CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Machine learning equalisation techniques for future passive optical networks
Authors	Murphy, Stephen
Publication date	2024
Original Citation	Murphy, S. L. 2024. Machine learning equalisation techniques for future passive optical networks. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2024, Stephen Murphy. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-10-31 19:24:00
Item downloaded from	https://hdl.handle.net/10468/17481

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Machine Learning Equalisation Techniques for Future Passive Optical Networks

Stephen Louis Murphy

Thesis submitted for the degree of Doctor of Philosophy



NATIONAL UNIVERSITY OF IRELAND, CORK

College of Science, Engineering and Food Science

Department of Physics

Tyndall National Institute

Photonics Systems Group

October 2024

Head of Department: Prof. Paul Callanan

Supervisors: Prof. Paul Townsend
Dr Cleitus Antony

Contents

List of Figures . . . iii
List of Tables . . . vii
Acknowledgements . . . ix
Abstract . . . xi

1 Introduction **1**
1.1 Evolution of optical access networks . . . 2
1.2 Neural network signal equalisation . . . 4
1.3 Thesis overview . . . 5

2 Future passive optical networks **8**
2.1 High-speed optical signalling . . . 9
2.1.1 Multi-level amplitude modulation . . . 11
2.1.2 Performance metrics . . . 13
2.1.3 System requirements . . . 16
2.2 Photonic system impairments . . . 19
2.2.1 Bandwidth and noise limitations . . . 19
2.2.2 Optical amplification . . . 21
2.2.3 Optical fiber channel effects . . . 23
2.3 Signal equalisation and neural networks . . . 26
2.3.1 Conventional equalisation . . . 27
2.3.2 Neural network theory . . . 29
2.3.3 State-of-the-art overview . . . 35
2.3.4 Methodology for signal equalisation . . . 39
2.4 Initial simulation investigations . . . 42
2.4.1 Bandwidth compensation . . . 42
2.4.2 Mitigating dispersion effects . . . 43
2.5 Summary . . . 45

3 Extending SOA preamplifier dynamic range with neural networks **46**
3.1 Context and research objective . . . 47
3.2 Semiconductor optical amplifiers . . . 47
3.2.1 Basic principles . . . 48
3.2.2 Gain dynamics . . . 49
3.2.3 Non-linear patterning effect . . . 51
3.2.4 Device under test . . . 55
3.3 Patterning in future PAM4 PON systems . . . 59
3.3.1 High-speed PAM4 testbed . . . 59
3.3.2 Achievable SOA input power dynamic range . . . 60
3.3.3 Analysis of FFE performance . . . 61
3.4 Designing a neural network equaliser . . . 63
3.4.1 Recurrent neural networks . . . 63
3.4.2 Performance and complexity analysis . . . 65
3.4.3 GRU gate mechanisms . . . 69
3.5 Summary . . . 71

4 Neural network equalisation in future 100G PON systems 72
4.1 Context and research objective 72
4.2 Non-linear equalisation techniques 73
4.2.1 Conventional FFE/DFE solutions 74
4.2.2 Volterra non-linear equalisation 75
4.2.3 Long short-term memory units 77
4.3 Upstream 100 Gbit/s PAM4 transmission 78
4.3.1 Exploiting SOA gain saturation for DR compression . . . 79
4.3.2 Experimental setup 81
4.3.3 Identifying non-linear impairment regions 83
4.3.4 Equalisation performance results 84
4.4 Complexity analysis 88
4.4.1 Parameters, multiplications, and pipelining 88
4.4.2 Comparison of non-linear equaliser complexities 91
4.4.3 Multi-symbol RNN equalisation 91
4.5 Summary 94

5 SkipNet: Adaptive Neural Network Equalisation for PON 96
5.1 Context and research objective 96
5.2 Burst-Mode Equalisation at the OLT 99
5.2.1 Existing adaptive equalisation solutions 100
5.2.2 Challenges facing upstream NNE 101
5.3 Adaptive NNE for burst-mode transmission 102
5.3.1 The neural network training barrier 102
5.3.2 SkipNet: Adaptive neural network equalisation 103
5.3.3 Multiple kernels for flexible non-linear equalisation . . . 106
5.4 Evaluating SkipNet performance for 100 Gbit/s PAM4 PON . . . 108
5.4.1 Experimental setup 108
5.4.2 Back-to-back and transmission performance 110
5.4.3 Achieving low-complexity, targeted NNE 113
5.4.4 Preamble training requirements 116
5.4.5 Achievable system dynamic range 118
5.5 Summary 119

6 Conclusion 122
6.1 Summary of contributions 123
6.2 Future developments 124

Acronyms 127

List of Publications 130

List of Figures

1.1 Optical access networks connect end users to the wider internet infrastructure using fiber optic cabling. 2
1.2 Equalisation allows distorted signals to be recovered at the Rx. . 4

2.1 Schematic of PON system components. 9
2.2 Comparison of (a) NRZ and (b) PAM4 modulation formats. . . . 11
2.3 A comparison of the spectral efficiency of NRZ, PAM4, and PAM8. 12
2.4 Plot of the error performances of NRZ, PAM4, and PAM8 for varying ϵ_b/N_0. 14
2.5 Example eye diagrams of (a) NRZ, (b) PAM4, and (c) PAM8. . . 16
2.6 Graphical definition of Rx dynamic range, as well as sensitivity and overload points. 18
2.7 Simulated 100 Gbit/s (a) NRZ and (b) PAM4 eyes resulting from a bandwidth-limited 25 GHz Rx, implemented using a 4$^{\text{th}}$ order Bessel filter. 20
2.8 Typical dispersion profile of a commercially available fiber. . . . 25
2.9 Eye diagrams show the effect of dispersion on a simulated 100 Gbit/s PAM4 signal at 1344 nm and using D and S parameters taken from Fig. 2.8. 26
2.10 (a) The structure of an FFE; (b) shows a simulated 100 Gbit/s PAM4 eye after a 10 GHz 4$^{\text{th}}$ order Bessel filter while (c) shows the same eye after equalisation using an FFE-11T. 27
2.11 Schematics showing (a) conventional NN structure, (b) neuron operation, (c) popular activation functions, and (d) the composition of an NNE. 31
2.12 Example NNE learning curves showing (a) no overfitting, and (b) overfitting 41
2.13 Simulation setup used for investigations into Rx bandwidth and chromatic dispersion impairments on 100 Gbit/s PAM4. 42
2.14 The sensitivity penalty of a bandwidth-limited receiver simulated using a 4th order digital Bessel filter. 43
2.15 Simulated chromatic dispersion penalty at 3.8×10^{-3} BER at (a) 1344 nm and (b) 1260 nm wavelengths. 44

3.1 Challenges facing SOA preamplification due to burst-mode upstream transmission at the OLT. 47
3.2 SOA structure, bandgap and optical amplification mechanism. . 48
3.3 (a) The PAM4 sequence used as input to the simulated SOA device above and below the saturation input power. The percentage change in carrier density and resulting output power of the device in both scenarios are shown for (b) a 50 Gbaud and (c) 12.5 Gbaud signal, with both showing clear signs of patterning. 53
3.4 The gain curve of the CIP SOA-S at a constant bias of 100 mA is shown, with the linear and non-linear operating regimes delineated by the 3-dB gain saturation point. 56

3.5 Experimental setup used to measure gain recovery time of SOA used. 56
3.6 Results of pump-probe experiment to measure the SOA-S gain recovery for a constant bias of 100 mA and a 1550 nm probe wavelength 57
3.7 Frequency response of SOA for different input powers. 58
3.8 Experimental setup used to emulate multi-rate upstream PAM4 transmission. 59
3.9 Increasing severity of SOA non-linear patterning on 100 Gbit/s PAM4 transmission as seen in eye diagrams and sampled symbol distributions. 60
3.10 Performance of high-speed PAM4 signals for a range of SOA input powers. Clearly above -8 dBm the signals undergo saturated SOA non-linear patterning effect. 61
3.11 Transmission performance of various baud rate PAM4 signals using an SOA+PIN Rx configuration, and FFE-80T post-equalisation. . 62
3.12 (a) Trained FFE taps showing the majority contribution of past symbols toward equalising the SOA patterning effect; (b) the strong dependence of linear FFE equalisation on the number of input taps for different baud rates. 62
3.13 (a) Shows a schematic of a GRU neuron with a gated feedback mechanism, and (b) is the corresponding state diagram of a GRU; (c) shows the backpropagation through time algorithm for training RNNs. 65
3.14 Schematic diagram of a GRU-based RNN equaliser. 66
3.15 BER results for (a) GRU-NNE (3T) at different baud-rates, and (b) a comparison of results for GRU-, FC-NNE, and FFE equalisers for 100 Gbit/s PAM4. 67
3.16 The performance dependence of (a) an FC-NNE on tap numbers is compared with that of a (b) GRU-NNE's performance dependence on training time steps, for varying SOA input powers. 68
3.17 Shows the simulated pulse fed to trained GRU-NNE models on saturated and unsaturated SOA data, and the resulting output activation response of the most significant GRU neuron. 70

4.1 Schematics of (a) an FFE/DFE equaliser, and (b) a 2^{nd}-order non-linear Volterra equaliser. 74
4.2 Schematic of an LSTM recurrent unit with memory cell c_t. . . . 77
4.3 (a) The effect of SOA gain compression on input dynamic range to the SOA due to gain saturation; (b) graph showing the differential output swing from the photoreceiver amplifier with the gain-switch point indicated along with 3% Total Harmonic Distortion threshold. 80

4.4 Experimental setup emulating 100 Gbit/s upstream PON transmission, using SOA gain saturation and neural network equalisation to enable > 28 dB dynamic range using only two electrical gain settings in the photoreceiver. 81
4.5 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. 83
4.6 BER performance of feedforward (FFE, VNLE, FC-NNE), and feedback (FFE/DFE, GRU-RNN, LSTM-RNN) equalisation applied to 100 Gbit/s PAM4 (a) back-to-back, and (b) transmission experiments. 85
4.7 The strong dependence of BtB 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. 86
4.8 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. 87
4.9 Multi-output GRU-based equaliser. 92
4.10 Graph showing GRU-RNN high dynamic range performance can be sustained when implementing multi-symbol output technique, at the cost of larger network size. 93

5.1 The burst-packet structure of upstream PON transmission according to the latest 50 Gibt/s HS-PON standards is shown. Only 150 ns preamble allocated for 25 Gbit/s upstream transmission, and 610 ns in a worst case scenario [12]. 99
5.2 (a) Outlines the design methodology behind creating an adaptive SkipNet equaliser, and how kernels are derived from the hidden layers of an ideally trained NNE before (b) being integrated into the SkipNet framework. Flexible and rapid adaptive NNE is then possible using a linear output neuron and the adaptive circuit shown in (c), which is equivalent to conventional adaptive FFE/DFE training. 104
5.3 (a) Shows the experimental setup which incorporates an SOA preamplifier 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). 109
5.4 (a) Shows the BtB performance of the SkipNet(K_1) equaliser 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 equalisers underlines the flexibility of the scheme to accommodate additional low-complexity kernels for greater non-linear capability. 111

5.5 These heatmaps in (a) and (b) show how K_1 and K_2's training conditions were chosen, by plotting kernel training power versus the BER performance of the resulting SkipNet equalisers. 114
5.6 The RMS values of the converged weight parameters in the SkipNet 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. 115
5.7 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 [116]. . . . 117
5.8 The SkipNet equaliser achieves > 25 dB DR for up to 68 ps/nm accumulated dispersion using just 250 symbols, clearly outperforming the adaptive FFE/DFE equaliser which cannot even support 19.5 dB PON DR requirement in the BtB scenario. 119

6.1 (a) Possible 400 Gbit/s PAM4 WDM PON scenario, with (b) adaptive SkipNet equaliser to mitigate SOA cross-channel patterning effects. 125

List of Tables

2.1 Gray encoding ensures that mismeasurement of neighbouring symbols leads to only a single bit error. 15

3.1 Comparison of NNE complexities in the context of total parameters, RVMs, and achievable DR performance. 69

4.1 Comparison of Non-Linear Equaliser DR Performance, Parameters, and Real Valued Multiplication Operations. 90

5.1 Packet-by-packet training complexities 113

I, Stephen Louis Murphy, certify that this thesis is my own work and has not been submitted for another degree at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.



Stephen Louis Murphy

Acknowledgements

This thesis brings together 6—sometimes long, sometimes short—years of research, and draws a line under an important chapter of my life. However, while I am proud of the work and accomplishments it represents, it fails to shine a light on the many wonderful people who have supported and enabled me throughout my PhD journey. A PhD can be a lonely place, but thanks to my family, friends, and soulmate, I have always had a shoulder to lean on, and lots more besides.

Thank you to my supervisors, Cleitus and Paul, for giving me the opportunity to join the Photonics Systems Group, and for their infinite patience guiding me throughout my work. Our last-minute conference deadline sessions will remain some of my fondest memories of my time at Tyndall, and hopefully we will continue to laugh about these over pints long into the future!

I count myself fortunate to have been a part of an amazing team in PSG. In particular, I cannot adequately express in words my appreciation for my friend Brian Murray, who taught me everything I know about the lab, and continued to encourage and support me long after he himself had left Tyndall. To Hilmi who was ever-present as we went through the trials of a PhD side-by-side, thank you for your friendship and humour: your never-ending faith in Arsenal football club is an inspiration! A special thanks to my friends and co-authors Fariba and Ehsan, who I was lucky enough to work closely with, and who always said yes when I needed help. Although there are too many to name, thank you to all the Tyndall friends and colleagues who have come and gone over the years, with some honorable mentions: Meysam, Eoin, Julie, Lorenzo, Xing, Conor, Gökhan, Mohammad, Samvit, Vanessa, Aritra, Fatima, Denis, Stefan, Giuseppe, Natalia, the Tyndall cycling gang, and many, many others.

Family and friends outside of work have more than played their part. I owe everything to my parents, Bernard and Kathy, who have given me everything so that I might be in the position I am in today: I can never thank them enough, but I promise to try my best. Siblings and best friends are experts at reminding us what is really important in life: thank you to Ben and Mark for their constant inquiries of "when will I get a real job, anyway?", and a massive thank you to Dr Emma O'Sullivan for always listening to my PhD complaints, and being the best baker in the greater Gurrane area—and probably elsewhere too. Thank you to Gary and Sylvia, Aidan and Igea, Sean, Ciara, Treasa, Declan and Ellen, Tom, Baiba and Aiden, Odí and Johannes, Bernie, Joe and Niall: you may not care about optical communications, but thank you for always asking how it was going. To Sini, you are the most important unspoken character in this thesis: thank you for never allowing me to doubt this finish line existed—as we begin our next chapter together, I cannot wait for all the adventures we are yet to have.

Finally, a special thank you to my grandparents, Stephen, Breda, Sheila, and Louis. Many years ago Prof. Louis Buckley guided me around UCC, and inspired me to achieve a doctorate myself.

Abstract

Passive Optical Networks (PON) have become crucial in delivering reliable, high-capacity, and low-latency fiber internet connectivity, largely replacing legacy copper cabling. Research into next generation 100 Gbit/s systems is ongoing to support future bandwidth demands, and the adoption of advanced modulation formats, such as 4-level pulse amplitude modulation (PAM4), will be key to achieving this. However, optical amplification will be required to support PAM4 transmission due to the strict optical loss budgets imposed by legacy network infrastructure. Semiconductor optical amplifiers (SOAs) have been widely proposed, but introduce non-linear signal distortions which can greatly restrict PAM4 performance. Moreover, fiber dispersion impairments are also a concern given PAM4's high linearity requirements. This thesis explores how neural network-based equalisation (NNE) can offer new solutions to these traditional physical limitations, and enable future 100 Gbit/s PAM4 PON systems.

Through rigorous experimental analysis, this work identifies SOA-induced non-linear patterning as a major barrier to achieving the required 19.5 dB dynamic range requirement for upstream PON transmission. Gated recurrent neural networks are proposed to overcome this effect, and are shown to fully recover PAM4 signals for a range of bit rates up to 128 Gbit/s. The proposed NNE achieves > 28 dB dynamic range in a 100 Gbit/s experimental PAM4 PON scenario, where additional non-linear effects such as fiber dispersion and electrical receiver saturation are also present. Meanwhile, conventional linear equalisation techniques are shown to be insufficient to meet the PON dynamic range requirement. Furthermore, complexity analyses of the proposed NNE solutions are included throughout, and multi-symbol equalisation techniques are investigated to alleviate the challenges of feedback-based NNEs in future hardware implementations.

The thesis' most significant contribution is the invention of SkipNet, an original adaptive neural network algorithm developed by the author. It overcomes the significant obstacle of complex NNE training, and enables real-time, packet-by-packet adaptation to dynamic transmission conditions in burst-mode PON systems. SkipNet is shown to match the performance of conventionally trained NNE solutions, while using as little as 250 training symbols to mitigate non-linear SOA patterning and fiber dispersion. The introduction of SkipNet opens new pathways for designing future PON systems that are scalable, cost-effective, and capable of meeting the ever-growing demand for high-speed connectivity.

Chapter 1

Introduction

Chapter 1 introduces neural network-based signal equalisation in the context of future passive optical networks, and outlines the structure of this thesis.

Fiber-optic networks communicate information by encoding signals onto light, which can then be transmitted through optical fiber cables over large distances. The remarkable reliability, bandwidth, and transmission distances of fiber-optic communications technology has been essential in ushering in the 'Information Age', and the vast trans-continental fiber cables that criss-cross the world form the "backbone" infrastructure of the modern internet [1], [2].

The ongoing adoption of bandwidth-intensive cloud services, and shift toward remote working is driving innovation at the network edge. This is evidenced by the ongoing worldwide roll out of optical access networks, which connect end-users via fiber to core telecommunications networks and are rapidly replacing slower, legacy copper cabling [3]. Upcoming technologies such as the Internet-of-Things [4], AR/VR, 8K video streaming, and 5/6G mobile connectivity [5]–[7], will continue to drive development of next-generation optical access technology alongside potential Industry 4.0 applications [8].

To sustain this growth and meet the bandwidth demands of an increasingly connected world, novel technologies such as advanced digital signal processing techniques are being considered, including those based on machine learning and the burgeoning field of artificial intelligence [9], [10].

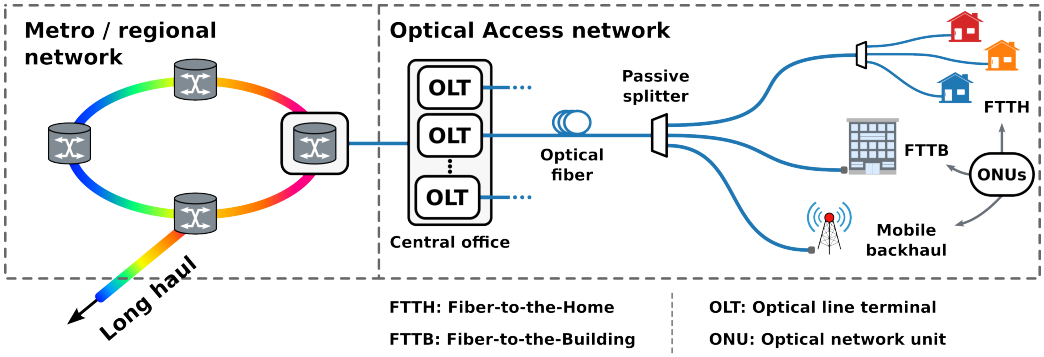


Figure 1.1: Optical access networks connect end users to the wider internet infrastructure using fiber optic cabling.

1.1 Evolution of optical access networks

Passive Optical Network (PON) technology dominates the optical access market due to its relatively low-cost and the need for mass deployment. These systems use passive power-splitters to connect 64 or more network subscribers via a single fiber link to an optical line terminal (OLT) located at a central office, which acts as a gateway to the core / metro telecommunication networks, as shown in Figure 1.1. This branching network architecture is known as the optical distribution network (ODN), and it makes PON ideal for both urban and rural deployment since users need not be co-located, but can be geographically disparate and up to 20 km from the central office. Further, there are no active components (e.g. amplifiers) installed in the ODN (thus "passive" optical network) which lowers energy and maintenance costs, at the cost of strict optical loss budgets for the ODN. Network capacity is shared amongst optical network units (ONUs) using the time division multiplexing (TDM) technique where ONUs take turns transmitting to the central office, and this results in the burst-mode nature of upstream network traffic specific to PON. Dealing with the varying power levels and signal conditions of different users arriving in "bursts" at the central office is one of the major technical challenges in implementing PON. Wavelength division multiplexing (WDM) can also be used to increase capacity even further [11].

Intensity modulation with direct detection (IM/DD) technology which encodes information using only amplitude modulation has been the basis of PON to date, due to its simplicity and the widespread availability of 10G class optics. Broadband PON (BPON) originated in 1995, and ultimately delivered 622 Mbit/s, while the latest Higher Speed PON (HS-PON) recommendations published by the ITU-T in September 2021 will achieve 50 Gbit/s [12]. In Ireland, the National Broadband Plan will initially deploy 10 Gbit/s XGS-PON [13]–[15] with scope to

upgrade in the future, while places such as Chattanooga, Tennessee are already offering 25 Gbit/s fiber links for businesses and domestic consumption[1] [16]. A sustained research effort is already underway to identify the technology which will deliver 100 Gbit/s/λ (per wavelength) line-rates and beyond so that PON technology can keep pace with bandwidth demand [17], [18].

However, fiber dispersion and limited device bandwidths are significant challenges to the continued use of binary, non-return-to-zero on-off keying (NRZ-OOK) modulation. Therefore, in a first for PON the 50 Gbit/s HS-PON standards concede that digital signal processing (DSP) will be necessary to support such a line-rate [19]. DSP will play an increasingly important role in future 100 Gbit/s and greater PON systems in order to continue scaling bit-rates with demand [20]. Proposals to date include: adopting higher order modulation formats such as 4-level pulse amplitude modulation (PAM4) [21], digital signal equalisation [22], coherent detection [23], and flexible rate systems [24], among others [25].

Of these, PAM4 is a preferred option for realising 100G, since it could be implemented using existing 25G class IM/DD technology, and avoids the significant costs involved in switching to coherent technology. Further, PAM4 is continuing to be developed in other application areas, such as data center interconnects (DCI) [26], and is a well-established technology. However, its disadvantages include an increased susceptibility to device non-linearities and noise, meaning optical amplification will be required to realise the strict 29 dB minimum HS-PON loss budget [12]. Avalanche photodiodes (APDs) are not readily available at 25G, and an alternative is to integrate a semiconductor optical amplifiers (SOA) with PIN photodiodes [27] to boost receiver (Rx) sensitivities and support PAM4 PON transmission. SOAs are seen as ideal optical preamplifiers thanks to their being low-cost, readily integrable, and operating in both C- and O-bands. However, the large PON dynamic range of 19.5 dB [12] is a challenge with this Rx configuration at the OLT, as high-power packets could cause SOA gain-saturation induced non-linearities, such as the patterning effect, i.e. self gain modulation (SGM) [28]. To overcome this, a range of advanced signal equalisation techniques have been proposed, including those based on neural networks.

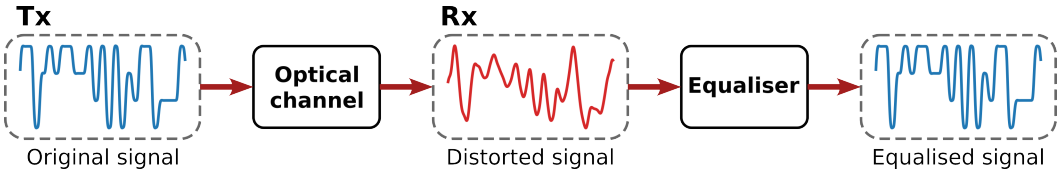


Figure 1.2: Equalisation allows distorted signals to be recovered at the Rx.

1.2 Neural network signal equalisation

The proliferation of artificial intelligence (AI) and machine learning (ML) techniques across science and engineering is having a profound effect on many research fields [29], [30]. In optical communications, ML is expected to enable more efficient and higher bandwidth networks, by enabling intelligent optical network management [31], optical signal monitoring [32], photonic device design [33], as well as advanced signal recovery (equalisation) algorithms [34] in the physical layer.

In this thesis, the focus is on the development of neural network-based equalisation algorithms for future PON systems. Equalisation is a form of signal processing which aims to optimise the recovery of a transmitted signal's data. Equalisation can boost signal-to-noise ratio (SNR) as well as compensating for inter-symbol interference (ISI) and possible distortions induced during signal transmission [35]. As such, it is seen as a potential solution to the severe fiber dispersion impairments and device bandwidth limitations expected in next generation PON, and could be key to enabling increased bit-rates and achievable transmission distances [36]. Linear feed forward equalisation (FFE) is a potential candidate due to its simplicity and effectiveness against bandwidth limited channels. However, more capable non-linear equalisers are also being investigated due to the susceptibility of advanced modulation formats such as PAM4 to non-linear distortions. In particular, non-linear fiber and device effects, including SOA patterning effects, are motivating research into such advanced equalisation. In this context, there have been numerous proposals for complex non-linear equalisation for PAM4 systems, including neural networks [34].

Neural network (NN) algorithms are responsible for some of the most famous successes within AI and ML, such as the transformer models underpinning ChatGPT [37], and the deep reinforcement learning models behind AlphaGo [38]. In the field of communications, NNs have been put forward as potential non-linear equalisers [39]. Recently, NNs are increasingly being proposed to overcome a

[1] "If we offer it, people will buy it" - Anonymous telecoms executive.

range of non-linear effects in both IM/DD and coherent optical systems. Originally developed to mimic the way neurons process information in the brain [40], these algorithms are universal approximators, meaning they can model a wide range of non-linear functions [41], [42]. Further, they are data driven and do not rely on analytical models of physical processes, but can learn from data. These characteristics of NNs make them incredibly flexible and powerful equalisers, since they can be trained to tackle a wide range of linear and non-linear distortions using only captured signals from a transmission link. As a result, a concerted research effort is underway to investigate which electrical and photonic non-linearities are susceptible to these neural network-based equalisers (NNEs), as well as significant work on adapting NN algorithms for use in high-speed optical communications.

PON systems in particular involve unique challenges for NNEs due to their dynamic burst-mode signalling, potentially multiple target impairments, and strict cost-budgets. Due to the potentially large number of ONUs per PON, there is a cost-imperative to make ONU transceivers as cheap as possible, while the OLT transceiver can be relatively more expensive [17]. A consequence of this is that advanced signal equalisation such as an NNE is most appropriately placed at the OLT, where its costs can be shared among all ONUs. For upstream transmission, a receiver-side NNE would therefore need to be robust to a wide range of possible signal impairments originating from ONUs at variable distances from the OLT, which is a potential problem for NNE algorithms which require large datasets and complex algorithms to adapt. Related to this, the feasibility of a given equalisation algorithm is ultimately constrained by limited computational resources and hardware throughput of the target hardware platform (e.g. ASIC, FPGA). NNEs can involve many multiplication and non-linear operations, as well as feedback loops, and are therefore complex and resource intensive to implement in digital logic with sufficient throughput for 100 Gbit/s PON. Identifying suitable NNE target impairments and developing low-complexity, effective NNE solutions appropriate for PON are therefore significant research challenges.

1.3 Thesis overview

In this thesis, the focus will be on developing an NNE algorithm which will enable 100 Gbit/s PAM4 signalling for the challenging PON upstream transmission case, where an SOA is utilised as a receiver preamplifier. This NNE solution will need to deal with SOA gain-saturation induced non-linear patterning effects, fiber

dispersion, and bandwidth limitations. Further, the final algorithm will need to be capable of rapid adaption to deal with the varying signal conditions inherent to burst signalling of upstream PON. The following chapters are designed with these goals in mind, and are detailed below.

Chapter 2 introduces the operating principles behind core concepts and technologies used in PON, as well as the background theory of NNEs. It motivates the adoption of PAM4 signalling for achieving future 100 Gbit/s PON, as well as introducing system performance metrics for evaluating these signals. Sources of signal degradation are introduced, including noise, bandwidth limitations, and fiber dispersion, as well as the effect of burst-mode signalling. Neural network equalisation is introduced and described in terms of its relation to existing FFE methods, before initial investigations into mitigating simulated bandwidth and dispersion impairments using NNEs are presented.

Chapter 3 introduces SOAs and the non-linear patterning effect they induce during gain saturation. Initial characterisation of the SOA used in this work is given, and an investigation of the effect of SOA patterning on 100 Gbit/s PAM4 and achievable dynamic range is reported, an important consideration when proposing SOAs as preamplifiers at the PON OLT. Equalisation solutions based on neural networks are then explored, and recurrent neural networks are introduced as optimal equalisation solutions. The experimental results discussed in this chapter are published in [A1].

In **Chapter 4** the NNE solution developed for isolated SOA patterning in Chapter 3 is evaluated in a 100 Gbit/s PAM4 system which further incorporates fiber dispersion and bandwidth limitations. It is shown that the NNE allows the gain saturation regime of the SOA to be exploited by compressing large optical dynamic ranges and therefore reduce requirements in subsequent burst-mode receiver electronics. Further, the NNE's performance and complexity is benchmarked against existing linear and non-linear equalisation solutions such as the Volterra non-linear equaliser, as well as alternative advanced NN structures. In particular multi-symbol output techniques are investigated to reduce equaliser complexity. The results in this chapter are published in [A2], [A3].

The thesis culminates in **Chapter 5** with the introduction of SkipNet, an original adaptive NNE algorithm proposed by the author, and which is capable of rapidly training on a packet-by-packet basis to deal with burst-mode signalling. First, the requirements for adaptive equalisation in upstream PON are discussed in the context of existing NNE training schemes, which are shown to be inappropriate

for such a task. Then, the SkipNet NNE algorithm is introduced as a solution for achieving adaptive neural network equaliser training, and its performance evaluated for 100 Gbit/s PAM4 PON. As far as the author is aware, SkipNet is the first demonstration of a truly adaptive NNE for PON, and has been published in [A4], [A5].

Chapter 6 concludes with a summary of the thesis contributions, and discusses possibilities for further research.

Chapter 2

Future passive optical networks

This chapter details the structure, operation, and current standards of passive optical networks, before discussing the potential of PAM4 to deliver next generation 100 *Gbit/s PON. Challenges in terms of device and fiber impairments are outlined, before equalisation is discussed as a key solution. Neural networks are introduced as possible non-linear equalisers and initial simulation results given which evaluate their performance on bandwidth limitations and fiber dispersion.*

Figure 2.1 illustrates the three main subdivisions of a PON system: the OLT located at the central office, the passive splitters and fiber cabling that comprise the ODN, and the subscriber termination points known as ONUs. These subsystems and how they interact are responsible for PON's cost-effectiveness, but also impose challenges to realising 100 Gbit/s line-rates. The first half of this chapter discusses the consequences of this design for component devices and technologies.

The structure shown in Figure 2.1 is highly cost-effective for several reasons related to the ODN. Importantly, the ODN is passive, and therefore no active components are deployed here, meaning maintenance and energy costs are low. Further, optical fiber attenuation is ≈ 0.2 dB/km which is extremely low relative to lossy coaxial cables [43], meaning network subscribers can be geographically distributed, and up to 20 km from the central office. Split ratios up to $1:64$ in existing PON standards ensure that costs are distributed among many users, and are economically viable for the service provider[1]. Time division multiplex-

[1]Ratios of up to $1:1024$ have been proposed in "Super PON" systems [44], [45].

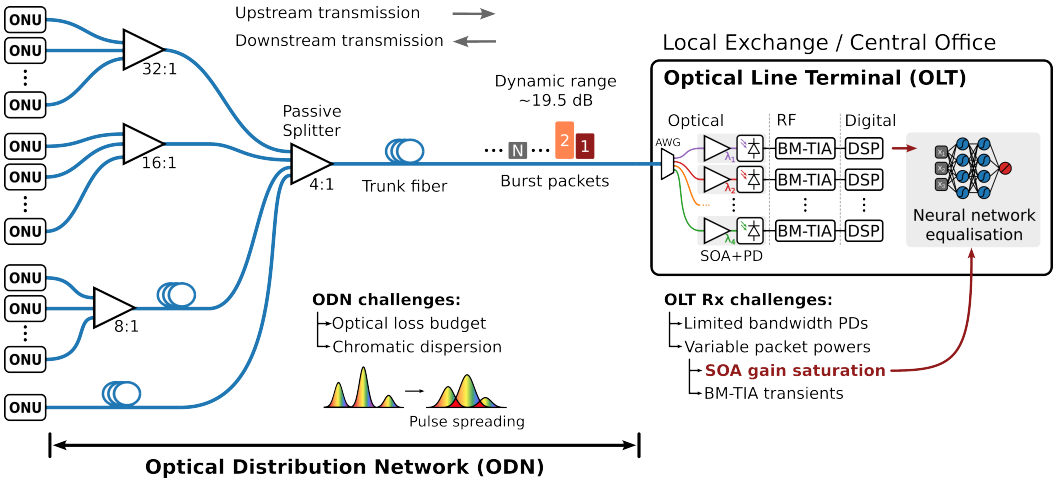


Figure 2.1: Schematic of PON system components.

ing (TDM) allows for the multiple network users to share the overall system bandwidth [46].

However, this network structure also makes it challenging to continually increase data rates in successive PON generations. Existing ODN infrastructure must continue to be supported between PON standards, since their installation represents up to 70% of initial PON investment [47]. This means higher data rate systems must adhere to the strict 29 dB optical loss budget that the existing fiber infrastructure imposes. Moreover, the burst-mode nature of upstream transmission involves data packets from different ONUs and with unique transmission paths arriving sequentially at the OLT receiver. This is a major challenge in OLT receiver design, since each packet's signal statistics and power levels can vary significantly, which gives rise to PON's challenging 19.5 dB dynamic range requirement.

Device bandwidth limitations and fiber dispersion are also bottlenecks to increased PON bandwidth capacity. The mass deployment of ONUs means high-volume, low cost devices must be used, which will result in the re-use of 25G class optics for 50 Gbit/s HS-PON [19]. Meanwhile, increasingly severe fiber dispersion impairments in IM/DD systems above 25 Gbit/s have already necessitated the switch from the C- to O-band transmission window in order to be closer to the zero-dispersion point and sustain the 20 km transmission reach [3].

2.1 High-speed optical signalling

Current PON standards, up to and including the latest 50 Gbit/s ITU-T HS-PON recommendations, are based on IM/DD technology, and specifically the non-

return-to-zero on-off keying modulation (NRZ-OOK, shortened to NRZ hereafter) [3]. This is a simple digital amplitude modulation format which uses two amplitude values to neatly encode the binary digits '0' and '1'. This can be achieved using an optical amplitude modulator switching between two distinct power levels P_0, P_1. An example of an NRZ data stream encoding binary information is shown in Figure 2.2 (a).

The strengths of NRZ include its noise tolerance, and implementation simplicity. Optical modulation is straightforward using low-cost directly modulated lasers (DML) or electro-absorption modulaters (EAM), while simple direct-detection methods can be used at the Rx. The sensitivity of avalanche photodiodes have readily sustained the large PON optical loss budget for NRZ up to 25 Gbit/s, and has meant optical amplification has not been needed in PON systems to date. However, NRZ is spectrally inefficient compared to more advanced modulation formats, and limited bandwidth PON devices will struggle to achieve beyond 50 Gbit/s line rates. The most recent 50 Gbit/s HS-PON standards are perhaps the last generation of PON which will use it [17].

There are several alternatives which have been widely proposed for ≥ 100 Gbit/s PON, including PAM4 and PAM8 [48], [49], quadrature phase-shift keying (QPSK) [50]–[52], and discrete multi-tone, also known as orthogonal frequency division multiplexing (OFDM) [53], [54]. Each of these can encode multiple bits per modulation symbol and are therefore more spectrally efficient than NRZ, however are typically more costly to implement.

The phase modulation of QPSK could offer significant increases in line-rate with the added benefit of high coherent receiver sensitivities to help meet and exceed the optical loss budget [55]. Dual-polarisation techniques could even further increase bit-rates. However, its adoption would represent a large increase in optical transceiver complexity and face significant challenges related to burst-mode reception at the OLT [56]. Similar issues exist for OFDM, where high peak-to-average-power ratios would further complicate transceiver design. Mass produced coherent pluggables targeting data center interconnects and long-haul communications could drive development of future 200 or 400 Gbit/s coherent PON [57], but limited research exists on their application in PON systems, where burst-mode transmission present unique challenges. As a result, it is unlikely that coherent PON will be realised in time for 100G PON.

The continued use of IM/DD technology in PON therefore appears the most cost-effective way to reach 100 Gbit/s line-rates. While no single format fully

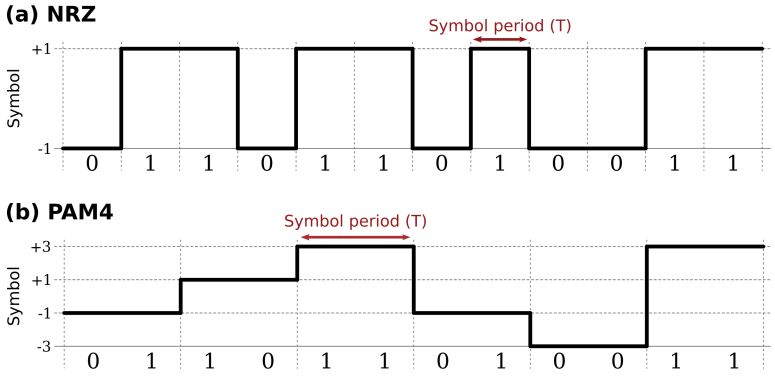


Figure 2.2: Comparison of (a) binary NRZ and (b) 4-level PAM4 modulation formats.

addresses the challenges posed by the PON optical loss budget and dynamic range requirements, multi-level pulse amplitude modulations such as PAM4 and PAM8 are attractive solutions when combined with advanced equalisation techniques. There is a strong case for PAM4 in particular, which has been instrumental in delivering 400G Ethernet standards [58] and for which a strong technological base exists. This thesis builds on recent work by proposing neural network-based equalisation as a solution for enabling spectrally efficient PAM4 with sufficient flexibility to adapt to real-world PON impairments.

2.1.1 Multi-level amplitude modulation

More spectrally efficient pulse amplitude modulation can be realised by using more than two amplitude levels to encode multiple bits per symbol. This allows for the the low-cost benefits of IM/DD technology to be retained and increases the bit rate while using the same bandwidth. In standard pulse amplitude modulation the number of symbols M required to encode N binary bits is given by:

$$M = 2^N$$

If two bits are to be encoded per symbol, i.e. $N = 2$, then a 4-level modulation is required and the resulting modulation format is called PAM4. Similarly PAM8 corresponds to $N = 3$. Figure 2.2 (b) shows that for half the same symbol rate, f_s, PAM4 achieves the same bit-rate as NRZ, while PAM8 requires just $f_s/3$. The bit rate f_b is calculated for a PAM-M signal using:

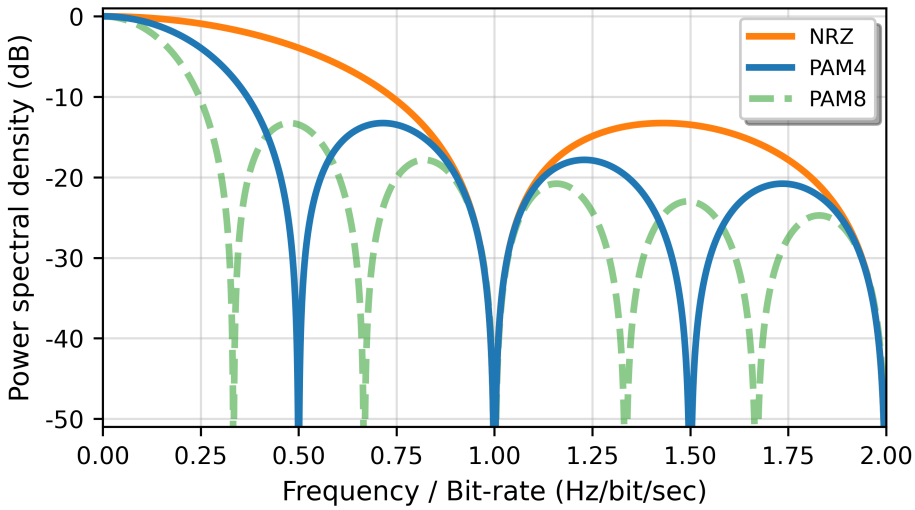


Figure 2.3: A comparison of the spectral efficiency of NRZ, PAM4, and PAM8.

$$f_b = f_s \cdot N = f_s \cdot \log_2 M. \tag{2.1}$$

Figure 2.3 illustrates the increased spectral efficiency of multi-level PAM modulations by plotting the power spectral density against that of NRZ. PAM4's main spectral lobe requires just half the spectrum bandwidth as NRZ to achieve the same f_b. Therefore switching to higher order PAM formats allows for data capacity to be increased in bandwidth limited scenarios, such as those expected in future 100G PON which may reuse low-cost 25G class optics. While the 50 Gbaud NRZ signalling of the latest HS-PON corresponds to 50 Gbit/s, a 50 Gbaud PAM4 signal would support 100 Gbit/s line-rates. It is worth noting however that the bit-rate grows logarithmically with M as per equation 2.1, while the gain in spectral efficiency decreases exponentially with M.

