/DFE techniques [23]. The standard LMS training algorithm can require thousands of training symbols to converge, which can be considered too slow for use in PON. An alternative is the recursive least squares (RLS) algorithm, which has shown FFE/DFE convergence in as little as 64 symbols for 10 Gbit/s signaling [24]. However, RLS is considered overly complex for low-cost implementation, and a preferred solution is gear shifted LMS (GS-LMS) where the LMS learning rate is reduced one or more times during training to fine-tune convergence. The authors of [25] achieve FFE/DFE convergence within 250 bits using GS-LMS at 10 Gbit/s, while in [12] an Rx sensitivity penalty of just 2 dB was demonstrated at 50 Gbit/s using 200 training symbols. For comparison, studies have estimated preamble requirements of just 72 ns for burst-mode 100G coherent signal detection and its associated DSP [26], while in [27] the authors estimate up to 416 training symbols to adapt coherent signal recovery algorithms for digital subcarrier modulation.

In contrast, adaptive NNE could offer a considerable increase in non-linear performance and capabilities over FFE/DFE solutions for high-speed PAM4 signaling, but unfortunately neural network training is substantially more complex. Proposed NNE solutions require datasets that are considerably larger than PON packet preamble lengths: usually consisting of tens if not hundreds of thousands of signal samples. During training, these large datasets are divided into "mini-batches" that are processed separately using stochastic optimization algorithms such as Adam [28] or RMSProp [29]. While the LMS algorithm can update an equalizer's parameters for each training symbol, an entire mini-batch must be processed for a single iteration of NNE parameter updates, with the caveat of small mini-batch sizes leading to poor training performance. Additionally, the backpropagation algorithm is required to update an NNE's hidden layer parameters [30], but this involves large numbers of multiplications, as discussed in [31], as well as being extremely memory intensive due to the caching of layer activations and multi-dimensional error gradients. For recurrent neural networks (RNNs), these issues are exacerbated by the backpropagation-through-time algorithm, which further propagates error signals along equalizer feedback connections [32]. Worse still, potentially hundreds of training iterations known as "epochs" are required to achieve NNE convergence. Therefore, even disregarding PON's stringent cost imperative, it is unlikely that conventional NNE training algorithms could be implemented on current available hardware at sufficient speed for 100G PAM4 PON.

Techniques to achieve adaptive NNE for short-reach and long-haul transmission have been proposed but are similarly unsuitable for PON. In [31] the authors propose "AdaNN," which is an adaptable NNE based on an online semi-supervised training method; however it requires 10^5 symbols to converge. Meanwhile in [33] it is demonstrated how adaptive digital signal processing can be incorporated into a non-linear NNE, but the method still relies on the complex backpropagation algorithm.

3. SKIPNET: AN ADAPTIVE NNE FRAMEWORK

A. Existing Adaptive FFE/DFE Solutions

The FFE/DFE is the most popular equalization solution proposed for overcoming bandwidth limitations and fiber dispersion in PON systems up to 50 Gbit/s due to its low complexity and easy adaption using the LMS algorithm. For the sake of clarity the focus here is limited to the adaptive operation of an FFE that is equivalent to that of an FFE/DFE. An FFE can be expressed as

$$\hat{y}_t = w \cdot x_t, \tag{1}$$

where x_t and $\hat{y}_t$ represent the sampled signal input values and the equalized sample output of the FFE at time t. The FFE tap weights w can be adapted packet-by-packet using the LMS algorithm, which calculates an error signal $\mathcal{E}$ (e.g., mean squared error) for each equalized training symbol and adapts the FFE tap weights using the gradient descent method:

$$w_{n+1} := w_n - \gamma \cdot \frac{d\mathcal{E}}{dw}, \tag{2}$$

where n is an index term used to denote successive FFE parameter values. The adaptive convergence rate can be tuned using the learning rate γ; however there is a trade-off to be made between convergence speed and performance/stability. Alternative algorithms such as GS-LMS try to overcome this by changing γ during equalizer adaption: a high γ is used to quickly converge toward the optimal weights, and after a set number of symbols γ is lowered one or more times to guarantee the best possible convergence. This has limited complexity overhead compared to LMS, but can lead to a significant increase in convergence speed [25].

B. Neural Network Training Barrier

The computational complexity of neural network training algorithms is a serious barrier to adaptive NNE deployment in a burst-mode PON setting. Although conventional NNE training schemes also rely on gradient descent, they require the backpropagation algorithm to propagate error through their multi-layer structure and obtain the error gradients for each NNE parameter [30], which greatly increases the computational and memory resources per parameter update compared to the LMS algorithm.

To illustrate this, consider a fully connected NNE (FC-NNE) composed of L layers of neurons, each of which is governed by the equations

$$z = w \cdot x + b,$$
$$a = \varphi(z), \tag{3}$$

where a is the activation output of the neuron, z is the weighted sum of the inputs, w and b are the trainable weight vector and bias parameters, and φ is a non-linear activation function such as the tanh or ReLU functions. In order to update a neuron's weights in the lth hidden layer of the FC-NNE, an error signal $\mathcal{E}$ generated from the forward pass of a training symbol must be backpropagated from the NNE's output layer, where the error signal is calculated, backwards through each intervening layer of the network using the chain rule of multi-variable calculus:

$$\frac{\partial \mathcal{E}}{\partial w^l} = \frac{\partial \mathcal{E}}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \cdots \cdot \frac{\partial z^{l+1}}{\partial a^l} \cdot \frac{\partial a^l}{\partial z^l} \cdot \frac{\partial z^l}{\partial w^l}, \quad \textbf{(4)}$$

where the output layer activation is labelled a^L. While backpropagation can be seen as an efficient implementation of Eq. (4), in practice this still involves at least one matrix multiplication per network layer per training symbol, which costs resources and latency to calculate. Further, the backpropagation calculations require the layer activations from the forward pass of the training symbol, meaning significant memory resources must also be employed. Reference [31] contains a more detailed discussion of backpropagation complexity.

However, simple stochastic gradient descent with backpropagation is not generally sufficient for optimizing neural networks. Instead, conventional neural network training relies on mini-batch training in conjunction with advanced parameter update algorithms such as Adam. The Adam algorithm calculates individual learning rates for every parameter in an NNE (potentially hundreds or thousands of parameters) using estimates of the first and second order statistical moments of calculated parameter gradients. Mini-batch training is required to do this: averaged error gradients are calculated for each NNE parameter using the forward and backward pass calculations of a subset (mini-batch) of the training data, typically of size 32, 64, 128, etc. Therefore, if a mini-batch size of N is used, each parameter update will require N forward and backward passes of equalizer training inputs involving the backpropagation algorithm to achieve gradient averaging, followed by the complex parameter update equations of Adam. Implementing these methods in high-speed PON is untenable, since equalizer adaption must be fast and hardware efficient.

C. SkipNet: Adaptive Neural Network Equalization

SkipNet is a hybrid adaptive NNE solution involving both fixed and adaptive components. It crucially removes the need for mini-batches and the backpropagation and Adam algorithms during online equalizer adaption and relies only on LMS once deployed. This makes its adaption complexity equivalent to that of previously studied FFE/DFE solutions.

