CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	High dynamic range 100G PON enabled by SOA preamplifier and recurrent neural networks
Authors	Murphy, Stephen L.;Jamali, Fariba;Townsend, Paul D.;Antony, Cleitus
Publication date	2023-02-28
Original Citation	Murphy, S. L., Jamali, F., Townsend, P. D. and Antony, C. (2023) 'High dynamic range 100G PON enabled by SOA preamplifier and recurrent neural networks', Journal of Lightwave Technology, 41(11), pp. 3522-3532. doi: 10.1109/JLT.2023.3249046
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1109/jlt.2023.3249046
Rights	© 2023, the Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/ - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-28 17:02:16
Item downloaded from	https://hdl.handle.net/10468/14940

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

High Dynamic Range 100G PON Enabled by SOA Preamplifier and Recurrent Neural Networks

Stephen L. Murphy, *Student Member, IEEE*, Fariba Jamali, Paul D. Townsend, *Member, IEEE*, and Cleitus Antony, *Member, IEEE*

***Abstract*—In recent years the PON research community has focused on future systems targeting 100 Gb/s/λ and beyond, with digital signal processing seen as a key enabling technology. Spectrally efficient 4-level pulse amplitude modulation (PAM4) is seen as a cost-effective solution that exploits the ready availability of cheaper, low-bandwidth devices, and Semiconductor Optical Amplifiers (SOA) are being investigated as receiver preamplifiers to compensate PAM4's high signal-to-noise ratio requirements and meet the demanding 29 dB PON loss budget. However, SOA gain saturation-induced patterning distortion is a concern in the context of PON burst-mode signalling, and the 19.5 dB loud-soft packet dynamic range expected by the most recent ITU-T 50G standards. In this article we propose a recurrent neural network equalisation technique based on gated recurrent units (GRU-RNN) to not only mitigate SOA patterning affecting loud packet bursts, but to also exploit their remarkable effectiveness at compensating non-linear impairments to unlock the SOA gain saturated regime. Using such an equaliser we demonstrate $>$ 28 dB system dynamic range in 100 Gb/s PAM4 system by using SOA gain compression in conjunction with GRU-RNN equalisation. We find that our proposed GRU-RNN has similar equalisation capabilities as non-linear Volterra, fully connected neural network, and long short-term memory based equalisers, but observe that feedback-based RNN equalisers are more suited to the varying levels of impairment inherent to PON burst-mode signalling due to their low input tap requirements. Recognising issues surrounding hardware implementation of RNNs, we investigate a multi-symbol equalisation scheme to lower the feedback latency requirements of our proposed GRU-RNN. Finally, we compare equaliser complexities and performances according to trainable parameters and real valued multiplication operations, finding that the proposed GRU-RNN equaliser is more efficient than those based on Volterra, fully connected neural networks or long short-term memory units proposed elsewhere.**

***Index Terms*—Digital signal processing, four-level pulse amplitude modulation, machine learning, optical fiber communications, passive optical network, semiconductor optical amplifier.**

I. Introduction

THE ITU-T Higher Speed Passive Optical Network (HS-PON) recommendations were published in September 2021, and outline industry requirements for next generation 50G Passive Optical Network (PON) [1], [2]. In a first for a PON standard, HS-PON will embrace the paradigm of Digital Signal Processing (DSP) to overcome the severe fiber dispersion impairment that will be encountered at this line rate. Meanwhile, the research community is already looking beyond this to future PON targeting single-channel 100 Gb/s and beyond using various technologies such as coherent detection [3] and flexible-rate PON [4], [5], as well as investigating how DSP could enable high-speed, intensity modulation with direct detection (IM/DD) systems. However, any future IM/DD solution will need to meet the challenging 29 dB PON optical loss budget necessary to support existing fiber infrastructure already installed by network operators.

In this article we summarise and extend our work in [6], where we explore a potential 100 Gb/s upstream IM/DD PON solution based on 50 Gbaud, 4-level Pulse Amplitude Modulation (PAM4). Compared to non-return-to-zero (NRZ) modulation, PAM4 is an attractive solution due to its reduced electro-optic bandwidth requirements, albeit at the cost of reduced receiver sensitivity. In order to boost sensitivity and achieve the 29 dB optical loss budget, Semiconductor Optical Amplifiers (SOAs) can be used as receiver preamplifiers, and have gained widespread interest for IM/DD PON since they are readily integrable, can operate in the C- and O-bands, and are relatively low-cost [7], [8]. However, the impact of SOA non-linearities such as the gain saturation-induced patterning effect is a concern, especially in PON upstream transmission due to the large loud-soft packet Dynamic Range (DR) inherent to PON burst-mode signalling. The 19.5 dB DR specified by HS-PON could therefore pose issues for realising 100G PAM4 in a scenario using an SOA preamplifier, due to PAM4's stringent linearity requirements.

Machine learning (ML) -based equalisation techniques have been proposed to compensate expected fiber dispersion and device impairments, including SOA preamplifier for future PON scenarios [9], [10], [11]. Work has previously focused on achieving the 29 dB optical loss budget, such as in [12] where recurrent neural networks are used in conjunction with an SOA preamplifier to realise a 30 dB loss budget for 100 Gb/s PAM4. However, the implications of the challenging 19.5 DR requirement have also drawn the attention of researchers using non-ML techniques, as in [13] where the authors investigate a look-up table pre-compensation technique on the transmitter side, and [7] which implements a non-linear Volterra equaliser (VNLE) to overcome SOA non-linearities and achieve 18 dB

Manuscript received 15 November 2022; revised 31 January 2023; accepted 19 February 2023. Date of publication 28 February 2023; date of current version 9 June 2023. This work was supported by Science Foundation Ireland (SFI) under Grant 12/RC/2276P2. *(Corresponding author: Stephen Murphy.)*

The authors are with the Tyndall National Institute, T12 R5CP Cork, Ireland (e-mail: stephen.murphy@tyndall.ie; fariba.jamali@tyndall.ie; paul.townsend@tyndall.ie; cleitus.antony@tyndall.ie).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2023.3249046.

Digital Object Identifier 10.1109/JLT.2023.3249046

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

DR for 100 Gb/s PAM4. In this work, we will use VNLE as a performance and complexity benchmark, since it is a popular alternative to ML-based equalisers in IM/DD systems and widely researched [14], [15], [16].

Previously, we proposed a recurrent neural network (RNN) equaliser architecture based on Gated Recurrent Units (GRUs) to extend the SOA input power dynamic range, and achieved over 28 dB for 100 Gb/s PAM4 [17]. The GRU-RNN equaliser achieved near complete recovery of modulated signals suffering from severe SOA patterning distortion while using only 3 symbol-spaced taps. The effectiveness of this GRU-RNN equaliser suggested an intriguing possibility which we explored in [6]: if such equalisers can enable the SOA to operate in gain saturation with tolerable patterning impairments, can we exploit the associated SOA gain suppression to reduce the input optical DR to the following photo-receiver? This could be particularly advantageous given the challenges of designing 50 Gbaud-, high dynamic range-capable burst mode receivers (BMRx) with sufficient linearity to support PAM4 modulation. The state of the art was recently presented at ECOC22, with the presentation of the first demonstration of 100 Gb/s PAM4 linear BMRx which achieves 15.4 dB dynamic range [18]. Exploiting SOA gain suppression in conjunction with our proposed GRU-based equaliser could potentially enable significant DR gains using such a BMRx in the future.

But to realise any potential gains in DR performance, GRU-RNN equalisation must be shown to be robust to multiple system impairments. Stringent component budgets will likely see 25G optics being utilised in future PON standards, and fiber dispersion will be a serious obstacle for 100 Gb/s PAM4. Therefore in [6] and here we investigate GRU-RNN equaliser performance against a combination of SOA patterning, fiber dispersion up to 91.8 ps/nm, and 25G bandwidth limitation. Exploiting SOA gain suppression, the optical dynamic range is reduced to such an extent that we realise 28 dB system dynamic range in continuous mode using just two Rx electrical gain settings, with the further introduction of some electrical saturation effects.