However, every modulator has a maximum achievable modulation amplitude, or depth, which preserves signal linearity without the introduction of distortions. Increasing the number of symbols in a PAM modulation means that more digital amplitude levels need to be squeezed into this limited modulation depth provided by the modulator device. Given a device which can achieve a maximum amplitude of A, the distance between symbol amplitude levels for a PAM-M signal can be expressed as:

$$\Delta = \frac{2A}{M-1} \tag{2.2}$$

This reduced distance between modulation symbols means that higher order PAM formats are more susceptible to noise, and induce serious Rx sensitivity penalties relative to NRZ which are discussed in the context of PON in the following sections.

2.1.2 Performance metrics

The signal-to-noise ratio (SNR) and bit error rate (BER) quantitative performance metrics are described below. Furthermore, eye diagrams can serve as helpful diagnostic tools to visualise impairments affecting an amplitude modulated signal.

Signal-to-noise ratio

Noise is unavoidable and effects digital communication signals by making it difficult to differentiate between signal symbols. The SNR performance metric quantifies the signal power relative to that of noise, and different modulation formats have higher SNR requirements to achieve the same transmission performance. SNR is typically expressed in decibels.

For a PAM4 signal in an additive white Gaussian noise (AWGN) channel, and with power levels P_i for $i \in \{0, 1, 2, 3\}$, the SNR can be calculated by dividing the average symbol power by the average noise powers using:

$$SNR = \frac{P_0 + P_1 + P_2 + P_3}{\sigma_0^2 + \sigma_1^2 + \sigma_2^2 + \sigma_3^2}, \tag{2.3}$$

where σ_i^2 is the noise variance of the i-th symbol. The SNR can be calculated similarly for any M-symbol PAM modulation format in an AWGN channel.

Bit error rate

The fundamental performance measure of a communication system is its ability to transmit information without making errors. The BER of a system quantifies this by measuring the number of errors n_e accumulated during transmission relative to the total number of bits transmitted n_b:

$$BER = \frac{n_e}{n_b}. \tag{2.4}$$

In this thesis, the counting method is used to estimate BER from transmitted

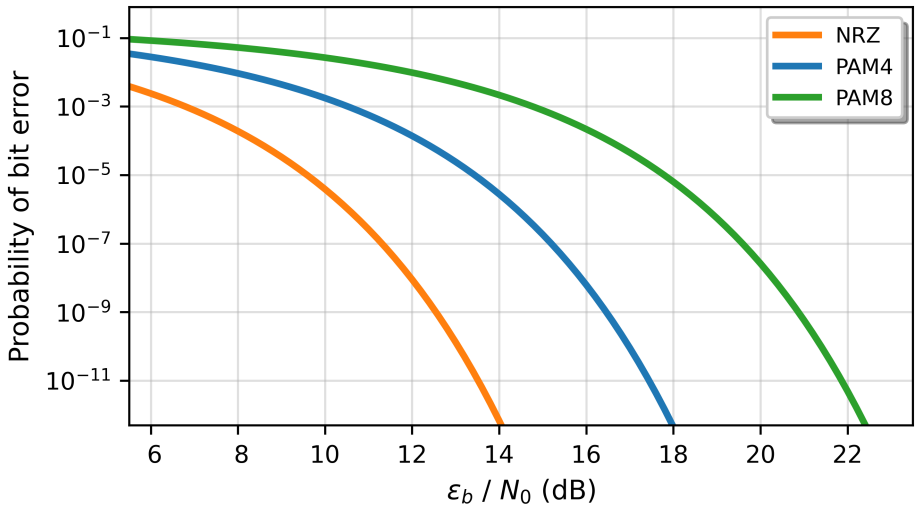


Figure 2.4: Plot of the error performances of NRZ, PAM4, and PAM8 for varying ϵ_b/N_0: clearly PAM4 requires more energy per bit to achieve the same error performance as NRZ.

bits in simulation and experiments: errors are counted for a sufficiently large number of bits, and the above equation is used. For higher order modulation formats where each symbol encodes multiple bits of information, a related quantity known as the symbol error rate (SER) can be calculated, from which the BER can then be found by decoding the incorrectly received symbols.

In order to highlight the higher SNR requirements of PAM4 and PAM8 relative to NRZ, it is possible to derive the SER performance of an M-level PAM format in an additive white gaussian noise (AWGN) channel using equations 2.2 and a general form of 2.3 [35]:

$$SER = \frac{M-1}{M} \operatorname{erfc}\left(\sqrt{\frac{3\log_2 M}{M^2-1} \cdot \frac{\epsilon_b}{N_0}}\right), \tag{2.5}$$

where ϵ_b is the energy per bit and N_0 is the noise power spectral density, i.e. the noise power per Hz bandwidth. The ratio ϵ_b/N_0 is a normalised form of SNR which does not take into account bandwidth limitations, and is therefore ideal for comparing modulation formats. The SNR can be related to this normalised value using:

$$SNR = \frac{\epsilon_b}{N_0} \cdot \frac{f_b}{B}, \tag{2.6}$$

where f_b is the channel bit rate as before, and B is the channel bandwidth.

Table 2.1: Gray encoding ensures that mismeasurement of neighbouring symbols leads to only a single bit error.

PAM4	Linear	Gray
0	00	00
1	01	01
2	10	11
3	11	10

Combined, the ratio of these two is referred to as the gross spectral efficiency of the channel, measured in (bits/s)/Hz.

Further, Gray encoding is a method for encoding binary bits to PAM symbols such that neighbouring symbols differ by only a single bit value [59]. This ensures minimal bit errors if a symbol is incorrectly received as one of its adjacent neighbouring symbols. Table 2.1 gives the Gray encoding for PAM4.

If Gray encoding is used and SNR is sufficiently high, it is possible to approximately relate the BER and SER values for a PAM signal [60]:

$$BER \approx \frac{1}{\log_2(M)} SER. \tag{2.7}$$

Figure 2.4 combines this equation with 2.5 and plots the dependence of BER on ϵ_b/N_0 for NRZ, PAM4, and PAM8. It is clear that higher order PAM formats require higher ϵ_b/N_0, and therefore SNR to achieve the same BER performance.

Therefore, the trade-off between higher data rates and increased SNR requirements must be considered carefully in the context of future high-speed PON. While PAM8 offers greater spectral efficiency and higher data rates than either of NRZ or PAM4 for the same bandwidth, it requires the highest SNR to achieve the same BER. For this reason, PAM4 appears the most suitable candidate. However, its degraded noise tolerance relative to that of NRZ is still a concern, and PAM4 will require optical amplification to overcome the strict PON optical loss budget.

Eye diagrams

Eye diagrams are a qualitative performance metric, which allow for signal integrity to be evaluated visually. They are used extensively throughout this thesis. An eye diagram is a statistical representation of a signal's overall quality which conveys information about the symbol transitions, eye heights and widths, as well

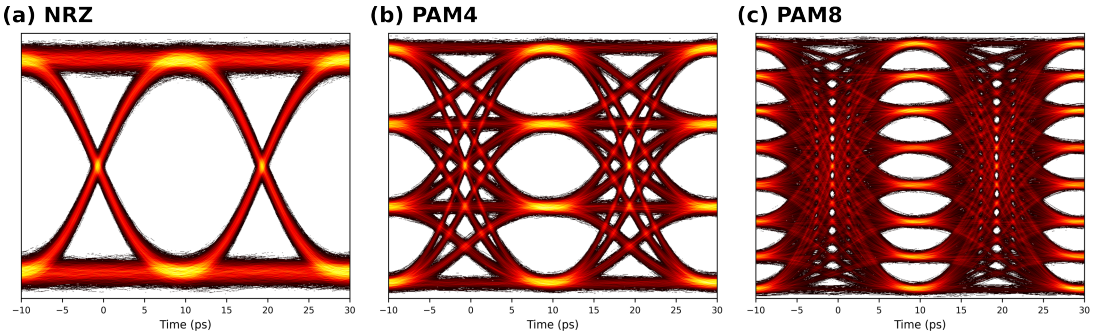


Figure 2.5: Example eye diagrams of (a) NRZ, (b) PAM4, and (c) PAM8.

as skew [61]. They are constructed by superimposing captured signal segments (typically 2 symbols in length) on top of each other, and plotting the result as a heatmap or density estimation graph.

Example eye diagrams can be seen in Figure 2.5. Note that in the PAM4 eye diagram there are 3 eye openings, and each of these has just $1/3$ the height of the single NRZ eye. If amplitude swing is constant, a rough calculation estimates an SNR loss of $20\log_{10}(1/3) \approx -9.5$ dB when switching from NRZ to PAM4.

Eye diagrams require long symbol sequences comprising an even distribution of modulation symbols and symbol transitions. Pseudo-random binary (PRBS) or quaternary (PRQS) sequences are ideal for this, and are discussed further in Section 2.3.4.

2.1.3 System requirements

Prospective 100 Gbit/s PAM4 signalling in future PON will need to adhere to the challenging minimum 29 dB optical loss budget, in addition to the 19.5 dB dynamic range requirement imposed by legacy PON infrastructure. These system metrics rely on BER measurements to determine the OLT receiver sensitivity and overload points in relation to the chosen forward error correction (FEC) threshold.

Forward error correction

An in-depth discussion of FEC requirements in future PON is beyond the scope of this thesis, and therefore only general points relevant to the choice of FEC used in this work are included below.

FEC methods introduce redundancy into a bit-stream at the transmitter, which allows the receiver to recover errors introduced during transmission. Each FEC

algorithm has an associated threshold, under which the FEC can ensure BER values of 10^{-12} or 10^{-15}, while FEC performance fails quite quickly above this threshold [35]. Hardware implementations of FEC can contribute significantly to overall DSP energy consumption [62], and reduce net data-rates by introducing redundant parity bits.

In the latest HS-PON standards, soft-decision FEC (SD-FEC) which has a threshold of 10^{-2} is proposed for 25 Gbit/s, and which involves a 20% bit-rate overhead, while the FEC requirements for 50 Gbit/s line-rate is still under consideration [12]. Meanwhile, the KP4 BER threshold of 10^{-4} is popular for PAM4 data center interconnects. The hard decision (HD) FEC threshold of 3.8×10^{-3} BER is used throughout this work, which has just 7% overhead. This FEC consumes less power than SD-FEC, and gives a balanced view on the potential performance of the neural network equalisers being considered.

Optical loss budget

A consequence of the branching structure of a PON ODN is that each OLT-ONU link can have a different optical path loss. This loss is largely determined by the number of passive splitters the optical signal needs to pass through, although fiber attenuation, connectors, and fiber splices also contribute loss. PON standards outline the minimum optical path loss that the optical signalling must be robust to, and the maximum loss that an optical signalling scheme can support is often referred to as its optical loss budget. A minimum optical loss budget of 29 dB is required in the HS-PON standards, and this must be strictly adhered to in order to guarantee service to all ONUs.

This is closely related to the employed receiver's sensitivity, which is defined as the minimum incident optical power for which a signal's measured BER is below the chosen FEC threshold. Once this is known, the optical loss budget for a given transmitter power can be estimated by simply subtracting the difference between the values. A Rx configuration involving a PIN photodiode integrated with an SOA preamplifier has been widely proposed to boost receiver sensitivity at the OLT [63], [64]. This is an attractive solution to meet the minimum PON loss budget for 100 Gbit/s PAM4 due to its degraded noise tolerance [65]. This is discussed further in Section 2.2.2.

While the dynamic range of such an SOA+PIN Rx is the primary focus of this thesis, the optical loss budget is also reported for completed experiments. This is estimated by assuming a mean optical launch power of $+8$ dBm, and subtracting

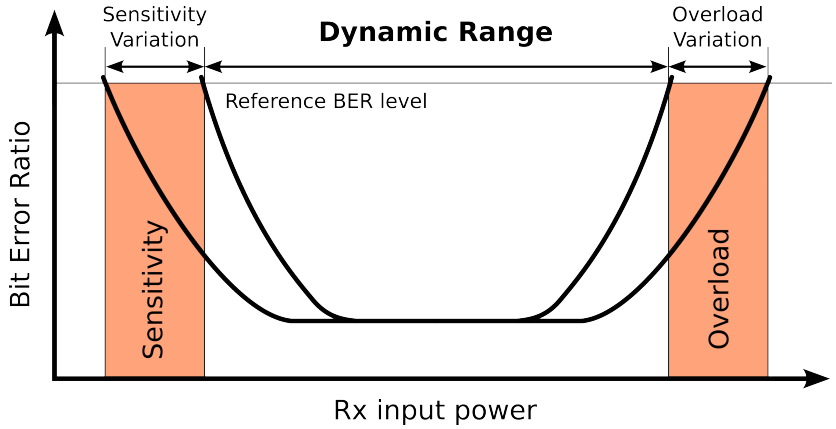


Figure 2.6: This graph illustrates how the operating dynamic range of a Rx is defined in terms of the Rx sensitivity and overload levels relative to a chosen BER threshold.

the measured Rx sensitivity achieved during experiments.

Dynamic range

There can be large power differences among the burst packets received by the OLT, due to the distinct optical paths each packet takes from its ONU, through the ODN. The dynamic range (DR) of an OLT Rx is therefore an important system metric, which specifies the difference between the lowest and highest packet powers that a Rx can tolerate while staying below the chosen BER threshold. Figure 2.6 shows how DR is defined in terms of Rx sensitivity and overload values.

In the 50 Gbit/s HS-PON standards, a minimum DR of 19.5 dB is outlined, meaning an OLT Rx must be capable of receiving packet powers $\approx 90\times$ greater than the lowest power packet. Typically burst-mode electronics are employed to deal with the packet power differences after optical to electrical conversion has taken place, and recent research has demonstrated linear burst-mode receivers (LBMRx) for 100 Gbit/s PAM4 [66], [67]. However, if an SOA preamplifier is used at the OLT Rx, then there is the potential for loud packets to cause SOA gain-saturation induced non-linear patterning effects, which can significantly impact the Rx overload point. Chapter 3 investigates such a scenario in detail.

2.2 Photonic system impairments

The focus so far has been on the spectral efficiency and noise tolerance of PAM4 signalling in relation to PON system requirements. In this section, the bandwidth limitations and noise mechanisms of relevant PON photonic device technology are discussed in more detail. Optical amplifiers are introduced for boosting OLT receiver sensitivity, and fiber dispersion is also introduced as a major limiting factor to scaling IM/DD PON technology.

2.2.1 Bandwidth and noise limitations

Photonic devices, also known as opto-electronic devices, are devices which generate, manipulate, or detect optical waves. Examples of photonic devices include laser diodes [68], semiconductor optical amplifiers [28], and photodiodes [1]. Every photonic device has a limited bandwidth, both electrical and optical, and active devices can contribute noise to a system.

Bandwidth limitations

When speaking of opto-electronic modulators and detectors, a device's bandwidth refers to the range of frequencies which it can support. This is determined from the devices frequency response, which quantifies the magnitude and phase of a device's output for a range of input frequencies [69]. The 3 dB bandwidth is reported as the point at which the magnitude of a device's frequency response drops by 3 dB.

At the time of writing, 25 GHz avalanche photodiodes are not yet produced in sufficient quantities to be cost-effective in next-generation PON, and 50 GHz devices are still under development. However, 25G class photodiode technology is more readily available, and could be integrated with SOA devices to support the required PON loss budget. Regardless, in future 100 Gbit/s PON, bandwidth constraints will be a factor, with low-bandwidth transmitter technology at the ONUs also a concern [48].

For a given bit-rate, PAM4's main frequency content occupies a smaller range of frequencies than that of NRZ (see Figure 2.3), and therefore PAM4 is less impacted by bandwidth limitations than NRZ. This can be seen in Figure 2.7, where the simulated 100 Gbaud NRZ signal is severely distorted in (a), while the 50 Gbaud PAM4 signal in (b) is less effected. Therefore, PAM4 is well-placed

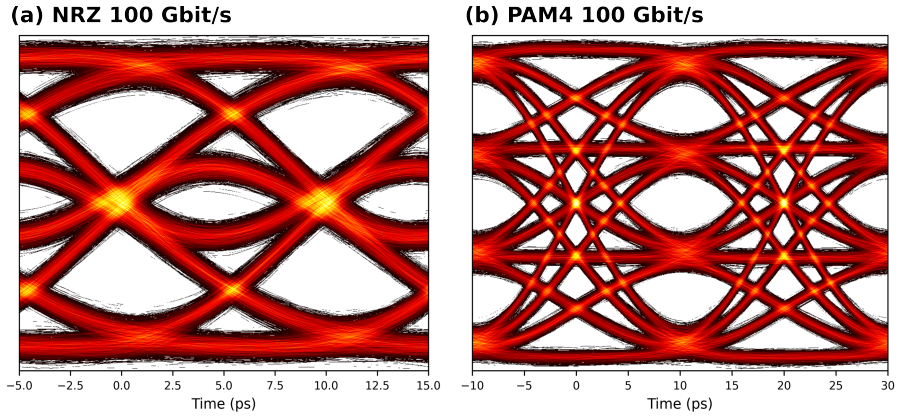


Figure 2.7: Simulated 100 Gbit/s (a) NRZ and (b) PAM4 eyes resulting from a bandwidth-limited 25 GHz Rx, implemented using a 4$^{\text{th}}$ order Bessel filter.

to enable higher bit-rates than NRZ using available technology, although signal equalisation may still be required to optimise signal transmission.

Noise mechanisms

Photodiodes convert optical signals into electrical ones. The photocurrent I generated by a photodiode is directly proportional to the incident optical power P_{in}:

$$I = R \cdot P_{in}, \tag{2.8}$$

where R is the photodiode responsivity, which is wavelength dependent and measured in A·W^{-1}. The photodiode, being a resistive element, will introduce thermal noise to the receiver whose power will depend on the temperature T and effective electrical bandwidth B_{eff} of the device. The thermal noise power σ_{th}^2 can be calculated using the equation:

$$\sigma_{th}^2 = \frac{4K_B T}{R_L} F_n B_{eff}, \tag{2.9}$$

where k_B is the Boltzmann constant, R_L is the value of the load resistor, and F_n is the amplifier noise figure [69]. It is standard practice for the noise equivalent power (NEP) of a photodiode to be stated on a device's datasheet, which is defined as the minimum optical input power per unit electrical bandwidth which will result in the output electrical signal having an SNR of 1 [69]. The thermal noise power for a photodiode can be simply calculated using:

$$\sigma_{th}^2 = R \cdot NEP \cdot \sqrt{B_{eff}}. \tag{2.10}$$

The above expression is used to simulate the noise characteristics of the photodiode device used in Section 2.4. A photodiode's output photocurrent can therefore be modelled as:

$$I = R \cdot P_{in} + \sigma_{th}^2. \tag{2.11}$$

In contrast, APDs exploit an internal gain mechanism to greatly boost their sensitivity [1].

2.2.2 Optical amplification

Optical amplifiers can improve Rx sensitvities and extend transmission distances by mitigating system power losses. As a result, they are fundamental components of modern optical communication systems. However, in the context of PON, amplification has not been required due to the relatively short transmission distances, and the high Rx sensitivities of systems using NRZ signalling in conjunction with APDs.

However, as discussed, APDs may not be an option for 100 Gbit/s PON, and bare photodiodes are unable to support the PON optical loss budget for PAM4. Instead, an alternative Rx design, which is based on integrated SOA and PIN photodiode devices, has been proposed. Such a design could achieve sufficiently high Rx sensitivity and bandwidth to support PAM4 signalling [65]. SOAs appear ideal, since they are readily integrable on chip with PIN photodiodes, can operate in both C- and O-bands, and are relatively low-cost. Alternative fiber-based amplifiers, such as the Erbium Doped Fiber Amplifier (EDFA) [70], do exist, and can offer greater optical gain and lower noise figures. However, these are bulky devices with large energy footprints, making them inappropriate for OLT deployment, and therefore SOAs are preferred.

Gain saturation

Optical amplification is achieved through the use of an active medium, which in the case of an SOA is a semiconductor junction which is electrically pumped to achieve a population inversion [28]. More details on the processes behind SOA amplification are given in Chapter 3. The gain G of an optical amplifier is

defined as the ratio between the output and input powers of the amplifier device, and is typically expressed in units of decibels (dB):

$$G_{dB} = 10\log_{10}(G) = 10\log_{10}\left(\frac{P_{out}}{P_{in}}\right). \tag{2.12}$$

However, every amplifier possesses limited energy with which it can amplify signals. Therefore, for input optical powers above a certain threshold, an amplifier's gain will decrease. This relation can be derived as [1]:

$$G = G_0 \exp\left[-\frac{G-1}{G}\frac{P_{out}}{P_S}\right]. \tag{2.13}$$

Here, G_0 is known as the small-signal gain, while G is the gain achieved to give the output power P_{out}. The input saturation power P_S is the input optical power for which the gain G is 3 dB lower than G_0. If $P_{in} > P_S$ then the amplifier is said to be operating in the non-linear gain-saturated regime.

For a fiber amplifier this is not an issue, as its gain recovery operates on the order of ms, and so does not noticeably effect signals modulated in the GHz range. However, the gain recovery action in SOAs is much faster, with recovery times ranging from a nanosecond to tens of picoseconds. For a high-speed optical signal traversing a saturated SOA, this can lead to non-linear self-gain modulation (SGM) effects, commonly known as patterning. Since PON systems require an OLT receiver to have a large DR, high-power packets could induce SOA gain saturation and therefore reduce the achievable DR of an SOA+PIN Rx. Extending the input power dynamic ranges of SOAs in such a scenario is the basis of Chapter 3, and is discussed further there.

Amplified spontaneous emission

Amplified spontaneous emission (ASE) noise is a fundamental side effect of optical amplification, and involves incoherent light being added to the amplified optical signal [1]. ASE is generated across the entire gain spectrum of an optical amplifier, due to it originating from the inverted active medium. This degrades the optical signal to noise ratio (OSNR) and can be detrimental to a systems performance. Since the gain spectrum of an SOA can be as wide as 50 nm, optical filters must be used to remove this out-of-band noise.

When an optical signal is detected by a photodiode and undergoes square-law

detection, both the signal and ASE terms will beat together and contribute to the output photocurrent. A signal with degraded OSNR will therefore see a significant ASE contribution to the final output:

$$I_{out} \propto \left|E_{sig} + E_{ASE}\right|^2 = \left|E_{sig}\right|^2 + 2\,\mathrm{Re}\{E_{sig} \cdot E_{ASE}^*\} + |E_{ASE}|^2. \tag{2.14}$$

Clearly the signal-noise beat term is linearly dependent on the signal power. A feature of this is that higher power modulated symbols will suffer greater ASE degradation, and this will be evident in eye diagrams in later sections.

2.2.3 Optical fiber channel effects

The optical fiber channel can induce linear and non-linear effects in optical communication systems. However, in this thesis the focus will be on chromatic dispersion and fiber losses, which are most relevant to PON.

Fiber propagation losses

Optical fiber is manufactured using ultra-pure fused silica glass, which can act as a waveguide for optical signals over a wide wavelength range [1], [71]. Signals in optical fiber experience attenuation at certain wavelengths due to absorption processes, which are related to the presence of silica and hydroxide molecules [71]. Rayleigh scattering is another source of losses, and is caused by microscopic fluctuations in the silica glass density leading to variations in the refractive index [1]. However, modern 'dry' fiber manufacturing technology minimises the presence of water vapour impurities, and can guarantee less than 0.2 dB·km^{-1} across a wide range of frequencies.

Over a 20 km ODN transmission path, fiber losses can amount to as little as 2 to 3 dB. While fiber bending losses also contribute, passive splitter technology is responsible for the majority of the PON optical loss budget[2]. The optical splitters employed in PON are usually based on planar lightwave circuits (or photonic integrated circuit) technology, which give high splitting efficiency while reducing reflections [73].

[2]An interesting topic which is outside the scope of this thesis is fiber splicing. Deployed fiber in an ODN can expect to be spliced many times over its lifetime due to accidental breaks, new equipment installation, etc. The reflections that result from the multiple splices and connectors can cause significant multi-path interference, which can degrade PAM4 performance in particular [72].

Chromatic dispersion

Chromatic dispersion (CD) is a major limiting factor to increasing PON bit-rates. Because the refractive index of an optical fiber is frequency dependent, the spectral components of an optical signal have different phase and group velocities as they travel through the fiber. This leads to pulse broadening (or compression), and the spectral components of an NRZ or PAM4 symbol will arrive at different times at the receiver, which is observed as ISI. However, the dispersive effects increase with the bandwidth of a signal, and therefore higher baud rates suffer from greater dispersion impairments. This is another major reason why spectrally inefficient NRZ is not favoured in the context of 100 Gbit/s PON. Further, while CD is approximately linear and effects only the phase of a signal, during square-law detection a signal's phase information is lost. Therefore, the CD impairment becomes a non-linear effect in a direct detection system which only recovers a signal's amplitude. For these reasons, non-linear signal equalisation techniques have been proposed to counter extreme dispersion effects in high-speed PAM4 systems [74]–[76].

The propagation constant $\beta(\omega)$ and its derivatives determine how the phase of a signal's spectral components evolve while travelling through an optical fiber. The most important of these for determining system performance limits is known as the Group Velocity Dispersion parameter:

$$\beta_2 = \frac{d^2\beta}{d\omega^2}. \tag{2.15}$$

In optical communications, this is typically expressed as a function of wavelength λ, rather than frequency. It is then known as the dispersion parameter $D\left(\lambda\right)$:

$$D = \frac{d^2\beta}{d\lambda^2} = -\frac{2\pi c}{\lambda^2}\beta_2. \tag{2.16}$$

This quantity is measured in ps/[nm·km], and is more commonly quoted in datasheets and technical documents. The temporal broadening ΔT of an optical pulse over a given distance L can then be determined using the equation:

$$\Delta T = D\left(\lambda\right) L \Delta\lambda, \tag{2.17}$$

where $\Delta\lambda$ is the spectral width of the pulse. From this, it follows that pulses with large spectral widths and / or which travel over long distances experience

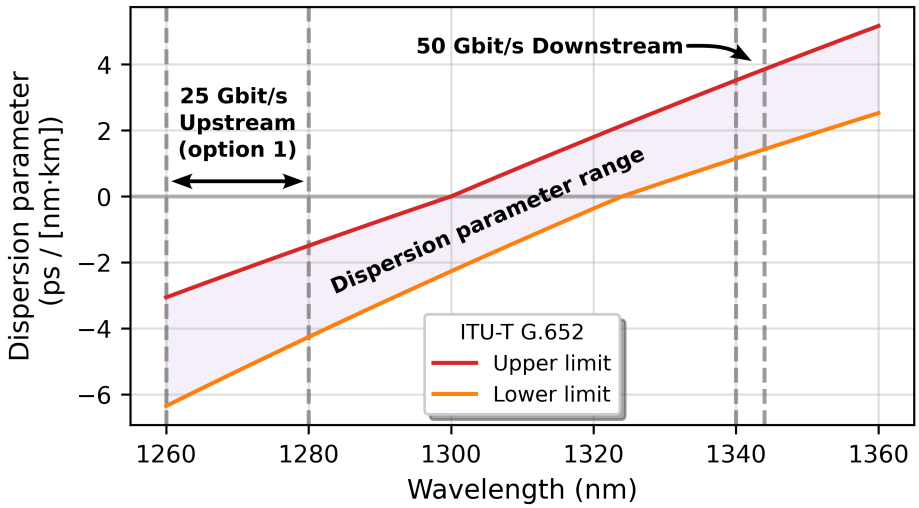


Figure 2.8: Typical dispersion profile of a commercially available fiber, generated using equation 6-1 in ITU-T G.652 [77], and parameters from Corning SMF-28+ product datasheet [78].

greater broadening effects, and thus ISI.

Figure 2.8 plots the dispersion parameter range in standard single mode fiber for the O-band, which is the wavelength range 1260 to 1360 nm. As can be seen, the zero-dispersion point occurs here, and as a result PON transmission windows use the O-band in order to minimise the effects of dispersion. The worst case dispersion scenarios in the 50 Gbit/s HS-PON standards occur for the 25 Gbit/s upstream, and 50 Gbit/s downstream transmission windows, as indicated in Figure 2.8.

However, even at the zero-dispersion point, dispersion cannot be eliminated completely. A signal's spectral content has a finite, non-zero bandwidth, and there will always be residual dispersion either side of the zero-dispersion point. Further, higher-order dispersive effects become relevant near this point, with the differential dispersion parameter S being the most important of these:

$$S = \frac{dD}{d\lambda} = \left(\frac{2\pi c}{\lambda^2}\right)^2 \beta_3 + \left(\frac{4\pi c}{\lambda^3}\right) \beta_2. \tag{2.18}$$

Figure 2.9 shows the simulated effects of accumulated dispersion for up to 30 km on 100 Gbit/s PAM4 signals. The resulting ISI is severe, even after just 20 km transmission, and will significantly limit system performance at these baud rates.

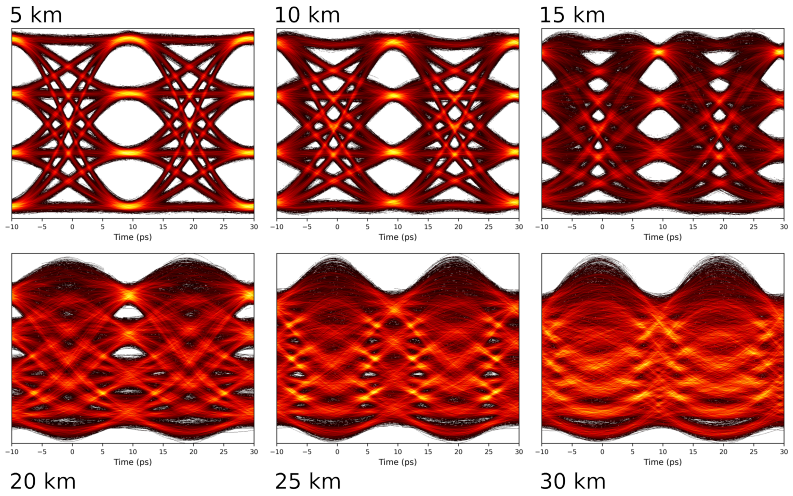


Figure 2.9: Eye diagrams show the effect of dispersion on a simulated 100 Gbit/s PAM4 signal at 1344 nm and using D and S parameters taken from Fig. 2.8.

Conventional equalisation techniques such as FFE and DFE have been proposed to counter this [79], [80], and in Section 2.4.2 their performance is compared with that of neural network techniques. The eye diagrams in Figure 2.9 were generated using the non-linear Schrödinger equation and the split-step Fourier method [1], and Section 2.4.2 includes more details of these simulations.

2.3 Signal equalisation and neural networks

Signal equalisation is a process used to mitigate the impact of ISI and distortion on a signal travelling through a communication channel. The goal of equalisation is for a received signal to match as closely as possible the transmitted signal, and it can be carried out at either the Tx (pre-distortion, pre-emphasis), the Rx (post-equalisation), or both [35].

Equalisation is set to become an important feature of future PON systems, with the latest HS-PON standards the first to endorse its use [19]. Linear feed forward equalisation (FFE) can compensate bandwidth limitations, and can be combined with decision feedback equalisation (DFE) [35] to mitigate chromatic dispersion. However, future 100 Gbit/s PAM4 systems will be more susceptible to non-linear signal distortions, such as SOA patterning effects, than traditional NRZ signalling. As a result, more advanced equalisers are being actively researched, including those based on neural network algorithms [34].

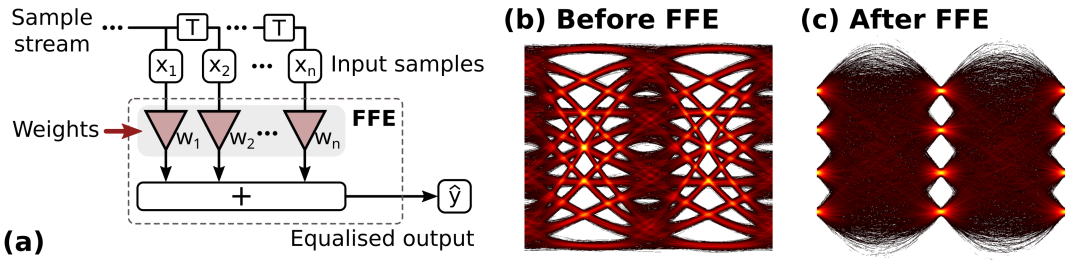


Figure 2.10: (a) The structure of an FFE; (b) shows a simulated 100 Gbit/s PAM4 eye after a 10 GHz 4$^{\text{th}}$ order Bessel filter while (c) shows the same eye after equalisation using an FFE-11T.

2.3.1 Conventional equalisation

The input to a digital equaliser[3] is a sequence of captured signal samples, typically originating from an analog-to-digital converter (ADC), and denoted here as the vector x. An equaliser model, f, aims to return an equalised sample $\hat{y}$ which approximates closely the transmitted sample y, and is free of ISI and distortion.

In this thesis, a training (or test) sample pair refers to a vector of equaliser input samples and their associated equaliser output: (x, y). Furthermore, a training (or test) dataset refers to a collection of many such sample pairs.[4]

Feed forward equalisation

The linear FFE technique can increase signal quality at the receiver. It achieves this by boosting the high-frequency components of a signal, leading to increased eye width and amplitude, and so is ideal for overcoming linear bandwidth-impairments, as well as limited chromatic dispersion effects. One of the disadvantages of an FFE, is that high-frequency noise is also amplified.

An FFE calculates the weighted sum of a series of signal samples x with a set of tap weights, as shown in Figure 2.10 (a). It is described by the equation:

$$\hat{y} = w \cdot x = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n, \tag{2.19}$$

where w is the vector of FFE weight parameters, and $\hat{y}$ is the equalised sample. Its output can be fed to a decision circuit to recover the transmitted symbol.

[3]Note that FFE techniques can be implemented using analog, or mixed signal circuits, however the focus here is on digital implementations.

[4]The notation used here is not the traditional one used in DSP, but instead follows that more generally associated with machine learning.

In this thesis, the time delay between the input samples fed to an FFE and other equalisers, is chosen to be 1 unit interval (UI), also referred to as a symbol period (T). This means that the input sampling rate equals the modulation symbol rate. Such an equaliser is referred to as a symbol-spaced FFE, and a 5 tap symbol-spaced FFE is denoted FFE-5T. Fractionally-spaced FFEs which use multiple samples per symbol are less susceptible to timing errors (jitter) [81]. However this increases the required sampling rate of the Rx at significant expense, while modern clock and data recovery circuits which can perform the same are becoming more commonplace [82], [83]. Therefore, fractional sampling is not considered further in this thesis.

Least mean squares and gradient descent

An FFE can be realised using least mean squares algorithms (LMS), which are also known as adaptive filters. Gradient descent can train the FFE weights using a known sequence of training sample pairs $\{(x_i, y_i)\}_{i=1}^{n}$, called a training dataset.

Optimisation is directed by a cost function $\mathcal{C}$, which quantifies the FFE performance based on its predictions $\hat{y} = f(x)$ and the actual target values y. The LMS algorithm derives its name from the mean square error (MSE) cost function:

$$\mathcal{C}(\hat{y}, y) = \text{MSE}(\hat{y}, y) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \,.^{5} \tag{2.20}$$

The above cost function defines an error surface in a $(k+1)$-dimensional space, where $k = \dim(w)$, i.e. the number of FFE taps. The goal of the LMS algorithm is to find w such that $\text{MSE}(\hat{y}, y)$ is minimised, i.e. at the minimum point of the error surface.

This is achieved using the gradient descent method, which iteratively updates the FFE weights using the gradient of the cost function with respect to the FFE weights:

$$\nabla_w \mathcal{C} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)\, x_i. \tag{2.21}$$

Note that the i term denotes the training sample index, and the gradient is k-dimensional, as is w. Once the gradient has been estimated, the weights can be

[5]The factor of 2 is a subjective choice, and will cancel in the derivative: it does not effect equaliser optimisation.

adapted in the opposite direction to the gradient, and toward the cost function minimum, using the following update equation:

$$w_{t+1} \coloneqq w_t - \alpha \nabla_w \mathcal{C}, \tag{2.22}$$

where the t index specifies the sequence of parameter updates, and α is an important hyperparameter known as the learning rate. This learning rate greatly effects the convergence properties of an FFE, and must be chosen carefully in order to guarantee convergence at an optimal rate [35], [81].

Burst-packet adaptive training

Note that the cost function in equation 2.20 is calculated as an average over all n samples in the training dataset for each parameter update. However, this is impractical in the context of upstream PON transmission, where channel conditions can vary rapidly from packet to packet, and equalisers must either adapt or be robust to a wide range of impairments [48], [84].

Therefore, proposed adaptive equalisers use an alternative, more viable approach to FFE optimisation known as stochastic gradient descent (SGD), where a single training sample $(x_i,\ y_i)$ is used per parameter update [85]. The sum over training samples in the cost function in equation 2.20 is therefore dropped, and the cost is no longer the MSE, but is effectively a quadratic cost function. A natural consequence of this is the training process becomes more noisy. But more importantly, SGD is implementable in hardware at the speeds necessary to realise an adaptive FFE within extremely short PON packet preambles of ≈ 150 ns [12]. This requirement is discussed in more detail, and with respect to NNEs, in Chapter 5.

2.3.2 Neural network theory

Neural networks (NN) are a collection of mathematical models which are highly effective at solving problems which can be intractable using analytical techniques, such as in computer vision [86] and natural language processing [87] tasks. They consist of non-linear processing units, known as neurons, arranged into layers to form a network structure, as shown in Figure 2.11 (a). Layer inputs can be raw data, or the processed output of a previous layer. The performance of an NN is effected by the number of layers and neurons it has, in addition to the

types of neurons used (e.g. recurrent, convolution neurons), while novel network structures can also contribute. Deep learning refers to the study of NNs which have many layers [88].

Neural network equalisation (NNE) should use small, specially designed NNs which are computationally efficient, meaning they can be implemented in existing digital hardware platforms, such as application specific integrated circuits (ASICs) or field programmable gate arrays (FPGAs). The benefits of the non-linear processing of an NNE should outweigh their associated implementation cost.

Below, a simple neural network architecture is introduced, along with its training process, before more sophisticated structures are discussed for equalisation.

Feedforward architectures

One of the most simple, general forms which neural networks take, are known as fully-connected (FC) feedforward neural networks [89]. These networks are made up of neurons which take a real-valued vector as input, and output a single value, as shown in Figure 2.11 (b). A basic neuron calculates the dot product of the input vector x with the neuron's weight parameters w, and adds a bias parameter b. The result of this calculation is then fed into a non-linear activation function, φ:

$$z = w \cdot x + b, \tag{2.23}$$

$$a = \varphi(z). \tag{2.24}$$

Note that this basic neuron can be seen as an extension of the operation of an FFE model, through the addition of a non-linear activation function and bias term (compare Figure 2.10 (a) with Figure 2.11 (b)).

