A Lightweight Probabilistic Sequence Model for Efficient Nonlinear LED Channel Equalization

Dylan Jones, Banghua Zhu, Lih Lin

Introduction

- Currently, **communication between nodes in modern AI datacenters is I/O bound**, especially when training foundation models that are distributed across many GPUs
- **MicroLEDs offer a potential high-throughput and energy-efficient (<pJ/bit) alternative to current data center pluggable optics** by leveraging hundreds of individual channels transmitted in parallel
- The achievable data rate is limited by the nonlinearity of the device and requires equalization signal processing
- This work demonstrates a **lightweight probabilistic channel model** that captures LED nonlinearity and noise, enabling end-to-end training of encoder and decoder models **to achieve faster, more efficient equalization in near-distance (~10 m) interconnects.**

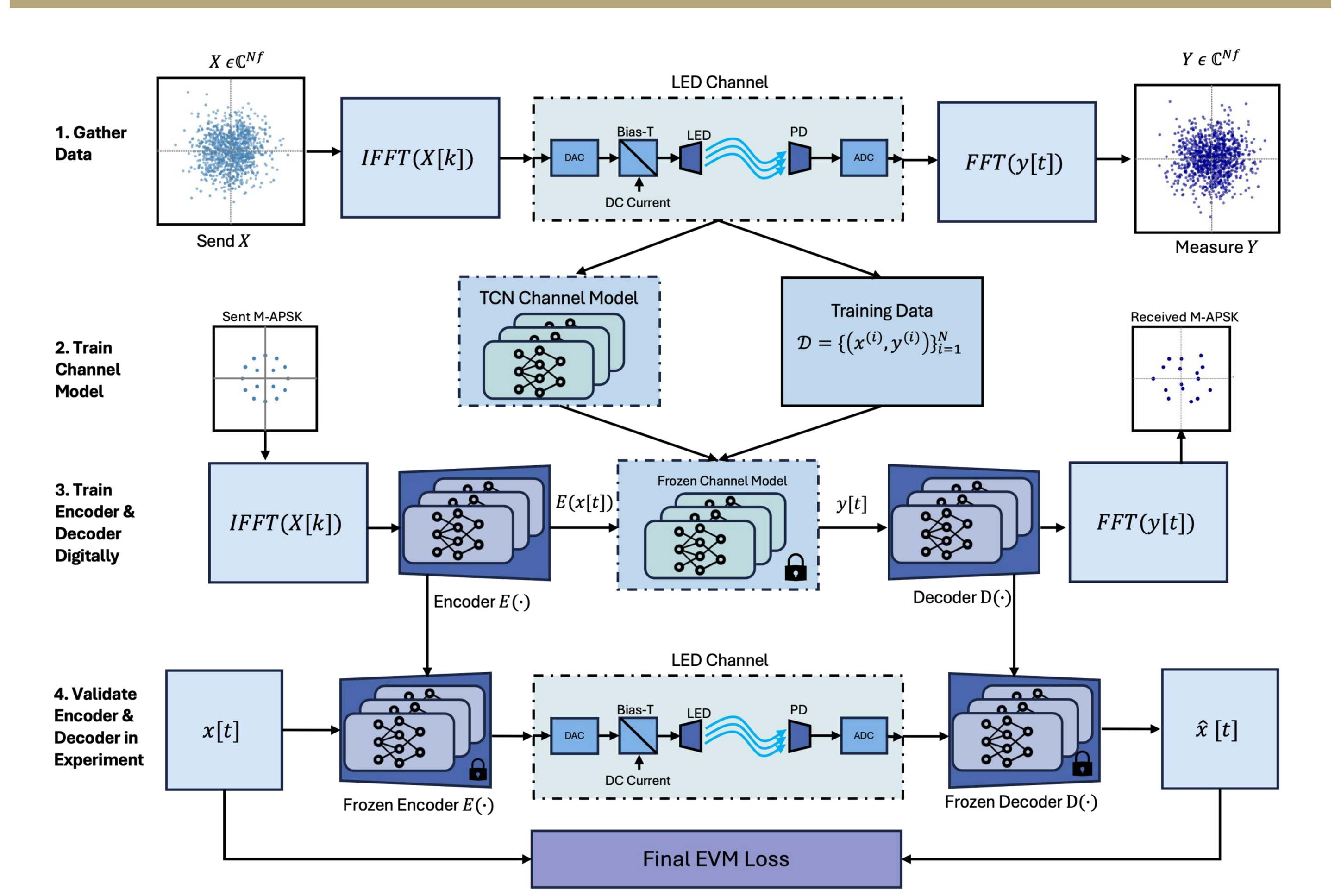
ABC Model

$$\frac{dN(t)}{dt} = \frac{I_{IN}(t)}{qV_{QW}} - AN(t) - BN^2(t) - CN^3(t)$$

$$\Phi_{\text{light}} = \mathrm{E}_{\text{phot}} V_{QW} B N^2(t)$$

- Classically, recombination in the LED quantum well is captured by the ABC equation
- A represents Shockley–Read–Hall recombination, B describes radiative recombination, and C accounts for nonradiative Auger recombination
- This equation dominates LED low-pass behavior for smaller, faster devices but **doesn't capture noise and other physics, such as temperature**

