Non-uniform Quantization and RUM for Optimizing Implementation of Real-Time FIR Equalization in Short-Reach Optical Links

Bohan Sang,¹ Kaihui Wang,¹ Luhan Jiang,¹ Chen Wang,¹ Yikai Wang,¹ Jiaxuan Liu,¹ Long Zhang,¹ Jingtao Ge,¹ Wen Zhou,¹ and Jianjun Yu^{1,*}

¹Key Laboratory for Information Science of Electromagnetic Waves, Fudan University, Shanghai, 200433, China *jianjun@fudan.edu.cn

Abstract: We propose a low-complexity equalization scheme with non-uniform quantization and rotational update mechanism (RUM). The scheme is verified in implementation of DDLMS equalization for 92-Gbaud 10-km offline and 14.7456-Gbaud 25-km FPGA-based real-time PAM4 IM/DD experimental transmission, results show up to 99.5% multiplications (DSP resource usage) are reduced comparing to normal DDLMS equalizer. A large number of equivalent taps can be achieved based on several active taps. © 2024 The Author(s)

1. Introduction

The growing data traffic has given rise to the continuous acceleration of short-reach optical interconnects. By now, 50G PON standard has been released [1], and many efforts have been taken for evaluating and optimizing short reach optical links [2–7]. In high-rate IM/DD transmission, dispersion, bandwidth limitation [3] and other effects brought by the non-ideality of the devices [8,9] can bring severe distortion to the signals, and real-time digital signal processing (DSP) can be unavoidable. DSP occupies around 50% of the power consumption of the entire optical module [5], and the adaptive equalization with large amount of multiplications in FIR filtering costs much complexity, especially in parallel situations.

Pruning and quantization are 2 main ways of reducing complexity in neural networks (NNs). These techniques have inspired the optimization of DSP algorithms in fiber transmission. We have proposed pruned equalization [10,11]. 'Zero-multiplexer' NN-based offline equalization with non-uniform quantization has also been proposed [12]. The multiplications can be turned into shifting operations thanks to the non-uniform quantization based on Powers of Two (PoT) [13] and its derivatives. However, the quantization is based on offline pre-trained weights, and the quantized values are hard to be updated due to the difficulty in additions. Therefore, this kind of quantized equalization is not real-time adaptive and can not track time-varying channels.

In this work, we propose optimized real-time equalization which includes real-time quantify and de-quantify modules and rotational update mechanism (RUM), supporting ultra-low complexity filtering and real-time updating. FIR equalizers with multiple taps can be compressed into equalizers with several active taps, while other taps are quantified and stored. The proposed scheme is implemented in both offline serial and real-time FPGA-based parallel DDLMS equalization, and sent into experimental evaluation. The resulting bit rate and distance (184 Gbit/s over 10 km offline and 29.4912 Gb/s over 25 km real-time) together with the complexity reduction (up to 75.3% in serial offline experiment and up to 99.5% in parallel real-time experiment) have proved its superiority.

2. DDLMS with PoT and RUM

The weights of DDLMS reflect the impulse response of the channel. The values of these weights are not uniform distributed. The blue bars in Fig. 1 (a) illustrates the distribution of the weights in a 15-taps DDLMS equalizer converged offline using ADC data captured in Sec. 4. Each of the blue bars has the same width. It can be seen that in fix-point quantization, the precision near 0 is not enough and the precision at middle range is wasted using 3-bit uniform quantization. In this case, PoT quantization marked as red bars with different widths is more in line with the distribution of weight values. Therefore, using PoT quantization can increase the utilization of quantization bits. There is no performance penalty when the weights are correct, and there is even gain in the case of low quantization bits. There is no performance penalty when the weights are correctly trained, and PoT performs even better in the case of low quantization bits.

However, the PoT quantized values is far less adaptable to addition and subtraction than to multiplication. Due to the little changes of values in each iteration and difficulty to cross uneven PoT levels, we propose RUM, which allows most weights fixed in PoT quantized status, and keeps a small full-precision working area. we decompose left and right half of the entire weight vector (except the middle weight) into *m* groups, and each group has *n* members. The whole number of taps is $n'_{taps} = 2mn + 1$. In this way, the full process of the proposed scheme can be expressed in Fig. 1 (b). After the sliding window of data flipped as signals to be multiplied and accumulated, only the input values at the indexes in working area are sent to be multiplied with the 2n = 4 full precision weights, and the rest are shifted according to the PoT weights. The output is obtained by accumulating the results of shifting and multiplications. The feed-back updating process in the working area is same as traditional DDLMS. Weights outside of working area are frozen and wait for



Fig. 1. (a) The histogram of weights after 3-bit uniform and PoT quantization. (b) The scheme of the proposed DDLMS with PoT and RUM (m = 3, n = 2).