Network layers are described using the same equations, and they differ from single neurons only in the dimensions of their output and parameter objects. For a network layer consisting of multiple neurons, w is a matrix of weight parameters constructed by stacking its neurons weight vectors and bias terms on top of each other. As a result, the dot product $w \cdot x$ becomes a matrix multiplication. In general terms, a network layer with n_d neurons, which takes n_i real-valued inputs, will have a weight matrix $w \in \mathbb{R}^{n_i \times n_d}$ and bias term $b \in \mathbb{R}^{n_d}$, and the output activation

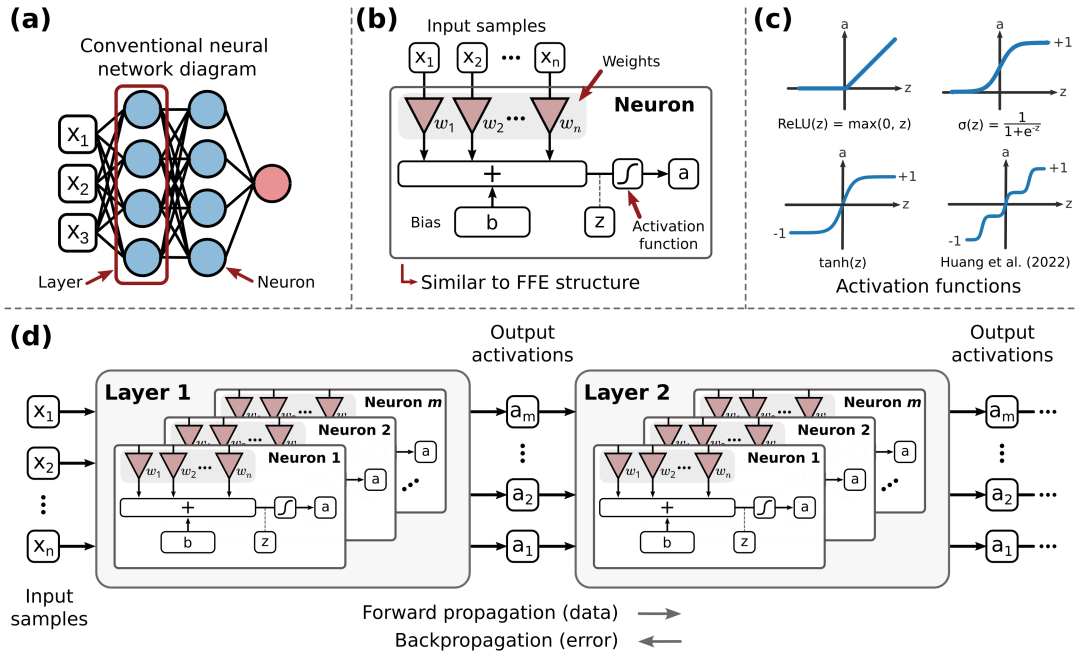


Figure 2.11: Schematics showing (a) conventional NN structure, (b) neuron operation, (c) popular activation functions, and (d) the composition of an NNE.

will be a vector, $a \in \mathbb{R}^{n_d}$. Data passes through the layers of a neural network sequentially, with the input layer being the first layer, where raw data is fed into the network, and the NN output being the output of its final layer. Intermediate layers are known as hidden layers (HL).

The choice of activation function is key to how multi-layer networks operate: if φ is linear, then a multi-layer NN will reduce to the equivalent of a single linear layer, since it would simply be a sequence of linear matrix multiplications. However, if the activation function is non-linear, a multi-layer NN has powerful function approximation properties. The most popular activation functions include the sigmoid, tanh, and Rectified Linear Unit (ReLU) functions shown in Figure 2.11 (c). The development of activation functions tracks closely innovations in NN architectures, and each function is designed to overcome specific challenges. While the sigmoid function was a popular choice in early NNs, its range is bounded between 0 and $+1$. The $\tanh$ function can output negative values between -1 and $+1$, while ReLU is designed to be less computationally complex, and is widely used in CNN models designed to compete in the ImageNet competition [90], [91].

Innovation in activation functions has continued within optical communications, and the final activation function in Figure 2.11 (c) was proposed by the authors

of [92] to optimise PAM4 NNE models. Further, the cross-entropy activation function [89] has been proposed to carry out joint NNE and symbol decision operations [93], but has high implementation complexity. In this thesis, proposed NNEs use linear output activations, which allows for a continuous, unbounded range for output samples from an NNE, which is equivalent to an FFE's output format.

Thanks to their structure, NNs are universal approximaters [41], [42], meaning they can model a wide range of non-linear functions, and potential equalisation tasks. However, training NNs to perform a specific task can be difficult, and there is no guaranteed method to achieve good NNE performance.

Gradient descent and backpropagation

The most popular algorithms used to train NNs and NNEs are based on extensions of the gradient descent algorithm discussed in the context of FFEs in Section 2.3.1. The backpropagation algorithm allows for error signals from a cost function to be propagated backwards through the layers of a neural network, and was famously popularised in [94], having previously been proposed elsewhere [95]. It achieves this by adapting the chain rule of differentiation [89].

As in FFE training, an error signal is generated using an appropriate cost function, $\mathcal{C}$, which is again chosen to be the MSE calculated over the entire training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, as described by:

$$\mathcal{C}\left(y, \hat{y}\right) = \text{MSE}\left(y, \hat{y}\right) = \frac{1}{2n} \sum_{(x,y)\in\mathcal{D}} \|y - \hat{y}\|^2. \tag{2.25}$$

Note that this definition of the MSE differs slightly from that of equation 2.20: here the vector norm is used, since $\hat{y}$ and y can be vectors. In order to train an NN using gradient descent, the gradients of $\mathcal{C}$ with respect to each of the network's parameters must be calculated:

$$\frac{\partial \mathcal{C}}{\partial w}, \frac{\partial \mathcal{C}}{\partial b}. \tag{2.26}$$

However, for w or b located in HLs, backpropagation must be used to calculate these derivatives. First, define the error of neuron j in layer l as:

$$\delta_j^l \equiv \frac{\partial \mathcal{C}}{\partial z_j^l}, \tag{2.27}$$

where z is as defined in equation 2.24. Therefore, the error in the output layer of an L layer network can be described by:

$$\delta_j^L = \frac{\partial \mathcal{C}}{\partial a_j^L} \frac{\partial a_j^L}{\partial z_j^L} = \frac{\partial \mathcal{C}}{\partial a_j^L} \varphi' \left(z_j^L \right). \tag{2.28}$$

The chain rule can now be applied to calculate the layer error for any intermediate HL, l:

$$\delta^l = \left(\left(w^{l+1} \right)^T \delta^{l+1} \right) \odot \varphi' \left(z^l \right), \tag{2.29}$$

where $\odot$ denotes the Hadamard product, or element-wise multiplication. It is clear that by combining equation 2.28 and 2.29 the error δ^l for any layer l in a neural network can be calculated.

Finally, it is then possible to calculate the partial derivative of the cost function with respect to individual network weights, and bias terms, using the layer error term as defined in equation 2.27:

$$\frac{\partial \mathcal{C}}{\partial b_j^l} = \frac{\partial \mathcal{C}}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} = \delta_j^l, \tag{2.30}$$

$$\frac{\partial \mathcal{C}}{\partial w_{jk}^l} = \frac{\partial \mathcal{C}}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l. \tag{2.31}$$

Together, the equations 2.28 to 2.31 describe the backpropagation algorithm, and allow for each network parameter to be updated via gradient descent. It is clear that any activation function φ must be differentiable for backpropagation to work.

These equations reveal the additional complexity involved in training NNEs compared to the adaptive LMS training of FFEs. In particular, the activations a^l and summed weights z^l need to be cached during the forward pass in order to subsequently calculate the layer error terms δ^l during the backward pass. This would require large amounts of memory and computational resources to implement in

real-time at the throughput speeds necessary for 100 Gbit/s PAM4, and it is not obvious that this is viable, or recommended. The ramifications of these training procedures for the application of NNE in upstream PON are explored in further detail in Chapter 5.

Mini-batch training and Adam optimisation

As discussed, NN structures are significantly more complex than FFE models, and this is further reflected in the optimisation strategies used to train them.

The large number of parameters in NNs mean that they are more susceptible to noisy training, and therefore the SGD described for FFE training in Section 2.3.1 is inadequate. In order to counter this, mini-batch training[6] is used, where parameter updates are calculated using subsets of the training dataset, usually some power of 2, such as $32, 64, 128, 256$, etc. Further, NN training is carried out for multiple epochs, where each epoch is a complete iteration of the entire training dataset. The order in which training samples are processed is randomised, or shuffled, between epochs, in order to guarantee diversity in the calculated error gradients and therefore optimal convergence.

The fixed learning rate in equation 2.22 is also an obstacle to optimal NN convergence, and a range of more advanced algorithms which extend gradient descent have been developed, such as Adam [96], RMSProp [97], and AdaGrad [98]. Adam stands for Adaptive Moment Estimation algorithm, and is the standard optimisation algorithm for NNs. It has been shown to reliably converge for a wide range of optimisation problems, which it achieves by computing individual learning rates for each NN parameter, similar to AdaGrad, and combines this with the adaptive learning rate technique of RMSProp.

During optimisation, Adam tracks the exponential moving averages of the gradient g_t and squared gradient g_t^2 values, denoted at time step t by $m_t,\ v_t$ respectively:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t, \tag{2.32}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2, \tag{2.33}$$

where the $\beta_1,\ \beta_2$ terms are the exponential decay parameters, with default values

[6]Batch training is where every parameter update is calculated using the entire training dataset, and is extremely computationally expensive; mini-batch training is a compromise solution between SGD and batch training.

of 0.9 and 0.999. These moving averages are estimates of the mean, and the uncentered variance of the gradient. Since these are initialised at 0, a further step is introduced to counter any initialisation bias towards 0 in the early time steps:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \tag{2.34}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}. \tag{2.35}$$

The bias-corrected estimates are therefore $\hat{m}_t$ and $\hat{v}_t$. The parameter updates are then calculated using these gradient estimates and the equation below:

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{2.36}$$

where θ can be any network weight or bias parameter, α is the initial learning rate. Furthermore, the variable ϵ is introduced to avoid division by 0, and has a default value of 10^{-8}.

Interestingly, Adam has been used to train more conventional equalisers, including FFEs in PAM-based optical interconnects, and the resulting equalisers have shown increased performance over those using standard LMS techniques [99]. However, while Adam remains the *de facto* optimisation algorithm for NNEs developed for optical communications, these NN optimisation processes involving mini-batch training, backpropagation, and Adam are overly complex for training equalisers within strict PON packet preamble times. Given that such solutions would need to be implemented on FPGA or ASIC platforms with limited memory and compute resources, and within the strict cost-benefit margins of PON, this raises a serious question over the viability of adaptive NNE for PON, which is addressed in Chapter 5.

2.3.3 State-of-the-art overview

The potential non-linear modelling power of NNEs is clear, but identifying appropriate optical network scenarios which can benefit from this is an active topic of research. Further, the multitude of algorithms and methods available from machine learning must be evaluated in the context of these optical transmission

scenarios, while taking into consideration their eventual deployment to high-speed, low-power hardware.

High-speed IM/DD technology

The increased susceptibility of advanced PAM4 and PAM8 modulation formats to non-linear effects has been a primary motivator for NNE research in next generation DCI and PON systems [34].

Linear fiber dispersion effects become non-linear distortions in IM/DD systems due to square-law detection, as discussed in Section 2.2.3. The resulting signal impairments are the primary limitation to increasing the bit-rate and transmission of IM/DD technology in short-reach systems, such as PON. As such, NNE has been widely investigated for compensating fiber dispersion [100]–[103]. In [104], 212 Gbit/s PAM4 transmission was demonstrated over 1 km in C-band using an NNE based on long short-term memory units (LSTMs), with notable benefits over conventional equalisers. However, this can be seen as an extreme dispersion case, and it is not clear from the literature whether NNEs would offer significant performance gains that justify their cost for 100 Gbit/s PAM4 over just 20 km, while using O-band devices. Section 2.4.2 investigates through simulations the potential benefits of NNEs for fiber dispersion compensation in the PON scenario.

Fiber non-linearities such as the Kerr effect, and four wave mixing, are unlikely to effect future PON systems thanks to their restricted distance and launch power combinations: the 50 Gbit/s HS-PON recommendations outline a maximum ONU launch power of $+9$ dBm [12]. However, high launch powers have been investigated in the context of PAM4 IM/DD systems and in conjunction with NNE solutions in [101], where the authors investigated launch powers up to $+19$ dBm.

A more relevant application of NNEs in future PON systems is the compensation of non-linear transceiver effects. Here, NNE solutions could enable lower-cost modulators, such as EAMs and DMLs with non-ideal linearity characteristics, to support advanced modulation formats [102], [105], [106]. In particular, these NNE solutions could counter impairments caused by the interaction between Tx-induced chirp, and fiber dispersion. Low-bandwidth, 10G class DMLs have also been shown to support up to 54 Gbit/s PAM4 and PAM8 transmission with the help of NNEs [107]. Further, non-linear effects could perhaps be exploited for system-level benefits if an NNE can ensure the recovery of transmitted data.

Chapter 4 investigates this concept further, in the context of the non-linear SOA patterning effect.

The increased attention on SOAs to boost Rx sensitivities has come with the caveat of potential gain-saturation induced non-linear effects. A range of equalisation solutions have been proposed, including those based on NNEs [108], [109]. However, these analyses are limited in their investigation of the SOA gain-saturation regime, and 'loud' burst-packets may be expected to induce more severe patterning effects than those considered. Chapter 3 investigates the patterning effect more closely for high-power PAM4 signals, and an NNE solution is presented which can completely recover patterned data streams.

Algorithmic developments

A standardised approach to designing NNE solutions has yet to be proposed within the optical communications community, and this is reflected in the multitude of published algorithms to date.

Feedforward neural network architectures have been widely proposed [110], and have the benefit of being readily pipelined when implemented on hardware [93]. Included in this category are CNNs, whose convolutional layers have been implemented as NNE input layers in order to reduce the number of NN parameters needed to process large numbers of equaliser taps [103], [111]. However, RNNs can also achieve this by using a feedback mechanism to reduce overall tap numbers. They appear a more natural choice for equalisation than CNN methods, since they avoid the complex convolution computation. Advanced RNNs using gated recurrent units (GRU) and LSTM units have been proposed as NNEs for IM/DD systems [104], and the NNE solution in [109] in particular uses RNNs to mitigate SOA non-linearities. However, while RNN algorithms are highly effective equalisers, their feedback mechanisms are a challenge to implement at high-speeds in hardware, as discussed in the next section.

Alternative paradigms such as end-to-end learning using NN-based autoencoder techniques are also being pursued, which use pre- and post- NNEs which are jointly optimised [112]. However, this relies on the optical channel being known sufficiently well to backpropagate error signals from the Rx to the Tx, and as such, the multi-point to point architecture of PON would appear a serious obstacle. However an initial attempt to to adapt this technique for coherent PON is described in [113].

When choosing an activation function, the majority of NNE papers use the standard ReLU or $\tanh$ activations. A primary concern is the cost of implementing these functions in hardware, making ReLU a popular solution for NNEs. This is discussed further in [93], where the authors describe an activation function specifically designed for PAM signals, and which is represented in Figure 2.11 (c).

However, more urgent in the context of PON is reconciling complex NNE training algorithms with the dynamic burst-packet conditions at the PON OLT. NNE solutions will be costly to implement relative to conventional equalisation techniques, and it would be most cost-effective to deploy NNEs at the OLT, where their benefits can be shared amongst all ONUs. Universal style NNEs which learn all possible combinations of ONU channel impairments have been investigated in [114], but they require large neural network models, are difficult to train, and fail to equalise impairments outside their training data. Adaptive FFE/DFE equalisation techniques do exist for dealing with burst-mode upstream transmission in PON [84], [115], [116], but conventional NNE training is orders of magnitude more complex. This is due to them having large dataset requirements, and using computationally intensive optimisation algorithms (Adam, backpropagation), which are not viable for online packet-by-packet training of an NNE in upstream PON transmission. Transfer learning techniques, which employ pre-trained NNEs, could reduce training requirements on similar channel impairments, as has been proposed for point-to-point IM/DD systems in [117]. However, existing transfer learning techniques simply reduce the required number of training epochs, while retaining the use of complex training algorithms, and are therefore not currently suited for PON. Chapter 5 discusses these topics in more detail, and describes an original adaptive NNE solution developed by the author, named SkipNet, which can achieve realistic adaptive NNE for burst-mode reception in PON.

Hardware and complexity challenges

The viability, and implementation cost of NNEs in hardware have also been topics of interest [118], and there have been several FPGA prototypes demonstrating NNEs in real-time systems [93], [119], including those targeting PON systems [120].

The costs in terms of hardware resources and energy consumption are key considerations when deploying NNEs in hardware, and there exists extensive literature comparing NNE complexity with that of conventional equalisation [121]. When

determining the overall complexity of an NNE, the numbers of quantisation bits required to represent its parameters and sample data, are key factors. If an NNE's parameters can be quantised without loss of model accuracy, then the total hardware footprint of the model can be minimised. This is important, since in order to achieve the required throughput for high-speed signals (e.g. 50 GHz for 100 Gbit/s PAM4), parallelisation and pipelining techniques must be used to compensate the limited DSP clock frequencies achievable [122]. In [123], the authors carry out a detailed investigation into quantisation for a 50 Gbit/s PAM4 NNE, and find that memory savings of up to 80 % can be realised by quantising a network, so that it relies on integer, rather than floating point arithmetic operations. The authors of [124] implement the pruning method from deep learning, which reduces the number of NNE parameters by removing those which are below a certain threshold value. They conclude that this is an effective way to lower overall NNE footprint. In this thesis, NNE size is minimised by using as few hidden layers and neurons as possible, while retaining optimal performance, however pruning is not implemented.

The feedback mechanism in RNNs is a recognised challenge and obstacle to hardware implementations of NNEs, since it cannot be easily parallelised in hardware, as explained in [93]. In order to counter this, multi-symbol output methods have been investigated in [125], [126]. This technique relies on equalising multiple symbol values simultaneously, by increasing the number of NNE outputs. As a result, the feedback mechanism need not operate at the throughput rate for the equaliser, but can be implemented at a fractional rate. This technique is implemented in Chapter 4 for GRU and LSTM NNEs to determine the tradeoffs between model size and the number of symbol outputs.

Although outside the scope of this thesis, it is worth mentioning that significant efforts are being made in the area of photonic neural networks, or neuromorphic photonics [127], [128]. If an NNE could be implemented as part of a photonic chip in the optical domain, it could avoid the energy costs and processing limitations of current digital platforms [128]. Spiking neural networks are also a novel technology that aims to achieve similar savings through analog computing [129], [130].

2.3.4 Methodology for signal equalisation

This section introduces the methodology used to investigate NNE throughout this thesis, including dataset generation, training and testing symbol sequences,

as well as the practical implementation of these in experiments.

Dataset generation and packet structures

The results reported in this thesis are generated from physical experiments, which are detailed in the experimental setup sections of Chapters 3, 4, and 5. Each of these experiments were conducted in continuous mode, since an appropriate burst-mode TIA was not available. However transmitted symbol sequences were still arranged into packets, and these formed the basis of the NNE training and performance analysis common to all experiments described throughout this work.

The limited memory of the Arbitrary Waveform Generator used, meant that just 131072 symbols could be loaded for transmission at any given time. The loaded sequence is referred to as a packet, and is divided between training, validation, and test symbol sequences, and further incorporated a guard band of 20 ns. During experiments, the AWG would transmit continuously by repeatedly looping over this packet structure. Large test datasets could be assembled by combining multiple acquisitions of this packet structure. An advantage of this was that the NNE training datasets were guaranteed to have the same signal statistics as the test datasets, having been transmitted together. Furthermore, this packet structure mirrored the burst-packet structures utilised for adaptive equalisation, where known training symbols are included in the packet preamble for the purposes of equaliser training.

Pseudo-random symbol sequences and overfitting

Pseudo-random binary sequences (PRBS) are traditionally used to evaluate the performance of an optical communications link, as they appear statistically random but can be generated efficiently in hardware using deterministic linear feedback shift registers [131]. Further, PRBS have near-zero DC offset, and provide uniform coverage of a wide range of binary sequence permutations that a communication channel would be expected to support. Pseudo-random quaternary sequences (PRQS) are an extension of PRBS to 4-level modulation systems, such as PAM4 [132], and have the same advantages. The length of a PRBS symbol sequence depends on the size of its state register: a register of size n can give rise to a sequence of $2^n - 1$ symbols before it starts to repeat.

It has been widely reported that training and testing an NNE on the same PRBS / PRQS pattern can lead to overestimating equaliser performance [133], [134]. This is encountered when the input taps of an NNE exceeds the register state

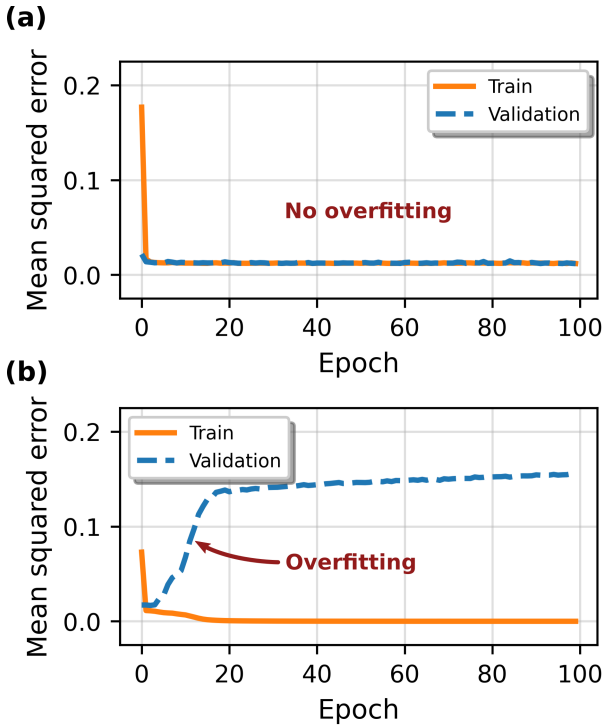


Figure 2.12: Example NNE learning curves showing (a) no overfitting, and (b) overfitting, where the model performance on the validation dataset diverges from that of the training dataset.

length, allowing it to learn the PRBS generation rule. To avoid such overfitting, random test patterns are used in Chapters 2 and 5, while in Chapter 3 and 4 a PRQS-14 training pattern is used in conjunction with limited NNE input taps, and a PRQS-15 test dataset ensure overfitting cannot occur.

Furthermore, overfitting can be detected using a validation dataset with the same characteristics as the training set. Validation sets are used to monitor a model's training, and to tune any optimisation hyperparameters [85]. A common symptom of overfitting is when a model's performance on a validation set stalls, or starts to degrade, while its performance on the training dataset continues to increase. In Figure 2.12 (a), the model performance on the validation dataset matches that of the training dataset, meaning the model is learning the targeted impairment equalisation. However, in (b) the validation MSE increases rapidly, while the training MSE converges to close to zero. This is a clear indication that overfitting has occurred. In this thesis, PRQS13 is used as a validation dataset for the purposes of monitoring overfitting, since it is orthogonal to both PRQS14 and PRQS15.

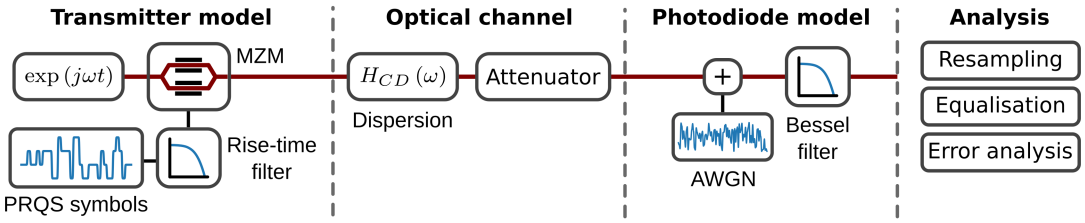


Figure 2.13: Simulation setup used for investigations into Rx bandwidth and chromatic dispersion impairments on 100 Gbit/s PAM4.

Bit error rate evaluation

In order to obtain reasonable estimates for BER levels as low as 10^{-6}, multiple data packets are combined to build test datasets consisting of over 500k symbols. The BER for a given experimental condition could then be calculated by counting the errors in the test symbol patterns in each captured packet, and dividing the aggregate number of errors by the total number of symbols considered. Equaliser training could either be carried out for each packet considered, as an adaptive equaliser might be expected to do in a real PON, or alternatively it could be trained on a single packet, and applied unchanged to all test datasets.

2.4 Initial simulation investigations

This section includes an initial investigation into the effectiveness of NNE for bandwidth limitations and chromatic dispersion: two major challenges facing next-generation PAM4 PON. Of particular interest is whether NNE offers advantages over linear FFE methods in these scenarios. These investigations are carried out using the simulation setup shown in Figure 2.13, which was programmed using Python.

2.4.1 Bandwidth compensation

A PIN photodiode is simulated with limited bandwidth using a 4$^{\text{th}}$-order Bessel filter implemented digitally. A Bessel filter has maximally flat phase response [135], and so is appropriate for modelling the bandwidth behaviour of linear devices. For each Rx bandwidth considered, the equalisers are trained on a training dataset consisting of 15000 randomly generated PAM4 symbols, and tested on repeated PRQS-15, comprising over 500k symbols. The sensitivity performance penalties reported below are calculated relative to the HD-FEC, with BER threshold of 3.8×10^{-3}.

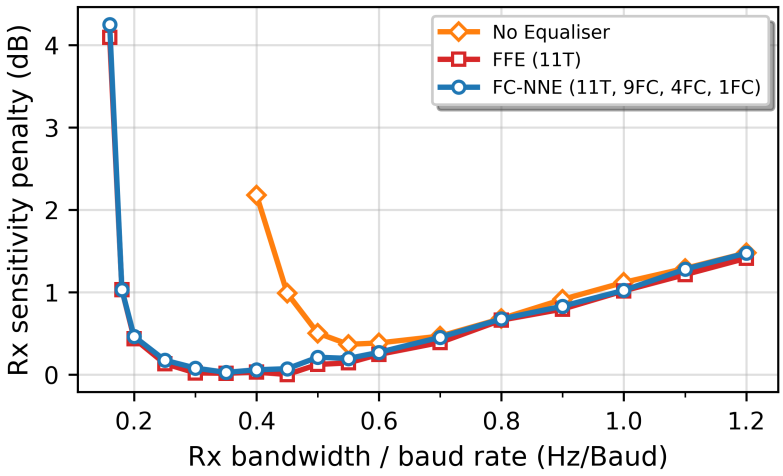


Figure 2.14: The sensitivity penalty of a bandwidth-limited receiver simulated using a 4th order digital Bessel filter.

For comparison, an FFE-11T (11 symbol spaced taps) is evaluated on the simulated data. It is trained with the LMS algorithm, as described in Section 2.3.1. An FC-NNE with 11 symbol-spaced taps (11T) is also evaluated, and consists of 3 fully-connected layers with structure (9FC, 4FC, 1FC). The two hidden layers use the non-linear tanh activation function, while the output layer has a linear activation with which to predict symbols.

The sensitivity penalties reported in Figure 2.14 show that the FFE and NNE have similar performance for all considered Rx bandwidths. This agrees with theory, since bandwidth limitations are linear impairments, and the FFE is not at a disadvantage relative to the non-linear FC-NNE. Conversely, this confirms that NNEs can be effective against severe bandwidth limitations, and therefore can at least match the equalisation capabilities of FFEs.

2.4.2 Mitigating dispersion effects

The effect of chromatic dispersion can be simulated by including the optical fiber channel response in the simulation setup [1]:

$$H_{CD}(\omega) = \exp\left[-i\left(\frac{\beta_2\omega^2}{2} + \frac{\beta_3\omega^3}{6}\right)\cdot L\right], \tag{2.37}$$

where only second and third order dispersion effects are taken into account using β_2, β_3 introduced in Section 2.2.3, and L is the transmission distance. In this simulation, the Rx bandwidth is set to 42 GHz, meaning dispersion impairments

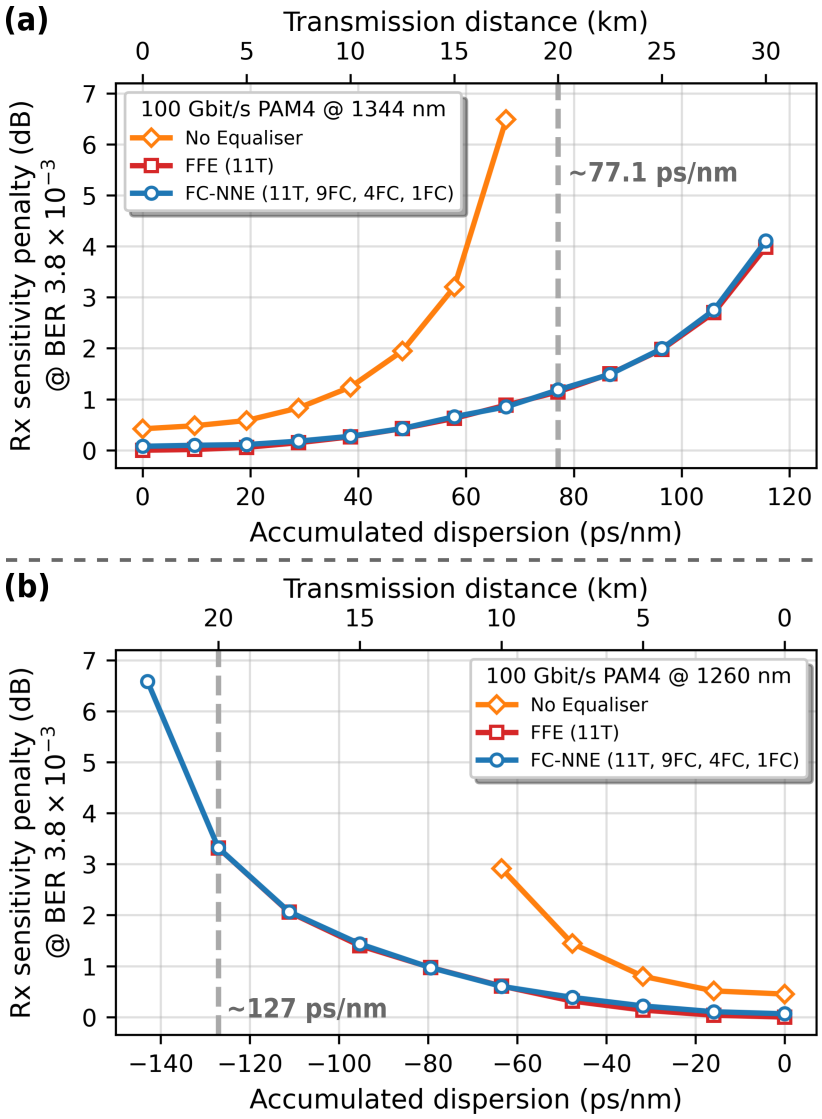


Figure 2.15: Simulated chromatic dispersion penalty at 3.8×10^{-3} BER at (a) 1344 nm and (b) 1260 nm wavelengths.

are being considered in isolation. Dispersion effects were simulated separately for 1344 nm and 1260 nm, which correspond to the worst case dispersion scenarios outlined in 50G HS-PON, as shown in Figure 2.8, and the values for D and S are taken from this graph. Figure 2.15 (a) and (b) show the Rx sensitivity penalty versus transmission distance results using the 100 Gbit/s PAM4 simulation setup in Figure 2.13 at these wavelengths.

In both cases, an FFE-11T is compared with the same FC-NNE considered for the limited bandwidth simulations. It is clear that the FC-NNE largely matches the performance of the linear FFE algorithm for both wavelengths. However, for dispersion below -120 ps·nm^{-1}, the FFE fails to achieve below the HD-FEC threshold, while the FC-NNE does so but with > 3 dB sensitivity penalty. From

this, we can conclude that for isolated chromatic dispersion impairment, an FC-NNE provides no obvious performance benefit over that of an FFE at lower dispersion values, but can extend transmission distance for extreme dispersion cases. Both the FFE and FC-NNE give large performance benefits compared to the no equaliser case.

2.5 Summary

This chapter introduces the critical challenges and system requirements facing future 100 Gbit/s PON systems. In particular, a key focus is on the use of PAM4 modulation to achieve higher data rates while maintaining cost-effectiveness, given the constraints of reusing 25G class optics. However, bandwidth limitations and increased noise susceptibility in PAM4 present challenges to maintaining signal integrity, especially as the optical loss budget must adhere to strict standards. Fiber dispersion is identified as a further potential obstacle to PAM4 due to its relatively high linearity requirements.

Equalisation is a potential solution to ISI and signal distortions effecting PAM4 in future 100 Gbit/s PON systems. However, while the FFE technique is effective against linear bandwidth impairments, it is ineffective at mitigating non-linear distortions such as fiber dispersion or SOA gain saturation. Equalisation based on neural networks is therefore introduced as a sophisticated non-linear extension of the FFE technique, and which is capable of modelling non-linearities and time dependencies in transmitted signals. Initial simulation results demonstrate the potential of NNEs to significantly improve signal recovery in bandwidth-limited and dispersion-impaired environments, and can match FFE performance, setting the stage for further exploration in subsequent chapters.

In Chapter 3, the focus shifts toward a deeper exploration of neural network-based equalisation techniques and their application in real-time signal recovery for high-speed PON systems. The chapter delves into the architecture of GRUs, and examines how their recurrent structure can effectively handle the memory effects and complex non-linearities introduced by SOA gain saturation. By leveraging the ability of GRUs to adapt to time-varying impairments, Chapter 3 provides a more detailed analysis of how these advanced techniques can outperform traditional methods like FFE, leading to significant improvements in system performance.

Chapter 3

Extending SOA preamplifier dynamic range with neural networks

In this chapter, SOA gain saturation induced non-linear patterning is investigated in the context of achievable PON dynamic range. The origin of the patterning effect is explained, and the relation between SOA gain recovery and the severity of the patterning effect is explored through simulation and experiment. Neural network equalisers are then proposed as a solution to overcoming the patterning effect in the context of burst-mode signalling in future 100 *Gbit/s PAM4 PON. In particular, recurrent neural networks which utilise gated recurrent units are shown to be both effective and efficient, and these results are published in [A1].*

Semiconductor optical amplifiers have been proposed as optical preamplifiers at the OLT in order to achieve the 29 dB minimum optical loss budget of PON while using noise susceptible modulation formats such as PAM4. However, the 19.5 dB DR requirement between high- ('loud') and low-power ('soft') burst packets in upstream PON transmission could lead to SOA gain saturation, and its associated non-linear effects. In particular, self gain modulation (SGM), which is commonly referred to as the patterning effect, can have a detrimental effect on high-speed PAM4 signals. Neural network equalisation offers a potential solution to overcome this non-linear effect.

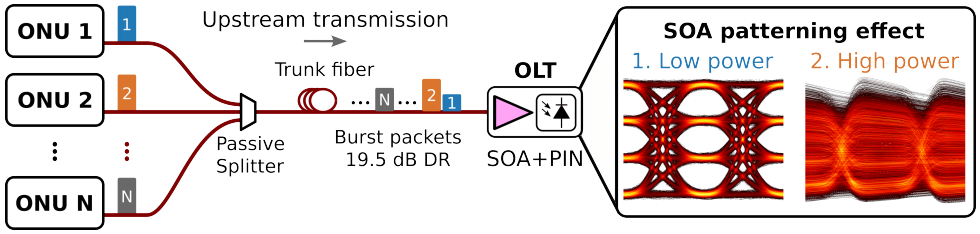


Figure 3.1: Challenges facing SOA preamplification due to burst-mode upstream transmission at the OLT.

3.1 Context and research objective

The challenging 29 dB PON optical loss budget means optical amplification will be required to compensate for PAM4's reduced SNR performance in any future 100 Gbit/s PAM4 PON. SOAs offer a compelling solution as pre-amplification units at the optical line terminal, being low cost, integrable, and operating in both C- and O- wavelength bands. But this raises another concern, as the SOA input power dynamic range may not meet the minimum required PON DR of 19.5 dB without equalisation, because of gain saturation induced patterning.

As shown in Figure 3.1, loud bursts received at the OLT could lead to SOA gain saturation, and a form of non-linear self gain modulation, commonly known as the patterning effect. This effect is a manifestation of the carrier dynamics in the gain medium of the SOA.

Recently, NNEs have been suggested as Tx-side pre-equalisation, to overcome SOA patterning effects in 50G PAM4 PON and extend system dynamic range [108]. Further, feedback-based RNNs have been proposed as efficient post-equalisers for 100 Gb/s PAM4 PON in conjunction with O-band SOA preamplifiers, and have been shown to achieve 30 dB optical loss budget [65]. However the impact of SOA saturation, which is critical for high dynamic range operation, was not considered.

3.2 Semiconductor optical amplifiers

An SOA device uses a semiconductor gain medium to achieve optical amplification. An appropriate bandgap is engineered in the gain medium which overlaps with a waveguide structure, and photons travelling through the SOA structure induce stimulated emission and therefore optical amplification. These devices operate in a similar way to a laser diode, although with anti-reflectivity coatings at the coupling edges to ensure photons travel through the gain material

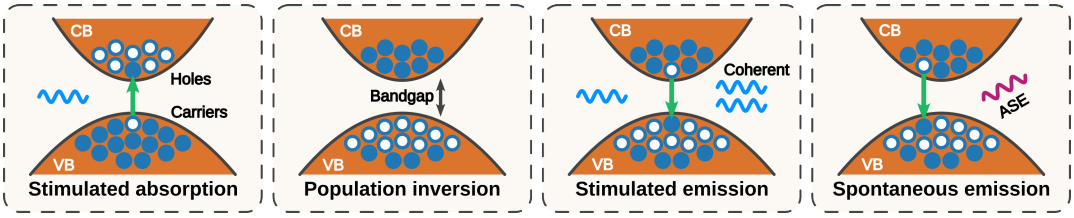


Figure 3.2: SOA structure, bandgap and optical amplification mechanism.

only once. The major advantages an SOA has over a fiber-based amplifier are its small form factor, and low drive currents, typically measured in hundreds of mA. These factors make them ideal for realising high-power, low-cost ONU and OLT transceivers, where they can act as booster or pre-amplifiers.

3.2.1 Basic principles

During operation, a drive current injects electrons into the active region of the SOA, where these carriers occupy energy states in the conduction band (CB). Once the drive current exceeds a certain threshold, the number of carriers in the CB exceeds that of the valence band (VB) and a population inversion is achieved. Under these conditions, an SOA is capable of amplifying optical signals traversing the active region.

There are three radiative mechanisms by which carriers move between the CB and VB, each of which involves the emission or absorption of a photon: stimulated emission, stimulated absorption, and spontaneous emission. Stimulated emission is responsible for optical amplification in SOAs. It involves a photon with energy equal to, or greater than the semiconductor bandgap, stimulating a carrier in the conduction band to transition to the VB. This results in the emission of an identical photon, travelling in the same direction, and with identical phase, wavelength, and polarisation. Conversely, stimulated absorption occurs when such an incident photon is absorbed by a carrier in the VB, which then transitions to the CB. Optical gain relies on the probability of stimulated emission being greater than that of stimulated absorption, for which a population inversion is required.

Spontaneous emission is the third mechanism, and is unavoidable in SOAs. This is because there will always exist a non-zero probability of CB carriers transitioning to the lower VB energy states in order to fill the holes vacated by the excited CB carriers. Photons which are emitted during spontaneous emission have random phase, energy, and polarisation, and constitute an SOAs ASE.

Detailed analysis of SOA operating mechanisms can be found in [1], [28], [136].

3.2.2 Gain dynamics

The gain of an SOA is directly related to the carrier density, N, in the CB. The greater the CB carrier density, the greater the number of carrier-hole pairs which are available for stimulated emission as an input optical signal travels through the SOA device. The transparency carrier density N_{tr} can be thought of as a threshold density: if the CB carrier density N is below this value, then an optical signal travelling through the SOA will experience attenuation due to internal loss processes, as well as spontaneous and non-radiative carrier-hole recombination mechanisms. When $N = N_{tr}$, then the active region will effectively be transparent to the incident signal. However, if $N > N_{tr}$, then the signal will instead experience optical gain. The SOA material gain coefficient g_m can be approximated at the input signal wavelength using this threshold:

$$g_m(N) = a_0(N - N_{tr}), \tag{3.1}$$

where $a_0 = dg_m/dN$ is the differential gain, which is assumed here to be constant. Using this equation, it is possible to express the power of an optical signal $P(z)$ propagating through an SOA along the axis z as:

$$\frac{dP(z)}{dz} = [\Gamma g_m(z) - \alpha_{int}] P(z), \tag{3.2}$$

where α_{int} is the internal attenuation coefficient due to propagation losses, and Γ is the confinement factor, which represents the overlap between the device active region and the propagating signal mode. For an SOA of length L it is possible to solve the above equation and obtain an expression for G, by considering $P(0) = P_{in}$ and $P(L) = P_{out}$:

$$G = \exp\left\{\int_0^L [\Gamma g_m(z) - \alpha_{int}]\, dz\right\}. \tag{3.3}$$

But the material gain coefficient, g_m, depends on N which can vary in time, since the signal power $P(z,t)$ can influence the carrier density through stimulated emission and other mechanisms. Therefore, to realise a simple model of SOA gain dynamics, a second differential equation to approximate the carrier density behaviour must be introduced, which takes into account injected carriers as well as the recombination mechanisms discussed above:

$$\frac{\partial N(z,t)}{\partial t} = \underbrace{\frac{I}{eV}}_{\substack{\text{injected}\\ \text{carriers}}} - \underbrace{\frac{N(z,t)}{\tau_s}}_{\substack{\text{spontaneous}\\ \text{recombination}}} - \underbrace{\frac{g_m(z,t)\,P(z,t)}{\hbar\omega A_{eff}}}_{\substack{\text{stimulated}\\ \text{emission}}}. \tag{3.4}$$

The first term on the right hand side represents the injected carrier density, where I is the device drive current, e is the electron charge, and V is the active region volume. The second term expresses losses due to spontaneous and non-radiative recombination processes, where τ_s is the spontaneous carrier lifetime, which can be approximated using:

$$\tau_s = \frac{1}{A + BN + CN^2}. \tag{3.5}$$

Here, A is the defect recombination coefficient, B represents the radiative recombination coefficient, and C is the contribution from non-radiative Auger recombination processes. These are described in more detail in [28], [137].