Figure 2(a) outlines the methodology involved in creating non-linear kernels for use in an adaptive NNE using the SkipNet framework. In this case, SOA patterning is identified as the target non-linearity, and an NNE with a linear output layer is trained offline on the desired worst-case patterning scenario using conventional "ideal" methods, i.e., the backpropagation algorithm in conjunction with Adam optimization and mini-batch training. The hidden layers of this

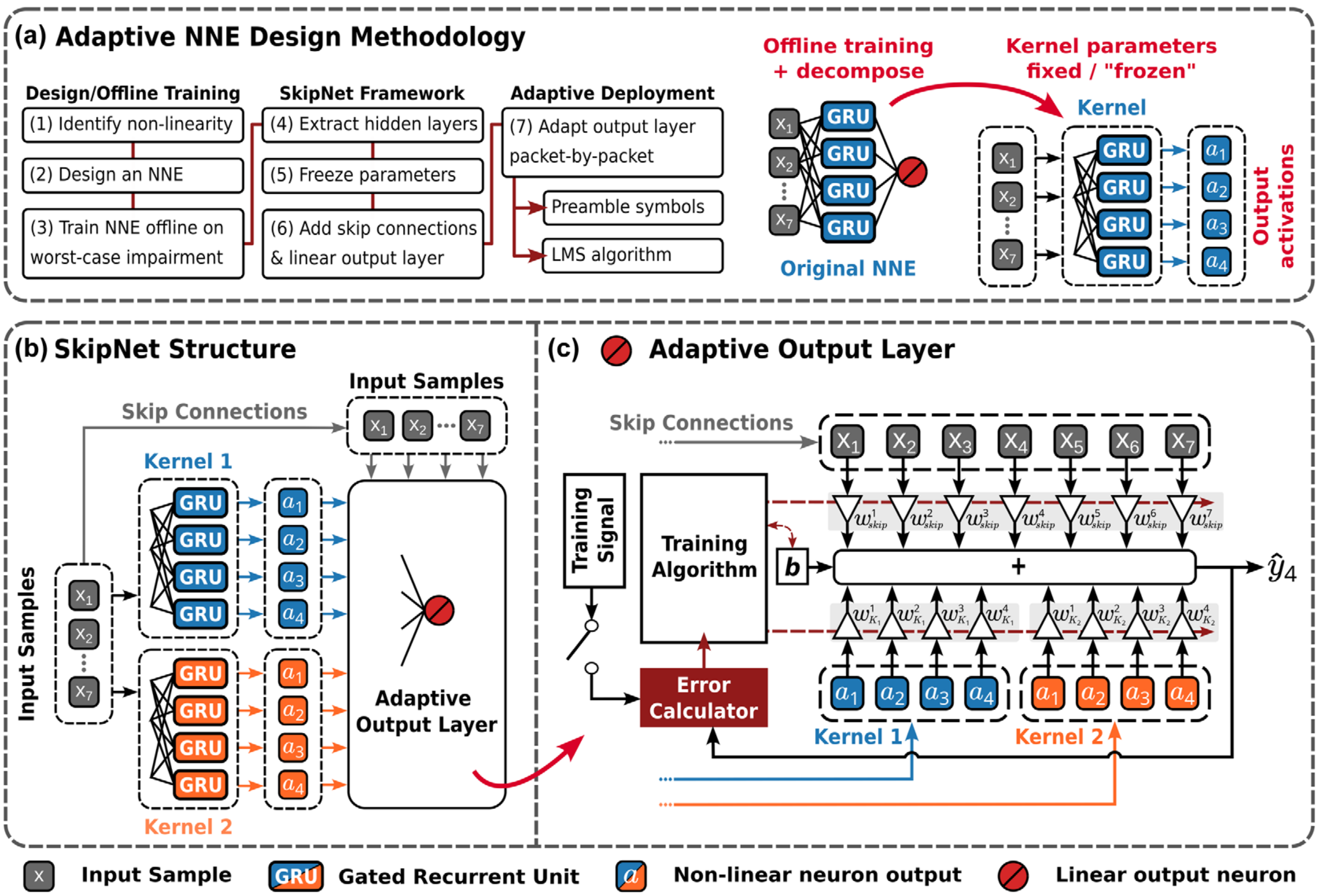

Fig. 2. (a) Outlines the design methodology behind creating an adaptive SkipNet equalizer and how kernels are derived from the hidden layers of an ideally trained NNE before being integrated into the SkipNet framework, as shown in (b). Flexible and rapid adaptive NNE is then possible using a linear output neuron and the adaptive circuit shown in (c), which is equivalent to how a conventional adaptive FFE/DFE operates.

ideally trained NNE are responsible for its non-linear equalization capabilities and so are extracted by removing and discarding the linear output layer. The parameters of these hidden layers are then "frozen," i.e., kept fixed thereafter and not adapted, after which they are referred to as a non-linear kernel. This non-linear kernel is denoted here as $K(\cdot)$ and takes as input raw signal samples just as the ideally trained NNE did before:

$$a_t^{L-1} = K(x_t), \tag{5}$$

where the kernel output, a_t^{L-1}, is the non-linear output activations of the final extracted hidden layer when acting on input signal samples x_t at time t. Note that the proposed adaptive scheme is agnostic to the kernel hidden layer structures, and so $K(\cdot)$ can represent fully connected, convolutional, and recurrent layers, or even combinations of these.

Once the kernel NNE has been trained, it can be integrated into the SkipNet framework. This involves combining the kernel NNE with an adaptive output layer that is chosen to be a single fully connected neuron as described by Eq. (3), with linear activation function $\varphi(x) = x$. Skip connections are also added at this point, which feed the raw equalizer input samples to the adaptive output layer by bypassing the NNE kernel, as seen in Fig. 2(b). Therefore, the adaptive output layer takes as input $K(x_t)$, the pre-trained kernel's output activations, as well as the raw input samples x_t via skip connections, as shown in Fig. 2(b). SkipNet's output layer can then be described by

$$\hat{y}_t = w_{\mathrm{skip}} \cdot x_t + w_k \cdot K(x_t) + b, \tag{6}$$

where $\hat{y}_t$ are equalized samples, and w_{skip} and w_k are the adaptive parameters for the raw signal samples x_t and non-linear kernel output $K(x_t)$, respectively. Figure 2(c) illustrates the operation of Eq. (6) and how this output layer can now be adaptively trained using the LMS optimization method and preamble symbols.

