

Residual LSTM Equalization versus Classical Equalizers in Linear, Polynomial, and Long-State Nonlinear Channels

Luiza C. Cortez, Cristhof J. Roosen Runge, Ulisses Dias

Abstract—This work investigates the performance of classical and neural equalization schemes under different channel memory and nonlinearity regimes. Four equalizers are compared: Zero Forcing (ZF), Least Mean Squares (LMS), diagonal Volterra equalization, and a residual Long Short-Term Memory (LSTM) equalizer. Three representative communication channel scenarios are considered: (i) linear-memory-dominant channels, (ii) polynomial nonlinear channels, and (iii) long nonlinear state-memory channels. These regimes emulate practical impairments observed in optical wireless links, RF front-ends, photonic systems, visible light communication (VLC), nonlinear power amplification systems, and coherent optical receivers. Simulation results show that classical linear equalizers are highly effective in predominantly linear channels, Volterra equalization is advantageous in polynomial nonlinear regimes, and LSTM equalization becomes preferable when the channel exhibits long nonlinear temporal dependencies and state dynamics.

Keywords—Channel Equalization, Volterra Filters, Long Short-Term Memory, Nonlinear Communication Channels, Recurrent Neural Network.

I. INTRODUCTION

Modern communication systems increasingly operate near hardware and spectral efficiency limits, making nonlinear distortion and long-memory effects unavoidable. Examples include high-power RF amplifiers, photonic integrated systems, visible light communication (VLC), coherent optical links, oversampled receivers, nonlinear front-end electronics, and optical wireless communication (OWC) systems. In such systems, equalization becomes essential for recovering the transmitted symbols under severe inter-symbol interference (ISI) and nonlinear channel distortion.

Classical equalization approaches such as Zero Forcing (ZF) and Least Mean Squares (LMS) remain attractive due to their low complexity and interpretability. However, their performance is limited in nonlinear channels with memory. Volterra equalizers partially overcome this limitation by approximating nonlinear systems through polynomial expansions [1]–[3]. However, when the channel exhibits complex, long-state, or non-polynomial dynamics, Volterra models require an impractically large number of coefficients, complicating real-time deployment [11].

Recently, machine learning approaches, specifically Recurrent Neural Networks (RNNs) and Long Short-Term Memory

(LSTM) networks, have emerged as strong candidates for post-equalization due to their ability to model deep temporal dependencies without explicit analytical mapping [4], [6], [10].

This work provides a focused comparative analysis of four architectures—ZF, LMS, diagonal Volterra, and a residual LSTM equalizer—across three distinct, well-defined physical regimes: (i) linear-memory-dominant, (ii) low-order polynomial nonlinear, and (iii) long nonlinear state-dependent memory channels. We isolate these mechanisms to identify the exact operating boundaries where machine learning outperforms classical parametric signal processing.

II. MOTIVATION AND PRACTICAL CHANNEL REGIMES

The channel models adopted in this work are behavioral models. They are not intended to reproduce a single standardized physical channel, but to isolate impairment mechanisms commonly observed in practical systems: linear dispersive memory, polynomial nonlinear distortion, and nonlinear state-dependent memory.

This modeling strategy is widely employed in nonlinear equalization and digital predistortion studies, where Volterra, memory-polynomial, and generalized memory-polynomial structures are used to represent nonlinear devices with memory, particularly RF power amplifiers and wideband front-ends [3], [8], [9]. Neural equalizers have also been investigated in optical and visible-light links where channel impairments combine bandwidth limitation, nonlinear conversion, and temporal memory [6], [7].

The three regimes considered here represent different dominant impairment mechanisms.

a) Regime 1 – Linear Memory with Weak Nonlinearities:

This regime comprises channels where linear memory (ISI) is dominant, but weak polynomial nonlinearities are also present. The nonlinear distortion exists yet remains small relative to the linear component, making the channel amenable to classical FIR-based equalizers. Examples include frequency-selective wireless links after partial RF-chain linearization; sub-6 GHz or mmWave 5G receivers with residual nonlinearity below saturation; and optical links where device nonlinearity is weak compared to multipath.

b) Regime 2 – Polynomial Nonlinear with Finite Mem-

ory: This regime encompasses channels dominated by polynomial nonlinearities with finite memory. The nonlinear distortion follows a memory-polynomial or low-order Volterra

structure, with multiple delayed terms contributing to the output. Examples include RF transmitters operating near power-amplifier compression; wideband radio equipment with frequency-dependent nonlinearity; and IM/DD optical or VLC systems with LED or modulator nonlinear transfer functions.

c) Regime 3 – Nonlinear State Memory with Slow Dynamics: This regime models channels where a slowly varying internal state (e.g., thermal temperature, bias voltage, AGC gain) nonlinearly interacts with the input signal. In addition to polynomial and memory terms, the output includes nonlinear transformations (such as trigonometric or hyperbolic functions) that emulate complex behaviors like phosphor relaxation, junction temperature effects, or gain recovery phenomena. Examples include RF power amplifiers with thermal memory; optical transmitters with slow laser dynamics; and VLC emitters with phosphor relaxation.