The third term in Equation 3.4 takes into account the decrease in N due to stimulated emission, i.e. the mechanism which is directly responsible for the SOA optical gain. Here, $\hbar$ is the reduced Planck's constant, and $A_{eff} = A/\Gamma$ is the effective amplifier mode cross-section area.

Gain saturation

An SOA deployed as a preamplifier at the OLT for upstream PON transmission could undergo gain saturation for high-power packets, as shown in Figure 3.1.

In the small signal approximation where the input power P_{in} is small, the third term in Equation 3.4 representing stimulated emission is much smaller than the net contribution from the injected carriers and recombination due to spontaneous and non-radiative processes. This leads to steady state operation where the overall gain G is high and constant, and is largely unaffected by the input signal. In this small-signal regime, the SOA dynamics are dominated by the ASE contribution of the gain medium and G is therefore determined by the ASE mechanisms. However, for high-power signals the stimulated emission term becomes significant, resulting in an overall reduction in the carrier density N, leading to a new steady state solution with a reduced value for G since there are less carrier-hole pairs available for stimulated emission.

For a continuous wave (i.e. constant power) input signal, it is possible to solve

Equation 3.4 to find the steady state value for the carrier density N. In this scenario both the carrier density $N(z)$ and power $P(z)$ are now only dependent on z, and are independent of time. Substituting in Equation 3.1, and setting $\partial N/\partial t = 0$ it is possible to obtain:

$$0 = \frac{I}{eV} - \frac{1}{\tau_s}\left[N_{tr} + \frac{g_m(z)}{a_0}\right] - \frac{g_m(z)P(z)}{\hbar\omega A_{eff}}. \tag{3.6}$$

Moreover, substituting equation 3.1 into the above equation and solving for N gives the unsaturated steady state carrier density N_{st}:

$$N_{st} = \frac{I\tau_s}{eV}. \tag{3.7}$$

The unsaturated gain coefficient g_0 can then be defined as:

$$g_0 = a_0\left(N_{st} - N_{tr}\right). \tag{3.8}$$

Using the above equations, it is possible to derive an expression for the material gain coefficient as a function of z:

$$g_m(z) = \frac{g_0}{a + \frac{P(z)}{P_{sat}}}, \tag{3.9}$$

where the saturation power of the SOA device with length L is given by:

$$P_{sat} = \frac{\hbar\omega}{\tau_s}\frac{A_{eff}}{a_0 L}. \tag{3.10}$$

If the input power to the SOA is too high, gain saturation may occur as in fiber-based amplifiers. However, in contrast to fiber amplifiers whose gain recovery operates on the order of milliseconds, an SOA's gain is much more responsive, with gain recovery times of just hundreds of picoseconds being typical.

3.2.3 Non-linear patterning effect

The fast gain recovery of SOAs gives rise to many of the non-linear effects of these devices [136]. In particular, if an SOA's gain recovery is similar to the symbol interval of an amplitude modulated signal, then patterning can occur. Patterning acts to distort high-power input signals to an SOA, and this distortion

is non-linear in nature and leads to the gain experienced by a symbol depending on the sequence of symbols directly prior to it [28]. The carrier rate equation in equation 3.4 is sufficient for understanding and modelling the patterning effect in SOAs.

For small signals, the stimulated emission term in equation 3.4 will not cause a large perturbation in the carrier density, due to the input signal's power $P(z)$ being small. However, for a sufficiently high-power signal, stimulated emission can become the dominant term of the equation. For a continuous wave signal, this can lead to a lower steady-state carrier density, resulting in reduced overall gain of the SOA, as already discussed. However, if the signal is a high power pulse, the carrier density will only be temporarily perturbed, after which it will recover according to the interaction between the constant rate of injected carriers into the gain medium and the rate of spontaneous recombination. The gain recovery of an SOA is typically defined as the time taken for the gain to recover from 10 to 90 % of its steady-state value. Typical gain recovery values for a bulk SOA are on the order of a nanosecond or hundreds of picoseconds.

A high power 100 Gbit/s PAM4 signal will be seen as a series of saturating pulses by the SOA gain medium. The short 20 ps symbol periods involved, mean that the carrier density is unable to reach a steady state, and the fluctuations in gain are dependent on the symbol pattern. This is known as the patterning effect. In general, the closer a modulated symbol period is to the SOA gain recovery time, the worse the patterning effect will become [28]. However, if the symbol period is much greater or less than the gain recovery time, a steady state can be reached according to the average power of the signal, and the patterning effect will be insignificant, as in the unsaturated case. This is what happens for fiber amplifiers, which have gain recoveries on the order of milliseconds.

Modelling non-linear patterning

The rate equations 3.2 and 3.4 can be solved numerically to simulate an SOA device. The results of such a simple simulation are included in Figure 3.3 below for a series of PAM4 symbols, i.e. pulses. This illustrate the interplay between pulse power and carrier density in an SOA device, and shows the resulting patterning effect for experienced by the PAM4 signal.

These simulations were carried out by dividing the SOA into 5 separate segments, as described in [28], [136]. For each of these segments, the rate equations were solved using the Dormand-Prince method [138]. The SOA device characteristics

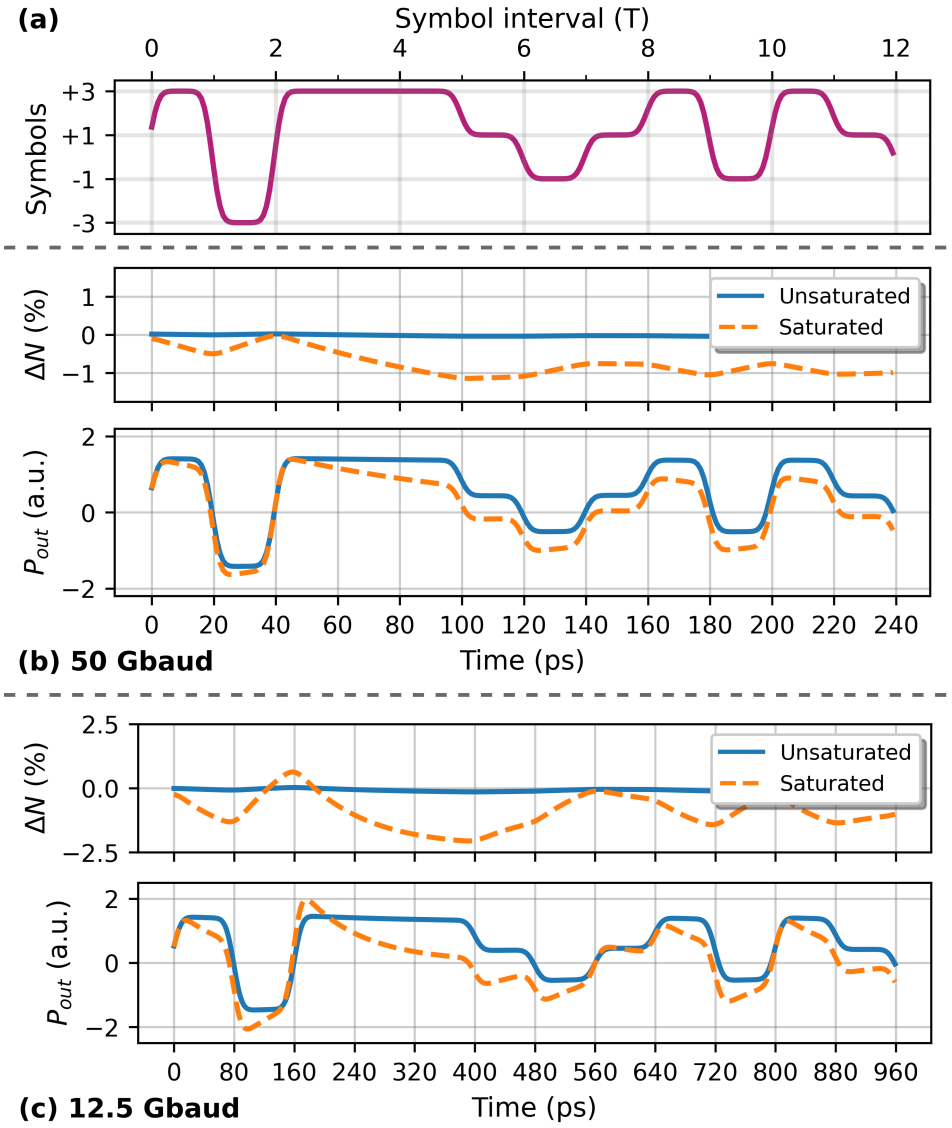


Figure 3.3: (a) The PAM4 sequence used as input to the simulated SOA device above and below the saturation input power. The percentage change in carrier density and resulting output power of the device in both scenarios are shown for (b) a 50 Gbaud and (c) 12.5 Gbaud signal, with both showing clear signs of patterning.

were taken from [28], since the objective is not to model a specific SOA, but to demonstrate the general mechanism behind patterning. As such, the effects of ASE have also been ignored for this simulation.

Figure 3.3 (a) shows the PAM4 symbol sequence that is input to the simulated SOA device. The modulated signal is simulated for a small signal scenario (-14 dBm) indicated by a solid line, and a high-power ($+4$ dBm) one, labelled using a dashed line, and which results in SOA saturation. In Figure 3.3 (b), the resulting

% change in averaged carrier density ΔN over the 5 segments of the SOA is plotted, alongside the resulting output power of the 50 Gbaud PAM4 symbols for both saturated and unsaturated scenarios. It is clear that the high power input signal causes significant deviations in the gain medium carrier density, and this results directly in the observed patterning effects in the device output power. Particularly noteworthy is that the sequence of 3 consecutive high-power symbols ($+3$) results in carrier density degradation which persists for multiple subsequent symbols. Meanwhile, for low input power below the saturation level, a steady state N is reached and patterning is not observed.

Further, when the simulated baud rate is lowered from 50 to 12.5 Gbaud, Figure 3.3 (c) shows that the patterning effect becomes more severe. This is because the saturating pulses are now longer in duration (100 ps), and therefore contain more energy with which to degrade the CB carrier density, N. Conversely, longer low-power symbols also allow more time for gain recovery, meaning N experiences greater variations, both above and below its mean value.

Mitigation strategies

It is possible to mitigate the patterning effect for upstream PON transmission by varying the gain recovery time of the SOA according to the received packet power. This can be achieved using either electrical or optical methods on a packet-by-packet basis, both of which have been investigated in [A6] for 100 Gbit/s PAM4 PON transmission.

Reducing the SOA drive current leads to a smaller steady-state carrier density in the SOA active region, which means that the achievable gain G is also reduced. However, a smaller drive current also leads to an increase in the gain recovery time, since the rate at which carriers are injected into the active region is reduced, and this can diminish the patterning effect, depending on the signal baud-rate. For upstream traffic in PON, a trade-off must be made between reducing the SOA drive current to alleviate patterning effects and providing sufficient gain to achieve a good Rx sensitivity. At a certain input signal power, the drive current cannot be reduced sufficiently to achieve signal equalisation, meaning the achievable dynamic range of this method is limited.

An alternative method to increase the gain recovery time and reduce the SOA patterning effects is to inject a continuous-wave holding beam of sufficient power into the SOA at a wavelength offset from the data carrier signal. The holding beam will deplete the available carriers in the active region, and reduce the gain

recovery rate with a corresponding decrease in patterning as before. A benefit of this method is that it induces less ASE noise compared to the reduced drive current method, as shown in [A6]. However, as with the drive current method, this will deplete the available gain of the amplifier for upstream data signals, and further the holding-beam signal power required for equalisation cannot be increased without limit to equalise the highest power data signals. These combined factors mean that the achievable dynamic range of this method is also limited.

Alternatively, equalisation techniques can avoid these issues, and support PAM4 transmission farther into an SOA's gain saturated regime than the above methods. In particular NNE-based solutions targeting SOA patterning are the focus of this thesis, and will be discussed in the following sections and chapters.

3.2.4 Device under test

In this thesis, the same SOA device is used as the device under test in each reported experiment. The graphs in figures 3.4 (a) and (b) show the continuous wave gain characteristics of this SOA, which are typical of commercially available SOAs. The SOA operates in the linear regime at 100 mA constant bias for input powers less than -8 dBm, and offers up to 20 dB gain. The output saturation power of the SOA is $+13$ dBm, and for input powers above -8 dBm it operates in the non-linear regime where patterning effects becomes significant.

Further, the signal OSNR is plotted against SOA bias current in Figure 3.4 (b). As mentioned, patterning can be mitigated by reducing the SOA bias current and therefore increasing the gain recovery time, but this leads to an associated decrease in the SOA's output OSNR performance due to ASE.

Pump-probe experiment

A preliminary investigation of the SOA gain recovery was conducted using a pump-probe experiment, as shown in Figure 3.5, and which was similar to the experimental setup described in [139]. In place of a pulsed laser (which was unavailable at the time), pulses with sufficient energy to saturate the SOA were created using a high-speed DAC and MZM configuration as shown. This was achieved by biasing the MZM at its null point, and driving it with an electrically amplified 391 MHz pulse train, where individual pulses had a full width at half maximum (FWHM) of 34 ps.

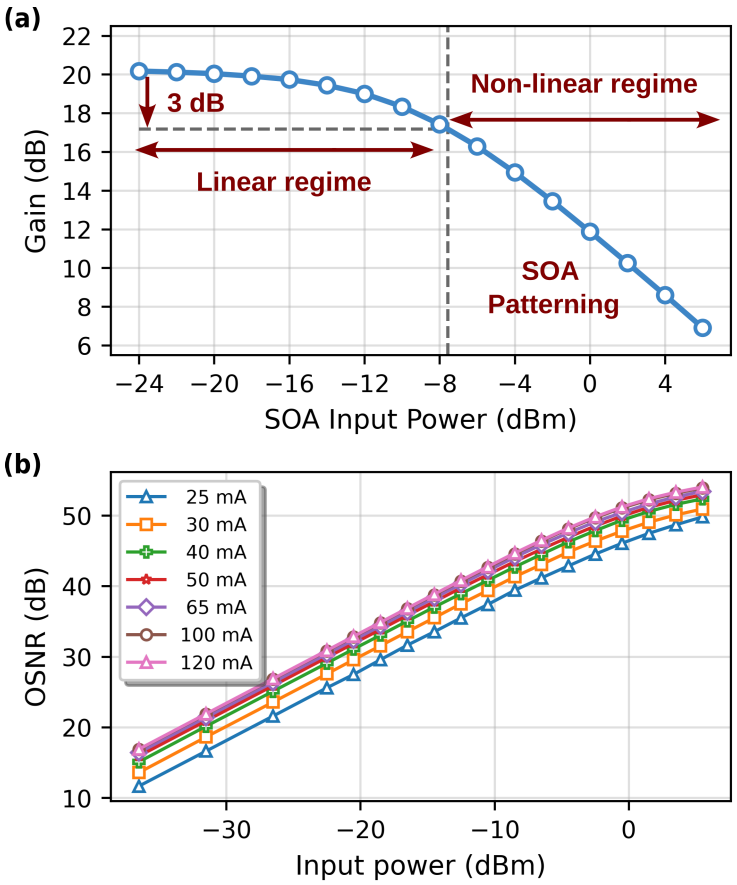


Figure 3.4: The gain curve of the CIP SOA-S at a constant bias of 100 mA is shown, with the linear and non-linear operating regimes delineated by the 3-dB gain saturation point.

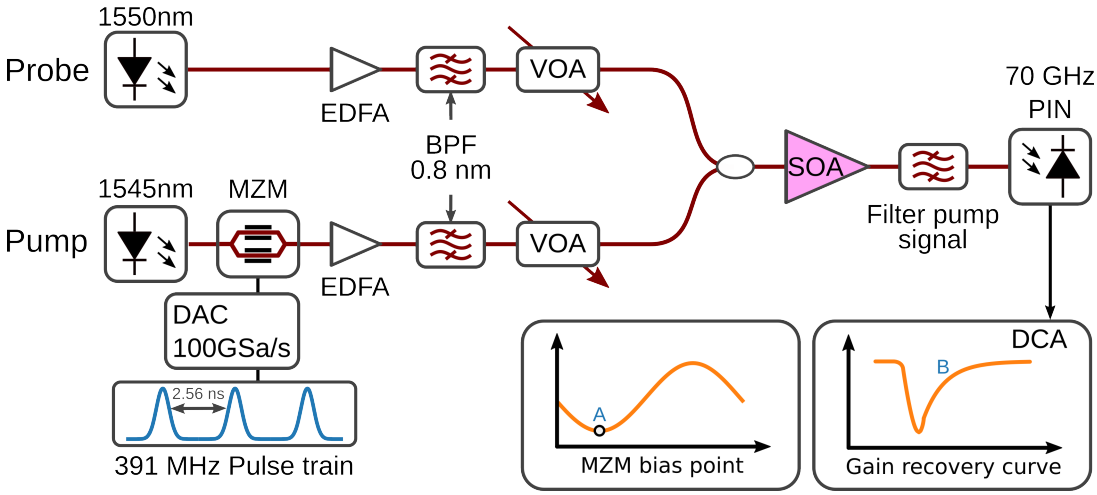


Figure 3.5: Experimental setup used to measure gain recovery time of SOA used.

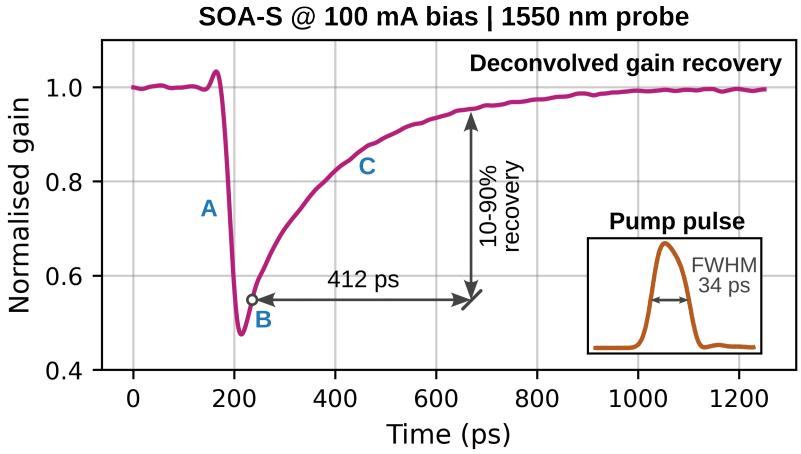


Figure 3.6: Results of pump-probe experiment to measure the SOA-S gain recovery for a constant bias of 100 mA and a 1550 nm probe wavelength

A continuous wave probe signal was combined with the pulsed pump signal before being input into the SOA. The pump signal wavelength were tuneable, and 1540, 1545, 1555, and 1560 nm wavelengths were tested. Each pulse acts to saturate the SOA, but the repetition rate was chosen such that it allowed sufficient time for the gain to fully recover before the next pulse arrived. After the SOA, an optical filter was used to remove the pump signal, allowing for the gain recovery of the probe signal to be detected by a suitably triggered digital communications analyser.

The results of this experiment are shown in Figure 3.6 for a 1545 nm pump signal, where the SOA is biased at a constant 100 mA. The gain recovery is measured to be 412 ps after the recovery curve is deconvolved with the pulse shape, which is in-line with expected gain recoveries of commercially available bulk SOAs. However, the optical power at the MZM null point may still have had a background saturating effect on the SOA, and the true recovery time may be shorter than that measured. Regardless, the 412 ps time is a good indicator of typical SOA performance.

From this result, we can predict that 100 Gbit/s PAM4 signals should suffer less from the non-linear patterning effect than lower baud rates such as 12.5 Gbaud. This is because the 412 ps SOA gain recovery time is closer to the 12.5 Gbaud symbol period of 80 ps than it is to the 20 ps of the higher baud-rate. However, while individual 50 Gbaud symbols are unlikely to induce severe carrier depletion, long sequences of high-power, or low-power symbols are likely to significantly effect the SOA carrier density, and distortions such as those modelled in Figure 3.3 (b) are probable. Conversely, 12.5 and 25 Gbaud signals, with 40 and 80 ps

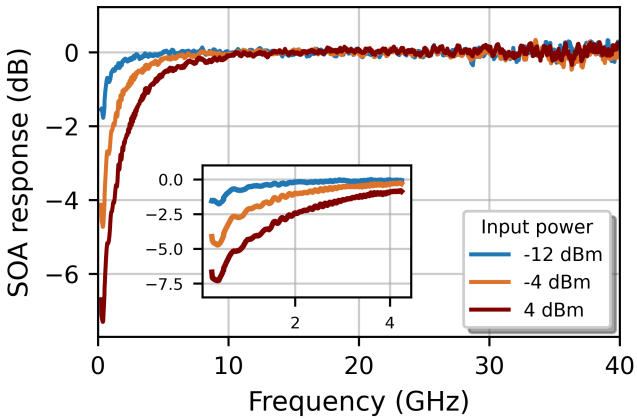


Figure 3.7: Frequency response of SOA for different input powers.

symbol intervals respectively, are expected to suffer greater patterning distortion during a single symbol period, similar to that demonstrated in the simulations in Figure 3.3 (c).

Frequency response

The SOA frequency response can be measured using the experimental testbed outlined in Section 3.3.1, and the Schroeder pulse technique [140]. This involves transmitting a chirped signal consisting of a defined series of sine waves through a device, and measuring their amplitude at the device output. The SOA frequency response is shown for input powers -12, -4, and $+4$ dBm in Figure 3.7.[1]

The SOA clearly acts as a high pass filter, with increased attenuation for frequencies below 3 GHz associated with increasing signal input power. This is evidence that the SOA is less able to sustain long sequences of high-power symbols the deeper into saturation it is driven, since such sequences cause more severe depletion of the carrier density and therefore gain. Further, the spectral content of lower speed PAM signals is more concentrated at lower frequencies which are the most effected by this high-pass filtering operation. Therefore, it would be expected that lower baud-rate modulation will suffer more significant symbol distortions than higher baud-rate modulation. This is in agreement with the simulated patterning results in Figure 3.3 (c).

[1]The noise in the response above 30 GHz can be attributed to the limited system bandwidth and resulting attenuation at high frequencies.

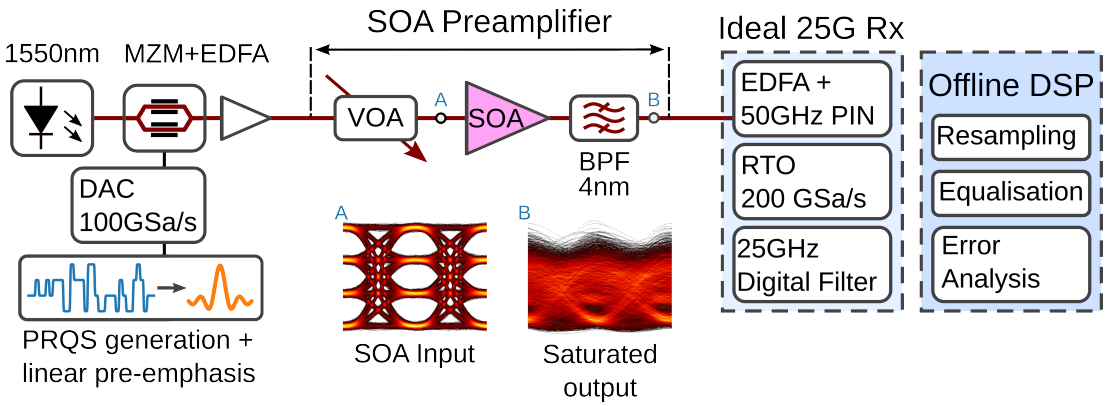


Figure 3.8: Experimental setup used to emulate multi-rate upstream PAM4 transmission.

3.3 Patterning in future PAM4 PON systems

In this section, the severity of the non-linear SOA patterning effect on high-speed PAM4 signals up to 128 Gbit/s is determined experimentally, and the system level ramifications of the achievable SOA input DR are discussed.

3.3.1 High-speed PAM4 testbed

The experimental setup shown in Figure 3.8 allows for high bit-rate PAM4 signals with 6-dB extinction ratio[2] to be generated in the C-band using a 100 GSa/s DAC. The differential output from the DAC drives an MZM in push-pull configuration, and together with a booster EDFA, this emulates an ideal high-power Tx. The system bandwidth is extended using linear digital pre-compensation based on the system frequency response, measured using the Schroeder pulse technique [140]. A photoreceiver with integrated TIA was unavailable for this study, and so a 50 GHz photodiode combined with an EDFA was substituted. This Tx/Rx combination allowed the performance of PAM4 signals to be investigated in both saturated and unsaturated SOA conditions. A variable optical attenuator placed before the SOA emulate the optical distribution network losses, as shown in Figure 3.8. Signals were captured using a 200 GSa/s real time oscilloscope, and a 25 GHz 4$^{\text{th}}$-order Bessel filter was applied digitally to mimic the bandwidth limitation that would arise from the use of 25G class opto-electronics. Figure 3.9 shows eye diagrams of the 100 Gbit/s PAM4 signals generated using this setup for different input powers to the SOA. For low input powers the signals are noise limited, while for higher SOA input powers patterning effects are clearly visible.

[2]The extinction ratio of a PAM4 signal is calculated by taking the ratio of the highest and lowest power symbols, i.e. P_{+3}/P_{-3} using the PAM4 alphabet described in Fig. 2.2 (b).

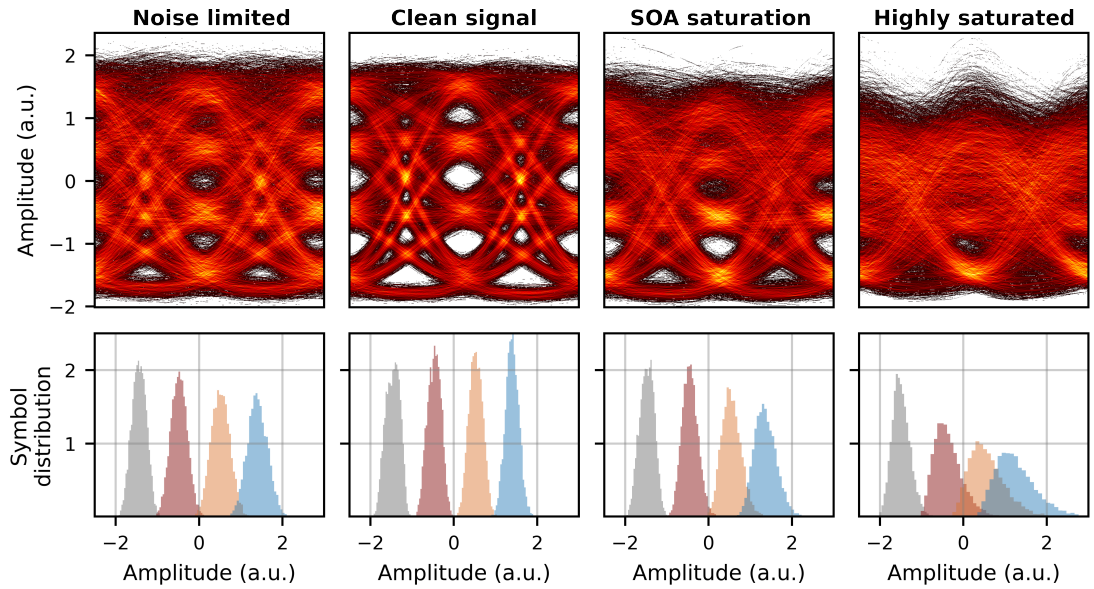


Figure 3.9: Increasing severity of SOA non-linear patterning on 100 Gbit/s PAM4 transmission as seen in eye diagrams and sampled symbol distributions.

The HD-FEC threshold of 3.8×10^{-3} BER is used to determine both the Rx sensitivity and Rx overshoot for a particular baud-rate. The SOA input dynamic range can then be determined as shown in 2.6. For each input power to the SOA, the equalisers are trained using a random symbol sequence, while BER values are estimated using the counting method and ≈ 458000 captured PRQS15 symbols.

3.3.2 Achievable SOA input power dynamic range

The BER performance of PAM4 signals for a range of input powers to the SOA and at various baud rates are shown in Figure 3.10 without any equalisation. In this analysis, a Bessel filter is digitally applied to each captured signal with a 3-dB bandwidth equal to the baud rate under consideration, i.e. for a 12.5 Gbaud signal, a 12.5 GHz filter is applied. This allows for isolated SOA patterning to be considered across multiple baud rates without bandwidth limitations effecting the results.

At input powers less than -14 dBm, the effect of thermal noise and ASE are apparent, and the Rx can be said to be noise limited with a measured sensitivity of -21.4 dBm for 100 Gbit/s (50 Gbaud) PAM4. Achievable SNR is higher for lower baud rates using this setup, and they therefore achieve higher Rx sensitivities than the high baud rate signals.

For powers above -14 dBm, the effects of non-linear SOA patterning become apparent for all baud rates. However the Rx overshoots exhibit an inverted

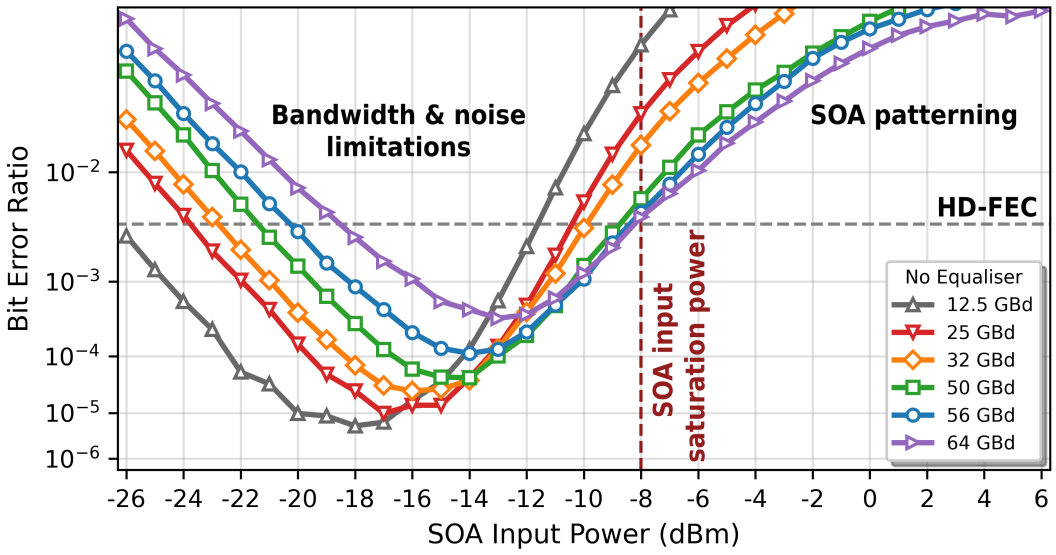


Figure 3.10: Performance of high-speed PAM4 signals for a range of SOA input powers. Clearly above -8 dBm the signals undergo saturated SOA non-linear patterning effect.

relationship with the signal baud rate, with the lower baud rate signals having lower Rx overshoot values. Indeed, this apparent relation between achievable Rx overshoot and baud rate agrees with the predictions based on simulations, SOA gain recovery, and frequency response measurements, described earlier in this chapter. Further, the susceptibility of PAM4 to SOA patterning appears a major obstacle to achieving the 19.5 dB DR necessary for upstream PON transmission.

3.3.3 Analysis of FFE performance

In order to extend the input power DR of the SOA, an FFE-80T is applied to each baud rate, and the results shown in Figure 3.11. The FFE does compensate some of the patterning effect and extend the operating regime of the SOA beyond the -8 dBm input saturation power of the SOA. The 50 Gbaud PAM4 signal with this equalisation achieves 19.6 dB DR, which is just above the DR requirements for the 50 Gbit/s HS-PON standards. However this is only taking into account isolated SOA patterning, and the lack of an error margin means this is impractical for a real-world system. The highest DR achieved is by the 12.5 Gbaud signal, however this is due to it having the best Rx sensitivity, which compensates for it having the worst achievable Rx overshoot.

However, by comparing these achievable dynamic ranges with the minimum 19.5 dB DR required by 50G HS-PON, it is clear that an FFE is not powerful enough

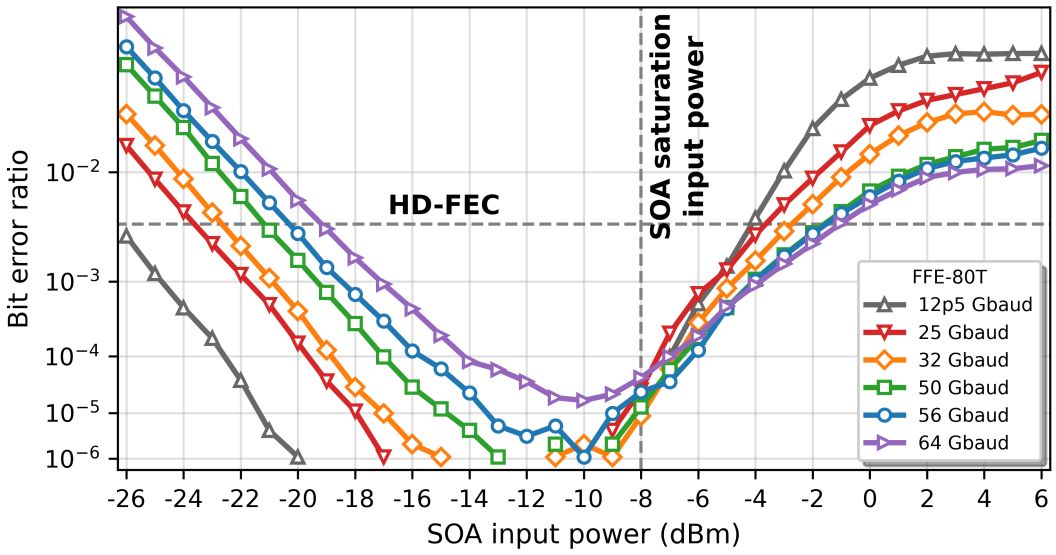


Figure 3.11: Transmission performance of various baud rate PAM4 signals using an SOA+PIN Rx configuration, and FFE-80T post-equalisation.

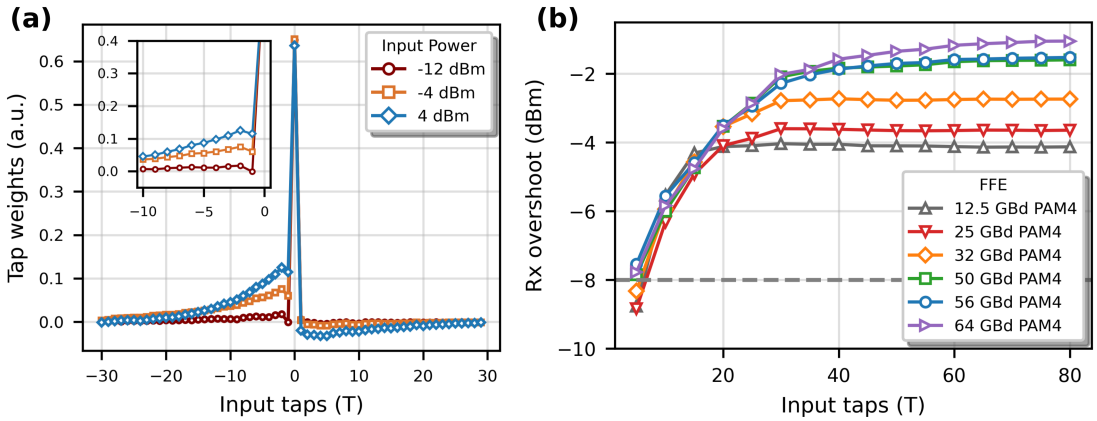


Figure 3.12: (a) Trained FFE taps showing the majority contribution of past symbols toward equalising the SOA patterning effect; (b) the strong dependence of linear FFE equalisation on the number of input taps for different baud rates.

to compensate the SOA patterning effect completely. Therefore it is insufficient to sustain high-DR 100 Gbit/s PAM4 upstream transmission.

Further, there is a strong correlation between the number of FFE input taps and the achievable Rx overshoot, as shown in Figure 3.12 (b). For each baud rate considered, the Rx overshoot for a given number of FFE input taps is plotted, for up to 80 symbol spaced taps. Increasing the number of FFE taps leads to saturation in achieved Rx overshoot after a certain point. This gives an indication of the dependence of the patterning effect on past symbols, and FFEs with larger input spans perform better since they take as input more past samples.

Figure 3.12 (a) shows an example of converged FFE-60T tap weights after training on a 100 Gbit/s PAM4 signal at SOA input power $+4$ dBm. An important feature of these converged weights is that equalisation depends only on the prior-symbol taps. The post-symbol tap weights are near zero, and therefore do not contribute to the equalisation process.

3.4 Designing a neural network equaliser

Neural networks have been proposed and shown effective for mitigating limited SOA non-linear impairments for PAM4 and PAM8 signals [108], [109], [141]. In particular, feedforward networks have been preferred, as they can be implemented in a pipelined manner on existing DSP hardware platforms (ASIC, FPGA) [93]. However, RNN structures appear as ideal solutions, as they can effectively model the long-memory effects of gain saturation induced patterning in an SOA preamplifier.

3.4.1 Recurrent neural networks

Recurrent neural networks incorporate a feedback loop into their operation, which allows them to process long time series data efficiently by tracking relevant signal features over time. Therefore, RNNs should not need large numbers of input taps to mitigate long-memory impairments such as SOA patterning effects. An RNN neuron is governed by the equation:

$$h_t = \varphi\left(W_x \cdot x_t + U_h \cdot h_{t-1} + b\right), \tag{3.11}$$

where h_t is the unit output at time t, x_t is the input at time t, and W_x, U_h are the input and feedback weights, while the bias term is labelled b. The non-linear activation is typically chosen to be either sigmoid or $\tanh()$, and is denoted as φ here. The feedback loop is realised by the $U_h \cdot h_{t-1}$ term, where h_{t-1} is the neuron output at the previous time step $(t-1)$.

Gated recurrent units

However, within the ML community standard RNNs are known to be unstable and difficult to train for long-term memory effects, due to issues such as vanishing and exploding gradients [142]. This has led to the development of more sophisticated

recurrent neurons such as GRUs [143], which can have 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:

$$\begin{aligned} z_t &= \sigma\left(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z\right), \\ r_t &= \sigma\left(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r\right), \\ \hat{h}_t &= \varphi\left(W_h \cdot x_t + U_h \cdot (r_t \circ h_{t-1}) + b_h\right), \\ h_t &= z_t \circ \hat{h}_t + (1 - z_t) \circ h_{t-1}, \end{aligned} \tag{3.12}$$

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 Figure 3.13 (a). This enables a GRU-RNN to determine the relevant feedback and input state information for the current symbol equalisation. A state diagram of a GRU unit is also shown in Figure 3.13 (b).

Backpropagation-through-time

The ability of RNN-based equalisers to implement feedback effectively is achieved using the backpropagation-through-time (BPtT) training algorithm [144], which can be used in conjunction with Adam, just like standard backpropagation. BPtT works by "unrolling" an RNN along its feedback connections, and then backpropagating the error along these feedback connections, effectively backwards in time. The number of time steps the RNN is unrolled along its feedback connections is therefore a training hyperparameter, which determines the RNN's "memory", i.e. ability to address ISI outside its immediate input taps.