The skip connections are vital for realizing equalization performance across a wide range of packet conditions: non-linearly impaired or otherwise. The non-linear NNE kernel K is trained to carry out non-linear processing for a specific impairment, and if it is applied to a different type of impairment, or even a clean signal, it will non-linearly scramble the signal rather than equalize it. In such a scenario, skip connections allow clean signals to skip straight to the adaptive output layer, thus bypassing $K(\cdot)$ and remaining unimpaired by the NNE operation. Further, skip connections can effectively act like an FFE, capable of dealing with bandwidth limitations, and even help mitigate fiber dispersion. These benefits involve only a linear increase in the number of adaptive output layer multiplication operations, equal to the number of skip connections, and therefore represent a small increase in overall SkipNet complexity. Skip connections are inspired by residual networks [34] and have previously been proposed as "cascade" connections to boost RNN equalizer performance for short-reach PAM4 transmission, and they lower the number of epochs required during NNE training [35]. The SkipNet structure shown in Fig. 2(b) is also similar to that proposed in [36]. Crucially however, neither of these works exploits skip connections and the kernel approach to achieve adaptive NNE training using the LMS algorithm over a single training epoch.

The SkipNet output layer described by Eq. (6) and adaptive FFE/DFE have equivalent complexity, since both use LMS and stochastic gradient descent to achieve fast adaption, as described in Eq. (2). Additionally, more sophisticated variants of LMS such as GS-LMS can just as easily be employed, as well as other methods developed to achieve faster LMS convergence such as in [23,37].

D. Multiple Kernels for Flexible Non-linear Equalization

However, in future PON there could exist multiple non-linear impairments that may be present in isolation or in combination for any given packet. For example, in this paper we consider gain saturation induced SOA patterning, which will only affect loud packets, as well as scenarios where a packet suffers combined SOA patterning and extreme dispersion effects. In this paper, separate NNE kernels are trained for an isolated non-linear SOA patterning effect, K_1, and for the combined patterning/dispersion scenario, K_2. These are incorporated into the adaptable output layer as

$$\hat{y}_t = w_{\mathrm{skip}} \cdot x_t + w_{K_1} \cdot K_1(x_t) + w_{K_2} \cdot K_2(x_t) + b, \tag{7}$$

where now there are two sets of adaptable parameters for the two non-linear kernels: w_{K_1} and w_{K_2}. Significantly however, Eq. (7) is not fundamentally changed from Eq. (6), and so the adaption mechanism remains unchanged and the flexibility of the proposed adaptive NNE is clear. The modular nature of the kernel NNEs avoids the exponential growth in parameters (model bloat) experienced by large standalone or fixed NNE algorithms designed for multi-impairment equalization, since these kernels target specific, isolated impairments that require only low-complexity NNE blocks. Each kernel is trained only once, and is deployed in a fixed state, thereby removing NNE training overheads. Additionally, transfer learning is intrinsic to Eqs. (6) and (7), since the adaptive output layer allows kernels to contribute to the equalization of channel impairments for which they may not have been specifically trained.

4. EXPERIMENTAL SETUP

The 100 Gbit/s PAM4 experimental setup shown in Fig. 3(a) is used to evaluate the proposed SkipNet adaptive equalization scheme in continuous mode. This setup can produce a wide range of possible packet impairments, including SOA gain saturation induced pattering effects, fiber dispersion, bandwidth limitations, as well as combinations of these, as shown in Fig. 3(c). An ideal high-power Tx is emulated using a Mach–Zehnder modulator in conjunction with a high-power EDFA booster amplifier. This is driven by a differential output DAC operating at 100 GSa/s to generate a 50 Gbd PAM4 signal with 6 dB extinction ratio. The limited bandwidth of the Tx opto-electronics is linearly compensated for up to 40 GHz prior to transmission.

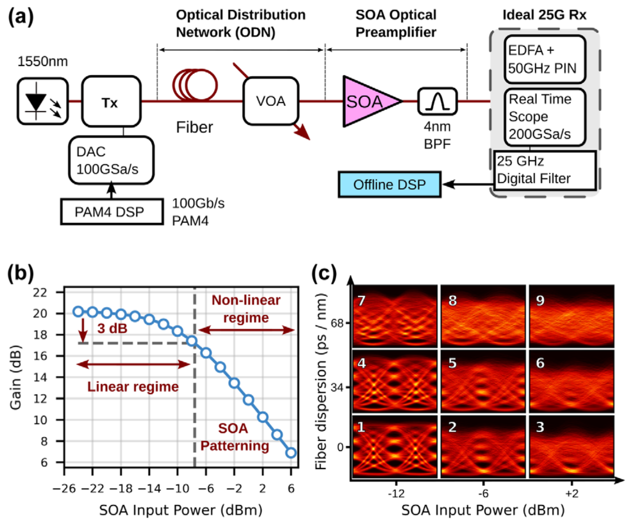

Fig. 3. (a) Shows the experimental setup, which incorporates an SOA pre-amplifier whose gain characteristics are shown in (b). This setup allows for a diverse range of packet impairment permutations to be investigated, as evidenced by the eye diagrams in (c). BPF, bandpass filter; DAC, digital-to-analog converter.

A variable optical attenuator (VOA) mimics the PON optical distribution network losses to create loud and soft packets in continuous mode, and the input power to the SOA + PIN Rx is swept from −26 dBm to +6 dBm. HS-PON specifies a tolerance for up to 77 ps/nm dispersion at 1342 ± 2 nm, corresponding to 20 km transmission in the O-band [14]. Due to a lack of available O-band devices, this experiment is carried out at 1550 nm, and dispersion impairments up to 68 and 81.6 ps/nm were realized by changing the fiber transmission distances to 4 and 4.8 km, respectively. These were the closest achievable dispersion values to the HS-PON 77 ps/nm value while using standard single mode fiber with a dispersion parameter of 17 ps/(nm · km). Therefore similar dispersion impairments could be investigated, although the dispersion slope will differ between the two transmission windows. Similarly, SOA operating principles are not fundamentally changed between the C- and O-bands.

The OLT comprises an SOA preamplifier kept at constant bias followed by an ideal Rx comprising an EDFA and 50 GHz PIN, with a 25 GHz fourth order Bessel filter digitally applied to emulate 25G class optics. The signals are captured using a 200 GSa/s real time scope and the measured Rx sensitivity is found to be −21.1 dBm. The SOA preamplifier gain characteristics for constant 100 mA bias current are shown in Fig. 3(b) and the SOA saturation input power is −8 dBm, above which SOA gain-saturation induced patterning effects are clearly evident, as shown in Fig. 3(c), eye diagrams 2 and 3. This poses a clear challenge to meeting the 19.5 dB dynamic range requirement outlined in the 50G HS-PON standards [22], before even the effects of fiber dispersion induced sensitivity penalties have been considered. The severe effect of fiber dispersion can be seen in Fig. 3(c) eye diagram 4, while eye diagram 9 shows the combined effect of 68 ps/nm dispersion and the SOA patterning effect induced by an input power of +2 dBm.

5. RESULTS

In this paper, the system DR and receiver sensitivity are calculated with respect to the SOA preamplifier input power and the hard decision forward error correction (HD-FEC) threshold of 3.8×10^{-3} bit error ratio (BER). Each reported BER value is estimated using the counting method applied to ∼520,000 captured symbols.