It is cost-effective to place expensive DSP at the optical line terminal rather than at the multiple locations of individual PON subscribers. However this requires OLT equalisation to be able to deal with varying levels of distortion on a packet-by-packet basis, and in the case of an SOA preamplifier this manifests as different levels of non-linear patterning distortion among loud and soft bursts. The proposed GRU-RNN equaliser using only 3 input taps therefore has an efficiency advantage over equalisation methods which rely on large numbers of taps to deal with severe impairment, such as the feed forward equaliser (FFE), since the excess taps required to compensate loud bursts are wasted on less severe soft-burst impairments. However, there exist difficulties around the implementation of an RNN-based equaliser in hardware, specifically the timing requirements of the RNN feedback mechanism, as noted in [12], [19]. Parallel multi-symbol output schemes for neural network-based equalisers have been investigated in [20], and also implemented on Field Programmable Gate Arrays (FPGA) in [12], [19], [21], with the authors of [19] realising NNE based on long short-term memory (LSTM) with competitive complexity for coherent transmission.

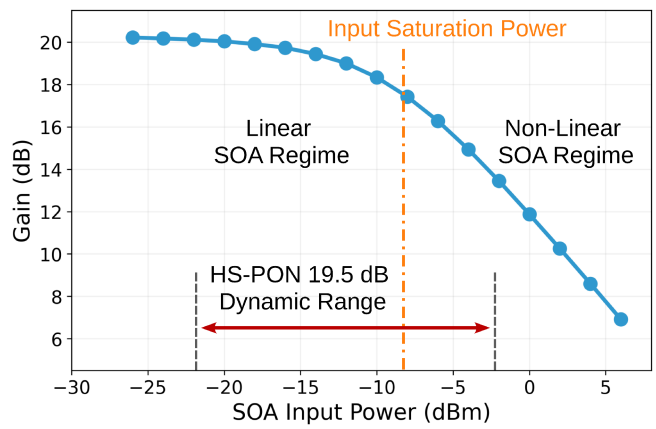

Fig. 1. The measured gain curve of the SOA device under investigation as PON OLT preamplifier is shown; the input saturation power is found to be -8 dBm.

Here we investigate such multi-symbol techniques and their impact on overall system DR and equaliser complexity in a PON context. Further, we test the limits of this method and whether the more complex feedback mechanism of LSTM-RNN offers greater performance than our GRU-RNN.

Section II discusses the origins and implication of SOA gain saturation for 100 Gb/s PAM4, as well as outlining the opportunity of exploiting SOA gain compression to boost achievable DR. Section III gives an overview of the non-linear equalisers being investigated: GRU-, LSTM-RNN, and multi-symbol equalisation methods, as well as introducing VNLE which is used as a conventional non-linear benchmark. The experimental setup for emulating 100 Gb/s PAM4 upstream PON transmission is detailed in Section IV, while results are presented in Section V, with a final discussion on equaliser complexity included in Section VI.

II. SOA Preamplifier for 100G PON

To meet the HS-PON 29 dB optical loss budget using 100 Gb/s PAM4, SOA preamplifiers combined with photodiodes have been proposed for the OLT [7]. However, the 19.5 dB PON dynamic range requirement will mean SOA preamplifiers potentially operating in the gain-saturated regime of the SOA for high-power, loud bursts, leading to non-linear pattern-dependent distortions of the modulated signal, known as patterning.

The optical input power at which the gain of an SOA decreases by 3 dB is known as the SOA's input saturation power, $P_{in,\,sat}$, and for inputs greater than this the SOA operates in its non-linear, gain-saturated regime. Fig. 1 shows the measured gain curve for the SOA (Model: CIP SOA-S) used in this work. The 19.5 dB DR requirement is superimposed on this, starting at -22 dBm which is the system receiver sensitivity (see Section V), and ending at -2.5 dBm $> P_{in,\,sat}$. Since the input saturation power of this device is measured to be -8 dBm, loud packets will be unable to avoid severe SOA patterning which ultimately limits the achievable DR performance. Fig. 2 clearly illustrates the detrimental effect this patterning has on a 100 Gb/s PAM4 signal.

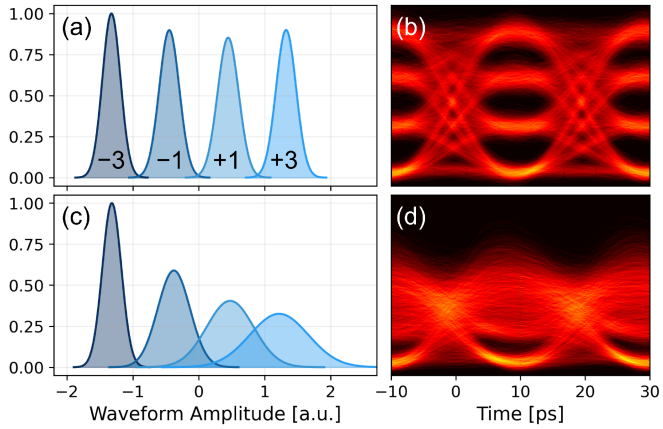

Fig. 2. (a) Normalised 50 Gbaud PAM4 symbol distributions and (b) eye diagram are shown for an SOA input power of -15 dBm; (c) and (d) illustrate the effect of severe gain saturation-induced patterning effect on such signals at an SOA input power of 0 dBm.

The origin of SOA gain-saturation-induced patterning can be understood by considering the carrier dynamics in the device active region [22], [23]. The carrier population available for stimulated emission is directly related to the gain of an SOA, and so carrier population can be seen as a proxy for gain.

Gain saturation can occur when the rate of stimulated emission in the active region due to input optical power is such that the rate of injected carriers due to the SOA bias current is insufficient to maintain the current steady-state carrier population density, resulting in a lower population density being established. Since the SOA gain is directly related to the carrier population density in the active region, increasing the input optical power, and thus the rate of stimulated emission, beyond a critical value $P_{in,\,sat}$ will lead to gain saturation. Crucially, when the SOA is operating in the gain saturation regime and the modulation baud rate approaches the carrier recovery rate, defined as $1/\tau_c$ where τ_c is the spontaneous carrier lifetime, a steady-state carrier density cannot be achieved, resulting in the gain seen by a given data symbol becoming strongly pattern-dependent on the symbols preceding it [23]. The extent of the patterning effect, i.e. how many past symbols the current symbol gain depends on, is related to the magnitude of the stimulated emission. This means we expect patterning to increase in severity with increased SOA input power, especially at the high baud rates required for 100 Gb/s PAM4. These patterning effects are clearly evident in Fig. 2(c) and (d), where the SOA is operating far into its non-linear gain saturated regime.

However, as we will show in this work, a PON system employing the GRU-RNN equaliser described in Section III can recover signals from severe SOA patterning effects, and so the SOA gain-saturation can be exploited to compress the optical dynamic range between the SOA input, and its output to the Rx. This DR compression is clearly illustrated in Fig. 4(b), where the system DR of 28 dB is reduced to 14 dB at the photoreceiver input. Such a scheme could potentially allow DR requirements to be relaxed for the PON burst-mode electronics used to equalise packet powers at the OLT.

III. Machine Learning and Non-Linear Equalisation Techniques

Neural Network Equalisers (NNEs) have been proposed as a solution to SOA non-linearities for future PON systems [9], [11], while Volterra equalisers are a competitive alternative, which are also non-linear [7], [16]. However, these mostly focus on achieving the optical loss budget set out in HS-PON, and DR considerations are not discussed. Here we discuss the merits of RNN-based equalisers over more traditional VNLE and fully connected neural network equalisers (FC-NNE), in terms of PON system DR and effectiveness against varying SOA patterning impairment.

A. Neural Network Equalisation for SOA Patterning

The Feedforward Equaliser (FFE) is the most well-known equalisation technique, widely used to compensate bandwidth limitations and lesser fiber dispersion. However, FFE performance is limited due to it being a linear equaliser governed by the equation:

$$\hat{y}_t = w \cdot x_t$$

Where the equalised sample $\hat{y}_t$ is simply a weighted sum of n_i input samples, x_t, at time t, and optimised tap weights w. Due to their linear design, FFEs are unsuitable for non-linear impairments such as SOA patterning. However, FC-NNEs offer a non-linear alternative, and can be seen as a non-linear extension of the FFE, with FC-NNE layer equation given by:

$$a_t = \varphi\left(W \cdot x_t + b\right)$$

Where $x_t \in \mathbb{R}^{n_i}$ is a vector of n_i input samples at time t, $W \in \mathbb{R}^{n_h \times n_i}$, $b \in \mathbb{R}^{n_h}$ are the trainable weights and bias terms for the n_h FC units in the layer, φ is a non-linear activation function, and $a \in \mathbb{R}^{n_h}$ is the output of the NNE neuron / layer. This output is either fed to another layer, or is the final equalised output of the FC-NNE. The number of trainable parameters N_{param} and real valued multiplication (RVM) operations N_{RVM} involved in each NNE considered in this work will be used as complexity metrics in Section VI. To calculate these values for a single FC layer in a NNE, we use the equations below:

$$N_{params} = n_h \times (n_i + 1)$$
$$N_{RVM} = n_h \times n_i$$

The total values for each equaliser considered in this work are calculated and reported in Table I.