Fig. 2. (a) The experimental setup of 92 GBaud VSB PAM4 transmission. (b) The offline performance comparison between normal and the proposed DDLMS.

working area switching. The switching process consists of 2 steps: PoT real-time quantization and de-quantization. We proposed real-time PoT quantization based on dichotomy, the process can be finished within $log_2(b)$ clocks (b = 16 in Fig. 1 (b)). As for de-quantization, this can be achieved by left-shifting 1 according to the PoT weights. The working area switches from middle to the side areas iteratively with a period of 1000 times updating.

RUM can bring penalty to the performance because rotational updating cannot achieve the best convergence, the performance degrades especially when m is large. The trade-off between performance and complexity can be achieved by adjusting m and n. The optimized number of multiplications using the proposed DDLMS with PoT and RUM is 2n in serial implementation. When it comes to parallel scenarios, only a small number of parallel paths need updating, and other parallel paths can sync and share the updated weights [14]. The advantage of the proposed scheme can be more obvious: only the updating paths need active full-precision weights, while other paths can all use PoT weights to save complexity, and their weights updates by copying the PoT weights in the updating paths after each period.

3. Offline 184 Gbit/s 10 km Transmission

With the aid of our previous published 92-GBaud 10-km vestigial side-band (VSB) transmission system [15], we test the proposed DDLMS scheme in offline 184 Gbit/s PAM4 transmission. The experimental setup and the offline DSP chain are shown in Fig. 2 (a). An external cavity laser (ECL) generates the optical carrier which is modulated by a 40 GHz 3-dB bandwidth LiNbO3 Mach-Zehnder modulator (MZM) driven by a 92 GSa/s arbitrary waveform generator (AWG) and amplified by an electrical amplifier with 22 dB gain. After the 10 km SSMF transmission, the signals are amplified by an Erbium-doped fiber amplifier (EDFA). A tunable optical filter (TOF) with 0.8 nm bandwidth is employed to generate VSB signals. At the receiver side (Rx), the optical signals are detected by a photodiode (PD) with 70 GHz 3-dB bandwidth. The detected electrical signals are amplified by an electrical amplifier and then captured by a 256 GSa/s digital oscilloscope with 59 GHz 3-dB bandwidth. All of the other DSP steps are kept the same for comparison in DDLMS.

In such a high-baudrate bandwidth-limited transmission, ISI can be severe and a large tap number is needed. For the impulse response of the channel, increasing the number of taps has a significant marginal effect on the gain of BER performance. This indicates that the multiplications in different taps do not have the same contribution to the final performance. In this case, we implement normal 73 taps with 73 multiplications in one shot, equivalent 73 taps $(m = 4, n = 9, 73 = 2 \times 4 \times 9 + 1)$ with the same number of taps and lower complexity (2n = 18), and equivalent 293 taps $(m = 2, n = 37, 297 = 2 \times 4 \times 37 + 1)$ with the similar number of multiplications (2n = 74). The BER versus received optical power (ROP) results are shown in Fig. 2 (b). It can be seen that reliable transmission with BER below 20% SD-FEC can be achieved at 2 dBm ROP using all 3 equalizers. The proposed low-complexity scheme can achieve similar performance with tolerable penalty using only 24.7% of the multiplications. The proposed same-complexity scheme can achieve around 0.5 dB ROP gain comparing to normal scheme, and can meet the SD-FEC threshold at 1 dBm ROP.



Fig. 3. (a) The experimental setup of 14.7456 GBaud real-time PAM4 transmission. (b) The photo of the FPGA and the implementation in it (including DDLMS with m=7, n=1). (c) The BER performance and resource utilization comparison figure.