Recurrent NNEs based on gated recurrent units (GRUs) [38] and that are similar to those investigated in [9] form the base NNE structures from which all reported "ideal," "fixed," and SkipNet kernel NNEs are derived. This common NNE structure is composed of a single hidden layer of four GRU units fed by seven symbol-spaced (−T) input taps as shown in Fig. 2. The output layer of all the NNEs consists of a single neuron with linear activation function, equivalent to Eq. (3) for the "fixed" and "ideal" NNEs $\varphi(x) = x$, while the operations of the adaptive SkipNet equalizers are equivalent to Eqs. (6) and (7). To avoid overfitting, all training datasets consist of randomly generated symbol sequences [39], while test datasets comprise repeated pseudo-random quaternary sequences (PRQS15).

A. BtB and Transmission Performance of SkipNet Adaptive Equalization Using Skip Connections

Figure 4(a) shows the 100G PAM4 system back-to-back (BtB) performance. For low SOA input powers noise degradation is evident, while the effects of the SOA gain saturation induced patterning are observed for powers greater than −12 dBm. Without signal equalization the Rx sensitivity is −21.1 dBm with only 12.7 dB DR under the HD-FEC threshold. For comparison, a 7-T FFE/DFE equalizer with two decision feedback taps (2-FB) is trained using the LMS algorithm and 15,000 symbols and achieves −22.6 dBm sensitivity and ∼16.0 dB DR. Both of these results are below the minimum 19.5 dB DR recommended by the HS-PON standards.

The performance of the proposed SkipNet scheme is benchmarked against that of an ideally trained NNE with similar structure, as described above. For each data point, the ideal NNE is trained using conventional neural network methods and a training dataset of ∼16,000 random training symbols grouped into mini-batches of size 512. This dataset is iterated over between 50 and 400 times (epochs) using the early-stopping method, which monitors performance on a separate validation set (PRQS13). This ideal scheme achieves the best receiver sensitivity reported here, of −23.1 dBm, with a total DR above 29.1 dB. This agrees with results in [9,22], which report that GRU-based equalizers can completely overcome the non-linear SOA patterning effect for 100 Gbit/s PAM4. Assuming a launch power of +8 dBm, this would imply a system optical loss budget of 31.1 dB, which exceeds the 29 dB requirement set out in HS-PON. For completeness, a version of the ideal NNE with skip connections is also evaluated. However as shown in Fig. 4(a), the ideal NNE results overlap with and without skip connections, meaning that there are no performance gains from including skip connections in an ideally trained NNE. However, when used in conjunction with an adaptive output layer, as in the SkipNet framework, skip connections are vital for adaptive equalization, as described

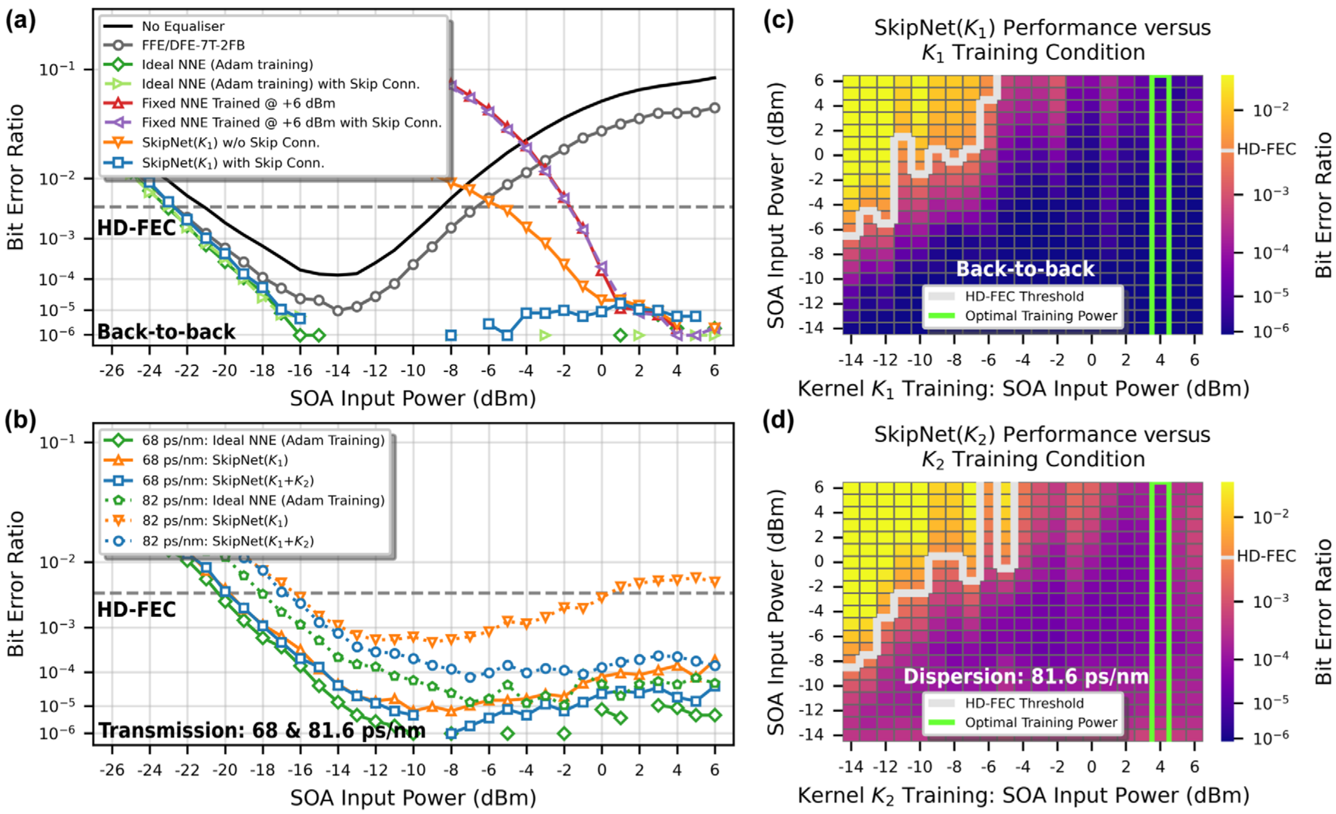

Fig. 4. (a) Shows the BtB performance of the SkipNet(K_1) equalizer with and without skip connections, versus ideally trained and fixed (non-adaptive) NNEs. In (b), the transmission performance of single (K_1) and double ($K_1 + K_2$) kernel SkipNet equalizers underlines the flexibility of the scheme to accommodate additional low-complexity kernels for greater non-linear capability. The heatmaps in (c) and (d) show how K_1's and K_2's training conditions were chosen, by plotting kernel training power versus the BER performance of the resulting SkipNet equalizers.

below. Similarly, adding skip connections did not improve on the ideal NNE performance reported in Fig. 4(b).