A drawback of both FFEs and FC-NNEs, is that the number of equaliser input taps needs to be calibrated according to the impairment being considered, and specifically that impairment's temporal characteristics. In [17], we showed the strong correlation between the number of input taps and FC-NNE performance compensating a range of gain-saturated SOA input powers, while the FC-NNE structure (number of neurons, hidden layers) was kept constant. From this it was clear that sufficient input taps are needed to unlock the non-linear equalisation capabilities of FC-NNE, with up to 40 taps being required to overcome severe SOA patterning distortion at $+5$ dBm input power.

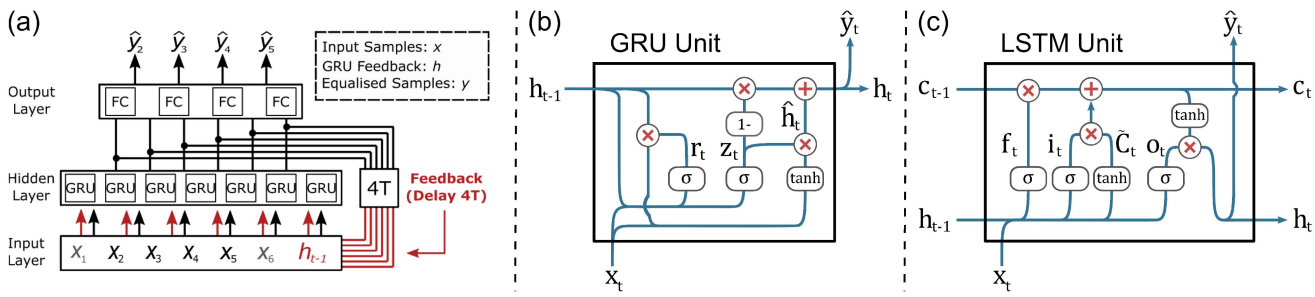

Fig. 3. Schematics of (a) the proposed GRU-RNN equaliser implementing 4-output multi-symbol equalisation and showing reduced RNN feedback timing requirements; (b) a single GRU unit showing how input and feedback information is controlled using gating mechanisms; and (c) similarly for a single LSTM unit, which has a memory cell state c_t which it can selectively update using input and feedback information.

This is a concern in the context of PON, since to equalise variable levels of signal distortion among burst packets (loud, soft, and everything in between), any proposed equaliser must be capable of equalising the worst-case scenario, which will require the most number of taps. However, for less stressed packets in PON systems, i.e. the majority, these additional taps will be surplus to requirements, and along with their associated multiplications will represent wasted energy and latency.

B. Volterra Non-Linear Equaliser as an Alternative

The Volterra equaliser architecture is widely seen as an alternative to machine learning based techniques. It can be thought of as a "super" polynomial fit to an equalisation problem, where subsets of the input taps, specified by the memory depth of the equaliser (m_1, m_2, m_3), are combined in all possible permutations up to some arbitrary order. Due to the exponential scaling of VNLE kernels, i.e. trainable parameters, in practice these equalisers are normally restricted to order 3. Note that memory depth m_1 corresponds to the number of VNLE input taps.

However, the issue discussed for FC-NNE in PON is also valid for VNLE. This is because its non-linear modelling ability is based on second and third order combinations of its input taps, and the number of kernels therefore scales exponentially with the increased higher order VNLE memory depth required for worst case loud packets in PON.

The equation which defines a 3rd order VNLE with memory depths $m_i = 2l_i + 1$ is given below:

$$\begin{aligned} y(n) \;=\; b &+ \sum_{j_1=-l_1}^{l_1} k_1(j_1)x(n-j_1) \\ &+ \sum_{j_1=-l_2}^{l_2} \sum_{j_2=j_1}^{l_2} k_2(j_1,j_2)x(n-j_1)x(n-j_2) \\ &+ \sum_{j_1=-l_3}^{l_3} \sum_{j_2=j_1}^{l_3} \sum_{j_3=j_2}^{l_3} k_3(j_1,j_2,j_3) \\ &\times\; x(n-j_1)x(n-j_2)x(n-j_3) \end{aligned}$$

Where b is a constant bias term, k_1, k_2, k_3 are the first, second, and third order kernels respectively, and $x(n), y(n)$ represent the nth input sample and equalised output of the VNLE respectively.

The complexity metrics, i.e. number of parameters / kernels and RVMs, can be calculated for such a VNLE using the following equations:

$$N_{kernels} = m_1 + \frac{m_2(m_2+1)}{2} + \frac{m_3(m_3+1)(m_3+2)}{6}$$

$$N_{RVM} = m_1 + m_2(m_2+1) + \frac{m_3(m_3+1)(m_3+2)}{2}$$

From these, and the VNLE defining equation above, it is clear that the calculation of a single ith order term, requires i RVM operations. Also evident is the exponential kernel scaling for 2nd and 3rd order terms. More details on VNLE principles and operation can be found in [14], [15].

C. Advantages of Recurrent Neural Networks

However, it is possible to make the number of taps required by an NNE agnostic to the degree of impairment. This is done using neural networks that incorporate feedback mechanisms, known as Recurrent Neural Networks (RNNs), which are designed to process time-series data [24]. In contrast to FC-NNEs whose equalisation performance is related to their input taps, RNN equalisers can effectively leverage their feedback mechanism to mitigate a range of non-linear inter-symbol interference extending beyond their immediate input taps. This makes it possible to compensate a variety of non-linear impairments using as little as 3 symbol spaced equaliser taps, as demonstrated in this work and in [6].

The simplest recurrent neural network (RNN) is defined by the layer equation:

$$h_t \;=\; \varphi\left(W \cdot x_t + U \cdot h_{t-1} + b\right)$$

This differs from the FC layer equation by the feedback term $U \cdot h_{t-1}$, which is the recurrent weight matrix U acting on the previous layer output h_{t-1}. RNN equalisers based on this simple recurrent layer have been used to equalise SOA non-linearities to improve Rx sensitivity to meet the PON optical loss budget with reduced equaliser tap numbers in [12].

However, within the ML community simple RNNs such as these are known to be unstable and difficult to train for long-term memory effects. For this reason, here and in [6] we propose using an RNN equaliser based on Gated Recurrent Units (GRUs) [25],