III. SYSTEM AND CHANNEL MODELS

The simulations employ Gray-mapped 16-QAM modulation, a widely used format in spectrally efficient communication systems. Its sensitivity to residual ISI and nonlinear distortion makes it computationally suitable for practical Monte Carlo BER estimation while still stressing the equalizers under nonlinear conditions.

The transmitted complex-valued 16-QAM sequence is denoted by $x[n]$. The linear channel output is given by

$$u[n] = \sum_{k=0}^{M-1} h[k]x[n-k] \quad (1)$$

where $u[n]$ is the linear channel response, $h[k]$ is the complex channel impulse response, and M is the channel memory length.

The channel taps follow an exponentially decaying model:

$$h[k] \propto \text{decay}^k \quad (2)$$

where the parameter *decay* controls the persistence of the channel memory. The channel impulse response is normalized to unit energy, such that $\sum_{k=0}^{M-1} |h[k]|^2 = 1$.

The nonlinear terms employed in this work are inspired by memory-polynomial and generalized memory-polynomial models commonly used in RF power amplifier modeling [3], [8], [9]. Terms such as

$$u[n]|u[n]|^2 \quad (3)$$

represent third-order nonlinear distortion and intermodulation effects. Delayed cross-memory terms emulate nonlinear memory mechanisms.

A. Regime 1: Linear Memory Dominant

The first regime models channels where ISI is dominant and nonlinearities are weak. The received signal is modeled as

$$y[n] = u[n] + 0.08\alpha u[n]|u[n]|^2 + 0.04\alpha u[n-1]|u[n-3]|^2 + w[n] \quad (4)$$

where α controls nonlinear intensity and $w[n]$ is additive white Gaussian noise (AWGN).

In this regime, the channel remains mostly linear, favoring classical FIR-based equalizers.

B. Regime 2: Polynomial Nonlinear

The second regime models channels dominated by polynomial nonlinearities:

$$\begin{aligned} v[n] = & u[n] + 0.35\alpha u[n]|u[n]|^2 \\ & + 0.20\alpha u[n-1]|u[n-1]|^2 \\ & + 0.18\alpha u[n]|u[n-2]|^2 \\ & + 0.12\alpha u[n-2]|u[n-3]|^2. \end{aligned} \quad (5)$$

A mild amplitude-dependent compression is subsequently applied as $\tilde{v}[n] = v[n]/(1+0.03\alpha|v[n]|^2)$, followed by AWGN addition, such that $y[n] = \tilde{v}[n] + w[n]$.

This structure resembles a low-order Volterra or memory-polynomial nonlinear system.

C. Regime 3: Long Nonlinear State Memory

The third regime introduces nonlinear temporal state dynamics through a slowly varying first-order recursive state variable:

$$q[n] = 0.990q[n-1] + 0.010|u[n]|^2 \quad (6)$$

The resulting state sequence is normalized to unit maximum, i.e., $q[n]/\max_n q[n]$, before being used in the channel model. The nonlinear channel output before compression and phase rotation is given by

$$\begin{aligned} v[n] = & u[n] + 0.015\alpha u[n]|u[n]|^2 \\ & + 0.08\alpha u[n-1]|u[n-4]|^2 \\ & + 0.26\alpha \tanh(1.6u[n-4]) \\ & + 0.30\alpha \sin(1.2u[n-7]u^*[n-2]) \\ & + 0.22\alpha \sin(1.1u[n-10]) \\ & + 0.55\alpha q[n]u[n]. \end{aligned} \quad (7)$$

For complex-valued arguments, the nonlinear functions in (7) are applied independently to the real and imaginary components, i.e., $f(z) = f(\Re\{z\}) + jf(\Im\{z\})$, for $f \in \{\tanh, \sin\}$.

A mild amplitude-dependent compression is first applied as $\tilde{v}[n] = v[n]/(1 + 0.02\alpha|v[n]|^2)$. A state-dependent phase rotation is then applied as $\bar{v}[n] = \tilde{v}[n] \exp\{j0.12\alpha \sin(|v[n]| + q[n])\}$. Finally, AWGN is added such that $y[n] = \bar{v}[n] + w[n]$.