In the context of SOA patterning, as an SOA is driven further into saturation, the memory dependence of the patterning effect increases, as seen in Figure 3.12 (b), with a corresponding increase in required equaliser input taps. But a GRU-RNN could be trained to overcome such varying levels of SOA patterning simply by increasing the number of time steps it processes during training. Importantly, the increased complexity of a large number of training time-steps is incurred

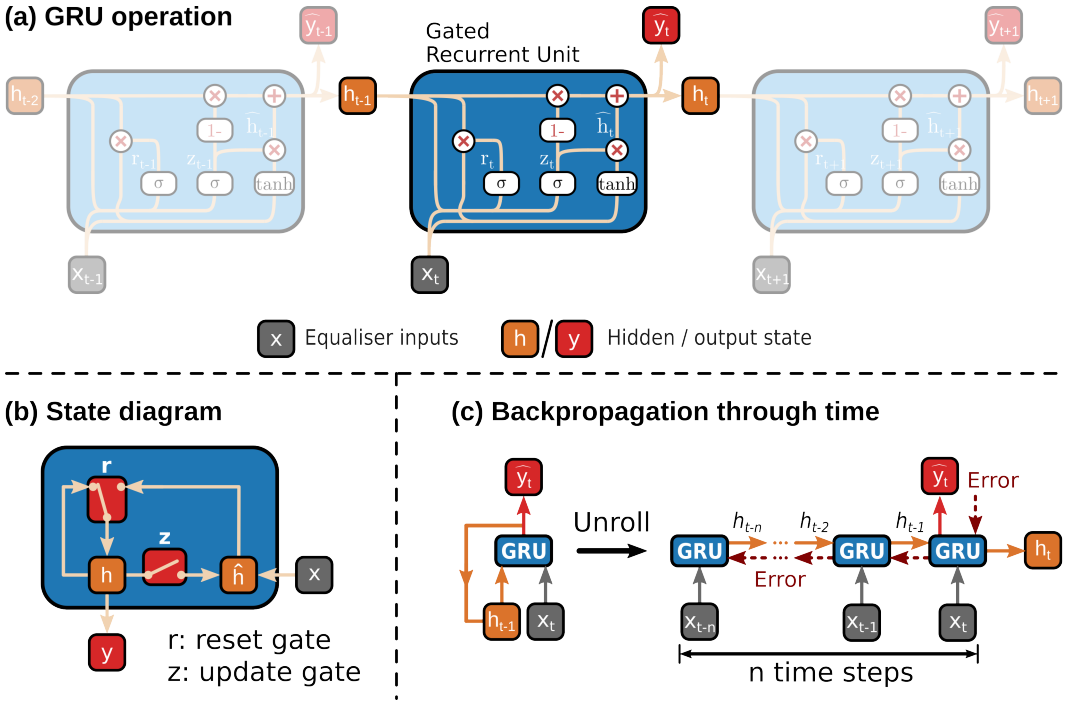


Figure 3.13: (a) Shows a schematic of a GRU neuron with a gated feedback mechanism, and (b) is the corresponding state diagram of a GRU; (c) shows the backpropagation through time algorithm for training RNNs.

only once at training time, while the effect of using a large number of taps in an FC-NNE is carried through to online inference, resulting in increased deployment costs.

Conversely, the BPtT algorithm is especially memory intensive, and further compromises already long neural network training times. If an equaliser was required to deal with varying SOA patterning impairment levels, such as in burst-mode PON, it would not be feasible to implement the complex BPtT algorithm to adapt the equaliser for each burst packet.

3.4.2 Performance and complexity analysis

Figure 3.14 shows the structure of the GRU-NNE with 3T taps under consideration. The performance results of this GRU-NNE are shown in Figure 3.15 (a) for each of the PAM4 baud-rates previously discussed. From this, it is clear that the GRU-NNE is highly effective against the non-linear SOA patterning, and for each baud rate considered, it achieves below the HD-FEC threshold up to and beyond the $+6$ dBm SOA input power considered. In terms of the achievable DR, this means that for 100 Gbit/s PAM4 signals, this GRU-NNE equaliser can sustain > 27 dB input power DR to an SOA. Note however, that the NNEs considered

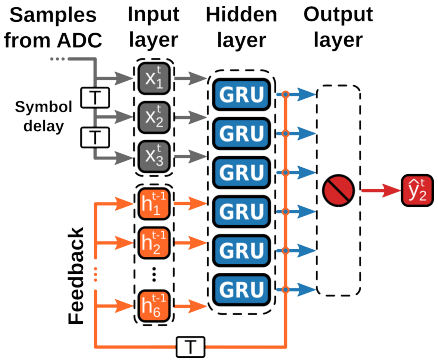


Figure 3.14: Schematic diagram of a GRU-based RNN equaliser.

must be trained separately for each SOA input power considered, a challenge which is tackled further in Chapter 5.

Figure 3.15 (b) focuses on the performance of the GRU-NNE for 100 Gbit/s PAM4 with a 25 GHz Rx bandwidth, emulated using a Bessel filter, and compares it to that of an FC-NNE with varying input taps. From Figure 3.15 (a), a baseline receiver sensitivity of -22 dBm is achieved at the HD-FEC threshold which would meet the required 29 dB PON minimum optical loss budget assuming a $+8$ dBm launch power. As the input power increases beyond -12 dBm the effect of SOA gain saturation and non-linear ISI due to patterning become apparent, with only 12.5 dB maximum achievable input power dynamic range with no equaliser, and 19.1 dB with an FFE-80T. An FC-NNE is also able to achieve large DR, however it requires 80T taps to have similar performance to the GRU-NNE. In contrast, the GRU-based equaliser uses only 3T taps to achieve achieve a BER $< 10^{-5}$ up to $+6$ dBm input power to the SOA, which is well below the HD-FEC limit and corresponds to > 28.2 dB SOA input power DR.

It is clear that the NNEs which are capable of non-linear equalisation are more effective than the conventional FFE, and can almost completely overcome the ISI due to SOA patterning over a large range of powers. However, the GRU-3T outperforms the FC-NNE equalisers, and uses the least number of input taps thanks to its gated feedback mechanism.

Feedback vs. feedforward networks

Similar to the FFE solutions in section 3.3.1, the performance of an FC-NNE applied to SOA patterning will have a strong dependence on the number of input taps used. However, the advantages of a non-linear FC-NNE are also clear, with the FC-NNE-80T extending the achievable dynamic range by up to 27.6 dB, which

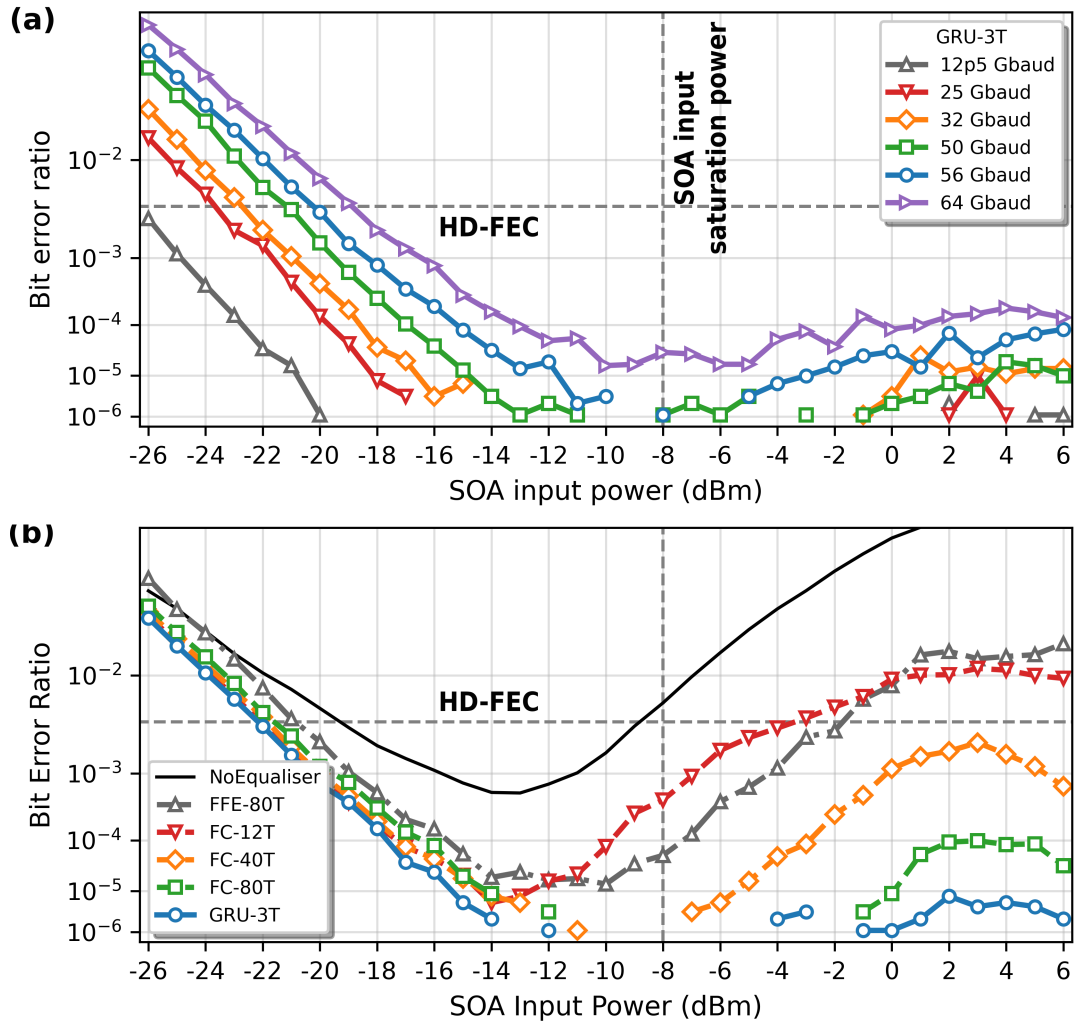


Figure 3.15: BER results for (a) GRU-NNE (3T) at different baud-rates, and (b) a comparison of results for GRU-, FC-NNE, and FFE equalisers for 100 Gbit/s PAM4.

is 8.4 dB greater that that of the FFE-80T solution shown in Figure 3.15.

This dependence on large numbers of input taps has important implications for an FC-NNE deployed to equalise upstream PAM4 transmission. Not all received packets will drive an SOA preamplifier into saturation, and therefore an NNE designed with 80T taps to overcome the worst possible patterning scenario will be highly inefficient. For most received packets, the equalisation requirements will be considerably less. The tap dependent performance of an FC-NNE is shown in Figure 3.16 (a), where it is clear that more taps are required to equalise higher SOA input powers. In particular, 80 taps are required at an SOA input power of 0 dBm to achieve the same performance as a 30 tap FC-NNE at the SOA input saturation power of -8 dBm.

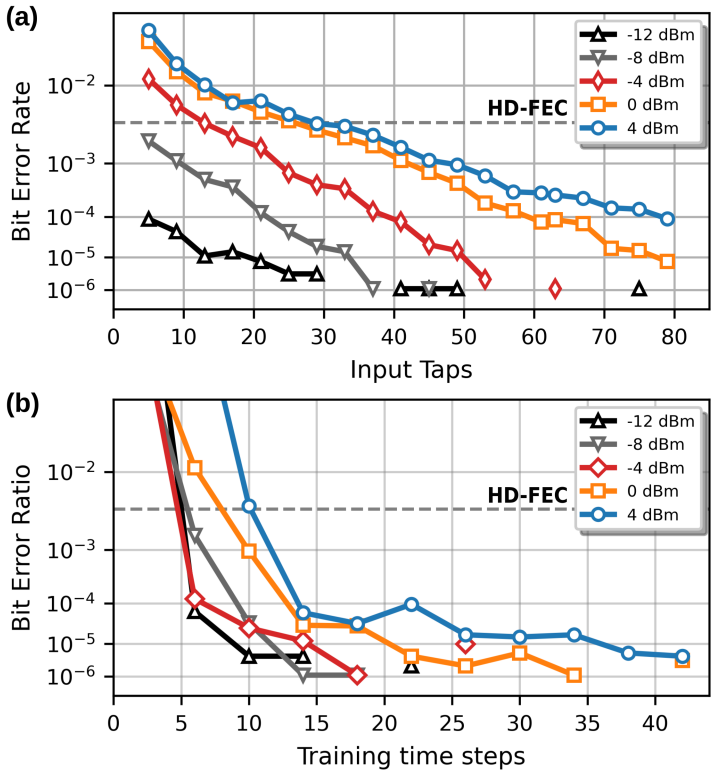


Figure 3.16: The performance dependence of (a) an FC-NNE on tap numbers is compared with that of a (b) GRU-NNE's performance dependence on training time steps, for varying SOA input powers.

The GRU-NNE overcomes this tap-dependence thanks to its feedback mechanism, which can track relevant information from prior symbols, and therefore requires only 3T to fully overcome the worst case patterning. Therefore, the cost overhead for a GRU-NNE is confined to its training in the form of increased time-steps during the BPtT algorithm, as shown in Figure 3.16 (b) which plots the performance of the GRU-NNE with constant 3T input taps, versus the training time steps used for the same SOA input powers. From this, the flexibility of a GRU-NNE is clear as varying degrees of patterning impairments can be addressed by simply tuning the equaliser training as opposed to its structure.

Complexity advantages of RNNs

The consequences of this tap dependence of an FC-NNE's performance are shown in Table 3.1, which compares the equaliser complexities according to the number of parameters and real valued multiplications (RVMs) they use per equalised symbol. Although a single GRU neuron is more complex than a standard neural network neuron, the tap dependence of the FC-NNE's performance leads to it requiring significantly more input taps to mitigate the worst-case SOA patterning.

Table 3.1: Comparison of NNE complexities in the context of total parameters, RVMs, and achievable DR performance.

Equaliser	Input Taps	Structure	Parameters	RVMs	SOA Input Power Dynamic Range (dB)
FC-NNE	40T	(9FC, 4FC, 1FC)	414	400	27.8
FC-NNE	80T	(9FC, 4FC, 1FC)	774	760	27.6
GRU-NNE	3T	(6GRU, 1FC)	187	168	28.2

Therefore, the FC-NNE-80T which has similar performance to the 3 tap GRU-NNE uses $4\times$ the number of trainable parameters, while the GRU uses just 22% of the RVMs to equalise a symbol. Even compared to the FC-NNE-40T, the GRU-NNE uses $< 50\%$ of the parameters and RVMs while achieving greater overall BER performance.

Since the level of SOA gain-saturation and patterning effecting a signal is dynamic, and would depend on the specific burst-packet under consideration, the tap-dependence of an FC-NNE is a major disadvantage. In order to guarantee performance for all levels of patterning, the number of taps required by an FC-NNE for the most severe patterning effect would need to be incorporated. However as seen above this corresponds to a large overall increase in equaliser cost and complexity, to guarantee performance for a possibly low-probability high-power burst. In such a case, the GRU-NNE has a clear advantage, as its complexity and tap requirements are not linked to its performance: only its training procedure needs to be adapted.

3.4.3 GRU gate mechanisms

In order to gain insight into the operation of the GRU-NNE, signal pulses are simulated, and fed as input to GRU-NNE models which have been trained on real experimental data taken at -14 and $+6$ dBm SOA input powers. The output activation of the GRU neuron in the HL with the largest contribution[3] to the NNE output layer is plotted for this stimulated pulse input, and plotted in Figure 3.17 (b) and (c), with the simulated pulse shown in (a). This activation is the h_t term from equation 3.12.

For the GRU-NNE trained on data where SOA patterning is not present (i.e. -14 dBm input power), it is clear that the neuron activation plotted in Figure 3.17

[3]Largest contribution can be quantified by comparing the output layer weight parameters assigned to each neuron, remembering the output layer is simply a weighted sum as in equation 2.24.

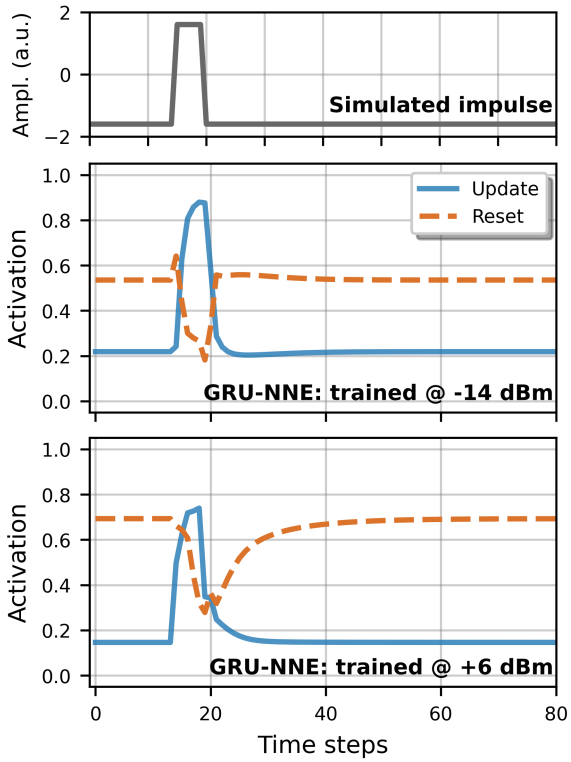


Figure 3.17: Shows the simulated pulse fed to trained GRU-NNE models on saturated and unsaturated SOA data, and the resulting output activation response of the most significant GRU neuron.

(b) reacts to the simulated input pulse, but immediately returns to its original value for the subsequent long sequence of low power symbols. Meanwhile, the activation for the GRU-NNE trained on the $+6$ dBm data which involves high levels of SOA saturation and patterning effect can be seen in Figure 3.17 (c), and the difference in the activation response is marked. In this scenario, the GRU activation reacts to the simulated high-power pulse as before, however the effect of the pulse persists for > 20 time steps, or equivalently, > 20 symbol periods / unit intervals.

Therefore, the effectiveness of the GRU-NNE at tracking the effects of a saturating symbol over multiple subsequent time steps are clear. Unlike chromatic dispersion, where the carrier phase information is lost, the SOA patterning effect is predominantly manifested in the amplitude of the modulated signal it effects[4]. Therefore the GRU-NNE can track the level of saturation experienced by each symbol in a sequence once it has been trained using sufficient time steps. It can then use this information to equalise and remove the saturation effects, in order

[4]The self-phase modulation due to SOA gain saturation is relatively much weaker than the SGM effect [28].

to recover the original signal.

3.5 Summary

In this chapter, the focus has been on the potential for NNE to overcome gain-saturation induced SOA patterning effects and enhance the achievable dynamic range of an SOA+PIN OLT-side Rx for PAM4 signals. Through experimental investigations, it was found that the non-linear patterning effect does severely limit PAM4 transmission, and in particular can severely limit the Rx overshoot for baud-rates ranging from 12.5, to 64 Gbaud. Further, linear FFE solutions are limited in their ability to recover PAM4 symbols from such distortions. Non-linear FC-NNEs were found to effectively overcome the isolated SOA patterning impairment, however their performance is highly dependent on the number of taps used, due to patterning being a long-memory effect. As a consequence, the required taps, and therefore complexity, of the FC-NNE solution increases with the severity of the patterning distortion.

Recurrent neural networks which incorporate a feedback mechanism were introduced as a solution to this, and in particular those based on GRU neurons. Using an appropriate number of time steps with the BPtT training algorithm, the required taps for a GRU-based equaliser is just 3 to mitigate the patterning effect for all SOA input powers. Significantly, the complexity of this GRU-NNE is shown to be less than that of a similarly capable FC-NNE. The proposed GRU-NNE is therefore an efficient solution, which allows for > 28 dB dynamic range 100 Gbit/s PAM4 transmission using an SOA+PIN Rx. This is substantially greater than the 19.5 dB required by the 50 Gbit/s HS-PON standards.

Chapter 4

Neural network equalisation in future 100G PON systems

This chapter expands the performance analysis of the proposed GRU-based equaliser in 100 *Gbit/s/*λ *upstream PON transmission scenario by considering multiple impairments, such as combined fiber dispersion and patterning effects. Alternative non-linear equalisation techniques are considered and compared based on achievable DR and complexity metrics. Further, it is shown that SOA preamplifier gain compression can reduce burst-mode receiver requirements. Multi-symbol equalisation solutions are also explored to reduce RNN implementation challenges. The analysis and results of this chapter are based on the published papers [A2], [A3].*

4.1 Context and research objective

The effectiveness of the GRU-NNE equaliser proposed in Chapter 3 for overcoming non-linear SOA patterning effects suggests an intriguing possibility: since the GRU-NNE equaliser allows an SOA preamplifier to operate in gain saturation and can recover severe patterning impairments, can the associated SOA gain suppression be exploited 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 [84], [145], [146]. The experimental system setup described in this chapter is designed with this in mind, and the implications for future BMRx DR requirements outlined.

However, to realise any potential gains in DR performance the GRU-NNE must be robust to multiple system impairments, not just SOA patterning effects. Stringent component budgets will likely see 25G optics being utilised in future PON standards, and fiber dispersion will be a serious obstacle for 100 Gbit/s PAM4. Therefore in this chapter the GRU-NNE's performance is evaluated against a combination of expected impairments in a future PON system, including SOA patterning, fiber dispersion up to 91.8 ps·nm^{-1}, 25G bandwidth limitation, and electrical Rx saturation effects.

Alternative NNEs to the GRU-NNE model proposed in the previous chapter have been explored for future IM/DD PON scenarios [108], [141], [147], as well as non-ML based ones. For fiber dispersion, FFE+DFE solutions are favored due to their simplicity, and have been widely researched in the context of burst-mode reception at the PON OLT [22], [115]. Further, researchers using non-ML equalisers have also investigated the challenging 19.5 dB PON DR requirement with an SOA, such as in [148] where the authors investigate a look-up table pre-compensation technique on the transmitter side, and [65] which implements a Volterra non-linear equaliser (VNLE) to overcome SOA non-linearities and achieve 18 dB DR for 100 Gbit/s PAM4. In this chapter, a VNLE will be used as a performance and complexity benchmark, since it is a popular alternative to ML-based equalisers in IM/DD systems and widely researched [149]–[151].

In the previous chapter, the proposed GRU-NNE was shown to be more efficient than an FFE, or FC-NNE equalisers thanks to its feedback mechanism. However, there exists difficulties around the implementation of such an RNN-based equaliser in hardware, specifically the timing requirements of the feedback mechanism, as noted in [92], [119]. Parallel multi-symbol output schemes for neural network-based equalisers have been investigated in [104], and also implemented on FPGA in [92], [93], [119], with the authors of [119] realising an NNE based on LSTMs with competitive complexity for coherent transmission. Therefore, these multi-symbol techniques are investigated for the proposed GRU-NNE in the context of upstream PON DR and equaliser complexity.

4.2 Non-linear equalisation techniques

In this section, FFE/DFE and non-linear Volterra equalisers are introduced, along with their underlying theory, as alternative equaliser solutions for 100 Gbit/s PAM4 PON. Additionally, a more capable RNN neuron known as the long short-

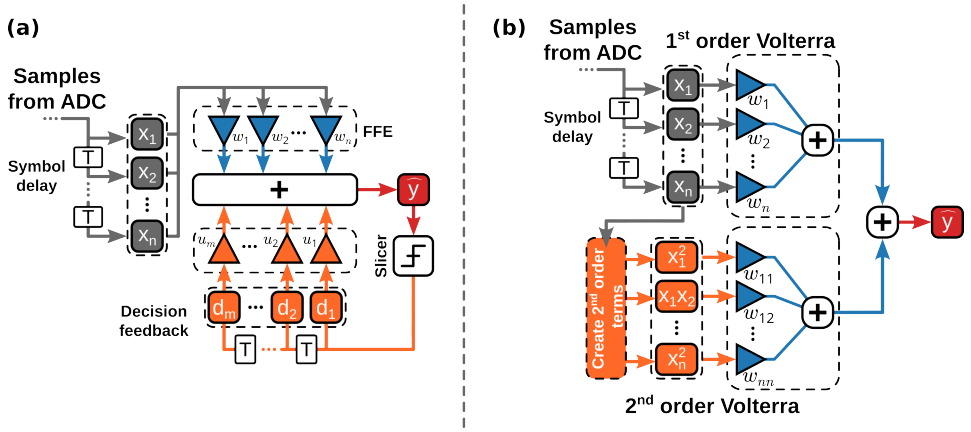


Figure 4.1: Schematics of (a) an FFE/DFE equaliser, and (b) a 2^{nd}-order non-linear Volterra equaliser.

term memory unit (LSTM) is described as an alternative to the GRU neurons introduced in Section 3.4.1. The application of each of these equalisation solutions to future high-speed PON is also reviewed, as well as complexity challenges associated with their hardware implementations.

4.2.1 Conventional FFE/DFE solutions

The FFE has already been introduced as a means to mitigate bandwidth limitations and limited fiber dispersion impairments, and is a favoured solution for adaptive equalisation of burst-packets in upstream PON [22], [67], [116]. However, FFE performance is limited due to it being a linear equaliser, which may be insufficient to guarantee PAM4 performance in future PON systems due to its high-linearity requirements. However, an FFE can be integrated with a DFE to extend its equalisation capabilities and tackle non-linear impairments. The FFE/DFE solution is popular within the PON research community for compensating more serious fiber dispersion impairments.

In a DFE, past symbol decisions are used as feedback to inform the current equalisation calculation. Symbol decisions are based on the action of a decision circuit, which takes as input the equaliser output sample $\hat{y}$, and returns a symbol decision $\hat{d}$ based on preset thresholds. For PAM4, the symbol decision $\hat{d}$ must represent one of the four modulation symbols in the PAM4 alphabet, i.e. $\hat{d} \in \{-3, -1, +1, +3\}$. The equation governing the DFE equaliser can therefore by written as:

$$\begin{aligned}\hat{y} &= w \cdot x_t + u \cdot \left[\hat{d}_{t-1} \dots \hat{d}_{t-m}\right], \\ &= w \cdot x_t + u \cdot D_t.\end{aligned} \tag{4.1}$$

In this equation, x_t are the raw equaliser input samples at time t, and w_x is the associated weight parameter vector associated with these input taps. The vector D_t consists of the previous m symbol decision outputs $\hat{d}_{t-1}, \cdots, \hat{d}_{t-m}$ from the decision circuit immediately prior to time t. These are combined in a weighted sum with the vector w_d which consists of the trainable weights for the m DFE feedback taps. The FFE/DFE schematic in Figure 4.1 (a) illustrates the operation of this equaliser for n sample inputs, and m decision feedback taps. DFE circuits with 2 to 3 feedback taps are not uncommon within the PON research literature [115], [116].

An FFE/DFE has low complexity in terms of RVMs and trainable parameters, with only $n + m$ parameters and associated multiplication operations per equalised sample. However, the feedback mechanism involved in the DFE is challenging to implement in hardware at sufficient speed for high-baud rate transmission, much like the feedback in an RNN. Modern communications DSP is usually implemented using techniques such as pipelining [122], which allows DSP algorithms to operate at many times the clock speed of the chosen hardware platform. However, feedback paths in DSP algorithms are difficult to pipeline, since each subsequent equalisation (for example) depends on the output of the previous equalisation operation in the case of a DFE. For this reason, mixed signal equaliser approaches are of recent interest [152]. This challenge is tackled in the context of the proposed GRU-NNE in Section 4.4.3.

4.2.2 Volterra non-linear equalisation

The Volterra non-linear equaliser (VNLE) is widely seen as an alternative to neural network based techniques. It can be thought of as a "super" polynomial fit to an equalisation problem, where subsets of the input taps are combined in all possible permutations, typically up to 2$^{\text{nd}}$ or 3$^{\text{rd}}$ order. The size of these subsets is decided according to the equaliser memory depth, which for 3$^{\text{rd}}$ order can be specified as (m_1, m_2, m_3), where the first order memory depth m_1 is the number of VNLE input taps, m_2 is the number of input taps from which 2$^{\text{nd}}$ order terms are formed, and so on. These equalisers are normally restricted to order 3, due

to the exponential scaling of VNLE kernels, i.e. trainable parameters, in practice.

Figure 4.1 (b) shows the schematic of a 2nd order VNLE, and conveys its essential operating principles. The equation which defines a 3rd order VNLE with memory depths $m_i = 2l_i + 1$ is given below:

$$\begin{aligned} y(n) \;=\; b + & \sum_{j_1=-l_1}^{l_1} k_1(j_1)x(n-j_1), \\ & + \sum_{j_1=-l_2}^{l_2} \sum_{j_2=j_1}^{l_2} k_2(j_1,j_2)x(n-j_1)x(n-j_2), \\ & + \sum_{j_1=-l_3}^{l_3} \sum_{j_2=j_1}^{l_3} \sum_{j_3=j_2}^{l_3} k_3(j_1,j_2,j_3)x(n-j_1)x(n-j_2)x(n-j_3), \end{aligned} \tag{4.2}$$

where b is a constant bias term, k_1, k_2, k_3 are the first, second, and third order kernels (i.e. weight parameters) respectively, and $x(n), y(n)$ represent the n^{th} input sample and equalised output of the VNLE respectively.

In the context of PON, Volterra-based equalisers have been proposed as an alternative to ML-based equalisers for compensating fiber non-linearities, such as in [151]. Further, some authors have posited that VNLE's can have lower complexity than neural network based equalisation.

However, similar to the FFEs and FC-NNEs considered in Chapter 3, the tap requirements of a VNLE strongly depends on the impairment it is applied to. This issue is compounded for a VNLE, whose non-linear modelling ability is based on second and third order combinations of its input taps, which means its trainable parameters scale exponentially with the increased memory depth required for worst case loud packets in PON. The number of parameters / kernels and RVMs for a 3rd order VNLE can be calculated using the following equations:

$$N_{kernels} = m_1 + \frac{m_2(m_2+1)}{2} + \frac{m_3(m_3+1)(m_3+2)}{6}, \tag{4.3}$$

$$N_{RVM} = m_1 + m_2(m_2+1) + \frac{m_3(m_3+1)(m_3+2)}{2}. \tag{4.4}$$

From these, and the VNLE defining equation above, it is clear that the calculation of a single i^{th} 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 [149], [150]. Pruning is a common

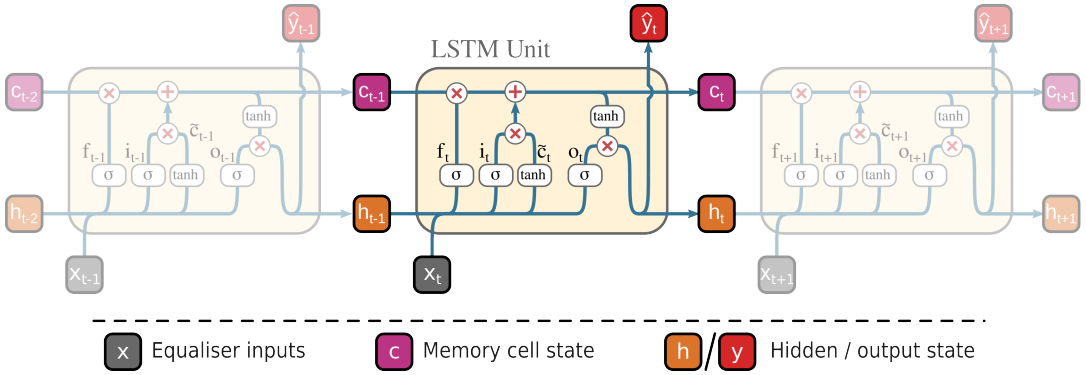


Figure 4.2: Schematic of an LSTM recurrent unit with memory cell c_t.

method employed in designing low-complexity VNLEs, as seen in [153].

4.2.3 Long short-term memory units

Recurrent neural networks which utilise feedback mechanisms to efficiently process time series data were introduced in Section 3.4.1, along with the GRU neurons on which the proposed GRU-NNE is based. However, another popular RNN unit is the long short-term memory unit, or LSTM [154], which has been investigated for equalising fiber dispersion in IM/DD systems [104], [147].

The GRU unit was originally designed as a less complex alternative to the LSTM unit which would be less resource intensive to process. 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 Figure 4.2. The equations governing LSTM operation are similar to those of the GRU in equation 3.12, and are given below:

$$\begin{aligned} \tilde{c}_t &= \tanh\left(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c\right), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t, \\ h_t &= o_t \circ \tanh\left(c_t\right). \end{aligned} \tag{4.5}$$

Here, the W, U matrices represent the trainable input and recurrent weights resepectively, and the b vectors are the associated biases. The gating activation σ is the sigmoid function, while $\circ$ represents the Hadamard operator. The three gating mechanisms operate in the same manner as that of the GRU units described in Chapter 3:

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right), \\ i_t &= \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right), \\ o_t &= \sigma\left(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o\right), \end{aligned} \tag{4.6}$$

where σ denotes the sigmoid activation function. The flow of information within an LSTM is clear from equation 4.5. For each input a candidate cell state $\tilde{c}_t$ is calculated. The LSTM cell memory is then updated as a linear combination of the previous cell state c_{t-1} and the candidate cell state, whose contributions to the new cell state c_t are modulated via the forget and input gates: f_t, i_t. The LSTM output is then calculated as a combiation of c_t and the output gate, o_t.

In terms of complexity, LSTM units have a larger footprint than GRUs, and involve more parameters and multiplication operations. The number of parameters and RVMs for a neural network layer with n_h LSTM units and n_i input samples can be calculated using:

$$N_{LSTM,params} = 4n_h \times (n_i + n_h + 1), \tag{4.7}$$

$$N_{LSTM,RVM} = n_h \times (4n_i + 4n_h + 3), \tag{4.8}$$

while similarly, for a GRU layer these can be calculated using the equations:

$$N_{GRU,params} = 3n_h \times (n_i + n_h + 1), \tag{4.9}$$

$$N_{GRU,RVM} = 3n_h \times (n_i + n_h + 1). \tag{4.10}$$

The next sections will investigate whether an LSTM-based NNE has an inherent performance advantage over the simpler GRU-NNE in the context of 100Gbit/s PAM4 PON scenario, where non-linear SOA patterning and fiber dispersion are a concern.

4.3 Upstream 100 Gbit/s PAM4 transmission

This chapter extends the investigation of GRU-NNE performance for 100 Gbit/s PAM4 by incorporating fiber dispersion and bandwidth limitation impairments alongside the isolated SOA patterning impairment explored in Chapter 3. Further, a novel optical DR compression scheme is investigated which could potentially

relax the DR requirements for subsequent BMRx electronics at the OLT. This scheme is based on exploiting the non-linear SOA gain saturation effect in conjunction with GRU-NNE to recover the resulting patterned signals. This effect is demonstrated here using an optical-electrical Rx configuration comprising an SOA and PIN photodiode in conjunction with an electrical Rx with adjustable gain is used.

4.3.1 Exploiting SOA gain saturation for DR compression

Due to the bursty nature of upstream PON transmission, a BMRx is typically used at the OLT [115]. These incorporate burst-mode transimpedance amplifiers (BM-TIA) which are designed to adapt their gain on a packet-by-packet basis to deal with the varying packet powers of TDM upstream PON. In this way these devices can achieve power leveling of packets with large incident power differences prior to signal processing. However, the design of a high-speed BMRx must balance large bandwidth, linearity, gain, and DR requirements. In the context of future 100 Gbit/s PAM4 transmission linearity is especially valued, while PON system requirements place emphasis on the large 19.5 dB DR. The state of the art was recently presented at ECOC'22, with the presentation of the first demonstration of 100 Gbit/s PAM4 linear BMRx which achieves 15.4 dB dynamic range [66]. However, this is less than the minimum 19.5 dB DR specified in 50 Gbit/s HS-PON, and the advantages of the proposed optical DR compression technique should be clear.

The experiment in this chapter uses a photoreceiver with adjustable electrical amplifier gain, in conjunction with an SOA preamplifier. It is demonstrated that the SOA preamplifier can be used to reduce the large incident optical dynamic range when driven into gain saturation, to such an extent that just two photoreceiver gain settings are required to cover a large optical DR of 28 dB. The GRU-NNE equalisation technique can then recover the transmitted signal from the resulting non-linear SOA patterning effects for loud packets. However, the use of just two electrical gain settings leads to non-linear Rx saturation effects also being observed, which the GRU-NNE must additionally mitigate.

Figure 4.3 shows the optical and electrical operating characteristics of the Rx configuration used in this chapter. In Figure 4.3 (a), the optical power incident on the photoreceiver (i.e. after the SOA preamplifier and ASE filter) is plotted with respect to the SOA input power when the SOA is kept at a constant bias of 100 mA. The SOA input DR is 28 dB, and extends from -22 dBm to $+6$ dBm

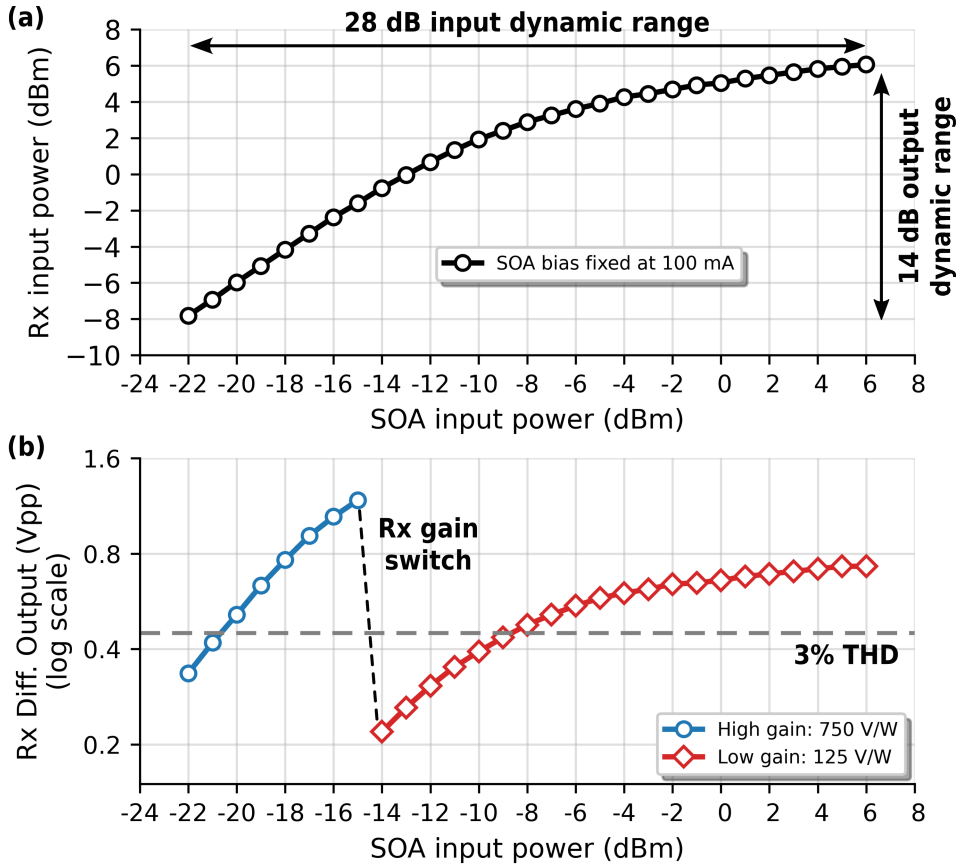


Figure 4.3: (a) The effect of SOA gain compression on input dynamic range to the SOA due to gain saturation; (b) graph showing the differential output swing from the photoreceiver amplifier with the gain-switch point indicated along with 3% Total Harmonic Distortion threshold.

optical power, while the photoreceiver input dynamic range after the SOA has been reduced by 50% to just 14 dB. This significant reduction in the DR is thanks to the non-linear SOA gain saturation, with no DR compression observed between -22 dBm and -15 dBm SOA input power, while from -14 dBm to $+6$ dBm is squeezed into just 7 dB photoreceiver input DR.

The degree of optical DR compression is such, that only two Rx gain settings at the photoreceiver can cover the entire 28 dB system DR. The Rx gain is changed from low to high setting at -15 dBm input to the SOA as shown in Figure 4.3 (b). The photoreceiver differential electrical output is also shown, representing the amplitude swing between the PAM4 outer symbols. The Rx has linear response ($< 3\%$ THD - Total Harmonic Distortion) up to a differential output swing of 450 mV, however after receiver gain switch from 750 V/W to 125 V/W, the Rx is again operating in its linear regime. However, as the SOA input power continues to increase, the SOA enters its gain saturated operating regime which introduces

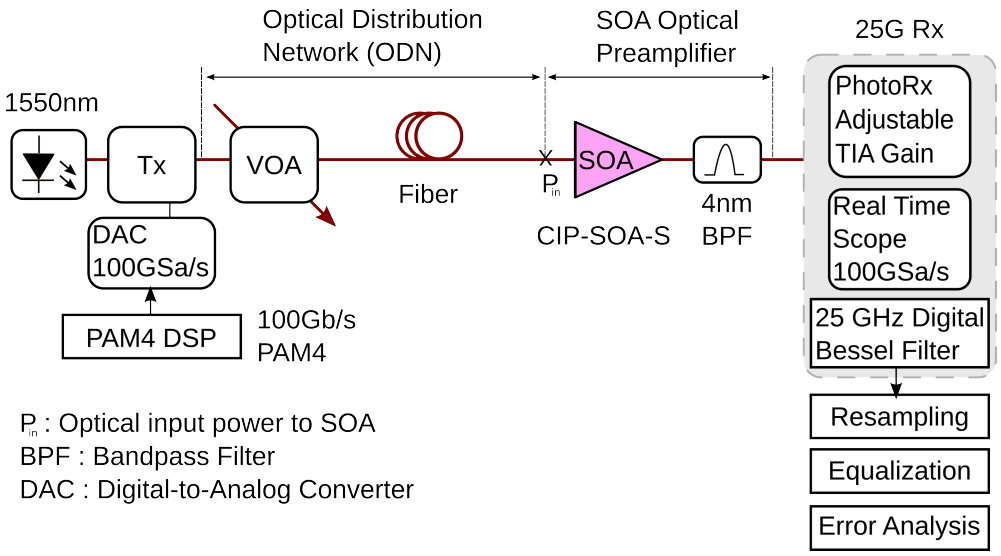


Figure 4.4: Experimental setup emulating 100 Gbit/s upstream PON transmission, using SOA gain saturation and neural network equalisation to enable > 28 dB dynamic range using only two electrical gain settings in the photoreceiver.

non-linear patterning effects, in addition to the Rx once more crossing the 3% THD threshold.