A SkipNet equalizer based on the kernel K_1 and integrated skip connections is also reported in Fig. 4(a). The kernel K_1 is pre-trained offline using ideal NNE training methods to overcome the severe SOA patterning effect at only +4 dBm SOA input power with zero dispersion, after which its parameters are fixed (i.e., not adapted). This adaptive scheme matches the performance of the "ideally" trained NNE, achieving −22.7 dBm sensitivity and 28.7 dB DR using only the LMS training algorithm with 15,000 training symbols to adapt its linear output layer for each distinct SOA input power. If the skip connections are removed, the adaptive SkipNet equalizer fails completely to equalize signals with input power below −6 dBm, which includes unimpaired signals that previously required no equalization. This underlines the vital importance of the skip connections for enabling true adaptive NNE, and this result can be understood by the argument made in Section 3.C: that the linear output layer cannot undo the non-linear processing of the kernel K_1 for unimpaired, "clean" signals. Similarly, naively applying a "fixed" NNE based on K_1 without adapting the output layer also fails, both with and without skip connections. These results agree with the similar analysis carried out in [19] where we first proposed using skip connections.

Figure 4(b) shows that, for dispersion up to 68 ps/nm the SkipNet(K_1) equalizer achieves close to the ideal NNE performance and achieves 25.8 dB DR, well above the HS-PON target despite the kernel K_1 not being exposed to dispersion effects during offline pre-training. However, the SkipNet(K_1) equalizer struggles to achieve below the HD-FEC limit for the combination of SOA patterning and 81.6 ps/nm accumulated dispersion. To overcome this, a second kernel K_2 can be incorporated, which is ideally pre-trained on combined SOA patterning and extreme fiber dispersion (+4 dBm, 81.6 ps/nm). For 81.6 ps/nm, this SkipNet($K_1 + K_2$) model achieves an impressive 22.9 dB dynamic range using only linear adaption, underlining the modular flexibility of the proposed scheme. This is in contrast to the single kernel SkipNet(K_1) equalizer that achieves just 16.7 dB DR in the same scenario. It should be noted, however, that 81.6 ps/nm dispersion at 100G is a significantly greater impairment than the 77 ps/nm at 50G outlined in the HS-PON recommendations [14]. Therefore, while the 19.5 dB DR target is met, the associated sensitivity penalty may be impractical in a real network setting.

Table 1 compares the per-packet training complexity of the ideal NNE with that of the SkipNet equalizers in terms of trainable parameters and required optimization algorithm. Since K_1 and K_2 each comprise a single hidden layer of four GRU units, the number of fixed parameters N for each kernel can be calculated using the equation [9]

Table 1. Packet-by-Packet Training Complexities

	Parameters		
Equalizer	Fixed	Trainable	Optimizer
Ideal NNE	0	149	Adam[a]
SkipNet(K_1)	144	12	LMS
SkipNet($K_1 + K_2$)	288	17	LMS

[a]Adam + backpropagation-through-time + mini-batches + early stopping.

$$N = 3 \cdot n_{\mathrm{GRU}} \cdot (n_{\mathrm{in}} + n_{\mathrm{GRU}} + 1)\,, \quad (8)$$

where n_{GRU} is the number of GRU units in the kernel hidden layer, and n_{in} is the number of kernel input samples, which is equal to 7-T. The number of parameters in the SkipNet output layer is determined from Eqs. (6) and (7) and equals the number of kernel outputs plus the number of skip connections and a bias term. Clearly the proposed SkipNet framework allows for reduced training complexity, with the SkipNet(K_1) equalizer having over 10× less trainable parameters compared with the ideally trained NNE, while using the simple LMS algorithm. The SkipNet($K_1 + K_2$) has additional fixed parameters due to integrating a second kernel, but still has only 17 trainable parameters.

B. Low-Complexity, Targeted Non-linear Equalization

The heatmaps in Figs. 4(c) and 4(d) show how the once-off kernel pre-training conditions for K_1 and K_2 were chosen. For each SOA input power between −14 and +6 dBm a SkipNet equalizer was created by pre-training kernel NNEs offline and evaluating the equalizer's adaptive performance over the same input power range. Interestingly, if kernel K_1 is pre-trained at any one SOA input power above −6 dBm, i.e., in the non-linear operating regime of the SOA, then the resulting SkipNet(K_1) equalizer can achieve below the FEC threshold for all non-linear SOA input powers considered. The final K_1 training power was chosen to be +4 dBm due to its good performance across the whole sweep. Similarly, an optimal training point for the K_2 kernel in the SkipNet(K_2) was found to be +4 dBm. However, in order to be effective against combined SOA patterning and fiber dispersion, the minimum input training power for K_2 was observed to be −5 dBm. Both kernels were then integrated into a single adaptive equalizer SkipNet($K_1 + K_2$) that is effective in both of the kernel training scenarios. These results show that model bloat can be avoided by pre-training small, low-complexity NNE kernels once on specific non-linearities, which can then be integrated with skip connections to adaptively equalize a wide range of packet impairments.

Figure 5 reveals how the SkipNet framework targets linear versus non-linear impairments using the kernels and skip connections. It does this by plotting the root mean squared (RMS) value of the converged weights w_{K_1}, w_{K_2}, and w_{skip} of the SkipNet($K_1 + K_2$) output layer for different SOA input powers and dispersion values. In the case of 0 ps/nm dispersion, Fig. 5(a) shows that the skip connections dominate the equalization process up to −8 dBm, above which the non-linear kernel K_1 that is pre-trained on SOA patterning (at +4 dBm) becomes the majority contributor to the SkipNet output. This aligns well with theory, as above −8 dBm input power,

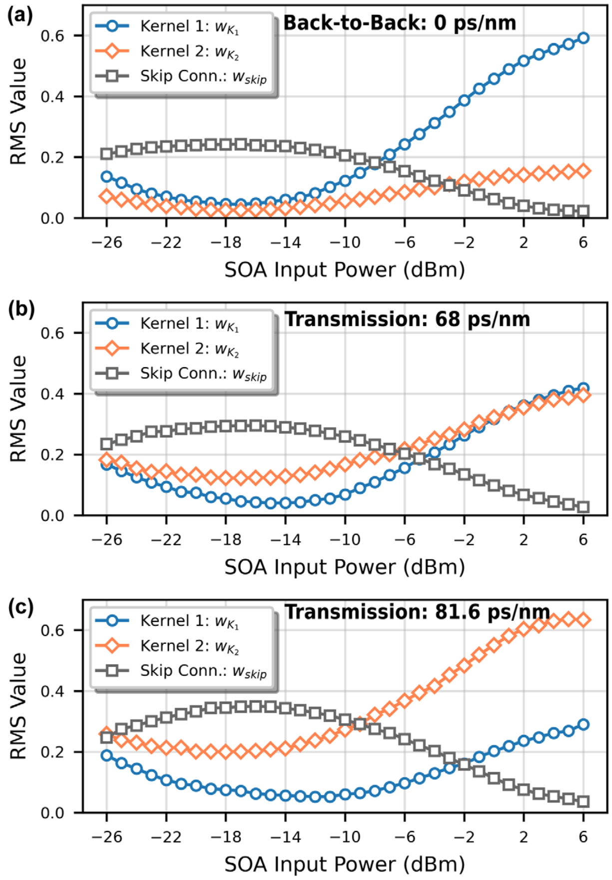

Fig. 5. The RMS values of the converged weight parameters in the SkipNet($K_1 + K_2$) output layer indicate kernel and skip connection relevance for (a) back-to-back (0 ps/nm), (b) 68 ps/nm, and (c) 81.6 ps/nm transmission scenarios. The LMS algorithm can clearly select between kernels and skip connections to target different packet impairments.

the SOA is known to be operating in its non-linear regime, as shown in Fig. 3(b). For the transmission cases of 68 and 81.6 ps/nm dispersion, Figs. 5(b) and 5(c) show K_2's growing contribution towards the final equalization outcome, and it is especially dominant for the scenario of a highly dispersed packet having undergone non-linear SOA patterning. Notice however that both kernels contribute to the equalization of non-linear channel conditions for which they were not trained, which shows that the SkipNet framework has the potential to exploit transfer learning for certain impairments.