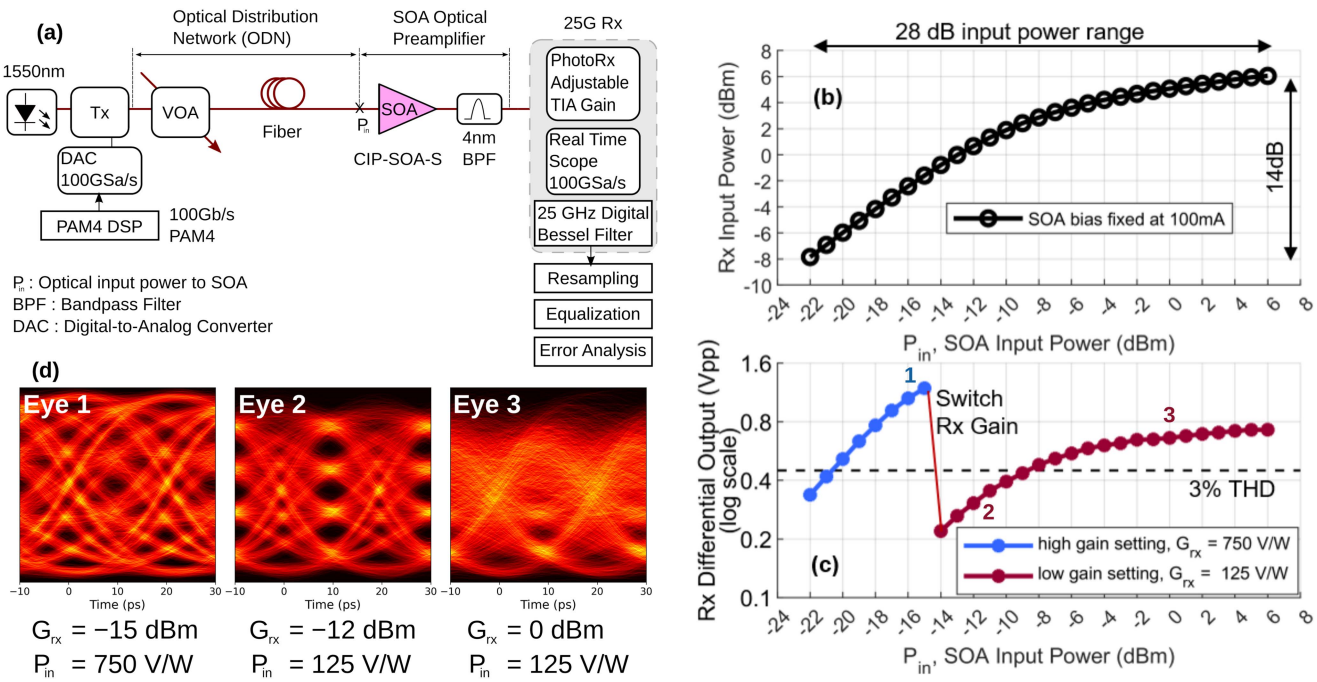

Fig. 4. (a) Experimental setup emulating upstream transmission in 100G PAM4 PON in continuous mode using an SOA preamplifier; (b) the effect of SOA gain compression on input dynamic range to the SOA due to gain saturation; (c) graph showing the differential output from the photoreceiver amplifier with the gain-switch point indicated along with 3% Total Harmonic Distortion threshold; and (d) eye diagrams of 100G PAM4 signals showing the effects of (1) Rx electrical gain saturation, (2) 25 GHz bandwidth limitation, and (3) combination of SOA patterning effect and Rx electrical gain saturation.

due to their superior training and inference stabilities while mitigating severe SOA patterning effects.

GRU-RNN equalisers implement a "gated" version of feedback mechanism, which is superior to that of standard recurrent units. GRU operation is defined by the equations:

$$\begin{aligned} z_t &= \sigma\left(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z\right) \\ r_t &= \sigma\left(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r\right) \\ \hat{h}_t &= \varphi\left(W_h \cdot x_t + U_h \cdot (r_t \circ h_{t-1}) + b_h\right) \\ h_t &= z_t \circ \hat{h}_t + (1 - z_t) \circ h_{t-1} \end{aligned}$$

Where $W \in \mathbb{R}^{n_h \times n_i}, U \in \mathbb{R}^{n_h \times n_h}, b \in \mathbb{R}^{n_h}$ are learned parameter matrices and bias vectors, σ is the sigmoid function, $\circ$ is the Hadamard product, and non-linear activation φ, which is set to $\tanh()$ in this work. The update and reset "gating" operation outputs are z_t and r_t which are constrained to values between 0 and 1 due to the sigmoid activation. They control the flow of information from the input x_t and feedback h_{t-1} taps to the candidate state, $\hat{h}_t$, and final output h_t, as illustrated in Fig. 3(b). This enables a GRU-RNN to determine the relevant feedback and input state information for the current symbol equalisation. The number of parameters and RVMs involved in a single GRU layer is calculated using the equations below:

$$\begin{aligned} N_{GRU,params} &= 3n_h \times (n_i + n_h + 1) \\ N_{GRU,RVM} &= 3n_h \times (n_i + n_h + 1) \end{aligned}$$

An alternative, more complex RNN unit is the long short-term memory unit, or LSTM [26], which has been investigated for equalising fiber dispersion in [10], [20]. The equations governing LSTM operation are similar to those of the GRU, and are given below:

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \\ i_t &= \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right) \\ o_t &= \sigma\left(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o\right) \\ \tilde{c}_t &= \tanh\left(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c\right) \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\ h_t &= o_t \circ \tanh\left(c_t\right) \end{aligned}$$

LSTMs also use "gated" calculations but further incorporate a "cell" state, c_t, which serves as an explicit form of network memory. This cell state can be selectively updated by the network with each calculation step, using the input, output, and forget gates: i_t, o_t, f_t, as well as a candidate cell state $\tilde{c}_t$, as in Fig. 3(c). Here, we investigate whether LSTM-RNN has an inherent performance advantage over GRU-RNN due to its explicit memory cell, for compensating SOA patterning effect in the context of 100 Gb/s PAM4 PON scenario. LSTM units involve more parameters and multiplication operations than GRUs, and this is reflected in the layer equations for N_{params} and N_{RVM} shown below:

$$\begin{aligned} N_{LSTM,params} &= 4n_h \times (n_i + n_h + 1) \\ N_{LSTM,RVM} &= n_h \times (4n_i + 4n_h + 3) \end{aligned}$$

The ability of RNN-based equalisers to implement feedback effectively is achieved using the backpropagation-through-time (BPtT) training algorithm [27], which is used in conjunction with the gradient descent optimisation algorithm Adam [28]. BPtT

works by "unrolling" an RNN along its feedback connection and then backpropagating the error calculated through a loss function along these feedback connections, effectively backwards in time. By tuning the hyperparameter controlling the number of times to unroll the RNN along its feedback, we can optimise the network's "memory", i.e. ability to address ISI outside its immediate input taps. Therefore, such a GRU-RNN can be trained to overcome varying levels of SOA patterning effect without changing its input structure, in contrast to the FC-NNE described in the previous section. However, this advantage of RNN does further compromise already long neural network training times, with the BPtT algorithm being especially memory intensive. How to train such an equaliser on a packet-by-packet basis in PON is an important research question outside the scope of the current article, although a potential solution may be to load pre-trained weights to the equaliser as required.

D. Multi-Symbol Equalisation

While RNN feedback structures remove the dependence of equaliser performance on input taps, this advantage comes with difficulties related to hardware (ASIC, FPGA) implementation. The feedback loops in RNN equalisers impose strict timing requirements, with the output of an RNN network layer h_{t-1} required to propagate around the feedback loop in time to act as input for the calculation of h_t in the next calculation cycle. As noted in [12], this limits DSP throughput, unlike in feedforward based equalisers such as FC-NNEs and FFEs which can exploit pipeline digital implementation [21]. Multi-symbol equalisation techniques, such as that investigated in [20], offer a way of alleviating this issue. This technique involves altering an RNN equaliser's structure to carry out parallel equalisation of multiple symbols in the same calculation cycle, as shown in Fig. 3(a). Whereas before, the output layer of such an equaliser consisted of a single FC neuron with linear activation, we now increase the output layer size to n-FC neurons, so that each equaliser calculation cycle will output n equalised samples. The feedback timing requirements are therefore reduced by a factor of $1/n$, thus potentially relaxing the hardware implementation requirements. Fig. 3(a) shows the structure of a GRU-RNN with 4 parallel outputs, referred to herein as GRU-PAR4. In future, we will append -PARn to an equaliser's name to indicate n parallel outputs.

Given an RNN-PARn equaliser, can it match the performance of the original single output version of itself? If yes, then key complexity metrics such as real RVM operations per equalised symbol can be improved thanks to this built-in parallelism. Therefore, according to this metric RNN-PARn could have overall better computational efficiency than single-output RNN equalisers, and even more efficient again than FC-NNE and VNLE. In the context of PON and variable impairment levels amongst burst packets, this sharing of equalisation calculations among symbols is even more persuasive, since less calculations per symbol corresponds to less calculations "wasted" on less stressed, soft packets.

But how robust are different RNNs to increasing levels of parallelisation? Once again, is the more complex LSTM-RNN better suited to implement multi-symbol equalisation for SOA patterning compensation? Is GRU-RNN equaliser size (i.e. number of units, hidden layers) a factor in supporting multi-symbol equalisation? These questions are discussed in Sections V and VI.

IV. Experimental Setup

Fig. 4(a) shows the experimental setup used to emulate an upstream 100G PON scenario in continuous mode. Although external or directly modulated lasers are preferred for upstream PON transmission, an ideal high-power Tx is realised in the C-band using an EDFA booster amplifier in conjunction with a Mach Zehnder modulator, in order to isolate the SOA preamplifier and Rx impairments for study. This is driven by a differential output DAC operating at 100 GSa/s, generating a 50 Gbd PAM4 signal with 6 dB extinction ratio. Linear pre-compensation corrects for system bandwidth limitations up to 33 GHz.