The state-dependent term $q[n]u[n]$ emulates slow nonlinear memory effects such as thermal dynamics, bias-network memory, or gain recovery phenomena. The nonlinear scaling parameter α is set to 0.10, 0.30, and 0.45 for Regimes 1, 2, and 3, respectively.

IV. EQUALIZATION SCHEMES

A. Zero Forcing (ZF)

The ZF equalizer estimates the transmitted symbol through a linear finite impulse response (FIR) filter:

$$\hat{x}[n] = \mathbf{w}^H \mathbf{r}[n] \quad (8)$$

where $\hat{x}[n]$ is the estimated transmitted symbol, $\mathbf{w}$ is the equalizer coefficient vector, $\mathbf{r}[n]$ is the received observation window, and $(\cdot)^H$ denotes the Hermitian transpose. The ZF coefficients are obtained from the known linear channel response so as to approximately invert the channel and suppress inter-symbol interference.

B. LMS Equalizer

The LMS equalizer adaptively updates its coefficients according to

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu e[n] \mathbf{r}^*[n] \quad (9)$$

where μ is the adaptation step size and $e[n]$ is the estimation error.

C. Diagonal Volterra Equalizer

The diagonal Volterra equalizer extends linear FIR filtering using nonlinear polynomial basis functions. The equalizer output is given by

$$\hat{x}[n] = \mathbf{w}^H \boldsymbol{\phi}[n] \quad (10)$$

where $\mathbf{w}$ is the equalizer coefficient vector and $\boldsymbol{\phi}[n]$ is the nonlinear regression vector defined as

$$\boldsymbol{\phi}[n] = [r[n-k], r[n-k]|r[n-k]|^2]_{k=0}^{K-1}. \quad (11)$$

where $K = 21$ is the equalizer memory length.

The first terms correspond to the linear FIR components, while the cubic terms model third-order nonlinear distortions. This structure efficiently models polynomial nonlinearities while avoiding the combinatorial growth associated with full Volterra expansions.

D. LSTM Equalizer

The proposed neural equalizer employs a residual LSTM structure. First, a linear least-squares (LS) baseline estimate is obtained. The LSTM then learns a nonlinear mapping $f_{\text{LSTM}}(\mathbf{r}[n])$ from the received observation window $\mathbf{r}[n]$ to a residual correction term:

$$\hat{x}_{\text{LSTM}}[n] = \hat{x}_{\text{LS}}[n] + f_{\text{LSTM}}(\mathbf{r}[n]). \quad (12)$$

where $\hat{x}_{\text{LS}}[n]$ is the LS baseline estimate and $f_{\text{LSTM}}(\cdot)$ represents the nonlinear function implemented by the recurrent neural network.

The residual structure stabilizes training and reduces the learning load by allowing the LSTM to focus primarily on nonlinear impairments, while the linear component is handled by the LS baseline.

The training and validation datasets contained 90,000 and 20,000 symbols, respectively. For BER estimation, Monte Carlo simulations were performed adaptively at each E_b/N_0 value. At least 200,000 16-QAM symbols were evaluated, and simulation continued until at least 100 bit errors were observed for each equalizer or a maximum of 30 million symbols was reached.

V. COMPLEXITY-AWARE COMPARISON

ZF and LMS employ 42 complex FIR coefficients, corresponding to 84 real-valued parameters. The diagonal Volterra equalizer uses 21 linear and 21 cubic complex coefficients, also corresponding to 84 real-valued parameters. The residual LSTM employs a single recurrent layer with 24 hidden units and a fixed observation window of 15 received complex samples. The samples are presented sequentially to the LSTM as 15 time steps, with the in-phase and quadrature components forming two real-valued input features per time step. An attention layer computes a weighted combination of the recurrent outputs, which is concatenated with the corresponding LS baseline estimate. The resulting representation is processed by a small fully connected network that generates the two-dimensional I/Q residual correction, which is added to the LS baseline estimate according to (12). The neural residual branch contains 3,687 trainable real-valued parameters. Including the 15-complex-coefficient LS baseline, the complete residual equalizer comprises 3,717 real-valued parameters, compared with 84 real-valued parameters for each of the ZF, LMS, and diagonal Volterra equalizers. Training was performed using the AdamW optimizer with early stopping based on validation loss. The network was trained using SmoothL1 loss, a learning rate of 1.2×10^{-3} , mini-batches of 512 samples, and AdamW weight decay regularization.