4.3.2 Experimental setup

Figure 4.4 shows the experimental setup used to emulate an upstream 100 Gbit/s PAM4 PON scenario in continuous mode in order to evaluate non-linear equaliser performances and the proposed DR compression technique. 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 an MZM, 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 Gbaud PAM4 signal with 6 dB extinction ratio. Linear pre-compensation corrects for system bandwidth limitations up to 33 GHz.

A VOA emulates variable ODN losses, allowing both loud and soft signal packets to be simulated in continuous-mode, with SOA input power varying between -26 and $+6$ dBm. This 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^{-1} was emulated using standard single mode fiber with dispersion parameter 17 ps/[nm · km] at 1550 nm. This exceeds the 77.1 ps·nm^{-1} dispersion

expected for 20 km transmission in the O-band using the upper-bound dispersion parameter value of 3.5 ps/[nm · km] calculated in Section 2.2.3, 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 is characterised in Chapter 3, Section 3.4, while the photoreceiver characteristics are shown in Figure 4.3 (b) and described in the previous section. In the context of the imagined OLT Rx configuration, to support the 19.5 dB dynamic range required the SOA will 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. But the non-linear SOA gain saturation acts to compress the system optical power DR from 28 dB to 14 dB at the SOA output, as shown in Fig 4.3 (a). The associated patterning effects introduced above the SOA saturation input power of -8 dBm, can be seen in the eye diagrams in Figure 4.5, eye 3, and these are later equalised using NNEs.

Waveforms are captured using a 100 GSa/s real time scope with 33 GHz bandwidth, while a 4^{th}-order Bessel filter imitates 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 chapter, equaliser training is carried out with a single pseudo random quaternary sequence (PRQS14) waveform [155], while multiple PRQS15 waveforms are used as test data and for BER estimation. For each SOA input power considered, the NNE solutions investigated are retrained with data captured for that specific input power, which guarantees optimal equalisation performance. Overfitting and performance overestimation as discussed in Section 2.3.4 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 were 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.

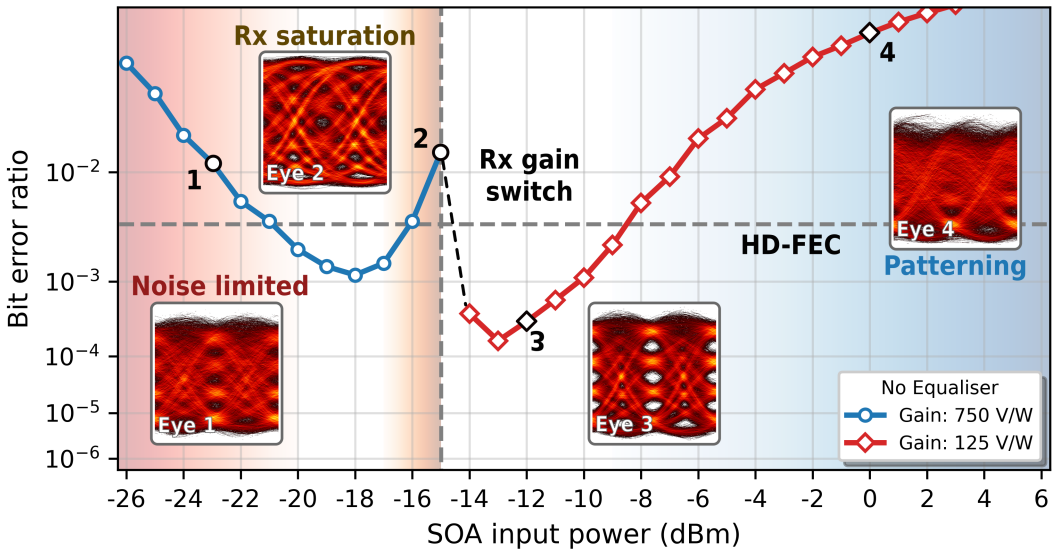


Figure 4.5: Eye diagrams of 100G PAM4 signals showing the effects of (1) Rx electrical gain saturation, (2) 25 GHz bandwidth limitation, and (3) combination of SOA patterning effect and Rx electrical gain saturation.

4.3.3 Identifying non-linear impairment regions

Figure 4.5 shows the BER performance without equalisation of the experimental setup described in the previous section, and using the proposed optical DR compression technique. The inset eye diagrams show the signal quality and various impairments effecting the 100 Gbit/s PAM4 signals over a wide range of SOA input powers.

The combined non-linearities of the SOA preamplifier and Rx electrical amplifier result in the distinctive "W" performance curve seen in Figure 4.5. For low powers below -18 dBm noise impairments related to SOA ASE and the Rx thermal noise dominate, as shown in Eye 1 in Figure 4.5. As the SOA input power increases beyond -18 dBm the effects of Rx saturation become apparent. 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 Figure 4.3 (b) and eye diagram 2 in Figure 4.5. After the Rx gain switch to 125 V/W the effects of Rx saturation are once again absent, as observed in Eye 3.

For high SOA input powers, the BER again exceeds the HD-FEC BER threshold. This occurs for input powers above the SOA input saturation power of -8 dBm, where the non-linear patterning effect begins to significantly distort the signal, as shown in Eye 4. However, Rx gain saturation effects are also present to a certain

degree, since the differential swing at the Rx exceeds the 3% THD threshold for SOA input powers above -9 dBm, as seen in Figure 4.3 (b).

The DR without equalisation is discontinuous between the two gain regimes, and clearly the proposed Rx configuration is unable to support the PON DR requirement. Further, non-linear equalisation is needed to support the DR compression scheme due to the effects of SOA patterning. The GRU-NNE non-linear equaliser is compared with other NNEs and non-linear Volterra equaliser in the next section.

4.3.4 Equalisation performance results

In this work, the HD-FEC limit of 3.8×10^{-3} BER is used to determine system sensitivity and achievable DR. BER is calculated using the counting method, and for each BER data point reported here, $\sim$130k symbols are used to estimate the error rate. Each equalisation solution must be retrained for each SOA input power considered, including the NNE.

The RNN equaliser structures evaluated here consist of a single hidden layer of 6 GRU, or LSTM units, followed by an output layer of 1 FC unit with linear activation which outputs the final equalised sample. 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, while its input taps are optimised for severely impaired high input power packets in our setup. The VNLE has memory depth structure $(41T, 13, 7)$, where the 41T input taps were chosen to mirror that of the FC-NNE, and the 2^{nd} and 3^{rd} order memory depths were optimised using brute force search.

Figure 4.6 (a) shows the 100 Gbit/s PAM4 system back-to-back (BtB) 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 the ONU, this corresponds to a total system loss budget of 30 dB, which exceeds the 29 dB outlined in 50 Gbit/s HS-PON.

The performances of the considered FFE and FC-NNE are also shown in Figure 4.6 (a), with each equaliser's tap numbers chosen to maximise performance. The maximum achievable BtB 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. Figure 4.6 (b) shows this achievable DR is reduced to only 11 dB when considering 81.6

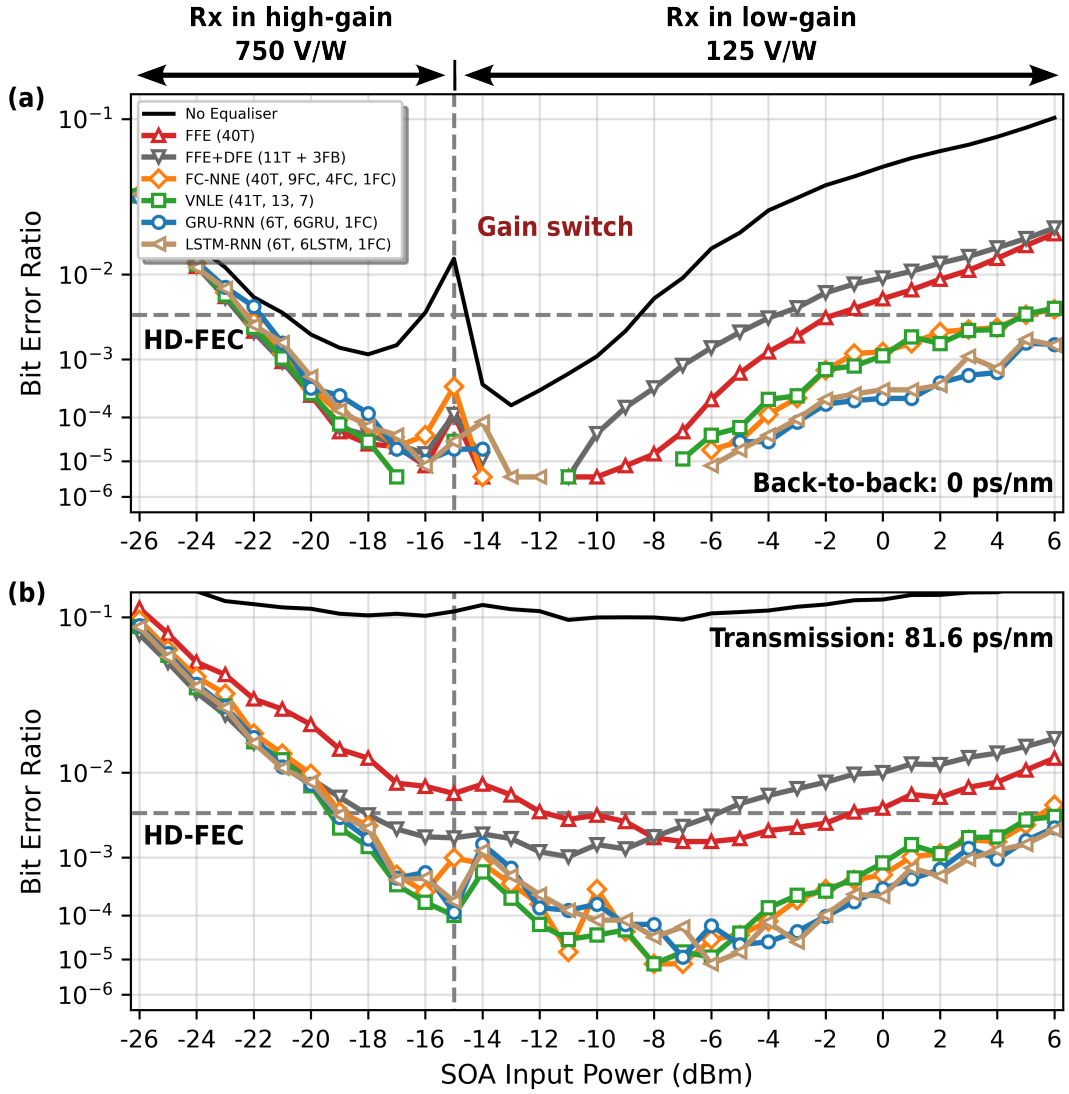


Figure 4.6: BER performance of feedforward (FFE, VNLE, FC-NNE), and feedback (FFE/DFE, GRU-RNN, LSTM-RNN) equalisation applied to 100 Gbit/s PAM4 (a) back-to-back, and (b) transmission experiments.

ps·nm^{-1} dispersion impairment, categorically ruling out linear equalisation as a solution for 100 Gbit/s PAM4 in this context. The FC-NNE using 40-T (symbol spaced) taps achieves 27 dB DR BtB, corresponding to a 6.5 dB increase in DR over optimal FFE. This is shown in Figure 4.7 and reflects the non-linear nature of the SOA patterning impairment, and corresponding equalisation capabilities of FC-NNE.

Optimal FC-NNE input taps were determined using Figure 4.7, where FC-NNE layer structure is kept constant while input taps are varied, and a strong correlation between input taps and BtB DR performance for both FFE, and FC-NNE is observed. This agrees with the results presented in Chapter 3, Section 3.4.2. Both

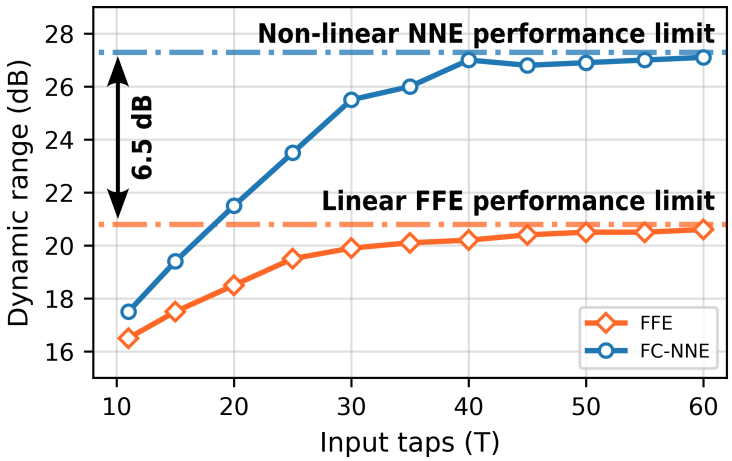


Figure 4.7: The strong dependence of BtB FC-NNE performance on the equaliser's number of input taps is shown, as well as the additional performance advantage a non-linear FC-NNE has over the linear FFE, resulting in an observed 6.5 dB dynamic range improvement.

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. As in Chapter 3, 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. Figure 4.6 (b) and 4.8 show the 40 tap FC-NNE matches the GRU-RNN equaliser DR up to 81.6 ps·nm^{-1} dispersion, indicating that FC-NNE learning capacity matches that of the GRU-RNN, but with the caveat of depending on a large number of input taps.

Also shown in Figure 4.6 (a) is an FFE combined with a DFE using 3 decision feedback taps. It achieves 18.5 dB dynamic range BtB, which is reduced to 12 dB with 81.6 ps·nm^{-1} 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 BtB DR.

The VNLE considered in this work has 41T taps and structure $(41T, 13, 7)$, for fair comparison with the FC-NNE. Figure 4.6 (a) shows it matches the performance of the FC-NNE, achieving $\sim$ 27 dB DR for BtB, and falls slightly short of the GRU-RNN performance. In the transmission case shown in Figure 4.6 (b), the VNLE has the same performance as the FC-NNE, while Figure 4.8 shows it also follows closely the same performance trend as the proposed GRU-RNN. From

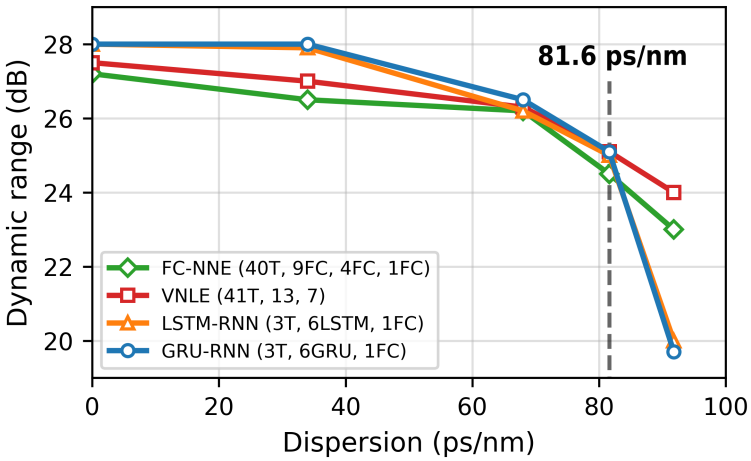


Figure 4.8: Achievable dynamic range of the Volterra, FC-NNE, GRU-RNN, and LSTM-RNN equalisers are shown to be similar when compared over a range of dispersion values, with no equaliser showing clear performance advantage.

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 BtB DR, and > 25 dB DR with up to 81.6 ps·nm^{-1} 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 Figure 4.6 demonstrates it overcoming 25G bandwidth limitations, fiber dispersion, and combinations of SOA patterning 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 the results in Chapter 3, 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 Figure 4.3 (b).

The more complex LSTM-RNN can only match the GRU-RNN DR performance in BtB and transmission experiments but cannot exceed it, as seen in Figure 4.8, 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 GRU feedback mechanism in the context of 100 Gbit/s PAM4 and the SOA patterning impairment considered here. Figure 4.8 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 requiring less input taps than FC-NNE.

4.4 Complexity analysis

Table 4.1 compares equaliser DR performance for BtB and transmission experiments in terms of the complexity metrics: trained parameters and RVM operations per symbol. The relevance of these metrics is described in the below section. A brief discussion on pipelining and the challenges presented by high-speed implementations of RNNs is also given.

4.4.1 Parameters, multiplications, and pipelining

In this thesis, offline software implementations of equaliser algorithms are investigated. However, the number of equaliser parameters, as well as the number of arithmetic operations needed per equalised symbol are important considerations for future hardware implementations.

Equaliser parameters can include the tap weights of an FFE, the weight and bias terms of the neurons of an NNE, as well as the kernel parameters of a VNLE. In hardware, these require memory resources to store, and must be transferred as signals across a DSP chip to the necessary adders and multipliers sufficiently fast so as to avoid timing delays in the digital circuit. These parameters and the raw signal samples are typically represented in fixed point format using anywhere from 6 to 12 bits for high speed DSP [156]. Proposed NNEs typically have many more parameters than conventional FFE solutions and it is not uncommon for proposed NNEs to comprise 1000s of parameters [104]. Therefore creating efficient NNE solutions with reduced parameters is an important design consideration when proposing solutions for eventual future real-time implementation. Further, lowering the number of bits used in sample and parameter quantisation without impacting an NNEs performance can significantly reduce arithmetic operation complexity and resource requirements [93], however this is best investigated through FPGA prototyping which is beyond the scope of this thesis.

The basic operations of a digital equaliser include addition and multiplication operations of fixed point numbers representing either input signal samples, or the equaliser parameters. These operations cost energy to perform, and contribute to the overall power usage of high-speed DSP chips implementing equalisation. Equaliser algorithms are typically compared according to the number of RVMs required per equalised output sample, since in terms of energy and resources these operations are the most expensive to carry out in hardware [118]. The lower the number of RVMs per equalised sample, the greater the computational efficiency of an equaliser.

Further, the pipelining technique is commonly implemented in digital hardware to obtain high frequency throughput [122]. This involves running multiple pipelines of sequential operations in parallel, which can increase DSP throughput. This is vital in achieving equalisation for 50 GHz signals such as 100 Gbit/s (50 Gbaud) PAM4, using lower frequency clock signals. However, the feedback mechanisms of RNN neurons such as GRUs are not susceptible to pipelining, since an RNN requires its output on the previous timestep in order to begin equalising input samples, meaning high bandwidth feedback channels are required. Multi-symbol equalisation aims to alleviate this requirement and is described in more detail below.

Table 4.1: 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	

4.4.2 Comparison of non-linear equaliser complexities

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 4.1. 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 1^{st}, 2^{nd}, and 3^{rd} order terms contribute 41, 91, and 84 kernels (i.e. weight parameters) 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 4.2.2 which exhibits exponential growth with respect to 2^{nd} and 3^{rd} 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 just 187 parameters, and 186 RVM operations per equalised symbol to achieve 28 and 25 dB DR performance for BtB and 81.6 ps·nm^{-1} 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.

4.4.3 Multi-symbol RNN 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 [92], this limits DSP throughput, unlike in feedforward based equalisers such as FC-NNEs and FFEs which can exploit pipeline digital implementation [93].

Multi-symbol equalisation techniques, such as that investigated in [104], 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 Figure 4.9. Whereas before, the output layer of such an equaliser consisted of a single FC neuron with linear activation, we now increase

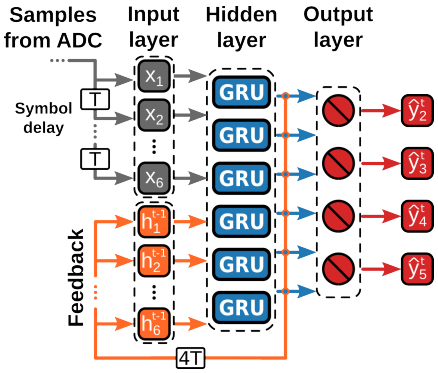


Figure 4.9: Multi-output GRU-based equaliser.

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 described in Section 4.4.1. Figure 4.9 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 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.

The results presented below compare the GRU- and LSTM-RNN equalisers performance using the multi-symbol equalisation technique, and benchmark them against the standard single symbol equalisation results already presented.

GRU versus LSTM multi-symbol equalisation

Figure 4.10 shows the achievable DR for BtB and 81.6 ps·nm^{-1} 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 BtB case, and 8 dB with 81.6 ps·nm^{-1} dispersion. The LSTM-RNN does not

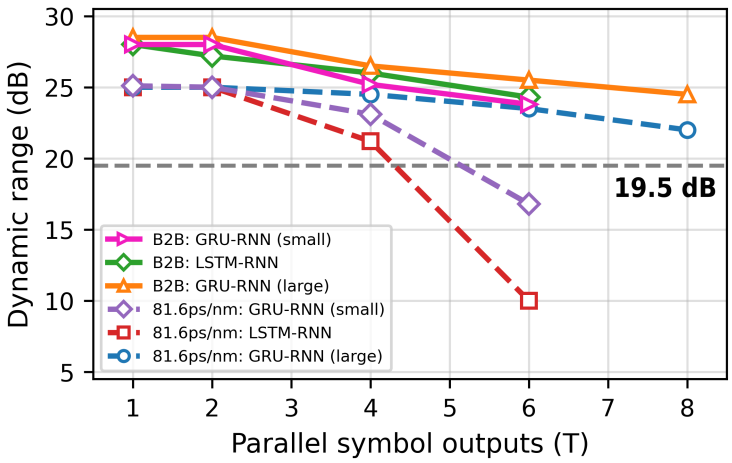


Figure 4.10: Graph showing GRU-RNN high dynamic range performance can be sustained when implementing multi-symbol output technique, at the cost of larger network size.

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 in the impairment scenarios considered.

Similarly, the larger GRU-RNN does not show significant performance gains in the BtB 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 4.1). 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 [119] successfully implementing an LSTM based equaliser on FPGA with 61 parallel outputs, and a 270 MHz clock frequency.

As discussed, the hardware implementation of RNN equalisers will be extremely challenging at 100 Gbit/s PAM4 data rate due to issues related to recurrent feedback mechanism latency. By implementing parallel outputs, we relax these timing requirements, and Table 4.1 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.

4.5 Summary

In this chapter, the proposed GRU-RNN is shown to be robust to a combination of device and fiber impairments, including up to 91.8 ps·nm^{-1} fiber dispersion, SOA patterning, 25G bandwidth restriction, and electrical Rx saturation effects, for 100 Gbit/s PAM4. In BtB this equaliser achieved a large 28 dB DR, and greater than 25 dB DR for up to 81.6 ps·nm^{-1} accumulated fiber dispersion.

In particular, the experimental setup used emulated upstream PON transmission in continuous mode, and allowed for an original DR compression technique to be demonstrated which exploits the SOA gain saturation effect. The GRU-RNN equaliser using only 3 symbol-spaced input taps allows the SOA to operate in its non-linear gain saturated regime which was exploited 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 computational complexity of any proposed neural network-based equalisation is a primary concern. Real-time operation requires low-latency processing, but neural networks, particularly those with recurrent architectures like GRUs, are computationally complex. This can increase power consumption and processing delays in practical systems, especially for high bit-rate applications like 100 Gbit/s PON. The GRU-NNE was therefore compared against several alternative

non-linear equalisers such as FC- and LSTM-based NNEs, as well as VNLE. It was found to outperform or match the performance of these equalisers while having lower overall complexity in terms of its total number of parameters and RVMs.

Further, multi-symbol equalisation techniques were 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 BtB DR for 4 parallel symbol output, and no significant advantage is observed using LSTM over GRU units, in spite of their increased complexity and more sophisticated memory mechanism. 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 BtB performance of a single output GRU-RNN equaliser. Therefore, multi-symbol equalisation could be vital in enabling real-time deployment in PON systems, on current hardware platforms.

Future work could focus on further optimising neural network architectures for real-time signal processing, on top of those investigated here. Techniques like network pruning or quantisation could help reduce the model size and inference time without sacrificing performance. However a more significant obstacle is the issue of NNE training in the context of burst-mode upstream PON transmission, where packet impairments can vary and the available time for NNE reconfiguration is constrained.

Chapter 5 seeks to address this, and introduces the novel SkipNet architecture as a solution. SkipNet is designed to allow for rapid, and low-complexity retraining of an NNE for real-time operation in burst-mode PON systems. It relies on a hybrid approach of fixed and adaptive components, SkipNet significantly reduces the computational burden and latency, making it better suited for real-time equalisation while maintaining high performance under challenging signal conditions. This innovation represents a key advancement in adaptive neural network-based equalisation for next-generation PON systems.

Chapter 5

SkipNet: Adaptive Neural Network Equalisation for PON

This chapter introduces SkipNet: an original algorithm, developed by the author, which enables adaptive neural network equalisation for future 100 *Gbit/s upstream PON transmission without the need for complex training algorithms such as Adam and backpropagation. It is based on novel skip connections which were first proposed by the author to enable PON packet-by-packet equalisation in [A4]. The algorithm's theory is introduced in the context of NNE training regimes proposed to date, and how these relate in turn to conventional adaptive equalisation methods in PON. Its performance is evaluated using a continuous mode* 100 *Gbit/s PAM4 link with an SOA preamplifier, where the main non-linear signal impairments are fiber dispersion and SOA patterning effects. This chapter is based on the expanded journal publication [A5].*

5.1 Context and research objective

Neural network equalisers have been widely proposed to achieve 100 Gbit/s and greater line speeds in future IM/DD [34], [104], [157] and PON systems [101]. This is due to NNEs ability to overcome non-linear signal impairments [158], and the expectation that such impairments will become more significant with the shift towards more advanced modulation formats (e.g. PAM4) implemented at ever higher baud rates. In the previous chapter, a GRU-NNE was demonstrated to have better performance than a range of conventional and alternative NNE

techniques, while having lower overall complexity in terms of parameters and RVMs. However, while challenges related to NNE complexity and hardware implementation have been addressed in the last chapter and received attention in the literature [92], questions remain concerning the suitability of NNE training complexity 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 ONUs, and where it can best influence overall network design. However, burst-mode signalling 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, reaching 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 [159]. Further, as described in detail in the past chapters: 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 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 equalisation is a possible solution to ensure such equaliser flexibility at the OLT and has been investigated for conventional DSP in both IM/DD [84] and coherent [160] PON systems. Here, the equaliser parameters are adapted on a packet-by-packet basis using training symbols embedded in each packet's preamble. But, as shown in Figure 5.1, the latest 50 Gbit/s HS-PON standards outline only 150 ns packet preamble time for OLT reconfiguration [12], which must accommodate BMRx settling time and clock and data recovery (CDR) alongside any equaliser adaption. Because of this, adaptive equalisers typically use variants of the simple LMS 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 [A3], [92], [101], [141], [157], 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" equalisation [161] which avoids adaptive training between packets also faces challenges. In theory, a sufficiently large NNE could be trained offline to equalise multiple ONU scenarios, allowing it to be deployed in a "fixed" state, such as in [162]. The authors of [114] propose such a "universal" NNE for 50 Gbit/s PAM4 PON which is pre-trained to compensate 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 equalise multiple ONU scenarios [114], and so consume more hardware resources and power [93]. Consequently, the cost of fixed NNE methods may be prohibitive given the numerous possible ONU impairment permutations.

This chapter introduces SkipNet: a novel adaptive NNE framework which enables the application of NNEs in PON, and which the author first proposed in [A4]. 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 equalisation 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 equalisation solution which intrinsically promotes transfer learning [163]. 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.

This chapter outlines the theory underpinning SkipNet and includes an investigation of its performance in a 100 Gbit/s/λ PAM4 PON scenario in continuous mode, with an SOA acting as an OLT preamplifier similar to Chapter 4. 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 DR performance of the SkipNet scheme is determined and

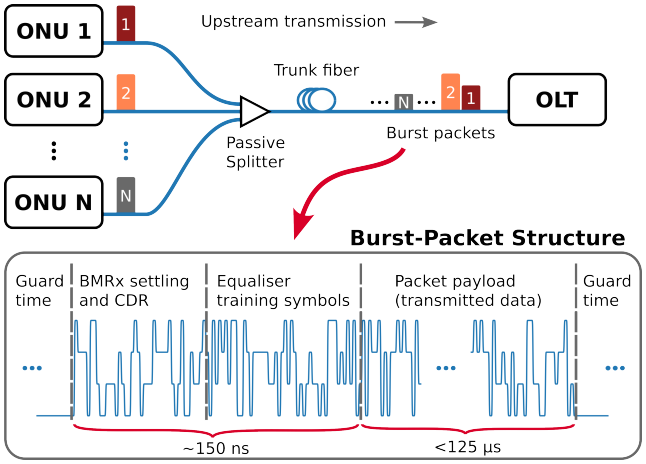


Figure 5.1: The burst-packet structure of upstream PON transmission according to the latest 50 Gibt/s HS-PON standards is shown. Only 150 ns preamble allocated for 25 Gbit/s upstream transmission, and 610 ns in a worst case scenario [12].

benchmarked against that of an “ideal” NNE which is trained using conventional methods. This "ideal" NNE solution is equivalent to the GRU-NNEs in Chapters 3 and 4 which have already been shown to extend SOA DR. Analysis of SkipNet’s operating mechanism and convergence properties is also provided.

5.2 Burst-Mode Equalisation at the OLT

Burst-mode signalling in high-speed PON presents a unique challenge for adaptive equalisation due to the extremely short packet preamble available for equaliser 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 [12], which corresponds to 7500 symbols for 100 Gbit/s PAM4. But, as previously mentioned, this preamble allocation must be shared among several OLT reconfiguration functions, as shown in Figure 5.1. Also, the frequency at which equaliser 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 equaliser adaption (e.g. ½, ¼, ...), 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 equalisation for PON is the training convergence rate, i.e. how many symbols it takes for the equaliser to converge. This must be balanced against the training algorithm’s complexity, which is intimately linked to power consump-

tion and clock cycles per parameter update. For example, an algorithm which halves the required training symbols but whose hardware implementation overly increases power consumption or requires more clock cycles per update may not be advantageous.

5.2.1 Existing adaptive equalisation solutions

The restrictive preamble time available for training has meant that existing PON adaptive equalisation research has focused on low-cost, low-complexity FFE/DFE techniques [164]. 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 signalling [115]. 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 [116] achieve FFE/DFE convergence within 250 bits using GS-LMS at 10 Gbit/s, while in [84] a 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 [165], while in [166] the authors estimate up to 416 training symbols to adapt coherent signal recovery algorithms for digital subcarrier modulation.

The FFE/DFE is the most popular equalisation 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 which is equivalent to that of an FFE/DFE. The FFE operating mechanism and training process using the LMS algorithm are discussed at length in Chapter 2, Section 2.3.1, but to clarify the discussion on adaptive NNE training the relevant equations are repeated here again. An FFE can be expressed as:

$$\hat{y}_t = w \cdot x_t, \tag{5.1}$$

where x_t and $\hat{y}_t$ represent the sampled signal input values and the equalised sample output of the FFE at time t. The FFE tap weights w can then be adapted packet-by-packet using the LMS algorithm in conjunction with an error signal

$\mathcal{E}$ for each equalised training symbol, and adapts the FFE tap weights using the gradient descent method:

$$w_{n+1} \coloneqq w_n - \gamma \cdot \frac{d\mathcal{E}}{dw}, \tag{5.2}$$

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 equaliser 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 [116].

5.2.2 Challenges facing upstream NNE

In contrast, neural network training is substantially more complex, and a detailed technical discussion of this can be found in Chapter 2, Section 2.3.2. In the context of PON, proposed NNE solutions require datasets which are considerably larger than available 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" which are processed separately using stochastic optimisation algorithms such as Adam [96] or RMSProp [97]. While the LMS algorithm can update an equaliser'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 [94], but this involves large numbers of multiplications, as discussed in [167], as well as being extremely memory intensive due to the caching of layer activations and multi-dimensional error gradients. For RNNs, these issues are exacerbated by the backpropagation-through-time algorithm which further propagates error signals along equaliser feedback connections [144]. 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 100 Gbit/s PAM4 PON.

Techniques to achieve adaptive NNE for short-reach and long-haul transmission have been proposed but are similarly unsuitable for PON. In [167], 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 [168] 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.

5.3 Adaptive NNE for burst-mode transmission

5.3.1 The neural network training barrier

Neural network training is covered in detail in Chapter 2 Section 2.3.2. The discussion here is designed to put this in the context of adaptive equalisation requirements for upstream PON, and the associated challenges facing NNE training schemes in such a scenario.

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 [94], which greatly increases the computational and memory resources per parameter update compared to the LMS algorithm.

To illustrate this, consider an FC-NNE composed of L layers of neurons, each of which is governed by the equations:

$$\begin{aligned} z &= w \cdot x + b, \\ a &= \varphi(z), \end{aligned} \tag{5.3}$$

where a is the activation output of the neuron, z is the weighted sum of the inputs, w, b are the trainable weight vector and bias parameters, and φ is a non-linear activation function. In order to update a neuron's weights in the l^{th} 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:

$$\frac{\partial \mathcal{E}}{\partial w^l} = \frac{\partial \mathcal{E}}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \ldots \cdot \frac{\partial z^{l+1}}{\partial a^l} \cdot \frac{\partial a^l}{\partial z^l} \cdot \frac{\partial z^l}{\partial w^l}, \tag{5.4}$$

where the ouput layer activation is labelled a^L. While backpropagation can be seen as an efficient implementation of equation 5.4[1], 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. Ref. [167] contains a detailed discussion of backpropagation complexity and derives the associated RVMs involved in parameter updates.

However, simple stochastic gradient descent with backpropagation is not generally sufficient for optimising 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 equaliser 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 equaliser adaption must be fast and hardware efficient.

5.3.2 SkipNet: Adaptive neural network equalisation

[1] See the backpropagation equations in Chapter 2.

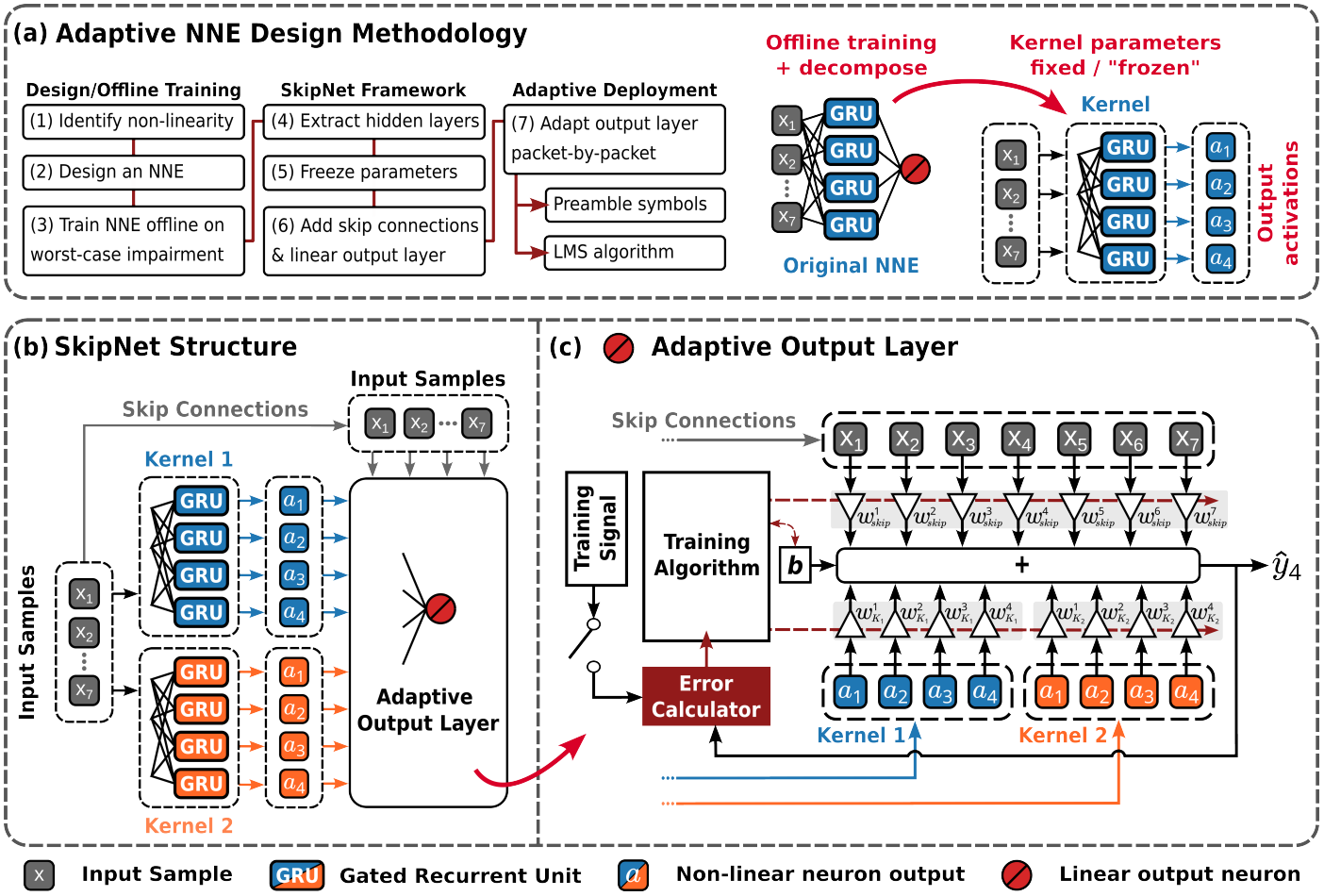


Figure 5.2: (a) Outlines the design methodology behind creating an adaptive SkipNet equaliser, and how kernels are derived from the hidden layers of an ideally trained NNE before (b) being integrated into the SkipNet framework. Flexible and rapid adaptive NNE is then possible using a linear output neuron and the adaptive circuit shown in (c), which is equivalent to conventional adaptive FFE/DFE training.

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 equaliser adaption, and relies only on LMS once deployed. This makes its adaption complexity equivalent to that of previously studied FFE/DFE solutions.

Figure 5.2 (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 linear output layer is trained offline on the desired worst-case patterning scenario using conventional "ideal" methods, i.e. backpropagation algorithm in conjunction with Adam optimisation and mini-batch training. The hidden layers of this ideally trained NNE are responsible for its non-linear equalisation capabilities, and so are extracted by removing and discarding the linear output layer. The parameters of these hidden layers are then "frozen", i.e. are kept fixed thereafter and are 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:

$$a_t^{L-1} = K(x_t), \tag{5.5}$$

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 which is chosen to be a single fully-connected neuron as described by equation 5.3, with linear activation function $\varphi(x) = x$. Skip connections are also added at this point, which feed the raw equaliser input samples to the adaptive output layer by bypassing the NNE kernel, as seen in Figure 5.2 (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 Figure 5.2 (b). SkipNet's output layer can then be described by:

$$\hat{y}_t = w_{skip} \cdot x_t + w_k \cdot K(x_t) + b, \tag{5.6}$$

where $\hat{y}_t$ are equalised samples, and w_{skip}, w_k are the adaptive parameters for the raw signal samples x_t and non-linear kernel output $K(x_t)$ respectively. Figure 5.2 (c) illustrates the operation of equation 5.6, and how this output layer can now be adaptively trained using LMS optimisation method and preamble symbols.

The skip connections are vital for realising equalisation 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 equalise 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 helping to 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 [169], and have previously been proposed as "cascade" connections to boost RNN equaliser performance for short-reach PAM4 transmission, and lower the number of epochs required during NNE training [170]. The SkipNet structure shown in Figure 5.2 (b) is also similar to that proposed in [121]. Crucially however, neither of these works exploit 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 equation 5.6 and adaptive FFE/DFE have equivalent complexity, since both use LMS and stochastic gradient descent to achieve fast adaption, as described in equation 5.2. 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 [164], [171].

5.3.3 Multiple kernels for flexible non-linear equalisation

However, in future PON there could exist multiple non-linear impairments which may be present in isolation or in combination for any given packet. For example, Chapter 4 considered gain saturation induced SOA patterning which only effects high-power packets, as well as scenarios where a packet suffers combined SOA patterning and extreme dispersion effects. In this chapter, separate NNE

kernels are trained for 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:

$$\hat{y}_t = w_{skip} \cdot x_t + w_{K_1} \cdot K_1(x_t) + w_{K_2} \cdot K_2(x_t) + b, \tag{5.7}$$

where now there are two sets of adaptable parameters for the two non-linear kernels: w_{K_1} and w_{K_2}. Significantly however, equation 5.7 is not fundamentally changed from equation 5.6, 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 equalisation, since these kernels target specific, isolated impairments which require only low-complexity NNE blocks. Each kernel is trained only once, and is deployed in a fixed state, thereby removing NNE training overheads.