The optical distribution network in a PON is responsible for introducing loss through passive splitters as well as fiber dispersion impairment up to 20 km transmission distance [1]. A Variable Optical Attenuator (VOA) emulates network losses, allowing both loud and soft signal packets to be simulated in continuous-mode, with SOA input power varying between -26 and $+6$ dBm. The experiment was carried out in the C-band using a 1550 nm laser since appropriate O-band devices were unavailable, however SOA operating principles are fundamentally unchanged between C- and O-band, and therefore patterning effects could still be studied. Fiber dispersion up to and beyond 81.6 ps/nm was emulated using standard single mode fiber with dispersion parameter 17 ps/(nm · km) at 1550 nm. This exceeds the 70 ps/nm dispersion expected for 20 km transmission in the O-band using the upper-bound dispersion parameter value of 3.5 ps/(nm · km) [29], with the caveat of disparate dispersion slope.

The OLT comprises an SOA preamplifier (Model: CIP SOA-S) kept at constant bias and a commercial photoreceiver with integrated adjustable conversion gain. The SOA preamplifier and photoreceiver characteristics are shown in Fig. 4(b) and (c). Fig. 1 shows the measured gain curve for the SOA-S used in these experiments at 100 mA drive current. In the context of the imagined OLT Rx configuration, to support the 19.5 dB dynamic range required by the ITU-T 50G standards, the SOA will need to be driven into saturation by loud packets. This is due to the sensitivity of the Rx configuration being -22 dBm, while the SOA input saturation power is -8 dBm. This compresses the system optical power DR from 28 dB to 14 dB at the SOA output, as shown in Fig. 4(b), but introduces patterning effects above the SOA saturation input power of -8 dBm, as evidenced in the eye diagrams in Fig. 4(d), eye 3, which are later equalised using NNEs.

This optical DR compression is such, that in this experiment we use only two Rx gain settings at the photoreceiver to cover the entire 28 dB system DR. The Rx gain is changed from low to high setting at -15 dBm input to the SOA OLT as shown in Fig. 4(c). The photoreceiver differential electrical output is also shown, representing the amplitude swing between the PAM4 outer symbols. The Rx has linear response ($< 3\%$

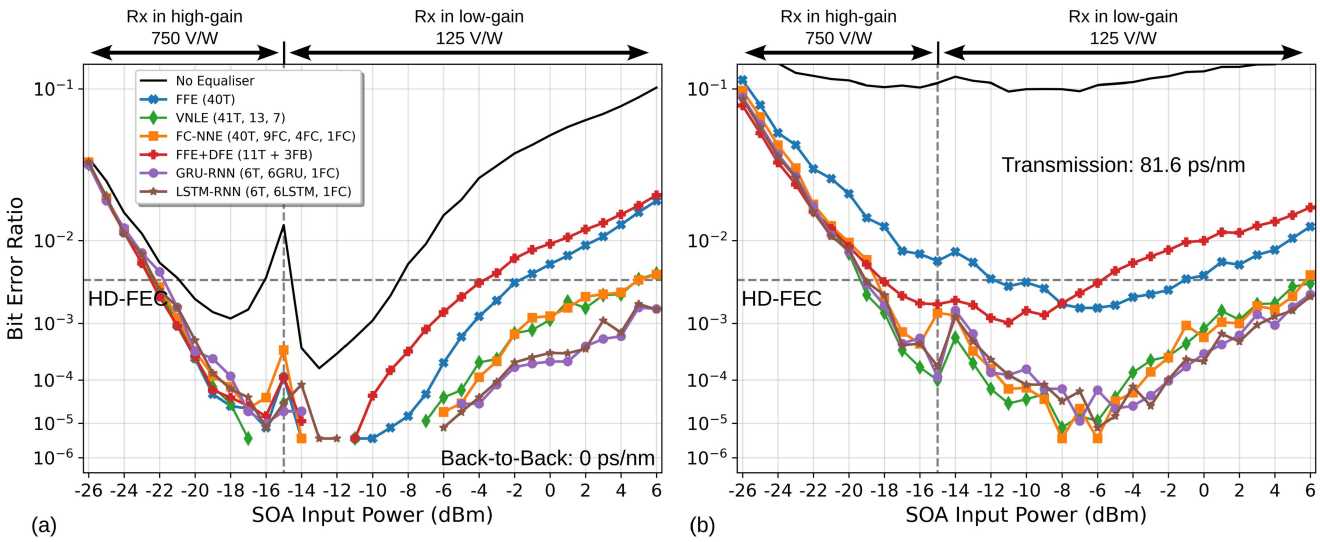

Fig. 5. Graphs showing bit error ratio (BER) performance for linear feedforward (FFE), non-linear feedforward (VNLE, FC-NNE), and non-linear feedback (FFE+DFE, GRU-RNN, LSTM-RNN) equalisation algorithms in 100 Gb/s PAM4 (a) back-to-back, and (b) transmission experiments. Rx sensitivity and dynamic range performances are inferred from these and other BER graphs.

THD - Total Harmonic Distortion) up to a differential output swing of 450 mV, and the eye diagram 1 in Fig. 4(d) shows clear degradation due to Rx electrical saturation at -15 dBm input to the SOA. After receiver gain switch from 750 V/W to 125 V/W, the signal in eye diagram 2 shows no evidence of such degradation at -12 dBm input.

Waveforms are captured using a 100 GSa/s real time scope with 33 GHz bandwidth, while a 4th-order Bessel filter is applied digitally to imitate 25G class opto-electronics. Offline processing is then carried out, before DSP using 1 sample per symbol is applied and final error analysis occurs.

In this work, equaliser training is carried out with a single pseudo random quaternary sequence (PRQS14) waveform [30], while multiple PRQS15 waveforms are used as test data and for BER estimation. Overfitting is a well studied phenomenon for NNEs [31], and performance overestimation is avoided by using independent PRQS-15, -14, and -13, patterns to generate the PAM4 symbols for testing, training, and validation respectively. Further, equaliser performance is monitored on these three datasets for signs of performance divergence which would indicate overfitting, but which was not observed. Each of the NNEs are trained using Adam optimisation for 1000 epochs using the early stopping method, meaning training is halted once performance improvement on the validation dataset stalls. This ensures parity among FC-NNE, GRU-RNN, and LSTM-RNN optimisation, while also ensuring excessive training does not lead to overfitting to the training data.

V. Results

In this work, we use the Hard Decision Forward Error Correction (HD-FEC) limit of 3.8×10^{-3} bit error ratio (BER) for determining system sensitivity and achievable dynamic range (DR). BER is calculated using the counting method, and for each BER data point reported here, $\sim$130 k symbols are used to estimate the error rate. The RNN equaliser structures evaluated here consist of a single hidden layer of 6 GRU, or LSTM units, followed by an output layer of n- FC units with linear activation which calculate the final n equalised samples. A larger version of GRU-RNN with 16 GRU units in the hidden layer is reported only in Fig. 8. The FC-NNE considered has two hidden layers of 9 and 4 FC units respectively using $\tanh()$ activation and single linearly activated FC as output. FC-NNE input taps is optimised for severely impaired high input power packets in our setup. The VNLE reported here and in Figs. 5 and 7 has memory depth structure $(41\mathrm{T}, 13, 7)$, where the 41T input taps were chosen to mirror that of the FC-NNE, and the 2nd and 3rd order memory depths are optimised using brute force search. Table I outlines selected VNLE and NNE equaliser structures used in this work, as well as their associated complexity in terms of trainable parameters and RVM operations per equalised symbol.

Fig. 5(a) shows the 100 Gb/s PAM4 system back-to-back (B2B) performance for FFE, NNE-FC, VNLE, GRU-RNN, and LSTM-RNN. The optical preamplified Rx sensitivity is measured to be -22 dBm at the HD-FEC BER threshold when using equalisation. Assuming $+8$ dBm launch power at ONU, this corresponds to a total system loss budget of 30 dB, which exceeds the 29 dB outlined in HS-PON.