The equalizers were configured or identified under the same channel conditions. The ZF coefficients were obtained from the known linear channel response, whereas the diagonal Volterra coefficients and the linear baseline of the residual LSTM were estimated through regularized least-squares optimization. The LMS equalizer employed iterative coefficient adaptation, while the residual LSTM was trained using gradient-based optimization with the AdamW algorithm.

This difference does not invalidate the comparison because the objective is not to claim strict equal-complexity optimality between heterogeneous model classes. Instead, the goal is to compare BER performance while explicitly accounting for model parameterization and to identify which structural bias is appropriate for each channel regime.

The results indicate that increasing model flexibility does not necessarily improve equalization performance. Although the LSTM possesses substantially greater representational capacity than the linear and Volterra equalizers, it is outperformed in Regimes 1 and 2 by models whose structure more closely matches the underlying channel characteristics. Conversely, when the channel exhibits nonlinear state-dependent memory, the additional modeling flexibility of the LSTM becomes advantageous. These observations suggest that structural compatibility between the equalizer and the channel may be more important than model complexity alone.

VI. RESULTS AND DISCUSSION

The simulations employ 16-QAM modulation under AWGN conditions. BER performance is evaluated over a range of E_b/N_0 values from 6 dB to 22 dB.

A. Regime 1 Results

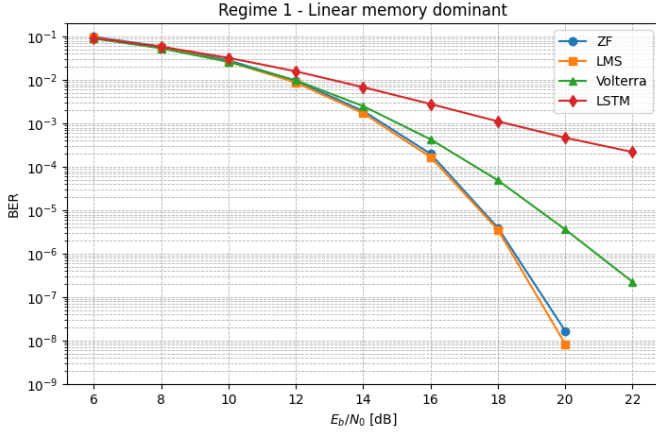


Fig. 1. BER performance for the linear-memory-dominant channel.

Fig. 1 shows the BER curves for the linear-memory-dominant channel. ZF and LMS achieve the best BER performance, while the LSTM clearly underperforms. The Volterra equalizer remains competitive at intermediate SNRs, but its performance is slightly inferior to ZF and LMS because part of its representational capacity is devoted to nonlinear basis functions that provide limited benefit in this predominantly linear regime.

These results also indicate that neural equalization is not advantageous when the dominant impairment is essentially linear ISI.

B. Regime 2 Results

Fig. 2 presents the results for the polynomial nonlinear channel. The Volterra and LSTM equalizers significantly outperform ZF and LMS, confirming that linear equalization is insufficient when nonlinear distortion dominates.

The Volterra equalizer performs particularly well because its basis functions closely match the polynomial structure of the channel. In fact, the Volterra equalizer achieves the lowest observed BER values in this regime, while the LSTM remains highly competitive. These results indicate that explicitly polynomial equalization structures remain highly effective when the dominant channel impairments are well approximated by low-order nonlinear polynomial models.

C. Regime 3 Results

Fig. 3 shows the BER curves for the long nonlinear state-memory channel. In this scenario, the LSTM achieves the best BER performance across the entire SNR range.

Unlike Volterra equalizers, which rely on static polynomial kernels, the LSTM can internally represent nonlinear temporal

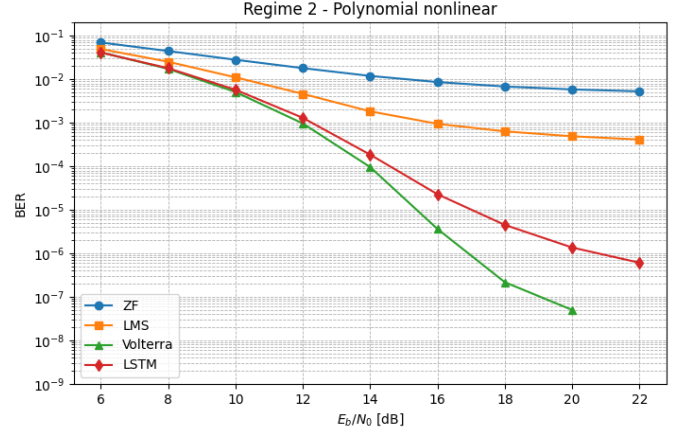


Fig. 2. BER performance for the polynomial nonlinear channel.