A note on transfer learning

Transfer learning is defined in [163] as follows:

> Transfer learning aims at improving the performance of target learners on target domains by transferring the knowledge contained in different but related source domains

As such, it could be said that transfer learning is intrinsic to equation 5.6 and 5.7, since the adaptive output layer allows kernels to contribute to the equalisation of channel impairments for which they may not have been specifically trained. Transfer learning has previously been proposed for short reach PAM4 links in order to reduce training symbol requirements when developing NNEs for varying fiber length and bit-rate scenarios [117]. However, this work still relies on conventional neural network training to update an NNEs hidden layer weights and biases, and targets reconfigurable DCI technology, rather than rapid adaption for upstream PON.

A full analysis of the potential of transfer learning between SkipNet kernels trained on disparate non-linear impairments is unfortunately beyond the scope of this work. However, results in the following sections support the idea that the adaptive output layer mechanism promotes such knowledge sharing to the benefit of equalisation performance.

5.4 Evaluating SkipNet performance for 100 Gbit/s PAM4 PON

As in previous chapters, the proposed algorithm's performance is evaluated in terms of system DR and receiver sensitivity, which are calculated with respect to the SOA preamplifier input power and the HD-FEC BER threshold of 3.8×10^{-3}. Each reported BER value is estimated using the counting method applied to ~ 520000 captured symbols.

The GRU-based NNEs explored in previous chapters form the base NNE structures from which all reported 'Ideal', 'Fixed' and SkipNet kernel NNEs are derived, and the proposed SkipNet equalisers can be seen as an evolution of the GRU-NNEs presented thus far. This common NNE structure is composed of a single hidden layer of 4 GRU units fed by 7 symbol-spaced (-T) input taps as shown in Figure 5.2. The output layer of all the NNEs consists of a single neuron with linear activation function, equivalent to equation 5.3 for the 'Fixed' and 'Ideal' NNEs $\varphi(x) = x$, while the operation of the adaptive SkipNet equalisers are equivalent to equation 5.6 and 5.7. As discussed in Chapter 2 Section 2.3.4, all training datasets consist of randomly generated symbol sequences in order to avoid overfitting, while test datasets comprise repeated PRQS15 patterns.

5.4.1 Experimental setup

The 100 Gbit/s PAM4 experimental setup shown in Figure 5.3 (a) is used to evaluate the proposed SkipNet adaptive equalisation scheme in continuous mode. This setup can produce a wide range of possible packet impairments, including SOA gain saturation induced patterning effects, fiber dispersion, bandwidth limitations, as well as combinations of these, as shown in Figure 5.3 (c). An ideal high-power Tx is emulated using an MZM in conjunction with a high power EDFA booster amplfier. 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 are linearly compensated up to 40 GHz prior to transmission.

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. Due to a lack of available O-band devices, this experiment is carried out at 1550 nm, and dis-

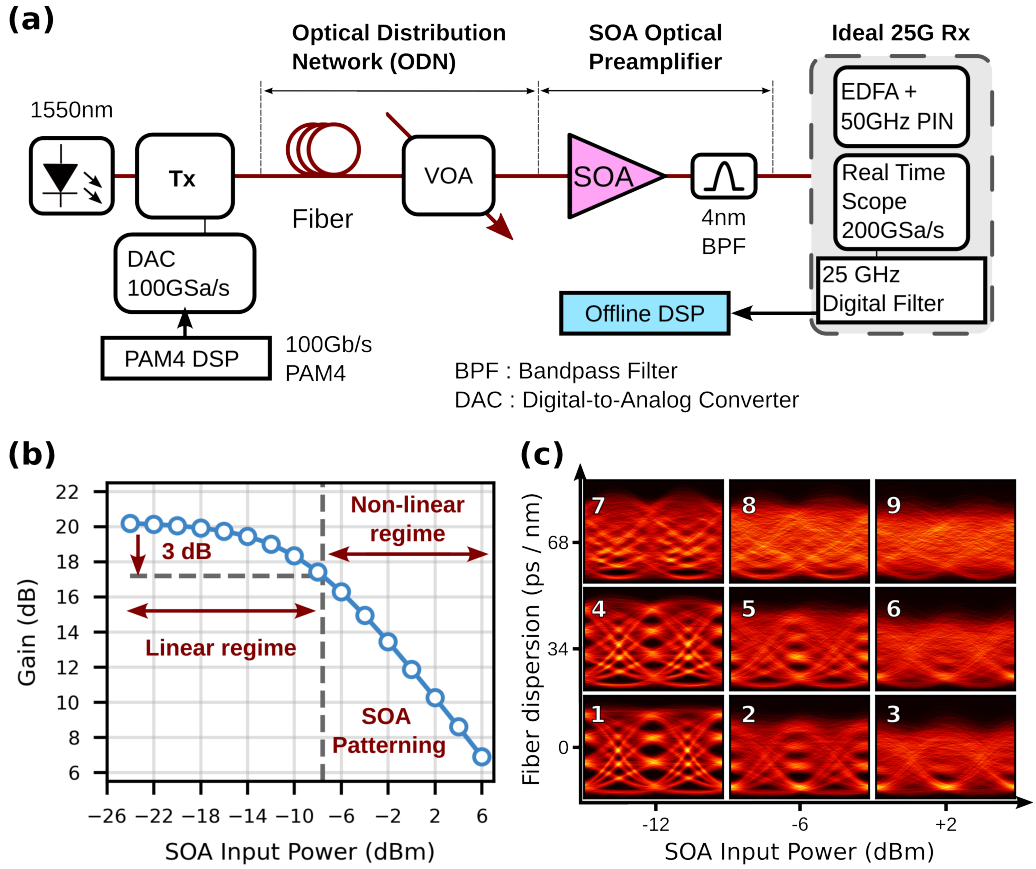


Figure 5.3: (a) Shows the experimental setup which incorporates an SOA pre-amplifier whose gain characteristics are shown in (b). This setup allows for a diverse range of packet impairment permutations to be investigated, as evidenced by the eye diagrams in (c).

persion impairments up to 68 and 81.6 ps/nm were realised by changing the fiber transmission distances to 4 and 4.8 km respectively. These were the closest achievable dispersion values to the HS-PON upper limit value of 77 ps/nm 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.

The OLT comprises an SOA preamplifier kept at constant bias followed by an ideal Rx comprising an EDFA and 50 GHz PIN, with 25 GHz 4th 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 Figure 5.3 (b), and the SOA saturation input power is -8 dBm. Above this power SOA gain-saturation induced patterning effects are clearly evident, as shown in Figure 5.3 (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 [A1], before even the effects of fiber dispersion induced sensitivity penalties have been considered. The severe effect of fiber dispersion can be seen in Figure 5.3 (c) eye diagram 4, while eye diagram 9 shows the combined effect of 68 ps/nm dispersion and SOA patterning effect induced by an input power of $+2$ dBm. Further, the OSNR values of signals output by the SOA range from 26 dB at -26 dBm input SOA power, all the way up to 41 dB at -12 dBm SOA input power, as seen in Figure 3.4 (b) in Chapter 3. This means that using this setup, the SkipNet equaliser convergence performance can be tested over a wide variety of signal and noise power levels, with thermal noise due to the Rx photodiode also a factor at low powers.

5.4.2 Back-to-back and transmission performance

Figure 5.4 (a) shows the 100 Gbit/s PAM4 system 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 equalisation 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 equaliser with two decision feedback taps (2-FB) is trained using the LMS algorithm and 15000 symbols, and achieves -22.6 dBm sensitivity and $\sim$ 16.0 dB DR. Both of these results are below the minimum 19.5 dB DR recommended by the HS-PON standards. The improved Rx sensitivity of the 7-T FFE compared to the no equalisation case is due to the bandwidth limitations of the Rx setup used in this experiment.

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 ~ 16000 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 [A1], [A3] which report that GRU-based equalisers 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,

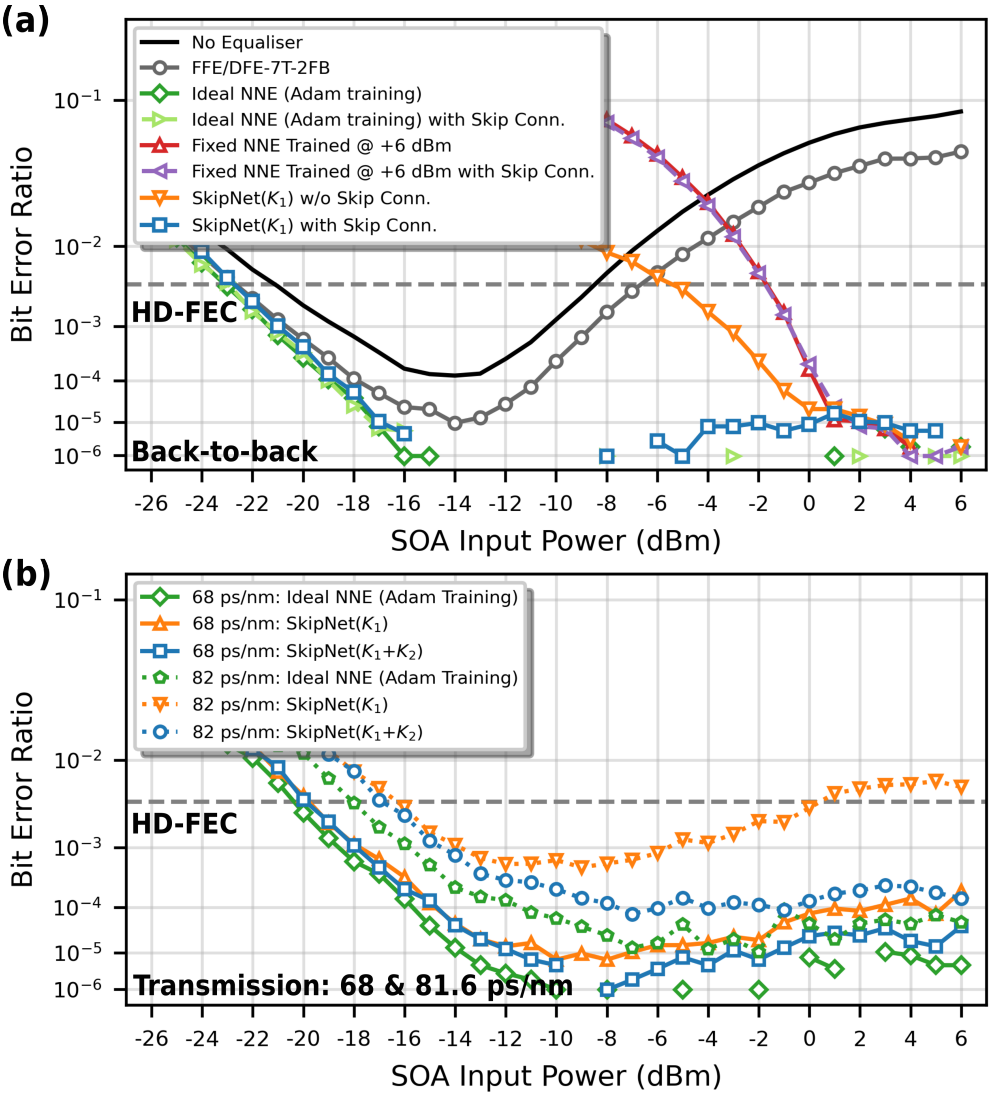


Figure 5.4: (a) Shows the BtB performance of the SkipNet(K_1) equaliser with and without skip connections, versus ideally trained and fixed (non-adaptive) NNEs. In (b), the transmission performance of single (K_1) and double ($K_1 + K_2$) kernel SkipNet equalisers underlines the flexibility of the scheme to accommodate additional low-complexity kernels for greater non-linear capability.

a version of the ideal NNE with skip connections is also evaluated. However as shown in Figure 5.4 (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 equalisation, as described below. Similarly, adding skip connections did not improve on the ideal NNE performance reported in Figure 5.4 (b).

A SkipNet equaliser based on the kernel K_1 and integrated skip connections is also reported in Figure 5.4 (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 15000 training symbols to adapt its linear output layer for each distinct SOA input power. If the skip connections are removed the adaptive SkipNet equaliser fails completely to equalise signals with input power below -6 dBm, which includes unimpaired signals that previously required no equalisation. 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 5.3 5.3.2: 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 [A4] where we first proposed using skip connections.

Figure 5.4 (b) shows that for dispersion up to 68 ps/nm the SkipNet(K_1) equaliser 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) equaliser 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) equaliser which 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

Table 5.1: Packet-by-packet training complexities

Equaliser	Parameters		Optimiser
	Fixed	Trainable	
Ideal NNE	0	149	Adam*
SkipNet(K_1)	144	12	LMS
SkipNet($K_1 + K_2$)	288	17	LMS

* Adam + backpropagation-through-time + mini-batches + early stopping

50G outlined in the HS-PON recommendations [12]. Therefore, while the 19.5 dB DR target is met, the associated sensitivity penalty may be impractical in a real network setting.

Table 5.1 compares the per-packet training complexity of the Ideal NNE with that of the SkipNet equalisers in terms of trainable parameters, and required optimisation algorithm. Since K_1 and K_2 each comprise a single hidden layer of 4 GRU units, the number of fixed parameters N for each kernel can be calculated using equation 4.9:

$$N = 3 \cdot n_{GRU} \cdot (n_{in} + n_{GRU} + 1)\,, \tag{5.8}$$

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 equation 5.6 and 5.7, 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) equaliser having over $10\times$ 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.

5.4.3 Achieving low-complexity, targeted NNE

The heatmaps in Figure 5.5 (a) and (b) 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 equaliser was created by pre-training kernel NNEs offline and evaluating the equaliser's adaptive performance over the same input power range. Interestingly, if kernel K_1 is pre-trained at any one SOA input power above

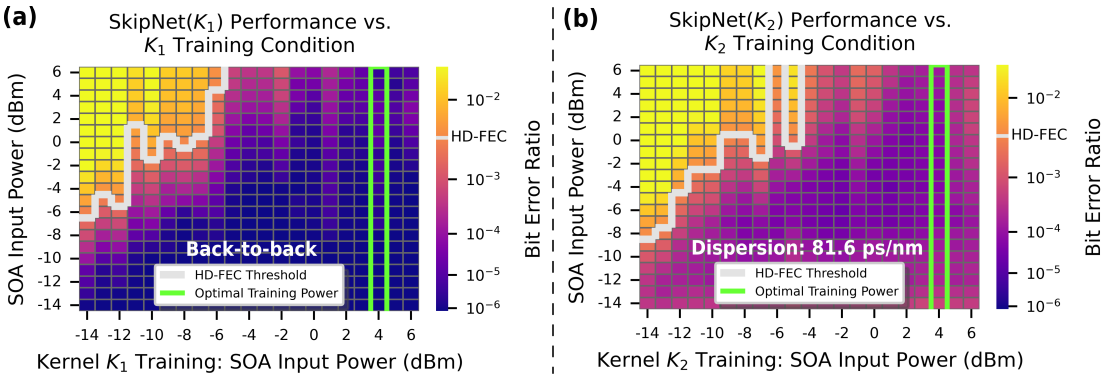


Figure 5.5: These heatmaps in (a) and (b) show how K_1 and K_2's training conditions were chosen, by plotting kernel training power versus the BER performance of the resulting SkipNet equalisers.

-6 dBm, i.e. in the non-linear operating regime of the SOA, then the resulting SkipNet(K_1) equaliser can achieve below the FEC threshold for all non-linear SOA input powers considered. This can be understood by considering the results in Chapter 3, where the GRU activations could be seen to have a relaxation time in Figure 3.17 (c) which is related to the SOA gain recovery time. Therefore, at the same bit-rate, the SkipNet equaliser must only modulate the GRU activations for different SOA input powers, but importantly the GRU gating mechanism which has learned the temporal characteristics of the patterning effect does not need to change.

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 equaliser SkipNet($K_1 + K_2$) which 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 equalise a wide range of packet impairments.

Figure 5.6 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}, w_{skip} of the SkipNet($K_1 + K_2$) output layer for different SOA input powers and dispersion values. If the RMS value of the converged weights is small, or close to zero, then the output activations of the associated kernel or skip connections are suppressed

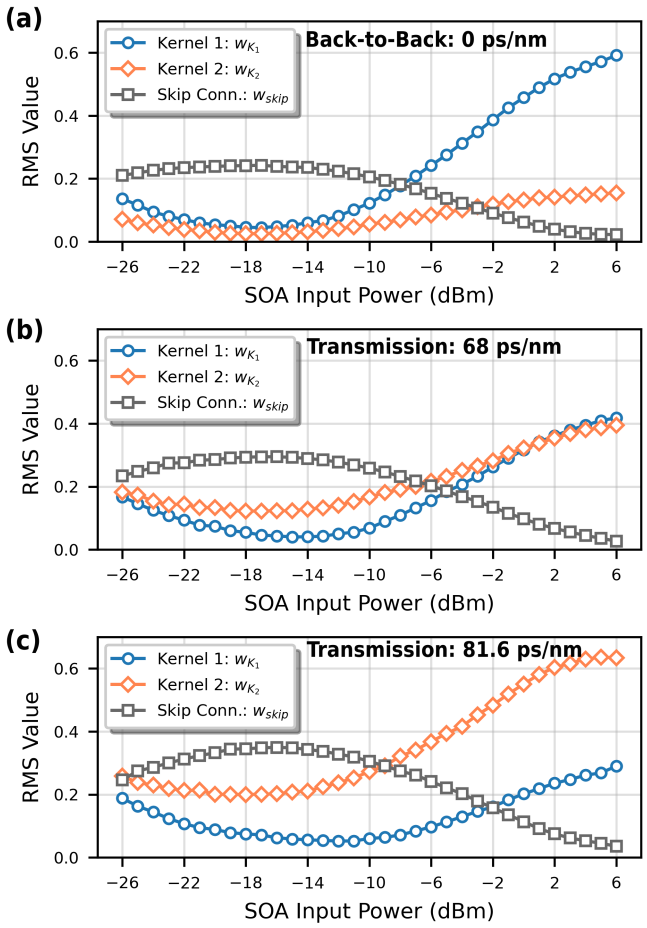


Figure 5.6: The RMS values of the converged weight parameters in the SkipNet output layer indicate kernel and skip connection relevance for (a) back-to-back (0 ps/nm), (b) 68 ps/nm, and (c) 81.6 ps/nm transmission scenarios. The LMS algorithm can clearly select between kernels and skip connections to target different packet impairments.

and have negligible effect on the SkipNet output. Similarly, a high RMS value indicates that the kernel outputs / skip connections will dominate the SkipNet output layer calculations. In the case of 0 ps/nm dispersion, Figure 5.6 (a) shows that the skip connections dominate the equalisation process up to -8 dBm, above which the non-linear kernel K_1 which 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 the SOA is known to be operating in its non-linear regime, as shown in Figure 5.3 (b). For the transmission cases of 68 and 81.6 ps/nm dispersion, Figure 5.6 (b) and (c) show K_2's growing contribution towards the final equalisation 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 equalisation 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 value, or energy, of both kernel's converged weights are significantly higher in the SOA patterning regions than the corresponding RMS value 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 equalisers, since the choice of learning rate for LMS-type optimisation algorithms is strongly dependent on the target signal distribution. Evidence of this can be seen in Figure 5.7.

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 optimise LMS convergence.

5.4.4 Preamble training requirements

The results in Figure 5.7 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 equalisation and coherent PON signal recovery preambles discussed in Section 5.2.

Figure 5.7 shows the minimum number of symbols required by the SkipNet($K_1 + K_2$) equaliser to achieve below the HD-FEC threshold with probability $> 95\%$ (assuming a normal distribution), for both the LMS and GS-LMS training algorithms. For each SOA input power and accumulated dispersion value considered, 8 acquisitions of the data "packet" comprising train, test, and validation sequence were captured during experiments. The statistics used in Figure 5.7 were compiled by independently training and testing the SkipNet($K_1 + K_2$) model 5 times

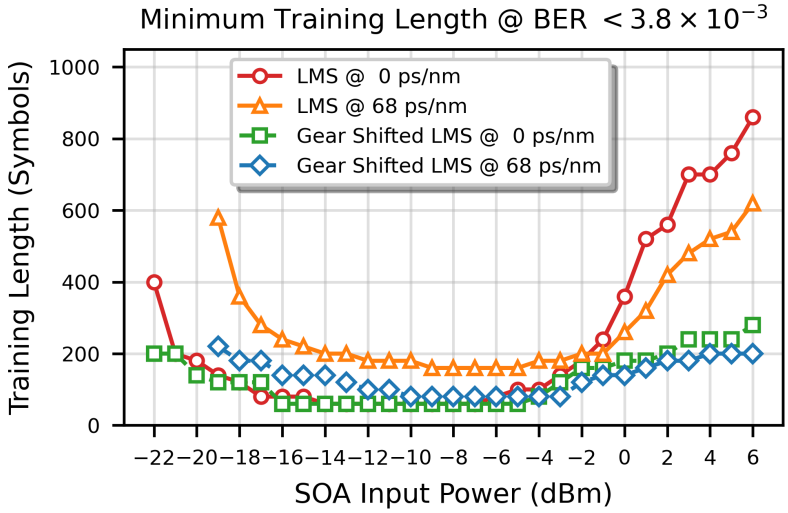


Figure 5.7: This graph shows the minimum number of training symbols required by SkipNet(K_1+K_2) to achieve BER under the HD-FEC threshold with probability $> 95\%$ across a range of SOA input powers. Rapid adaption can be achieved using the GS-LMS algorithm, comparable to that of existing FFE/DFE techniques [116].

for each of these 8 captured packets, resulting in 40 error rates with which to estimate the true model error rate. For each training, the output layer weights of the SkipNet model were initialised to zero, and random shuffling of the training input ensured each repeated training operation was unique. After training, the equaliser was applied to the associated packet payload comprised of ~ 65000 PRQS symbols, and the mean number of errors accumulated for each input power, dispersion value, and number of training symbols were recorded. This was repeated for various equaliser training lengths between 50 and 1000 symbols, and the results used to estimate how reliably the equaliser 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 Figure 5.6. 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 but which is outside the scope of this work.

The GS-LMS algorithm greatly reduces the training symbols required in such scenarios where the kernels are guiding the equalisation, 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($K_1 + K_2$) equaliser in conjunction with GS-LMS and less than 250 symbols.

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

5.4.5 Achievable system dynamic range

Figure 5.8 summarises the implications of adaptive NNE for PON by plotting the achievable DR versus accumulated dispersion of the adaptive SkipNet($K_1 + K_2$) equaliser for 100G PAM4 using an SOA preamplifier at the OLT.

While the HS-PON standards call for 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 15000 symbols, as seen in Figure 5.4 (a) and (b). This is in contrast to the conventional FFE/DFE with 7-T taps and 2 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.

Figure 5.8 also shows how the DR values achieved by the adaptive SkipNet($K_1 +$

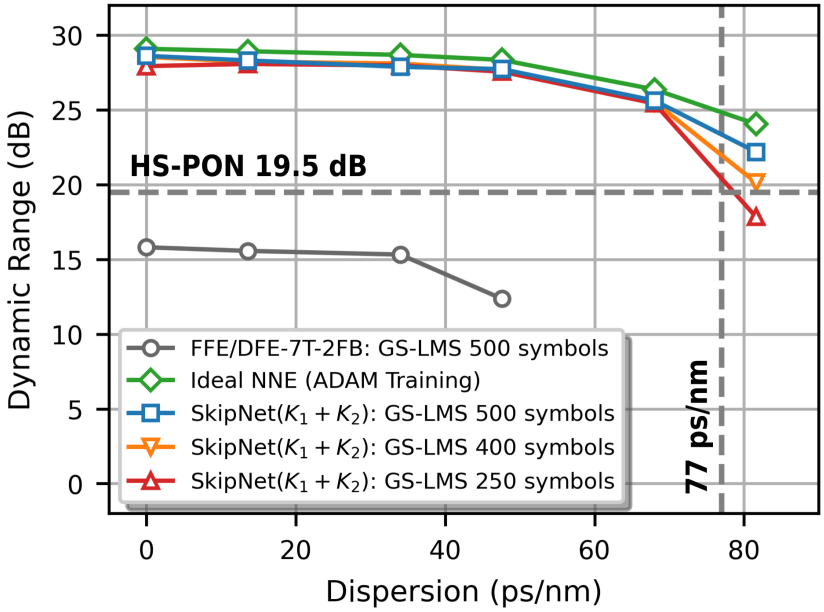


Figure 5.8: The SkipNet equaliser achieves > 25 dB DR for up to 68 ps/nm accumulated dispersion using just 250 symbols, clearly outperforming the adaptive FFE/DFE equaliser which cannot even support 19.5 dB PON DR requirement in the BtB scenario.

K_2) equaliser using just 400 training symbols approach those of an ideally trained NNE. The adaptive equaliser 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 behaviour it will be possible to close this gap to the ideal NNE performance using short training preamble.

5.5 Summary

In this chapter, an original adaptive framework for NNEs is introduced which is designed specifically for burst-mode PON systems. While previous chapters explored GRU-based NNEs, which successfully mitigated non-linear SOA patterning effects in 100 Gbit/s PAM4 transmissions, these methods did not adequately account for the complexities associated with neural network training in burst-mode scenarios. SkipNet represents a significant advance, achieving comparable dynamic range performance to GRU-NNEs while drastically simplifying the training process. The SkipNet architecture is the first of its kind to offer adaptive, packet-by-packet equalisation for future 100 Gbit/s PON deployments. Its design avoids the computational overhead typically associated with neural network

training methods by leveraging pre-trained kernel NNEs and a lightweight LMS adaptation algorithm.

It avoids the extensive training demands of traditional neural network training algorithms such as backpropagation and Adam optimisation, which often require hundreds of thousands of symbols for to converge. Instead, SkipNet achieves rapid convergence within as few as 250 symbols, even under varying signal impairments and burst-mode transmission conditions. This result is on par with conventional FFE/DFE training overheads. SkipNet therefore offers a more practical solution for real-time signal processing in optical communication systems.

The SkipNet equalisers presented here achieve similar large system DR without complex training, and show only small performance penalties when compared to the GRU-NNEs presented in the earlier chapters. As far as we are aware, SkipNet is the first adaptive neural network-based equalisation scheme targeting future 100G PON which 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. The results presented in this chapter are summarised below.

The proposed SkipNet adaptive NNE framework 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, 22.9 dB DR is achieved for 81.6 ps/nm dispersion, which is above the 19.5 DR requirement outlined in the 50G HS-PON standard. The SkipNet equaliser avoids complex neural network training since it's kernel NNEs undergo once-off pre-training prior to deployment. The simple LMS algorithm is shown sufficient to adapt SkipNet equalisers 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 which need tens or even hundreds of thousands of training symbols, and are wholly unsuitable for adaptive equalisation in upstream burst-mode PON. It is expected that SkipNet's convergence can be made even faster using methods developed for

LMS-like optimisation in FFE/DFE solutions.

In contrast to contemporary NNE solutions, which rely on complex, resource-intensive training algorithms unsuitable for adaptive equalisation in upstream burst-mode PON, SkipNet's pre-training approach and rapid adaptability mark a significant departure from existing methods. Furthermore, SkipNet's adaptive framework is not limited to PON systems; its potential extends to environments characterised by dynamic channel conditions due to environmental variations or system degradation, as well as applications demanding plug-and-play capabilities.

In conclusion, SkipNet offers a breakthrough in adaptive neural network equalisation, combining high performance with computational efficiency. Its ability to deliver rapid convergence with minimal computational overhead positions it as a viable solution not only for next-generation PON systems but also for a broader range of optical and wireless communication scenarios where real-time adaptability and low power consumption are paramount.

Chapter 6

Conclusion

In this final chapter, a brief summary of the novelty and original contributions of this work is provided, before possible future directions are discussed.

Future > 100 Gbit/s PON systems will require advanced signal equalisation techniques to meet challenging system metrics, such as the optical loss budget and dynamic range requirements. Advanced modulation in the form of PAM4 is an attractive candidate to deliver such high bit-rates while reusing 25G class optics, and SOAs can be integrated in order to compensate PAM4's reduced signal-to-noise performance. However, non-linearities in the form of fiber impairments and pattern-dependent saturation effects pose serious obstacles to such systems.

Neural network-based equalisation techniques are popular within the wider optical communications community due to their ability to model and compensate highly non-linear system impairments. However, the burst-mode nature specific to upstream PON transmission places severe restrictions on the size and training of NNEs which are not present in point-to-point systems.

This thesis investigates NNE solutions for upstream 100 Gbit/s PAM4 transmission using an SOA preamplifier. It concludes with the development of the novel SkipNet algorithm which enables low-complexity, adaptive neural network-based equalisation. SkipNet is the first such NNE algorithm to the best of our knowledge which could realistically be trained packet-by-packet in burst-mode PON systems while preserving performance and having low-computational complexity.

6.1 Summary of contributions

The research described in this thesis has led to the publication of 3 conference papers and 2 major journal publications, including an invited paper in the *Journal of Lightwave Technology* [A1]–[A5].

In Chapter 3, an analysis of potential 100 Gbit/s PAM4 PON is made in the context of an SOA+PIN Rx configuration at the OLT, and which forms the basis of [A1]. From this, SOA gain-saturation induced patterning is identified as a limiting factor in achieving the challenging minimum 19.5 dB dynamic range required by the OLT-side Rx. In order to counter this, we proposed a novel equaliser based on GRU neurons which achieves > 28 dB DR, far surpassing current 50 Gbit/s HS-PON requirements. Further, we showed that because of the long-memory effect of SOA patterning, this feedback based NNE requires less parameters and real valued multiplications than alternatively proposed feedforward NNEs, and is therefore less complex and more efficient [A1].

Chapter 4 extends this analysis by investigating the performance of such a GRU-NNE on a range of non-linear impairments which might be expected in future PON systems, including chromatic dispersion, bandwidth limitations, and electrical Rx saturation. These results led to both a conference and journal publication [A2], [A3]. It was shown that the GRU-NNE can sustain the impressive > 28 dB DR for 100 Gbit/s PAM4 in the presence of bandwidth limitation of 25 GHz. Further, up to 81.6 ps/nm accumulated dispersion, this equaliser can achieve 25 dB DR. The performance of the GRU-NNE was compared to that of a non-linear Volterra equaliser and FFE/DFE, and found to match or surpass all of these while using less RVMs and parameters. The feedback mechanism is a major obstacle to the digital implementation of a the GRU-NNE, and so the multi-symbol output techniques was investigated, and up to 8 parallel symbol outputs were achieved using the proposed GRU-NNE with no advantage seen in using the more complex alternative LSTM neurons. Finally, it was noted that because of the effectiveness of the GRU-NNE, the SOA non-linear gain compression could potentially be exploited to reduce the requirements on the electrical burst-mode Rx DR, which is typically balanced against linearity and bandwidth.

Chapter 5 describes the original SkipNet algorithm, which is the first such NNE which allows for adaptive equalisation on a packet-by-packet basis. SkipNet represents a significant step in realising low-complexity rapidly adaptive NNE for real-time deployment in PONs, and was published at the 2023 *European Confer-*

ence on Optical Communications [A4] as well as being accepted for publication in the *Journal of Optical Communications and Networking* [A5]. This original NNE solution is capable of being trained in just 250 symbols using the simple LMS algorithm, which is used for existing adaptive FFE/DFE solutions. This represents a significant decrease in the training requirements for NNEs in PON, and for the first time allows for viable adaptive NNE solutions at the PON OLT. The mechanism that allows our adaptive NNE solution to equalise such a wide variety of linear and non-linear impairments is the skip connection. The addition of these connections represents only a minor increase in equaliser complexity, while enabling true adaptive equalisation. Further, multiple low-complexity NNEs trained on different non-linear impairments can be integrated into the SkipNet structure to achieve a wide range of non-linear equalisation capability for 100 Gbit/s PAM4 transmission. It is expected that the SkipNet algorithm could prove an important milestone toward realising low-complexity, low-latency, adaptive NNE solutions outside of PON also, such as in the data center and long-haul transmission scenarios.

6.2 Future developments

There exists challenges to implementing RNN-based NNEs in hardware, and the SkipNet framework should be investigated further against non-linear effects other than those considered here. As such, there is broad scope to build on the results presented in this thesis, and further obstacles must be overcome before real-time implementations of the SkipNet equalisation algorithm can be realised on hardware.

Throughout this work the focus has been on a single wavelength carrier system with 100 Gbit/s PAM4 line-rate. However, the potential exists to extend this work to include multiple wavelengths and the system complexities that accompany this. Figure 6.1 (a) illustrates a potential 400 Gbit/s TWDM OLT Rx configuration, where a high-gain SOA preamplifier is used prior to wavelength de-multiplexing with an arrayed waveguide grating (AWG) and further single wavelength SOA preamplification. In this case, complex gain-saturation induced patterning effects linked to the combined burst-packet dynamics of the 4 carrier wavelengths may be observed. The single carrier GRU-NNE explored in this thesis would be ineffective in this scenario, since the level of SOA saturation per symbol could depend on multiple data streams.

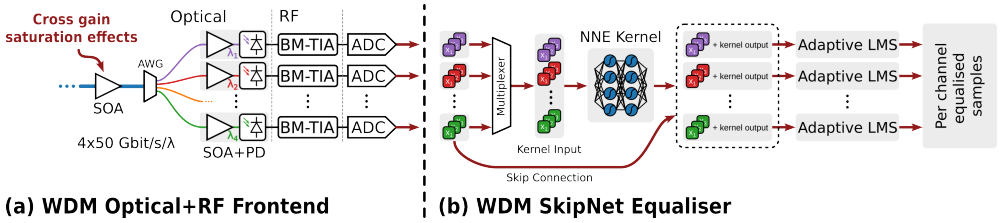


Figure 6.1: (a) Possible 400 Gbit/s PAM4 WDM PON scenario, with (b) adaptive SkipNet equaliser to mitigate SOA cross-channel patterning effects.

A potential solution to this is shown in Figure 6.1 (b), where a modified GRU-based SkipNet structure is shown which combines information from each of the sampled data-streams before splitting into 4 separate output equaliser stages. Each of these output stages could be dynamically adapted using the LMS algorithms. Such an equaliser scheme could extend the large DR benefits of the SkipNet equaliser to > 200 Gbit/s WDM PON systems based on the shown Rx configuration. WDM PAM4 PON systems are already of interest within the PON research community [11], [49], [109]. The proposed experiment would be a valuable contribution toward understanding the limits of SOA preamplifiers at the OLT, as well as developing a novel NNE solution to overcome these limits.

Digital implementations of feedback-based GRU-NNEs clearly face significant challenges related to available clock speeds, hardware resources, and energy consumption, and ultimately limit the throughput of real-time implementations. Therefore, within our research group there is an ongoing effort to be the first to realise a SkipNet-like equaliser on a real-time FPGA platform for 100 Gbit/s PAM4 transmission [A7]. This will require the multi-symbol equalisation technique explored in Chapter 4 to be integrated with the LMS adaption mechanism of SkipNet. Additionally, hardware optimisations such as quantisation and network pruning which lower NNE computational complexity are also a focus. These should sufficiently reduce the equaliser hardware footprint and allow for a cost-effective, real-time FPGA implementation at 100 Gbit/s, without compromising on the excellent SkipNet performance reported in this thesis.

Finally, it is worth noting that alternative NNE hardware solutions exist based on analog computing paradigms, such as neuromorphic photonics [127] and photonic reservoir computing [128]. Photonic chips which can implement optical equalisation and processing have been widely proposed, and there are a myriad of approaches to realise neural network architectures. There would be many benefits to such photonic solutions, which have been shown to offer significant reductions in energy footprint while delivering the benefits of non-linear, neural

network-like equalisation techniques [172]. Photonic implementations of NNEs could be readily integrated on a PIC with a Rx's optical frontend, and their bandwidth / throughput would naturally scale with that of future PON optics, since the electrical clock bottleneck could be avoided. Reservoir computing is especially interesting: typically non-linear signal processing is carried out in the optical domain on a PIC "reservoir", with a digital LMS filter equalising the reservoir's multiple output signals after direct detection using PDs [128]. An intriguing opportunity might exist if a correspondence between the SkipNet equaliser kernel structures and PIC-based optical reservoirs could be established, with an adaptive digital LMS filter allowing for burst-packet equalisation.

Acronyms

Adam	Adaptive moment estimation
AI	Artificial intelligence
APD	Avalanche photodetector
ASE	Amplified spontaneous emission
ASIC	Application specific integrated circuit
AWG	Arrayed waveguide grating
BER	Bit error rate
BMRx	Burst-mode receiver
BPON	Broadband passive optical network
BPtT	Backpropagation through time
BtB	Back-to-back i.e. 0 km transmission
CB	Conduction band
CDR	Clock and data recovery
DCI	Datacenter interconnect
DFE	Decision feedback equaliser
DML	Directly modulated laser
DSP	Digital signal processing
EAM	Electro-absorption modulator
FC	fully-connected

FEC	Forward error correction
FFE	Feedforward equaliser
FPGA	Field programmable gate array
FWHM	Full width at half maximum
GRU	Gated recurrent unit
HD-FEC	Hard decision forward error correction
HL	Hidden layer
HS-PON	Higher-speed passive optical network
ISI	Inter-symbol interference
LBMRx	Linear burst-mode receiver
LMS	Least mean squared
LSTM	Long short term memory unit
ML	Machine learning
MSE	Mean squared error
NBP	National Broadband Plan
NEP	Noise equivalent power
NN	Neural network
NNE	Neural network equaliser / equalisation
ODN	Optical distribution network
OLT	Optical line terminal
ONU	Optical network unit
OSNR	Optical signal-to-noise ratio
PON	Passive optical network
RLS	Recursive least squares
RMS	Root mean squared
RNN	Recurrent neural networks
RVM	Real valued multiplications

Rx	Receiver
SD-FEC	Soft decision forward error correction
SER	Symbol error rate
SGD	Stochastic gradient descent
SGM	Self gain modulation
SNR	Signal-to-noise ratio
SOA	Semiconductor optical amplifier
TDM	Time division multiplexing
VB	Valence band
VNLE	Volterra non-linear equaliser
VOA	Variable optical attenuator
WDM	Wavelength division multiplexing

List of Publications

[A1] **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*, Optica Publishing Group, 2022, SW4E.1. DOI: 10.1364/CLEO_SI.2022.SW4E.1.

[A2] **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 *European Conference on Optical Communication*, Optica Publishing Group, 2022, Th1C.6.

[A3] **S. L. Murphy**, F. Jamali, P. D. Townsend, and C. Antony, "High Dynamic Range 100G PON Enabled by SOA Preamplifier and Recurrent Neural Networks," *Journal of Lightwave Technology*, vol. 41, no. 11, pp. 3522–3532, 2023. DOI: 10.1109/JLT.2023.3249046.

[A4] **S. L. Murphy**, F. Jamali, P. D. Townsend, and C. Antony, "Adaptive Neural Network Equalisation Using Skip Connections for Future 100 Gbit/s/λ Passive Optical Networks," in *European Conference on Optical Communications*, Optica Publishing Group, 2023.

[A5] **S. L. Murphy**, P. D. Townsend, and C. Antony, "SkipNet: an adaptive neural network equalization algorithm for future passive optical networking," *J. Opt. Commun. Netw.*, vol. 16, no. 11, pp. 1082–1092, Nov. 2024. DOI: 10.1364/JOCN.528490.

[A6] F. Jamali, **S. L. Murphy**, C. Antony, and P. D. Townsend, "SOA-Based Power Equalisation for 100 Gb/s Passive Optical Network," in *2023 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)*, 2023, pp. 1–1. DOI: 10.1109/CLEO/Europe-EQEC57999.2023.10231954.

[A7] E. Roshanshomal, **S. L. Murphy**, S. O. Ayat, F. Jamali, P. D. Townsend, and C. Antony, "Low-Complexity FPGA-Accelerated NN-based Adaptive Equalizer for 100 Gb/s IMDD PON," in *IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, (Accepted), May 2025.

References

[1] G. Agrawal, *Fiber-Optic Communication Systems*. Wiley, 2021.

[2] TeleGeography, *Submarine Cable Map*, https://www.submarinecablemap.com/, Accessed: 2024-07-02.

[3] T. Horvath, P. Munster, V. Oujezsky, and N.-H. Bao, "Passive Optical Networks Progress: A Tutorial," *Electronics*, vol. 9, no. 7, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9071081.