The combined non-linearities of the SOA preamplifier and Rx electrical amplifier result in the distinctive "W" performance curve seen in Fig. 5(a), when no equalisation is applied. At the gain switch point of -15 dBm, the BER exceeds the HD-FEC limit due to Rx electrical non-linearities for high gain setting, as the Rx is operating well above the 3% Total Harmonic Distortion operating point, as seen in Fig. 4(c) and eye diagram 1 in Fig. 4(d). For SOA input powers above the SOA input saturation power of -8 dBm, the BER again exceeds HD-FEC as the SOA becomes gain saturated and the patterning effect begins to distort the signal significantly. The DR without equalisation is

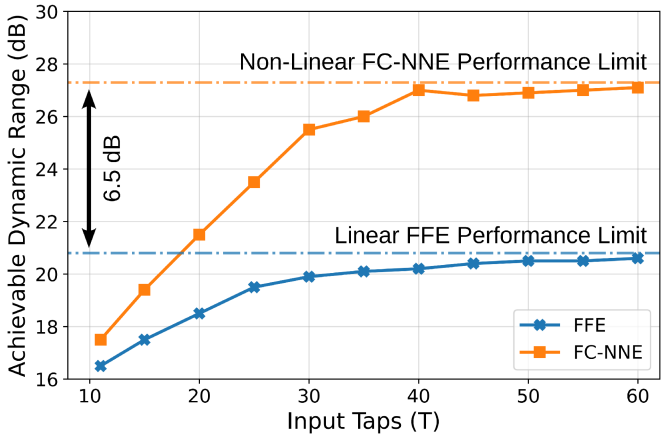

Fig. 6. The strong dependence of B2B FC-NNE performance on the equaliser's number of input taps is shown, as well as the additional performance advantage a non-linear FC-NNE has over the linear FFE, resulting in an observed 6.5 dB dynamic range improvement.

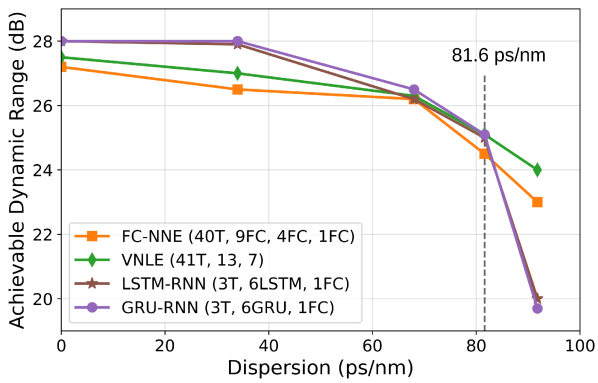

Fig. 7. Achievable dynamic range of the Volterra, FC-NNE, GRU-RNN, and LSTM-RNN equalisers are shown to be similar when compared over a range of dispersion values, with no equaliser showing clear performance advantage.

discontinuous between the two gain regimes, and it is therefore clear from Fig. 5(a) that equalisation is required to approach 19.5 dB.

The performances of FFE and FC-NNE are also shown in Fig. 5(a), with each equaliser's tap numbers chosen to maximise performance. The maximum achievable B2B dynamic range of the linear FFE is measured to be 20.5 dB, which is greater than the target 19.5 dB, but with only 1 dB tolerance. Fig. 5(b) shows this achievable DR is reduced to only 11 dB when considering 81.6 ps/nm dispersion impairment, categorically ruling out linear equalisation as a solution for 100 Gbit/s PAM4 in this context. The FC-NNE using 40 symbol spaced (T) taps achieves 27 dB DR, corresponding to a 6.5 dB increase in DR over optimal FFE, reflecting the non-linear nature of the SOA patterning impairment, and corresponding equalisation capabilities of FC-NNE.

Optimal FC-NNE input taps were determined using Fig. 6, where FC-NNE layer structure is kept constant while input taps are varied, and a strong correlation between input taps and B2B DR performance for both FFE, and FC-NNE is observed. Both equalisers achieve the -22 dBm receiver sensitivity reported above regardless of input tap number, and gains in DR due to increased input taps are achieved in the non-linear SOA regime above $P_{in,\,sat} = -8$ dBm. Therefore, it can be inferred that the increasing tap number requirements are due to the increasing severity of SOA patterning distortion for high SOA input powers. Beyond 40 taps, there is no improvement in FC-NNE performance, which approaches GRU-RNN performance. This suggests 40 tap FC-NNE is sufficient for the maximum extent of SOA patterning seen at $+6$ dBm SOA input power. Figs. 5(b) and 7 show the 40 tap FC-NNE matches the GRU-RNN equaliser DR up to 81.6 ps/nm dispersion, suggesting that FC-NNE learning capacity matches that of the GRU-RNN, but with a strong dependence on number of input taps.

Also shown in Fig. 5(a) is an FFE combined with decision feedback equaliser (DFE) using 3 decision feedback taps. It achieves 18.5 dB dynamic range B2B, which is reduced to 12 dB with 81.6 ps/nm dispersion impairment. No apparent performance advantage is seen by including the 3 feedback taps, and simply increasing FFE tap number to 40 outperforms the FFE+DFE B2B DR.

The VNLE considered in this work has 41T taps and structure $(41\mathrm{T}, 13, 7)$, for fair comparison with the FC-NNE. Fig. 5(a) shows it matches the performance to the FC-NNE, achieving ~ 27 dB DR for B2B, and falls slightly short of the GRU-RNN performance. In the transmission case shown in Fig. 5(b), the VNLE has the same performance as the FC-NNE, while Fig. 7 shows it also follows closely the same performance trend as the proposed GRU-RNN. From these results it is clear that the VNLE can emulate the non-linear modelling capabilities of the different NNEs when applied to SOA patterning and dispersion impairments, but does not exceed them in terms of performance.

The GRU-RNN equaliser using only 3 symbol spaced input taps and with structure (6GRU, 1FC) achieves an impressive > 28 dB B2B DR, and > 25 dB DR with up to 81.6 ps/nm dispersion, greatly exceeding the 19.5 dB outlined in HS-PON. This clearly illustrates the power of the GRU-based feedback mechanism to overcome SOA patterning impairment, while simultaneously avoiding the complexity bottleneck of excessive input taps required by the FC-NNE solution to achieve the same result. The proposed GRU-RNN is also clearly robust to multiple impairments, as Fig. 5 demonstrates it overcoming 25G bandwidth limitations, fiber dispersion, and combinations of SOA saturation and Rx electrical saturation. This suggest that GRU-RNN equalisers could be suitable to exploit SOA gain suppression in a system setting, to achieve extremely large > 28 dB system dynamic range. It is worth noting that in contrast to our work in [17], the GRU-RNN equaliser does not fully recover optimum BER at high SOA input powers, due to the fact that Rx electrical saturation impairment is also present at these powers, as seen in Fig. 4(c).

The more complex LSTM-RNN can only match the GRU-RNN DR performance in B2B and transmission experiments but cannot exceed it, as seen in Fig. 7, which also shows the FC-NNE had similar performance over all dispersion values considered. There appears to be no inherent advantage associated with the explicit memory "cell" state of LSTM units, and therefore LSTM-RNN does not offer any advantage over the less complex

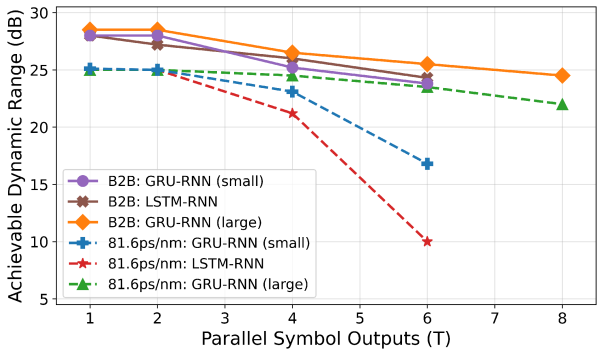

Fig. 8. Graph showing GRU-RNN high dynamic range performance can be sustained when implementing multi-symbol output technique, at the cost of larger network size.

GRU feedback mechanism in the context of 100 Gb/s PAM4 and the SOA patterning impairment considered here. Fig. 7 suggests that the equalisation capacities of each of FC-NNE, GRU-RNN, and LSTM-RNN are similar, although the GRU-RNN is less complex, with simpler feedback mechanism than LSTM-RNN, and less required input taps than FC-NNE.