The RMS values, or energies, of both kernels' converged weights are significantly higher in the SOA patterning regions than the corresponding RMS values for the skip connection weights in the unimpaired and pure dispersion signal regions (below −8 dBm). A possible reason for this is that here the signal distributions of the kernel outputs are diminished relative to those of the skip connections, and therefore the LMS algorithm converges to higher energy weight parameters in these regions to compensate. Such differences could have important implications for the convergence properties of SkipNet equalizers, since the choice of learning rate for LMS-type optimization algorithms is strongly dependent on the target signal distribution. Evidence of this can be seen in Fig. 6.

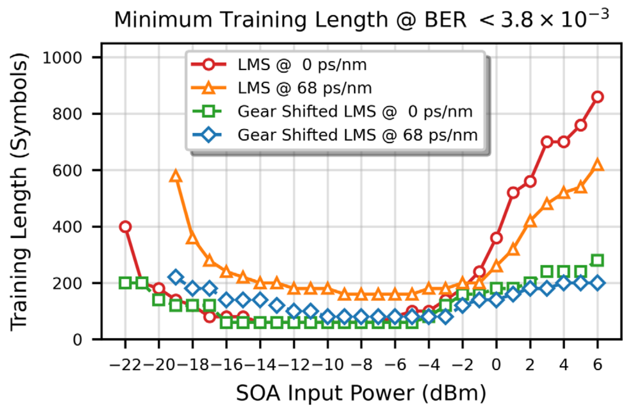

Fig. 6. This graph shows the minimum number of training symbols required by SkipNet($K_1 + K_2$) to achieve BER under the HD-FEC threshold with probability >95% across a range of SOA input powers. Rapid adaption can be achieved using the GS-LMS algorithm, comparable to that of existing FFE/DFE techniques [25].

Moreover, a major contributing factor to the observed discrepancies in signal distributions between the skip connections and kernel outputs is the kernel's output activation function. In this work, the kernel outputs are constrained to ±1 by the tanh activation function characteristic of the GRU units, while the distribution of values from the skip connections are not limited in such a way by the data preprocessing used. Future work could jointly consider the effects of varying packet power, analog-to-digital converter dynamic range, and fixed point sample representation on skip connection distributions and explore alternative kernel activation functions to potentially further optimize LMS convergence.

C. Analysis of Preamble Training Requirements

The results in Fig. 6 show that the proposed adaptive SkipNet scheme can achieve below the HD-FEC using as little as 200 training symbols, which matches the performance of reported FFE/DFE adaptive equalization and coherent PON signal recovery preambles discussed in Section 2.

Figure 6 shows the minimum number of symbols required by the SkipNet($K_1 + K_2$) equalizer to achieve below the HD-FEC threshold with probability >95% (assuming a normal distribution), for both the LMS and GS-LMS training algorithms. The statistics used in this graph were compiled from 40 independent trainings of the SkipNet($K_1 + K_2$) model on the 8 distinct "packets" available from experiments (5 trainings per packet) for each SOA input power and accumulated dispersion considered. For each training, the output layer weights of the SkipNet model were initialized to zero, and random shuffling of the training input ensured each repeated training operation was unique. After training, the equalizer was applied to the associated packet payload comprised of ~65,000 PRQS symbols, and the mean numbers of errors accumulated for each input power, dispersion value, and number of training symbols were recorded. This was repeated for various equalizer training lengths between 50 and 1000 symbols, and the results used to estimate how reliably the equalizer achieved below the HD-FEC BER threshold. The learning rate for the LMS algorithm was chosen to be 0.01, while two learning rates [0.05, 0.01] were used for the GS-LMS, with the higher learning rate being applied for the first half of the training symbols, before the switch to the lower rate being made.

The higher training requirements (>200 symbols) for packets undergoing severe SOA patterning and dispersion correspond to where the kernels K_1 and K_2 have the greatest contribution according to Fig. 5. As discussed in the previous section, these longer training lengths may be required to boost the w_{K_1} and w_{K_2} to compensate for diminished kernel output distributions relative to the skip connections. Interestingly however, less symbols are required in the transmission case at high SOA input powers, with 240 less symbols needed to achieve under the FEC threshold at +6 dBm input power and 68 ps/nm dispersion, than the BtB packet at the same power using LMS. This suggests a notable difference in adaptive NNE training between these two impairments, which could possibly be understood with a thorough investigation of the signal distributions involved.

The GS-LMS algorithm greatly reduces the training symbols required in such scenarios where the kernels are guiding the equalization, with only a small associated increase in training complexity. For BtB signals suffering high levels of SOA patterning, the GS-LMS algorithm requires on average only 33% of the training length used by simple LMS for powers above 0 dBm, while at +6 dBm it uses 580 less symbols. A similar trend is observed in the transmission case, where a 25 dB DR under the FEC threshold can be supported using the SkipNet($K_1 + K_2$) equalizer in conjunction with GS-LMS and less than 250 symbols.

Optimizations previously proposed to accelerate convergence in adaptive FFE/DFE solutions may also be applied to SkipNet and further reduce training lengths and thus the required preamble overhead. In particular, techniques that load pre-determined weights packet-by-packet and prior to adaption could lead to significant convergence improvements.

D. High-Dynamic-Range PAM4 PON Using Adaptive NNE

Figure 7 summarizes the implications of adaptive NNE for PON by plotting the achievable DR versus accumulated dispersion of the adaptive SkipNet($K_1 + K_2$) equalizer for 100G PAM4 using an SOA preamplifier at the OLT.

While the HS-PON standards call for a 19.5 dB DR for up to 77 ps/nm in the O-band, the 7-T tap SkipNet($K_1 + K_2$) can support >28 dB DR BtB, with >25 and 22.9 dB DR for 68 ps/nm and 81.6 ps/nm, respectively, using 15,000 symbols, as seen in Figs. 4(a) and 4(b). This is in contrast to the conventional FFE/DFE with 7-T taps and two decision feedbacks, which achieves 16.0 dB DR BtB and can only sustain 12.5 dB DR up to 48 ps/nm. Further, the proposed adaptive NNE can sustain these dynamic ranges up to 68 ps/nm dispersion using as little as 250 training symbols in conjunction with the GS-LMS algorithm and can meet the 19.5 dB DR requirement up to 81.6 ps/nm using just 400 symbols. These low training symbol requirements are equivalent to preamble lengths proposed in recent FFE/DFE studies, and therefore SkipNet is an appropriate and viable adaptive NNE solution for enabling future PON systems.