4. Real-time 29.4912 Gb/s 25 km Transmission

We demonstrated a single-lane PAM4 IM/DD transmission system shown in Fig. 3 (a). At transmitter end, an ECL working at 1550 nm is used for providing optical carrier, which is then modulated by an MZM with 20 GHz bandwidth. The MZM is driven by electrical signals generated by the 29.4912-GSa/s DAC (ADA06S032G) with a resolution of 6 bits and amplified by an electrical amplifier (EA). The FPGA provides the 14.7456-GBaud PAM4 signals for the DAC by mapping the loaded pseudorandom binary sequence, which has a pattern length of 1.92×10^5 symbols. The modulated optical signal is fed into a 25-km standard single-mode fiber (SSMF) link with a loss of 0.2 dB/km. At receiver end, the optical signal is first attenuated by a variable optical attenuator (VOA) to control the received optical power (ROP), and then detected by a 15-GHz PD. After that, a 6-bit ADC converter (AAD06S032G) operating at 29.4912 GSa/s is used for real-time receiving the electrical signals. The real-time parallel signal equalization is implemented in the XCVU9P-FLGB2104-2-I Xilinx FPGA (see Fig. 3 (b)), which mainly contains 15-tap 128-path parallel constant modulus algorithm with T/2 spacing [14] and the multiple 64-path parallel DDLMS modules for comparison.

We use ILA to capture signal data after DDLMS and calculate the BER performance. The results are shown in Fig. 3 (c). 4 modules are implemented and the BER performance is tested with multiple ROPs after 25 km SSFM transmission. At low ROP scenarios, The proposed PoT and RUM scheme is effective as 3 equivalent schemes achieve reliable transmission with BERs under HD-FEC when ROP at -5 dBm. However, when ROP gets higher, traditional DDLMS performs better as its optimization ability outperforms that of the RUM. It can be further compensated as RUM offers more equivalent taps (see equivalent 73 taps with m=9, n=4). Overall, all schemes achieve BER below KP4 threshold with ROP at -3 dBm, and have acceptable performances with ultra-low complexity comparing to the traditional 15-taps scheme in [14]. The equivalent 15 taps scheme (with m=7, n=1) using 1 path updating reduces 99.5% of multiplications.

Also, three main resources consumed in FPGA are countered in the right part of Fig. 3 (c): lookup tables (LUTs), DSPs (DSP48E2), and registers (flip-flops). The proposed scheme can sharply reduce up to 99.5% DSP usage comparing to traditional scheme (98.4% for equivalent 73 taps). The cost of LUTs is increased because of PoT quantization and the high-frequency working area access in RUM.

5. Conclusions

We bring PoT quantization and RUM technique to real-time equalization and verified in both offline and experimental form using DDLMS. In real-time DSP, a balance of performance and complexity can be found to outperform the traditional DDLMS. With the proposed scheme, a real-time 29.4912 Gbit/s PAM4 transmission experiment is achieved with BER below KP4 threshold, in which the DDLMS module reduces up to 99.5% multiplications comparing to traditional scheme. Our work is well-expandable as PoT can save complexity when weights are fixed or periodic updated and RUM gives a solution to train the whole equalizer in a partial iterative way. The real-time design provides valuable reference for achieving high-speed, low-power PON implementation. *This work is partially supported by the Natural Science Foundation of China (No. 62305067, No. 61935005, and No. 61835002, No. 62375219, No. 62331004)*.

References

- 1. ITU-T G.9804.3 (2021),
- https://www.itu.int/rec/T-REC-G.9804.3
- 2. D. van Veen et al., OFC, Tu1G.1 (2023).
- 3. W. Wang et al., JLT **41**(14), 4655–4662 (2023).
- 4. O. Ozolins et al., OFC, Tu3I.1 (2023).
- 5. R. Nagarajan et al., JLT 39(16), 5221-5231 (2021).
- 6. J. Li et al., OFC, T3K.6 (2020).
- 7. D. Pilori et al., IPC 10(2) (2018).

- 8. K. Wang et al., JLT 40(4), 979–986 (2022).
- 9. X. Chen et al., TCAD 39(5), 977–990 (2020).
- 10. B. Sang et al., JLT 40(9), 2890–2900 (2022).
- 11. Y. Wang et al., OL 48, 4562-4565 (2023).
- 12. T. Koike-Akino et al., ECOC, Tu2C.2 (2021).
- 13. A. Zhou, arXiv:1702.03044 (2017).
- 14. K. Wang et al., OL 48, 1514-1517 (2023).
- 15. C. Wang et al., PTL 34(18), 941–944 (2022).