[4] S. Aleksic, "A survey on optical technologies for iot, smart industry, and smart infrastructures," *Journal of Sensor and Actuator Networks*, vol. 8, no. 3, 2019, ISSN: 2224-2708. DOI: 10.3390/jsan8030047.

[5] C. Ranaweera, J. Kua, I. Dias, E. Wong, C. Lim, and A. Nirmalathas, "4G to 6G: disruptions and drivers for optical access [Invited]," *J. Opt. Commun. Netw.*, vol. 14, no. 2, A143–A153, Feb. 2022. DOI: 10.1364/JOCN.440798.

[6] A. Fayad, T. Cinkler, and J. Rak, "Toward 6G Optical Fronthaul: A Survey on Enabling Technologies and Research Perspectives," *IEEE Communications Surveys & Tutorials*, pp. 1–1, 2024. DOI: 10.1109/COMST.2024.3408090.

[7] T. Pfeiffer, P. Dom, S. Bidkar, F. Fredricx, K. Christodoulopoulos, and R. Bonk, "PON going beyond FTTH [Invited Tutorial]," *Journal of Optical Communications and Networking*, vol. 14, no. 1, A31–A40, 2022. DOI: 10.1364/JOCN.439241.

[8] R. Sabella, P. Iovanna, G. Bottari, and F. Cavaliere, "Optical transport for Industry 4.0 [Invited]," EN, *Journal of Optical Communications and Networking*, vol. 12, no. 8, pp. 264–276, Aug. 2020, Publisher: Optica Publishing Group, ISSN: 1943-0639. DOI: 10.1364/JOCN.390701.

[9] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, 2019. DOI: 10.1109/JLT.2019.2897313.

[10] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk, and S. J. Savory, "Machine learning for optical fiber communication systems: An introduction and overview," en, *APL Photonics*, vol. 6, no. 12, p. 121 101, Dec. 2021, ISSN: 2378-0967. DOI: 10.1063/5.0070838.

[11] A. Banerjee, Y. Park, F. Clarke, *et al.*, "Wavelength-division-multiplexed passive optical network (WDM-PON) technologiesfor broadband access: a review [Invited]," *J. Opt. Netw.*, vol. 4, no. 11, pp. 737–758, Nov. 2005. DOI: 10.1364/JON.4.000737.

[12] ITU-T Recommendation G.9804.3, *50-Gigabit-capable passive optical networks (50G-PON): Physical media dependent (PMD) layer specification*, Sep. 2021.

[13] ITU-T Recommendation G.9807 Series, *10 Gigabit capable symmetric passive optical network (XGSPON)*, 2016.

[14] A. Mason, *FTTP or 4G/5G for Ireland's NBP?* https://s3-eu-west-1.amazonaws.com/govieassets/8531/d40f531ae0194d01af3c48420d7db2cb.pdf, Accessed: 2024-07-02, Dec. 5, 2018.

[15] A. Mason, *XGS PON VfM analysis*, https://s3-eu-west-1.amazonaws.com/govieassets/8532/fe34b28728ff4fcebabcd4fbbf809a22.pdf, Accessed: 2024-07-02.

[16] *EPB Launches America's First Community-wide 25 Gig Internet Service*, https://epb.com/newsroom/press-releases/epb-launches-americas-first-community-wide-25-gig-internet-service/, "Accessed: 2024-07-02", Aug. 24, 2022.

[17] D. van Veen and V. Houtsma, "Strategies for economical next-generation 50G and 100G passive optical networks [Invited]," *Journal of Optical Communications and Networking*, vol. 12, no. 1, A95–A103, 2020. DOI: 10.1364/JOCN.12.000A95.

[18] A. E. Abejide, P. Duarte, R. Patel, *et al.*, "PON-FTTX Architecture and Bandwidth Analysis for Future Broadband Communications," in *5G and 6G Enhanced Broadband Communications*, D. I. A. Alimi and D. J. J. Popoola, Eds., Rijeka: IntechOpen, 2024, ch. 6. DOI: 10.5772/intechopen.114274.

[19] R. Bonk, D. Geng, D. Khotimsky, *et al.*, "50G-PON: The First ITU-T Higher-Speed PON System," *IEEE Communications Magazine*, vol. 60, no. 3, pp. 48–54, Mar. 2022, ISSN: 1558-1896. DOI: 10.1109/MCOM.001.2100441.

[20] P. Torres-Ferrera, H. Wang, V. Ferrero, and R. Gaudino, "100 Gbps/λ PON downstream O- and C-band alternatives using direct-detection and linear-impairment equalization [Invited]," EN, *Journal of Optical Communications and Networking*, vol. 13, no. 2, A111–A123, Feb. 2021, Publisher: Optica Publishing Group, ISSN: 1943-0639. DOI: 10.1364/JOCN.402437.

[21] G. Caruso, I. N. Cano, D. Nesset, G. Talli, and R. Gaudino, "Real-Time 100 Gb/s PAM-4 for Access Links With up to 34 dB Power Budget," *Journal of Lightwave Technology*, vol. 41, no. 11, pp. 3491–3497, 2023. DOI: 10.1109/JLT.2023.3244028.

[22] P. Torres-Ferrera, F. Effenberger, M. S. Faruk, S. J. Savory, and R. Gaudino, "Overview of high-speed TDM-PON beyond 50 Gbps per wavelength using digital signal processing [Invited Tutorial]," *Journal of Optical Communications and Networking*, vol. 14, no. 12, pp. 982–996, 2022. DOI: 10.1364/JOCN.468920.

[23] N. Suzuki, H. Miura, K. Matsuda, R. Matsumoto, and K. Motoshima, "100 Gb/s to 1 Tb/s Based Coherent Passive Optical Network Technology," *Journal of Lightwave Technology*, vol. 36, no. 8, pp. 1485–1491, 2018. DOI: 10.1109/JLT.2017.2785341.

[24] J. Zhou, J. He, X. Lu, *et al.*, "100G fine-granularity flexible-rate passive optical networks based on discrete multi-tone with PAPR optimization," *Journal of Optical Communications and Networking*, vol. 14, no. 11, pp. 944–950, 2022. DOI: 10.1364/JOCN.463671.

[25] J. Zhou, J. He, X. Lu, *et al.*, "100G fine-granularity flexible-rate passive optical networks based on discrete multi-tone with PAPR optimization," *Journal of Optical Communications and Networking*, vol. 14, no. 11, pp. 944–950, 2022. DOI: 10.1364/JOCN.463671.

[26] C. Xie and B. Zhang, "Scaling Optical Interconnects for Hyperscale Data Center Networks," *Proceedings of the IEEE*, vol. 110, no. 11, pp. 1699–1713, 2022. DOI: 10.1109/JPROC.2022.3178977.

[27] R. Bonk, H. Schmuck, W. Poehlmann, and T. Pfeiffer, "Beneficial olt transmitter and receiver concepts for ng-pon2 using semiconductor optical amplifiers [invited]," *Journal of Optical Communications and Networking*, vol. 7, no. 3, A467–A473, 2015. DOI: 10.1364/JOCN.7.00A467.

[28] M. J. Connelly, *Semiconductor Optical Amplifiers*. Springer Science & Business Media, 2007.

[29] OECD, *Artificial Intelligence in Science*. 2023, p. 300. DOI: https://doi.org/https://doi.org/10.1787/a8d820bd-en.

[30] D.-G. for Research and Innovation, "Quarterly research and innovation literature review: The impact of AI on R&I," European Commission, Tech. Rep., Jul. 11, 2023.

[31] L. Ruan, E. Wong, and H. Zhu, "Reinforcement learning-based bandwidth decision in optical access networks: A study of exploration strategy and time with confidence guarantee," *IEEE Transactions on Communications*, pp. 1–1, 2024. DOI: 10.1109/TCOMM.2024.3407195.

[32] D. K. Tizikara, J. Serugunda, and A. Katumba, "Machine learning-aided optical performance monitoring techniques: A review," *Frontiers in Communications and Networks*, vol. 2, 2022, ISSN: 2673-530X. DOI: 10.3389/frcmn.2021.756513.

[33] P. Manfredi, A. Waqas, and D. Melati, "Stochastic and multi-objective design of photonic devices with machine learning," en, *Scientific Reports*, vol. 14, no. 1, p. 7162, Mar. 2024, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-024-57315-4.

[34] C. Bluemm, B. Liu, T. Rahman, B. Li, U. Schlichtmann, and S. Calabrò, "Towards Neural Network Equalizer Implementations for IM/DD Transceivers," in *2023*

Opto-Electronics and Communications Conference (OECC), 2023, pp. 1–6. DOI: 10.1109/OECC56963.2023.10209762.

[35] P. Massoud Salehi and J. Proakis, *Digital Communications*. McGraw-Hill Education, 2007, ISBN: 9780072957167.

[36] D. Zhang, D. Liu, X. Wu, and D. Nesset, "Progress of ITU-T higher speed passive optical network (50G-PON) standardization," en, *Journal of Optical Communications and Networking*, vol. 12, no. 10, p. D99, Oct. 2020, ISSN: 1943-0620, 1943-0639. DOI: 10.1364/JOCN.391830.

[37] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017.

[38] D. Silver, A. Huang, C. J. Maddison, *et al.*, "Mastering the game of Go with deep neural networks and tree search," en, *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016, Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/nature16961.

[39] G. Gibson, S. Siu, and C. Cowan, "Application of Multilayer Perceptrons as Adaptive Channel Equalisers," en, *IFAC Proceedings Volumes*, vol. 23, no. 1, pp. 573–578, Apr. 1990, ISSN: 14746670. DOI: 10.1016/S1474-6670(17)52779-2.

[40] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65 6, pp. 386–408, 1958.

[41] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8.

[42] V. Maiorov and A. Pinkus, "Lower bounds for approximation by MLP neural networks," *Neurocomputing*, vol. 25, no. 1, pp. 81–91, 1999, ISSN: 0925-2312. DOI: https://doi.org/10.1016/S0925-2312(98)00111-8.

[43] H. Johnson and M. Graham, *High-speed Signal Propagation: Advanced Black Magic* (Prentice Hall PTR Signal Integrity Library). Prentice Hall PTR, 2003.

[44] C. DeSanti, L. Du, J. Guarin, J. Bone, and C. F. Lam, "Super-PON: an evolution for access networks [Invited]," *J. Opt. Commun. Netw.*, vol. 12, no. 10, pp. D66–D77, Oct. 2020. DOI: 10.1364/JOCN.391846.

[45] IEEE 802.3cs-2022, *IEEE Standard for Ethernet Amendment 2: Physical Layers and Management Parameters for Increased-Reach Point-to-Multipoint Ethernet Optical Subscriber Access (Super-PON)*, Nov. 2022.

[46] J. S. Sobolewski, "Data transmission media," in *Encyclopedia of Physical Science and Technology (Third Edition)*, R. A. Meyers, Ed., Third Edition, New York: Academic Press, 2003, pp. 277–303, ISBN: 978-0-12-227410-7. DOI: https://doi.org/10.1016/B0-12-227410-5/00165-4.

[47] "Challenges in Next-Gen PON Deployment," Viavi, Tech. Rep., Jan. 2006, Accessed: 07-07-2024.

[48] R. Bonk, E. Harstead, R. Borkowski, *et al.*, "Perspectives on and the road towards 100Gb/s TDM PON with intensity-modulation and direct-detection," *J. Opt. Commun. Netw.*, vol. 15, no. 8, pp. 518–526, Aug. 2023. DOI: 10.1364/JOCN.489228.

[49] J. Martínez, "Performance Analysis of A 200 Gb/s PAM-4 PAM-8 WDM-PON," *International Journal of Engineering Trends and Technology*, vol. 71, Feb. 2023. DOI: 10.14445/22315381/IJETT-V71I2P238.

[50] J. Zhang and Z. Jia, "Coherent Passive Optical Networks for 100G/λ-and-Beyond Fiber Access: Recent Progress and Outlook," *IEEE Network*, vol. 36, no. 2, pp. 116–123, 2022. DOI: 10.1109/MNET.005.2100604.

[51] Y. Zhu, L. Yi, B. Yang, *et al.*, "Comparative study of cost-effective coherent and direct detection schemes for 100 Gb/s/λ PON," *Journal of Optical Communications and Networking*, vol. 12, no. 9, pp. D36–D47, 2020. DOI: 10.1364/JOCN.390911.

[52] M. S. Faruk, X. Li, D. Nesset, I. N. Cano, A. Rafel, and S. J. Savory, "Coherent Passive Optical Networks: Why, When, and How," *IEEE Communications Magazine*, vol. 59, no. 12, pp. 112–117, 2021. DOI: 10.1109/MCOM.010.2100503.

[53] T. Takahara, T. Tanaka, M. Nishihara, *et al.*, "Discrete Multi-Tone for 100 Gb/s Optical Access Networks," in *Optical Fiber Communication Conference*, Optica Publishing Group, 2014, p. M2I.1. DOI: 10.1364/OFC.2014.M2I.1.

[54] A. Dochhan, T. Drenski, H. Griesser, M. H. Eiselt, and J.-P. Elbers, "Real-time discrete multi-tone transmission for passive optical networks in C- and O-band," in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–3. DOI: 10.1049/cp.2019.1110.

[55] X. Li, M. S. Faruk, and S. J. Savory, "Bidirectional Symmetrical 100 Gb/s/λ Coherent PON Using a Simplified ONU Transceiver," *IEEE Photonics Technology Letters*, vol. 34, no. 16, pp. 838–841, 2022. DOI: 10.1109/LPT.2022.3168294.

[56] J. Zhang, S. Xing, G. Li, and N. Chi, "High-Performance and Robust Burst Reception in Coherent PON," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, 2023, pp. 1–3. DOI: 10.1364/OFC.2023.W1I.6.

[57] Infinera, *Pluggables, Embeddeds, and the Three Vectors of Coherent Evolution*, https://www.infinera.com/wp-content/uploads/Pluggables-Embeddeds-and-the-Three-Vectors-of-Coherent-Evolution-0287-WP-RevA-0421.pdf, Accessed: 2024-08-01.

[58] "IEEE Standard for Ethernet," *IEEE Std 802.3-2018 (Revision of IEEE Std 802.3-2015)*, pp. 1–5600, 2018. DOI: 10.1109/IEEESTD.2018.8457469.

[59] H. M. Lucal, "Arithmetic Operations for Digital Computers Using a Modified Reflected Binary Code," *IRE Transactions on Electronic Computers*, vol. EC-8, no. 4, pp. 449–458, 1959. DOI: 10.1109/TEC.1959.5222057.

[60] G. C. Nevio Benvenuto, *Algorithms for Communication Systems and their Applications*. Wiley, 2002. DOI: 10.1002/0470855509.

[61] *Agilent 71501D Eye-Diagram Analysis User's Guide*, https://www.keysight.com/ie/en/assets/9018-40773/user-manuals/9018-40773.pdf?success=true, Accessed: 2024-08-01.

[62] Z. Zheng, C. Yang, C. Yang, and Z. Wang, "Energy consumption modeling and analysis for short-reach optical transmissions using irregular LDPC codes," *Opt. Express*, vol. 30, no. 24, pp. 43 852–43 871, Nov. 2022. DOI: 10.1364/OE.472564.

[63] R. Rosales, K. Atra, Y. Lin, *et al.*, "50G-PON Upstream With Over 36dB Link Budget Using an SOA-PIN Based Receiver," *IEEE Photonics Technology Letters*, vol. 34, no. 22, pp. 1222–1225, Nov. 2022, Conference Name: IEEE Photonics Technology Letters, ISSN: 1941-0174. DOI: 10.1109/LPT.2022.3207996.

[64] L. Breyne, C. Caillaud, T. Gurne, *et al.*, "50G Burst-Mode Receiver Using Monolithic SOA-UTC and Burst-Mode TIA," in *Optical Fiber Communication Conference (OFC) 2024*, Optica Publishing Group, 2024, Tu3H.1. DOI: 10.1364/OFC.2024.Tu3H.1.

[65] Z. Li, Y. Li, S. Luo, F. Yin, Y. Wang, and Y. Song, "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, p. 342, Feb. 2022. DOI: 10.3390/mi13030342.

[66] T. Gurne, G. Coudyzer, B. V. Lombergen, *et al.*, "First Demonstration of a 100 Gbit/s PAM-4 Linear Burst-Mode Transimpedance Amplifier for Upstream Flexible PON," en, *European Conference on Optical Communication*, Th3A.1, 2022.

[67] G. Coudyzer, M. Verplaetse, B. Van Lombergen, *et al.*, "100 Gbit/s PAM-4 Linear Burst-Mode Transimpedance Amplifier for Upstream Flexible Passive Optical Networks," *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3652–3659, 2023. DOI: 10.1109/JLT.2023.3262319.

[68] M. L. M. Larry A. Coldren Scott W. Corzine, "A phenomenological approach to diode lasers," in *Diode Lasers and Photonic Integrated Circuits*. John Wiley & Sons, Ltd, 2012, ch. Two, pp. 45–90, ISBN: 9781118148167. DOI: https://doi.org/10.1002/9781118148167.ch2. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118148167.ch2.

[69] R. Hui and M. O'sullivan, *Fiber-Optic Measurement Techniques (Second Edition)*, Second Edition. Academic Press, 2023, ISBN: 978-0-323-90957-0. DOI: https://doi.org/10.1016/B978-0-323-90957-0.09996-3.

[70] D. Payne, "Low-noise erbium-doped fibre amplifier operating at 1.54µm," English, *Electronics Letters*, vol. 23, 1026–1028(2), 19 Sep. 1987, ISSN: 0013-5194.

[71] A. Al-Azzawi, *Advanced Manufacturing for Optical Fibers and Integrated Photonic Devices*, First Edition. CRC Press, 2016. DOI: https://doi.org/10.1201/b18749.

[72] C. Huang, H. Song, L. Dai, *et al.*, "Optical Multipath Interference Mitigation for High-Speed PAM4 IMDD Transmission System," *Journal of Lightwave Technology*, vol. 40, no. 16, pp. 5490–5501, 2022. DOI: 10.1109/JLT.2022.3181695.

[73] A. Himeno, K. Kato, and T. Miya, "Silica-based planar lightwave circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 4, no. 6, pp. 913–924, 1998. DOI: 10.1109/2944.736076.

[74] N. Kaneda, J. Lee, and Y.-K. Chen, "Nonlinear equalizer for 112-Gb/s SSB-PAM4 in 80-km dispersion uncompensated link," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.

[75] X. Wu, J. Zhang, A. P. T. Lau, and C. Lu, "C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization," *Opt. Lett.*, vol. 47, no. 19, pp. 5144–5147, Oct. 2022. DOI: 10.1364/OL.473091.

[76] T. Wettlin, S. Pachnicke, T. Rahman, J. Wei, S. Calabro, and N. Stojanovic, "Complexity Reduction of Volterra Nonlinear Equalization for Optical Short-Reach IM/DD Systems," in *Photonic Networks; 21th ITG-Symposium*, 2020, pp. 1–6.

[77] ITU-T Recommendation G.652, *Characteristics of a single-mode optical fibre and cable*, Nov. 2016.

[78] *Smf-28e+ optical fiber product information*, Corning.

[79] P. Zhu, Y. Yoshida, A. Kanno, K. Akahane, and K.-i. Kitayama, "Dispersion Compensation over C-band WDM Grid for 100Gb/s PAM4 System by Low-complexity Optoelectronic Feedforward Equalization (OE-FFE)," in *Optical Fiber Communication Conference (OFC) 2023*, Optica Publishing Group, 2023, W4E.3. DOI: 10.1364/OFC.2023.W4E.3.

[80] X. Tang, Y. Qiao, Y.-W. Chen, Y. Lu, and G.-K. Chang, "Digital Pre- and Post-Equalization for C-Band 112-Gb/s PAM4 Short-Reach Transport Systems," *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4683–4690, 2020. DOI: 10.1109/JLT.2020.2993997.

[81] S. Porto, "Burst-mode electronic dispersion compensation in long reach PONs," Ph.D. dissertation, University College Cork, 2014.

[82] J. Zhang, X. Xiao, J. Yu, J. S. Wey, X. Huang, and Z. Ma, "Real-Time FPGA Demonstration of PAM-4 Burst-Mode All-Digital Clock and Data Recovery for Single wavelength 50G PON Application," in *Optical Fiber Communication Conference*, Optica Publishing Group, 2018, M1B.7. DOI: 10.1364/OFC.2018.M1B.7.

[83] J. Zhang, Q. Zhou, M. Zhu, *et al.*, "Demonstration of all-digital burst clock and data recovery for symmetrical 50 Gb/s/λ PON based on low-bandwidth optics," *Optics Communications*, vol. 516, p. 128 266, 2022, ISSN: 0030-4018. DOI: https://doi.org/10.1016/j.optcom.2022.128266.

[84] G. Coudyzer, P. Ossieur, L. Breyne, *et al.*, "Study of burst-mode adaptive equalization for >25G PON applications [Invited]," *Journal of Optical Communications and Networking*, vol. 12, no. 1, A104–A112, 2020. DOI: 10.1364/JOCN.12.00A104.

[85] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer series in statistics). Springer, 2001.

[86] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, and D. Andina, "Deep Learning for Computer Vision: A Brief Review," *Intell. Neuroscience*, vol. 2018, Jan. 2018. DOI: 10.1155/2018/7068349.

[87] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6000–6010, ISBN: 9781510860964.

[88] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," en, *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. DOI: 10.1038/nature14539.

[89] M. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.

[90] O. Russakovsky, J. Deng, H. Su, *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015.

[91] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., ser. Proceedings of Machine Learning Research, vol. 15, Fort Lauderdale, FL, USA: PMLR, Apr. 2011, pp. 315–323.

[92] 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," *Journal of Lightwave Technology*, vol. 40, no. 5, pp. 1353–1359, Mar. 2022. DOI: 10.1109/JLT.2021.3128579.

[93] 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," en, *Journal of Lightwave Technology*, vol. 40, no. 7, pp. 1972–1980, Apr. 2022, ISSN: 0733-8724, 1558-2213. DOI: 10.1109/JLT.2021.3133723.

[94] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, ISSN: 1476-4687. DOI: 10.1038/323533a0.

[95] J. Schmidhuber, *Who invented backpropagation?* 2014.

[96] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, arXiv:1412.6980, arXiv:1412.6980, Dec. 2014. arXiv: 1412.6980 [cs.LG].

[97] T. Tieleman and G. E. Hinton, *Lecture 6e rmsprop: Divide the gradient by a running average of its recent magnitude*, https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf#page=26, 2014.

[98] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, vol. 12, no. 61, pp. 2121–2159, 2011.

[99] H. Wang, J. Zhou, W. Liu, *et al.*, "BGD-based Adam algorithm for time-domain equalizer in PAM-based optical interconnects," *Opt. Lett.*, vol. 45, no. 1, pp. 141–144, Jan. 2020. DOI: 10.1364/OL.45.000141.

[100] J. Estaran, R. Rios-Mueller, M. A. Mestre, *et al.*, "Artificial Neural Networks for Linear and Non-Linear Impairment Mitigation in High-Baudrate IM/DD Systems," in *ECOC 2016; 42nd European Conference on Optical Communication*, 2016, pp. 1–3.

[101] L. Yi, T. Liao, L. Huang, L. Xue, P. Li, and W. Hu, "Machine Learning for 100 Gb/s/λ Passive Optical Network," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1621–1630, 2019. DOI: 10.1109/JLT.2018.2888547.

[102] P. Li, L. Yi, L. Xue, and W. Hu, "100Gbps IM/DD Transmission over 25km SSMF using 20G-class DML and PIN Enabled by Machine Learning," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1–3.

[103] C.-Y. Chuang, L.-C. Liu, C.-C. Wei, *et al.*, "Convolutional Neural Network based Nonlinear Classifier for 112-Gbps High Speed Optical Link," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1–3.

[104] 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," *Journal of Lightwave Technology*, vol. 40, no. 9, pp. 2890–2900, 2022. DOI: 10.1109/JLT.2022.3146863.

[105] C. Ye, D. Zhang, X. Huang, H. Feng, and K. Zhang, "Demonstration of 50Gbps IM/DD PAM4 PON over 10GHz Class Optics Using Neural Network Based Nonlinear Equalization," in *2017 European Conference on Optical Communication (ECOC)*, 2017, pp. 1–3. DOI: 10.1109/ECOC.2017.8346196.

[106] L. Zhang, F. Yang, H. Ming, *et al.*, "160-Gb/s Nyquist PAM-4 Transmission with GeSi-EAM using Artificial Neural Network Based Nonlinear Equalization," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[107] A. G. Reza, M. Troncoso-Costas, C. Browning, F. J. Diaz-Otero, and L. P. Barry, "Single-Lane 54-Gbit/s PAM-4/8 Signal Transmissions Using 10G-Class Directly Modulated Lasers Enabled by Low-Complexity Nonlinear Digital Equalization," *IEEE Photonics Journal*, vol. 14, no. 3, pp. 1–9, 2022. DOI: 10.1109/JPHOT.2022.3168683.

[108] L. Xue, L. Yi, R. Lin, L. Huang, and J. Chen, "SOA pattern effect mitigation by neural network based pre-equalizer for 50G PON," *Optics Express*, vol. 29, no. 16, pp. 24 714–24 722, Aug. 2021. DOI: 10.1364/OE.426781.

[109] A. G. Reza, M. Troncoso-Costas, C. Browning, S. O'Duill, and L. P. Barry, "Mitigation of SOA-Induced Nonlinearities With Recurrent Neural Networks in 75 Gbit/s/λ PAM-4 IM/DD WDM-PON Transmission Systems," *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3967–3975, 2023. DOI: 10.1109/JLT.2023.3278981.

[110] S. Gaiarin, X. Pang, O. Ozolins, *et al.*, "High Speed PAM-8 Optical Interconnects with Digital Equalization Based on Neural Network," in *2016 Asia Communications and Photonics Conference (ACP)*, 2016, pp. 1–3.

[111] J. Zhang, L. Yan, L. Jiang, *et al.*, "Convolutional Neural Network Equalizer for Short-reach Optical Communication Systems," in *2020 Asia Communications and Photonics Conference (ACP) and International Conference on Information Photonics and Optical Communications (IPOC)*, 2020, pp. 1–3.

[112] B. Karanov, M. Chagnon, F. Thouin, *et al.*, "End-to-End Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018. DOI: 10.1109/JLT.2018.2865109.

[113] S. Xing, Z. Li, C. Huang, *et al.*, "End-to-end deep learning for a flexible coherent PON with user-specific constellation optimization," *J. Opt. Commun. Netw.*, vol. 16, no. 1, pp. 59–70, Jan. 2024. DOI: 10.1364/JOCN.500500.

[114] V. Houtsma, E. Chou, and D. van Veen, "92 and 50 Gbps TDM-PON using Neural Network Enabled Receiver Equalization Specialized for PON," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1–3.

[115] P. Ossieur, C. Melange, T. De Ridder, *et al.*, "Burst-Mode Electronic Equalization for 10-Gb/s Passive Optical Networks," *IEEE Photonics Technology Letters*, vol. 20, no. 20, pp. 1706–1708, 2008. DOI: 10.1109/LPT.2008.2003417.

[116] 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]," *Journal of Optical Communications and Networking*, vol. 7, no. 1, A118–A125, 2015. DOI: 10.1364/JOCN.7.00A118.

[117] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Feedforward and Recurrent Neural Network-Based Transfer Learning for Nonlinear Equalization in Short-Reach Optical Links," *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 475–480, Jan. 2021.

[118] P. J. Freire, A. Napoli, B. Spinnler, *et al.*, "Reducing Computational Complexity of Neural Networks in Optical Channel Equalization: From Concepts to Implementation," *Journal of Lightwave Technology*, vol. 41, no. 14, pp. 4557–4581, 2023. DOI: 10.1109/JLT.2023.3234327.

[119] P. J. Freire, M. Anderson, B. Spinnler, *et al.*, "Towards FPGA Implementation of Neural Network-Based Nonlinearity Mitigation Equalizers in Coherent Optical Transmission Systems," *European Conference on Optical Communication*, We1C.2, 2022.

[120] N. Kaneda, Z. Zhu, C.-Y. Chuang, *et al.*, "Fpga implementation of deep neural network based equalizers for high-speed pon," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[121] M. Schädler, G. Böcherer, F. Diedolo, and S. Calabrò, "Nonlinear Component Equalization: A Comparison of Deep Neural Networks and Volterra Series," in *2022 European Conference on Optical Communication (ECOC)*, 2022, pp. 1–4.

[122] K. K. Parhi, *VLSI Digital Signal Processing Systems: Design and Implementation*. Wiley-Interscience, 1999.

[123] Z. Xu, Q. Wu, W. Lu, *et al.*, "Weight-adaptive joint mixed-precision quantization and pruning for neural network-based equalization in short-reach direct detection links," *Opt. Lett.*, vol. 49, no. 12, pp. 3500–3503, Jun. 2024. DOI: 10.1364/OL.527293.

[124] L. Ge, W. Zhang, C. Liang, and Z. He, "Compressed Neural Network Equalization Based on Iterative Pruning Algorithm for 112-Gbps VCSEL-Enabled Optical Interconnects," *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1323–1329, 2020. DOI: 10.1109/JLT.2020.2973718.

[125] B. Sang, J. Zhang, C. Wang, *et al.*, "Multi-Symbol Output Long Short-Term Memory Neural Network Equalizer For 200+ Gbps IM/DD System," en, p. 4,

[126] 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. Lightwave Technol.*, vol. 40, no. 5, pp. 1353–1359, Mar. 2022.

[127] B. J. Shastri, A. N. Tait, T. F. de Lima, *et al.*, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photonics*, pp. 102–114, 15 2021. DOI: 10.1038/s41566-020-00754-y.

[128] M. Yan, C. Huang, P. Bienstman, P. Tino, W. Lin, and J. Sun, "Emerging opportunities and challenges for the future of reservoir computing," *Nature Communications*, 15 2024. DOI: 10.1038/s41467-024-45187-1.

[129] E. Arnold, G. Böcherer, E. Müller, *et al.*, "Spiking Neural Network Equalization on Neuromorphic Hardware for IM/DD Optical Communication," in *2022 European Conference on Optical Communication (ECOC)*, 2022, pp. 1–4.

[130] G. Böcherer, F. Strasser, E. Arnold, *et al.*, "Spiking Neural Network Linear Equalization: Experimental Demonstration of 2km 100Gb/s IM/DD PAM4 Optical Transmission," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, 2023, pp. 1–3. DOI: 10.1364/OFC.2023.W4E.1.

[131] *O.150 : General requirements for instrumentation for performance measurements on digital transmission equipment*, Jun. 1996.

[132] I. Lyubomirsky, *PRQS Test Patterns for PAM4*, Accessed: 2024-08-06.

[133] T. A. Eriksson, H. Bülow, and A. Leven, "Applying Neural Networks in Optical Communication Systems: Possible Pitfalls," *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2091–2094, Dec. 2017, ISSN: 1941-0174. DOI: 10.1109/LPT.2017.2755663.

[134] J. Kim and H. Kim, "PRBS orders required to train ANN equalizer for PAM signal without overfitting," *Opt. Express*, vol. 30, no. 14, pp. 25 486–25 497, Jul. 2022. DOI: 10.1364/OE.461199.

[135] W. Thomson, "Delay networks having maximally flat frequency characteristics," English, *Proceedings of the IEE - Part III: Radio and Communication Engineering*, vol. 96, 487–490(3), 44 Nov. 1949, ISSN: 0369-8947.

[136] A. Sobhanan, A. Anthur, S. O'Duill, *et al.*, "Semiconductor optical amplifiers: Recent advances and applications," *Advances in Optics and Photonics*, vol. 14, no. 3, pp. 571–651, Sep. 2022, Publisher: Optica Publishing Group. DOI: 10.1364/AOP.451872.

[137] M. Power, "All-Optical Signal Processing using Semiconductor Optical Amplifiers for Next Generation Optical Networks," Ph.D. dissertation, University College Cork, 2014.

[138] J. Dormand and P. Prince, "A family of embedded Runge-Kutta formulae," *Journal of Computational and Applied Mathematics*, vol. 6, no. 1, pp. 19–26, 1980, ISSN: 0377-0427. DOI: https://doi.org/10.1016/0771-050X(80)90013-3.

[139] F. Girardin, G. Guekos, and A. Houbavlis, "Gain recovery of bulk semiconductor optical amplifiers," *IEEE Photonics Technology Letters*, vol. 10, no. 6, pp. 784–786, Jun. 1998, Conference Name: IEEE Photonics Technology Letters, ISSN: 1941-0174. DOI: 10.1109/68.681483.

[140] M. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.)," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 85–89, 1970. DOI: 10.1109/TIT.1970.1054411.

[141] K. Wang, C. Wang, J. Zhang, Y. Chen, and J. Yu, "Mitigation of SOA-Induced Nonlinearity With the Aid of Deep Learning Neural Networks," *Journal of Lightwave Technology*, vol. 40, no. 4, pp. 979–986, Feb. 2022. DOI: 10.1109/JLT.2021.3126765.

[142] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13, Atlanta, GA, USA: JMLR.org, 2013, III–1310–III–1318.

[143] K. Cho, B. van Merriënboer, C. Gulcehre, *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. DOI: 10.3115/v1/D14-1179.

[144] P. Werbos, "Backpropagation through time: What it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990. DOI: 10.1109/5.58337.

[145] G. Coudyzer, P. Ossieur, J. Bauwelinck, and X. Yin, "A 25Gbaud PAM-4 Linear Burst-Mode Receiver With Analog Gain- and Offset Control in 0.25μm SiGe:C BiCMOS," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 8, pp. 2206–2218, 2020. DOI: 10.1109/JSSC.2020.2987680.

[146] X. Yin, G. Coudyzer, P. Ossieur, L. Breyne, B. Van Lombergen, and J. Bauwelinck, "Linear Burst-Mode Receivers for DSP-Enabled Passive Optical Networks," en, in *Optical Fiber Communication Conference (OFC) 2021*, Washington, DC: OSA, 2021, M3G.1, ISBN: 978-1-943580-86-6. DOI: 10.1364/OFC.2021.M3G.1.

[147] X. Dai, X. Li, M. Luo, Q. You, and S. Yu, "LSTM networks enabled nonlinear equalization in 50-Gb/s PAM-4 transmission links," en, *Applied Optics*, vol. 58, no. 22, p. 6079, Aug. 2019, ISSN: 1559-128X, 2155-3165. DOI: 10.1364/AO.58.006079.

[148] K. Wang, J. Zhang, L. Zhao, X. Li, and J. Yu, "Mitigation of Pattern-Dependent Effect in SOA at O-Band by Using DSP," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 590–597, Feb. 2020, ISSN: 1558-2213. DOI: 10.1109/JLT.2019.2946188.

[149] J. Zhou, H. Wang, J. Wei, *et al.*, "Adaptive moment estimation for polynomial nonlinear equalizer in PAM8-based optical interconnects," *Optics Express*, vol. 27, no. 22, pp. 32 210–32 216, Oct. 2019. DOI: 10.1364/OE.27.032210.

[150] 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," *Journal of Lightwave Technology*, vol. 40, no. 16, pp. 5395–5406, 2022. DOI: 10.1109/JLT.2022.3177446.

[151] L. Huang, W. Jiang, Y. Xu, W. Hu, and L. Yi, "Low-complexity Volterra-inspired neural network equalizer in 100-G band-limited IMDD PON system," *Optics Letters*, vol. 47, no. 21, pp. 5692–5695, Nov. 2022. DOI: 10.1364/OL.474900.

[152] A. Awny, A. Thiede, M. Elkhouly, J. Borngräber, F. Korndörfer, and J. C. Scheytt, "Mixed-signal techniques in mm-wave range for 100 Gbit decision feedback equalizer," in *2010 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, 2010, pp. 80–83. DOI: 10.1109/SMIC.2010.5422959.

[153] C.-Y. Chuang, W.-F. Chang, C.-C. Wei, *et al.*, "Sparse Volterra Nonlinear Equalizer by Employing Pruning Algorithm for High-Speed PAM-4 850-nm VCSEL Optical Interconnect," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1–3.

[154] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," en, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.1997.9.8.1735.

[155] H. Badaoui, F. Yann, R. Petros, B.-E. Benkelfat, and M. Feham, "De Bruijn Pseudo Random Sequences Analysis For Modeling of Quaternary Modulation Formats," *International Journal of Computer Science Issues*, vol. 7, Jul. 2010.

[156] C. Laperle and M. O'Sullivan, "Advances in High-Speed DACs, ADCs, and DSP for Optical Coherent Transceivers," *J. Lightwave Technol.*, vol. 32, no. 4, pp. 629–643, Feb. 2014.

[157] L. Huang, Y. Xu, W. Jiang, L. Xue, W. Hu, and L. Yi, "Performance and Complexity Analysis of Conventional and Deep Learning Equalizers for the High-Speed IMDD PON," *Journal of Lightwave Technology*, vol. 40, no. 14, pp. 4528–4538, 2022. DOI: 10.1109/JLT.2022.3165529.

[158] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk, and S. J. Savory, "Machine learning for optical fiber communication systems: An introduction and overview," *APL Photonics*, vol. 6, no. 12, p. 121 101, Dec. 2021, ISSN: 2378-0967. DOI: 10.1063/5.0070838.

[159] L. A. Neto, J. Maes, P. Larsson-Edefors, J. Nakagawa, K. Onohara, and S. J. Trowbridge, "Considerations on the Use of Digital Signal Processing in Future Optical Access Networks," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 598–607, 2020. DOI: 10.1109/JLT.2019.2946687.

[160] K. Matsuda and N. Suzuki, "Hardware-Efficient Signal Processing Technologies for Coherent PON Systems," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1614–1620, Mar. 2019.

[161] G. Simon, F. Saliou, P. Chanclou, L. A. Neto, and H. Hallak Elwan, "50Gb/s TDM PON Digital Signal Processing Challenges: Mining current G-PON Field Data to Assist Higher Speed PON," in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333167.

[162] C. Ye, X. Hu, D. Zhang, and K. Zhang, "A Versatile NN-Equalization for 50Gbps TDM PON Burst Uplink," in *2020 European Conference on Optical Communications (ECOC)*, Dec. 2020, pp. 1–3. DOI: 10.1109/ECOC48923.2020.9333360.

[163] F. Zhuang, Z. Qi, K. Duan, *et al.*, "A Comprehensive Survey on Transfer Learning," English (US), *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, ISSN: 0018-9219. DOI: 10.1109/JPROC.2020.3004555.

[164] P. Torres-Ferrera, H. Wang, V. Ferrero, M. Valvo, and R. Gaudino, "Optimization of Band-Limited DSP-Aided 25 and 50 Gb/s PON Using 10G-Class DML and APD," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 608–618, 2020. DOI: 10.1109/JLT.2019.2946959.

[165] J. Zhang, Z. Jia, M. Xu, H. Zhang, and L. Alberto Campos, "Efficient preamble design and digital signal processing in upstream burst-mode detection of 100G TDM coherent-PON," *Journal of Optical Communications and Networking*, vol. 13, no. 2, A135–A143, 2021. DOI: 10.1364/JOCN.402591.

[166] H. Wang, J. Zhou, Z. Xing, *et al.*, "Fast-Convergence Digital Signal Processing for Coherent PON Using Digital SCM," *Journal of Lightwave Technology*, vol. 41, no. 14, pp. 4635–4643, 2023. DOI: 10.1109/JLT.2023.3243828.

[167] Q. Zhou, F. Zhang, and C. Yang, "AdaNN: Adaptive Neural Network-Based Equalizer via Online Semi-Supervised Learning," *Journal of Lightwave Technology*, vol. 38, no. 16, pp. 4315–4324, Aug. 2020.

[168] Q. Fan, C. Lu, and A. P. T. Lau, "Combined Neural Network and Adaptive DSP Training for Long-Haul Optical Communications," *Journal of Lightwave Technology*, vol. 39, no. 22, pp. 7083–7091, 2021. DOI: 10.1109/JLT.2021.3111437.

[169] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[170] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Cascade recurrent neural network-assisted nonlinear equalization for a 100 Gb/s PAM4 short-reach direct detection system," en, *Optics Letters*, vol. 45, no. 15, p. 4216, Aug. 2020, ISSN: 0146-9592, 1539-4794. DOI: 10.1364/OL.394048.

[171] S. Yao, X. Tang, R. Zhang, *et al.*, "Accelerating LMS-Based Equalization With Correlated Training Sequence in Bandlimited IM/DD Systems," *Journal of Lightwave Technology*, vol. 40, no. 13, pp. 4268–4275, 2022. DOI: 10.1109/JLT.2022.3164031.

[172] S. Masaad and P. Bienstman, "Opto-electronic machine learning network for Kramers -Kronig receiver linearization," *Opt. Express*, vol. 32, no. 13, pp. 23 561–23 574, Jun. 2024. DOI: 10.1364/OE.522673.