Multi-Symbol Equalisation

Fig. 8 shows the achievable DR for B2B and 81.6 ps/nm dispersion cases for increasing numbers of parallel outputs. The equalisers considered are the GRU-RNN and LSTM-RNN as before, in addition to a larger version of the GRU-RNN which uses 16 GRU units in the hidden layer. As the number of parallel outputs is increased to 6 for small GRU-RNN, DR performance drops off by 3 dB for B2B case, and 8 dB with 81.6 ps/nm dispersion. The LSTM-RNN does not improve upon the GRU-RNN parallel performance, and so we conclude that the added complexity of LSTM feedback and cell state memory does not correspond to increased parallel performance.

Similarly, the larger GRU-RNN does not show significant performance gains in the B2B equalisation task up to 6-parallel output symbols, despite having increased learning capacity due to it having ~ 5 times the number of parameters and RVMs (see Table I). However, in the transmission case the large GRU-RNN achieves > 22 dB DR for up to 8 parallel outputs, outperforming the smaller GRU-RNN by ~ 7 dB.

While 8 parallel outputs may not sufficiently lower the RNN feedback frequency for real-time implementation on existing hardware, these results indicate that the multi-symbol equalisation technique is viable even for relatively small RNN equalisers compensating severe impairments such as SOA patterning and fiber dispersion. Additionally, evidence here and in the literature suggests that the technique can be extended beyond the 8 parallel outputs considered, with [19] successfully implementing an LSTM based equaliser on FPGA with 61 parallel outputs, and 270 MHz clock frequency.

VI. Complexity Analysis

Table I compares equaliser DR performance for B2B and transmission experiments in terms of the complexity metrics: trained parameters and RVM operations per symbol. The implications of the FC-NNE dependence on large numbers of taps to compensate the SOA preamplifer patterning effect can be clearly seen in Table I. The increase in FC-NNE input taps from 11 to 40 allows the equaliser to achieve 27 dB DR performance, and approach the performance of GRU-RNN, however this comes at the cost of an 170% increase in trainable parameters, and 187% increase in RVM operations per equalised symbol. Meanwhile, the VNLE 1st, 2nd, and 3rd order terms contribute 41, 91, and 84 kernels respectively, for a total of 217 including bias term. This is appreciably less than the FC-NNE but still greater than that of the GRU-RNN. However, the VNLE requires 475 RVMs, more than the FC-NNE, and significantly more than the GRU-RNN, underlining the importance of the VNLE kernel equation in Section III which exhibits exponential growth with respect to 2nd and 3rd order memory depths.

The GRU-RNN (small) uses only 3 symbol-spaced taps, and has reduced complexity and thus increased efficiency compared to the 40 tap FC-NNE. It uses only 187 parameters, and 186 RVM operations per equalised symbol to achieve 28 and 25 dB DR performance for B2B and 81.6 ps/nm dispersion scenarios respectively. Additionally, the LSTM-RNN equaliser offers no tangible benefit over the GRU-RNN in terms of DR performance, while using ~ 1.3 times the number of trainable parameters and RVM operations.

As discussed, the hardware implementation of RNN equalisers will be extremely challenging at 100G PAM4 data rate due to issues related to recurrent feedback mechanism latency. By implementing parallel outputs, we relax these timing requirements, and Table I highlights the reduction in RVM operations per symbol associated with this technique. There is an increase in total GRU-RNN parameters from the original value of 187 to 262 when implementing GRU-PAR4 equaliser, but the number of multiplication operations per equalised symbol drops from 186 to just 64.5, a reduction of 65%. This represents a significant increase in equalisation efficiency over that of FC-NNE, which would be especially important for burst-mode PON where the majority of burst-packets will not require the full equalisation capabilities of the GRU-RNN or FC-NNE which are designed around the worst-case, loud burst-packet.

The large GRU-RNN with 16 GRU units in its hidden layer has increased learning capacity over the smaller GRU-RNN considered, and as such can support > 22 dB DR for up to 8 parallel outputs. This would alleviate the latency issues faced when trying to implement such a GRU-RNN equaliser in hardware, but comes at the cost of significant increases in parameters and thus equaliser memory footprint, with GRU-PAR8 using 1432 parameters and needing 178 multiplication operations per equalised symbol.

VII. Conclusion

We achieve large 28 dB PON optical dynamic range using 100 Gb/s PAM4 system with SOA preamplifier, which is designed to emulate a future PAM4 PON scenario in continuous mode. A GRU-RNN equaliser using only 3 symbol-spaced input taps is proposed to recover modulated data signals from the

TABLE I
COMPARISON OF NON-LINEAR EQUALISER DR PERFORMANCE, PARAMETERS, AND REAL VALUED MULTIPLICATION OPERATIONS

Equaliser	Input Taps	Structure	Parallel Outputs	Parameters	RVM Operations per Equalised Symbol	SOA Input Power Dynamic Range [dB]			Observations
						0 ps/nm	81.6 ps/nm	91.8 ps/nm	
FC-NNE	11T	(9FC, 4FC, 1FC)	1	153	139	17.5	14.5	7	Inefficient equalisation due to strong performance dependence on input taps
	40T	(9FC, 4FC, 1FC)	1	414	400	27	26.5	23	
VNLE	41T	(41, 13, 7)	1	217	475	27.5	25.1	24	Matches GRU-RNN performance, but using 255% more RVMs per symbol
GRU-RNN (small)	3T	(6GRU, 1FC)	1	187	186	28	25	19.5	Compensates wide range of PON impairments with 3 fixed input taps
	4T	(6GRU, 2FC)	2	212	105	28	25	19.5	
	6T	(6GRU, 4FC)	4	262	64.5	25	23	16	Robust equalisation of SOA patterning, fiber dispersion, bandwidth limitation, electrical Rx saturation
	8T	(6GRU, 6FC)	6	312	51	24	17	2.5	
GRU-RNN (large)	3T	(16GRU, 1FC)	1	977	976	28.5	25	-	Supports > 19.5 dB DR multi-symbol equalisation up to 8 parallel outputs for B2B, and 81.6 ps/nm dispersion
	4T	(16GRU, 2FC)	2	1041	520	28.5	25	-	
	6T	(16GRU, 4FC)	4	1169	292	26.5	24.5	-	Uses less multiplications per equalised symbol than FC-NNE and LSTM-RNN for same performance
	8T	(16GRU, 6FC)	6	1302	216	25.5	23.5	-	
	10T	(16GRU, 8FC)	8	1432	178	24.5	22	-	
LSTM-RNN	3T	(6LSTM, 1FC)	1	247	240	28	25	20	Explicit memory cell state does not offer performance advantage over GRU-RNN
	4T	(6LSTM, 2FC)	2	278	135	27	25	20	
	6T	(6LSTM, 4FC)	4	340	82.5	26	21	15	Requires more multiplication operations due to more complex feedback mechanism
	8T	(6LSTM, 6FC)	6	402	65	24.5	10	0	

extreme patterning effect present in high-power, loud packets received at the OLT. This allows us to operate the SOA in its non-linear, gain-saturated regime which we exploit at the system level to reduce optical dynamic range from 28 dB at the SOA input, to just 14 dB incident on the photoreceiver; which is within reach of current state of the art LBMRx technology. The GRU-RNN is shown to be robust to a combination of device and fiber impairments, including up to 91.8 ps/nm fiber dispersion, SOA patterning, 25G bandwidth restriction, and electrical Rx saturation effects. The performance of VNLE, conventional FC-NNE using 40 input taps, and LSTM-RNN equaliser with sophisticated explicit memory cell state, are shown to match that of the GRU-RNN for up to 81.6 ps/nm of dispersion, but do not offer any DR performance advantage in the experimental scenario realised here.

Multi-symbol equalisation techniques are investigated for GRU-, and LSTM-RNN as a means to alleviate the strict timing requirements of RNN equalisers for future hardware implementations. Both types of RNN achieve > 25 dB B2B DR for 4 parallel symbol output, and no significant advantage is observed using LSTM over GRU units. However increasing the number of gated units in the GRU-RNN from 6 to 16 allows for up to 8 parallel symbol outputs with transmission performance approaching the B2B performance of a single output GRU-RNN equaliser.

GRU-RNN requires significantly less input taps than a FC-NNE or VNLE, and uses a simpler feedback mechanism than the LSTM-RNN, resulting in it needing less multiplication operations per equalised symbol and hence greater computational efficiency. Based on our results, we believe that GRU-RNN architecture is well placed amongst neural network equaliser solutions to support the equalisation requirements of future, high DR, 100G PON.