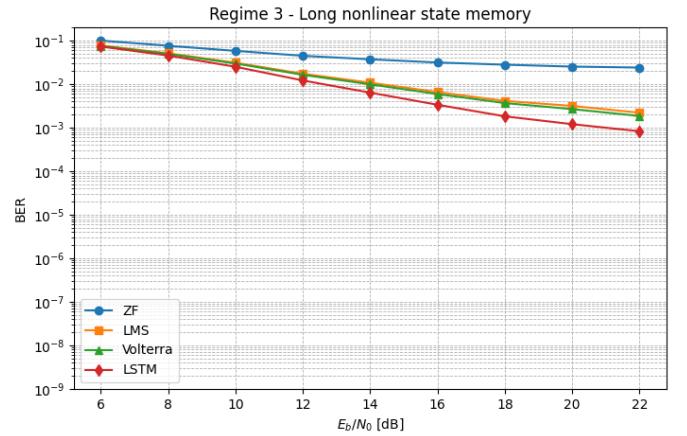


Fig. 3. BER performance for the long nonlinear state-memory channel.

state evolution. Consequently, the LSTM becomes more effective as the channel departs from purely polynomial behavior and transitions toward dynamic nonlinear systems.

The advantage becomes more visible at medium and high SNR values, where residual nonlinear memory rather than AWGN becomes the dominant impairment.

VII. CONCLUSION

This work compared classical and neural equalization approaches under three representative nonlinear channel regimes.

The results reinforce the idea that no equalization strategy is universally optimal. Rather, the most effective equalizer depends on the underlying channel characteristics and on the type of distortion that dominates system performance. Linear models excel in low-distortion paths, Volterra structures provide efficient parametric fits for low-order polynomial systems, and residual LSTMs become advantageous when dealing with long, non-polynomial state dependencies. These findings suggest that recurrent neural equalizers are particularly attractive for communication systems exhibiting dynamic nonlinear temporal behavior, where linear and static polynomial models may become insufficient. Examples include photonic communication systems, visible light communication (VLC) links,

RF front-ends with long-term memory effects, and nonlinear wireless communication systems.

An important observation is that increased model complexity does not automatically translate into improved BER performance. In several scenarios, simpler structures matched to the channel characteristics outperform more flexible neural models. Future work includes complexity reduction strategies and hybrid Volterra-neural architectures.

REFERENCES

- [1] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, vol. 8, no. 3, pp. 10–26, July 1991.
- [2] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York, NY, USA: Wiley, 2000.
- [3] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [5] D. Tian, P. Miao, H. Peng, W. Yin, and X. Li, "Volterra-aided neural network equalization for channel impairment compensation in visible light communication system," *Photonics*, vol. 9, no. 11, Art. no. 845, 2022.
- [6] X. Lu, C. Lu, W. Yu, L. Qiao, S. Liang, A. P. T. Lau, and N. Chi, "Memory-controlled deep LSTM neural network post-equalizer used in high-speed PAM VLC system," *Optics Express*, vol. 27, no. 5, pp. 7822–7833, 2019.
- [7] P. Miao, B. Zhu, C. Qi, Y. Jin, and C. Lin, "Deep hybrid neural network-based channel equalization in visible light communication," *IEEE Communications Letters*, vol. 26, no. 1, pp. 1–5, Jan. 2022.
- [8] Y.-J. Liu, J. Zhou, W. Chen, and B.-H. Zhou, "A robust augmented complexity-reduced generalized memory polynomial for wideband RF power amplifiers," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2389–2401, May 2014.
- [9] G. Sun, C. Yu, Y. Liu, S. Li, and J. Li, "A modified generalized memory polynomial model for RF power amplifiers," *Progress In Electromagnetics Research Letters*, vol. 47, pp. 97–102, 2014.
- [10] Y. Zhao, X. Huan, and L. Zhang, "Recurrent Neural Networks for Memory-Rich Nonlinear Channel Equalization in High-Speed Links," *IEEE Transactions on Communications*, vol. 72, no. 2, pp. 1102–1115, Feb. 2024.
- [11] M. Jeraj and G. L. Sicuranza, "On the Complexity and Modeling Limits of Volterra versus Deep Neural Equalizers," *IEEE Signal Processing Letters*, vol. 30, pp. 415–419, May 2023.