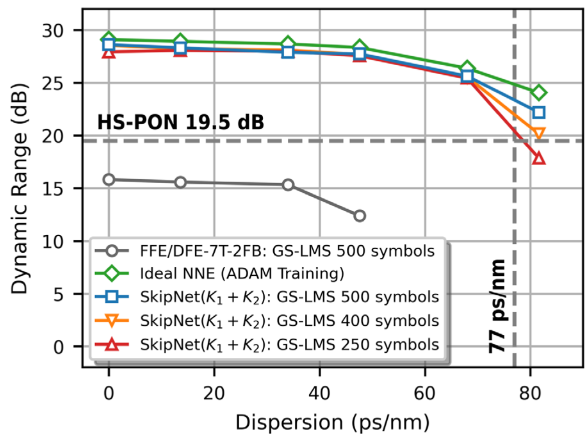

Fig. 7. The SkipNet equalizer achieves >25 dB DR for up to 68 ps/nm accumulated dispersion using just 250 symbols, clearly outperforming the adaptive FFE/DFE equalizer, which cannot even support the 19.5 dB PON DR requirement in the BtB scenario.

Also shown in Fig. 7, the DR values achieved by the adaptive SkipNet($K_1 + K_2$) equalizer using just 400 training symbols approach those of an ideally trained NNE. The adaptive equalizer incurs a 0.5 dB DR penalty relative to the ideal NNE in the BtB scenario, which increases to 2 dB penalty at 81.6 ps/nm dispersion. It is expected that, with further study of the SkipNet scheme's convergence behavior, it will be possible to close this gap to the ideal NNE performance using a short training preamble.

6. CONCLUSION

In this paper, an original adaptive NNE framework (SkipNet) is proposed for future 100 Gbit/s PAM4 upstream PON transmission using an SOA preamplifier at the OLT. This novel equalization scheme achieves >29 dB DR back-to-back, as well as >25 dB DR in transmission scenarios involving up to 68 ps/nm dispersion impairment. Additionally, a 22.9 dB DR is achieved for 81.6 ps/nm dispersion, which is above the 19.5 DR requirement outlined in the recent ITU-T 50G HS-PON standard. The SkipNet equalizer avoids complex neural network training since its kernel NNEs undergo once-off pre-training prior to deployment. The simple LMS algorithm is shown sufficient to adapt SkipNet equalizers to varying packet powers, as well as combinations of non-linear SOA patterning and fiber dispersion impairments. Further, using the GS-LMS algorithm, SkipNet is shown to converge in as little as 250 symbols for a wide range of packet scenarios, which is equivalent to conventional FFE/DFE training symbol requirements proposed in the literature. In contrast, proposed NNE solutions to date rely on extremely complex and resource heavy machine learning training algorithms that need tens or even hundreds of thousands of training symbols and are wholly unsuitable for adaptive equalization in upstream burst-mode PON. It is expected that SkipNet's convergence can be made even faster using methods developed for LMS-like optimization in FFE/DFE solutions.

As far as we are aware, SkipNet is the first adaptive neural-network-based equalization scheme targeting future 100G PON that could potentially be trained on a packet-by-packet basis for burst-mode reception at the OLT. Further, it is expected that SkipNet can find use outside of PON applications where complex conventional neural network training overheads are a barrier to the use of NNE. These could include scenarios where varying channel characteristics due to environmental factors or device degradation are a concern, or where plug-and-play capabilities are required.

Funding. Science Foundation Ireland IPIC Research Center (12/RC/2276P2).

REFERENCES

1. J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, *et al.*, "Machine learning for optical fiber communication systems: an introduction and overview," APL Photon. **6**, 121101 (2021).
2. L. Yi, T. Liao, L. Huang, *et al.*, "Machine learning for 100 Gb/s/λ passive optical network," J. Lightwave Technol. **37**, 1621–1630 (2019).
3. L. Huang, Y. Xu, W. Jiang, *et al.*, "Performance and complexity analysis of conventional and deep learning equalizers for the high-speed IMDD PON," J. Lightwave Technol. **40**, 4528–4538 (2022).
4. B. Sang, W. Zhou, Y. Tan, *et al.*, "Low complexity neural network equalization based on multi-symbol output technique for 200+ Gbps IM/DD short reach optical system," J. Lightwave Technol. **40**, 2890–2900 (2022).
5. C. Bluemm, B. Liu, T. Rahman, *et al.*, "Towards neural network equalizer implementations for IM/DD transceivers," in *Opto-Electronics and Communications Conference (OECC)* (2023).
6. L. Xue, L. Yi, R. Lin, *et al.*, "SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON," Opt. Express **29**, 24714–24722 (2021).
7. K. Wang, C. Wang, J. Zhang, *et al.*, "Mitigation of SOA-induced nonlinearity with the aid of deep learning neural networks," J. Lightwave Technol. **40**, 979–986 (2022).
8. A. G. Reza, M. Troncoso-Costas, C. Browning, *et al.*, "Mitigation of SOA-induced nonlinearities with recurrent neural networks in 75 Gbit/s/λ PAM-4 IM/DD WDM-PON transmission systems," J. Lightwave Technol. **41**, 3967–3975 (2023).
9. S. L. Murphy, F. Jamali, P. D. Townsend, *et al.*, "High dynamic range 100G PON enabled by SOA preamplifier and recurrent neural networks," J. Lightwave Technol. **41**, 3522–3532 (2023).
10. X. Huang, D. Zhang, X. Hu, *et al.*, "Low-complexity recurrent neural network based equalizer with embedded parallelization for 100-Gbit/s/λ PON," J. Lightwave Technol. **40**, 1353–1359 (2022).
11. L. A. Neto, J. Maes, P. Larsson-Edefors, *et al.*, "Considerations on the use of digital signal processing in future optical access networks," J. Lightwave Technol. **38**, 598–607 (2020).
12. G. Coudyzer, P. Ossieur, L. Breyne, *et al.*, "Study of burst-mode adaptive equalization for >25G PON applications [Invited]," J. Opt. Commun. Netw. **12**, A104–A112 (2020).
13. K. Matsuda and N. Suzuki, "Hardware-efficient signal processing technologies for coherent PON systems," J. Lightwave Technol. **37**, 1614–1620 (2019).
14. "50-gigabit-capable passive optical networks (50G-PON): physical media dependent (PMD) layer specification," ITU-T Recommendation G.9804.3 (2021).
15. G. Simon, F. Saliou, P. Chanclou, *et al.*, "50 Gb/s TDM PON digital signal processing challenges: mining current G-PON field data to assist higher speed PON," in *European Conference on Optical Communications (ECOC)* (2020).
16. C. Ye, X. Hu, D. Zhang, *et al.*, "A versatile NN-equalization for 50 Gbps TDM PON burst uplink," in *European Conference on Optical Communications (ECOC)* (2020).
17. V. Houtsma, E. Chou, and D. van Veen, "92 and 50 Gbps TDM-PON using neural network enabled receiver equalization specialized for PON," in *Optical Fiber Communication Conference (OFC)* (2019), paper M2B.6.