REFERENCES

[1] *50-Gigabit-Capable Passive Optical Networks (50G-PON): Physical Media Dependent (PMD) Layer Specification*, ITU-T Recommendation G.9804.3, International Telecommunications Union, Geneva, Switzerland, Sep. 2021.

[2] R. Bonk et al., "50G-PON: The first ITU-T higher-speed PON system," *IEEE Commun. Mag.*, vol. 60, no. 3, pp. 48–54, Mar. 2022.

[3] N. Suzuki, H. Miura, K. Matsuda, R. Matsumoto, and K. Motoshima, "100 Gb/s to 1 Tb/s based coherent passive optical network technology," *J. Lightw. Technol.*, vol. 36, no. 8, pp. 1485–1491, Apr. 2018.

[4] R. Borkowski et al., "FLCS-PON–A 100 Gbit/s flexible passive optical network: Concepts and field trial," *J. Lightw. Technol.*, vol. 39, no. 16, pp. 5314–5324, Aug. 2021. [Online]. Available: https://opg.optica.org/jlt/abstract.cfm?URI=jlt-39-16-5314

[5] J. Zhou et al., "100G fine-granularity flexible-rate passive optical networks based on discrete multi-tone with PAPR optimization," *J. Opt. Commun. Netw.*, vol. 14, no. 11, pp. 944–950, 2022.

[6] S. Murphy, F. Jamali, P. D. Townsend, and C. Antony, "High dynamic range 100 Gbit/s PAM4 PON with SOA preamplifier using gated recurrent neural network equaliser," in *Proc. Eur. Conf. Opt. Commun.*, 2022, Art. no. Th1C.6.

[7] Z. Li et al., "SOA amplified 100 Gb/s/λ PAM-4 TDM-PON supporting PR-30 power budget with >18 dB dynamic range," *Micromachines*, vol. 13, no. 3, Feb. 2022, Art. no. 342. [Online]. Available: https://www.mdpi.com/2072-666X/13/3/342

[8] J. Zhang et al., "SOA pre-amplified 100 Gb/s/λ PAM-4 TDM-PON downstream transmission using 10 gbps O-band transmitters," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 185–193, Jan. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8853279/

[9] L. Xue et al., "SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON," *Opt. Exp.*, vol. 29, no. 16, pp. 24714–24722, Aug. 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-16-24714

[10] X. Dai et al., "LSTM networks enabled nonlinear equalization in 50-Gb/s PAM-4 transmission links," *Appl. Opt.*, vol. 58, no. 22, Aug. 2019, Art. no. 6079. [Online]. Available: https://www.osapublishing.org/abstract.cfm?URI=ao-58-22-6079

[11] K. Wang, C. Wang, J. Zhang, Y. Chen, and J. Yu, "Mitigation of SOA-Induced nonlinearity with the aid of deep learning neural networks," *J. Lightw. Technol.*, vol. 40, no. 4, pp. 979–986, Feb. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9609624/

[12] X. Huang, D. Zhang, X. Hu, C. Ye, and K. Zhang, "Low-complexity recurrent neural network based equalizer with embedded parallelization for 100-Gbit/s/λ PON," *J. Lightw. Technol.*, vol. 40, no. 5, pp. 1353–1359, Mar. 2022.

[13] K. Wang, J. Zhang, L. Zhao, X. Li, and J. Yu, "Mitigation of pattern-dependent effect in SOA at O-band by using DSP," *J. Lightw. Technol.*, vol. 38, no. 3, pp. 590–597, Feb. 2020.

[14] J. Zhou et al., "Adaptive moment estimation for polynomial nonlinear equalizer in PAM8-based optical interconnects," *Opt. Exp.*, vol. 27, no. 22, pp. 32210–32216, Oct. 2019. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-22-32210

[15] Y. Xu, L. Huang, W. Jiang, L. Xue, W. Hu, and L. Yi, "Automatic optimization of volterra equalizer with deep reinforcement learning for intensity-modulated direct-detection optical communications," *J. Lightw. Technol.*, vol. 40, no. 16, pp. 5395–5406, Aug. 2022.

[16] L. Huang et al., "Low-complexity volterra-inspired neural network equalizer in 100-G band-limited IMDD PON system," *Opt. Lett.*, vol. 47, no. 21, pp. 5692–5695, Nov. 2022. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-47-21-5692

[17] S. Murphy, P. D. Townsend, and C. Antony, "Recurrent neural network equalizer to extend input power dynamic range of SOA in 100 Gb/s/λ PON," in *Proc. Conf. Lasers Electro-Opt.*, 2022, pp. 1–2. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=CLEO_SI-2022-SW4E.1

[18] T. Gurne et al., "First demonstration of a 100 Gbit/s PAM-4 linear burst-mode transimpedance amplifier for upstream flexible PON," in *Proc. Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.

[19] P. J. Freire et al., "Towards FPGA implementation of neural network-based nonlinearity mitigation equalizers in coherent optical transmission systems," in *Proc. Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.

[20] B. Sang et al., "Low complexity neural network equalization based on multi-symbol output technique for 200 Gbps IM/DD short reach optical system," *J. Lightw. Technol.*, vol. 40, no. 9, pp. 2890–2900, May 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9697352/

[21] N. Kaneda et al., "Fixed-point analysis and FPGA implementation of deep neural network based equalizers for high-speed PON," *J. Lightw. Technol.*, vol. 40, no. 7, pp. 1972–1980, Apr. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9645333/

[22] G. Agrawal and N. Olsson, "Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers," *IEEE J. Quantum Electron.*, vol. 25, no. 11, pp. 2297–2306, Nov. 1989.

[23] M. J. Connelly, *Semiconductor Optical Amplifiers*. Berlin, Germany: Springer, 2007.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: http://www.nature.com/articles/nature14539

[25] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. 2014 Conf. Empirical Methods Natural Lang. Process. ACL*, 2014, pp. 1724–1734, doi: 10.3115/v1/d14-1179. [Online]. Available: https://dblp.org/rec/conf/emnlp/ChoMGBBSB14.bib

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://direct.mit.edu/neco/article/9/8/1735-1780/6109

[27] P. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 7-9, vol. abs/1412.6980, 2015. [Online]. Available: https://dblp.org/rec/journals/corr/KingmaB14.bib

[29] *Characteristics of a Single-Mode Optical Fibre and Cable*, ITU-T Recommendation G.652, International Telecommunications Union, Geneva, Switzerland, Nov. 2016.

[30] H. Badaoui et al., "De Bruijn pseudo random sequences analysis for modeling of quaternary modulation formats," *Int. J. Comput. Sci. Issues*, vol. 7, pp. 35–38, 2010.

[31] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: Possible pitfalls," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2091–2094, Dec. 2017.

Stephen L. Murphy (Student Member, IEEE) received the B.Sc. degree in mathematics and physics from University College Cork, Cork, Ireland, in 2014, and the M.Sc. degree in mathematical physics from the University of Edinburgh, Edinburgh, U.K. in 2015. He is currently working toward the Ph.D. degree with the Tyndall National Institute, Cork, Ireland, as part of the Photonic Systems Group, supervised by Dr Cleitus Antony and Prof. Paul Townsend. His research interests include 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 with the European Conference in Optical Communications (ECOC'22), and the Conference on Lasers and Electro-Optics (CLEO'22).

Fariba Jamali received the B.Sc. degree in electrical engineering-electronics from Zanjan University, Zanjan, Iran, in 2012, and master's degree in the field of electrical engineering-communications from Tarbiat Modares University, Tehran, Iran, in 2016. Since 2021, she has been working toward the Doctoral degree with Tyndall National Institute, University College Cork (UCC), Cork, Ireland, as a Member of the Photonics System Group under the supervision of Dr Cleitus Antony and Prof. Paul Townsend. Her research interests include semiconductor optical amplifiers, and techniques to support high dynamic range transmission in future high-speed passive optical networks using power equalization methods in the optical domain.

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

Cleitus Antony (Member, IEEE) 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 with Hitachi's World-Class Fibre Optic Division, Opnext Japan Inc., Yokohama, Japan. 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 five patents and has authored more than 50 peer-reviewed publications. His research interest includes exploiting machine learning to advance next-generation optical communication systems and circuits.