18. N. Kaneda, C.-Y. Chuang, Z. Zhu, *et al.*, “Fixed-point analysis and FPGA implementation of deep neural network based equalizers for high-speed PON,” J. Lightwave Technol. **40**, 1972–1980 (2022).
19. S. L. Murphy, F. Jamali, P. D. Townsend, *et al.*, “Adaptive neural network equalisation using skip connections for future 100 Gbit/s/λ passive optical networks,” in *European Conference on Optical Communications* (2023).
20. F. Zhuang, Z. Qi, K. Duan, *et al.*, “A comprehensive survey on transfer learning,” Proc. IEEE **109**, 43–76 (2021).
21. Z. Li, Y. Li, S. Luo, *et al.*, “SOA amplified 100 Gb/s/λ PAM-4 TDM-PON supporting PR-30 power budget with >18 dB dynamic range,” Micromachines **13**, 342 (2022).
22. S. Murphy, P. D. Townsend, and C. Antony, “Recurrent neural network equalizer to extend input power dynamic range of SOA in 100Gb/s/λ PON,” in *Conference on Lasers and Electro-Optics* (2022), paper SW4E.1.
23. P. Torres-Ferrera, H. Wang, V. Ferrero, *et al.*, “Optimization of band-limited DSP-aided 25 and 50 Gb/s PON using 10G-class DML and APD,” J. Lightwave Technol. **38**, 608–618 (2020).
24. P. Ossieur, C. Melange, T. De Ridder, *et al.*, “Burst-mode electronic equalization for 10-Gb/s passive optical networks,” IEEE Photon. Technol. Lett. **20**, 1706–1708 (2008).
25. S. Porto, C. Antony, A. Jain, *et al.*, “Demonstration of 10 Gbit/s burst-mode transmission using a linear burst-mode receiver and burst-mode electronic equalization [Invited],” J. Opt. Commun. Netw. **7**, A118–A125 (2015).
26. J. Zhang, Z. Jia, M. Xu, *et al.*, “Efficient preamble design and digital signal processing in upstream burst-mode detection of 100G TDM coherent-PON,” J. Opt. Commun. Netw. **13**, A135–A143 (2021).
27. H. Wang, J. Zhou, Z. Xing, *et al.*, “Fast-convergence digital signal processing for coherent PON using digital SCM,” J. Lightwave Technol. **41**, 4635–4643 (2023).
28. D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in *3rd International Conference on Learning Representations (ICLR)*, San Diego, California, USA, 7–9 May 2015.
29. T. Tieleman and G. E. Hinton, “Lecture 6E rmsprop: divide the gradient by a running average of its recent magnitude” (2014), https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf#page=26.
30. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature **323**, 533–536 (1986).
31. Q. Zhou, F. Zhang, and C. Yang, “AdaNN: adaptive neural network-based equalizer via online semi-supervised learning,” J. Lightwave Technol. **38**, 4315–4324 (2020).
32. P. Werbos, “Backpropagation through time: what it does and how to do it,” Proc. IEEE **78**, 1550–1560 (1990).
33. Q. Fan, C. Lu, and A. P. T. Lau, “Combined neural network and adaptive DSP training for long-haul optical communications,” J. Lightwave Technol. **39**, 7083–7091 (2021).
34. K. He, X. Zhang, S. Ren, *et al.*, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 770–778.
35. Z. Xu, C. Sun, T. Ji, *et al.*, “Cascade recurrent neural network-assisted nonlinear equalization for a 100 Gb/s PAM4 short-reach direct detection system,” Opt. Lett. **45**, 4216–4219 (2020).
36. M. Schädler, G. Böcherer, F. Diedolo, *et al.*, “Nonlinear component equalization: a comparison of deep neural networks and Volterra series,” in *European Conference on Optical Communication (ECOC)* (2022).
37. S. Yao, X. Tang, R. Zhang, *et al.*, “Accelerating LMS-based equalization with correlated training sequence in bandlimited IM/DD systems,” J. Lightwave Technol. **40**, 4268–4275 (2022).
38. K. Cho, B. van Merrienboer, C. Gulcehre, *et al.*, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” arXiv (2014).
39. T. A. Eriksson, H. Bülow, and A. Leven, “Applying neural networks in optical communication systems: possible pitfalls,” IEEE Photon. Technol. Lett. **29**, 2091–2094 (2017).

Stephen L. Murphy (Student Member, IEEE) completed his B.Sc. in mathematics and physics from University College Cork (UCC) in 2014 and his M.Sc. in mathematical physics from the University of Edinburgh in 2015. Currently, he is working toward his Ph.D. degree at the Tyndall National Institute as part of the Photonic Systems Group, supervised by Dr. Cleitus Antony and Prof. Paul Townsend. His research focuses on applications of novel machine learning algorithms for optical access networks, and specifically on developing low-complexity digital signal processing techniques using neural network algorithms for future passive optical networks. He has presented his work at the European Conference on Optical Communications (ECOC’23), as well as the Conference on Lasers and Electro-Optics (CLEO’22).

Paul D. Townsend (Member, IEEE) received the Ph.D. degree in physics from the University of Cambridge, UK, in 1987. Between 1987 and 1990, he held Research Fellowship positions at St John’s College, Cambridge, UK, and at Bell Communications Research (Bellcore), Red Bank, New Jersey, USA, where he worked on non-linear optical switching phenomena in polymeric semiconductors. In 1990, he joined British Telecom (BT) Laboratories, Martlesham Heath, UK, as a Senior Staff Researcher to lead activities on quantum communications and cryptography. In 2001, he joined the Corning Research Centre, Martlesham Heath, UK, as Head of Group for Access Network Research. Since 2003, he has been with the Tyndall National Institute, University College Cork (UCC), Ireland, where he is Head of Photonics Research, Director of the Science Foundation Ireland IPIC National Photonics Research Centre, and Professor of Photonic Systems Research in the School of Physics, UCC. He is a Fellow of the Institute of Physics (UK and Ireland) and has (co)authored more than 250 publications, including 35 invited conference papers, and holds granted patents in 16 families. His research interests focus on photonic systems and enabling technologies for optical communications.

Cleitus Antony received the Ph.D. degree in physics from the Tyndall National Institute, Cork, Ireland, in 2011, which contributed to record-breaking experimental demonstrations in advanced high-capacity fibre to the home networks. From 2001 to 2006, he was an engineer at Hitachi’s World-Class Fibre Optic Division, Opnext Japan Inc., Yokohama. He has 15 years of experience as a researcher in optical communications and worked on several EU and national projects on research topics that span novel optical network architectures, optical amplification strategies, and digital signal processing techniques. He is currently a Staff Senior Researcher with the Photonics Systems Group, Tyndall. He holds 5 patents and has authored more than 50 peer-reviewed publications. His recent research interest includes exploiting machine learning to advance next-generation optical communication systems and circuits.