Experimental and Modeling Flow



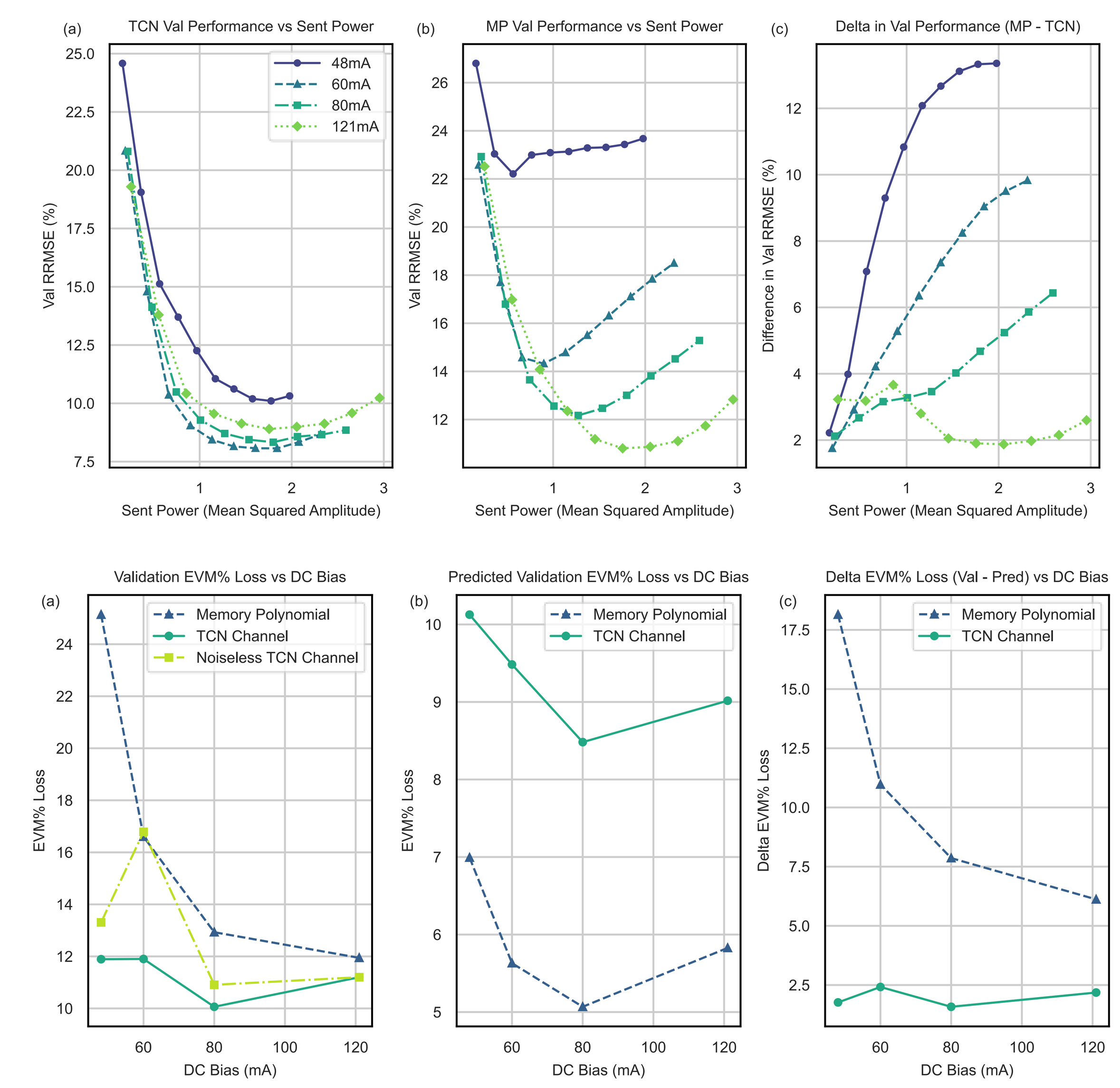
Probabilistic TCN Model Formulation

$$(\mu[n], \sigma[n], \nu[n]) = f_\theta(x[n - R + 1 : n])$$

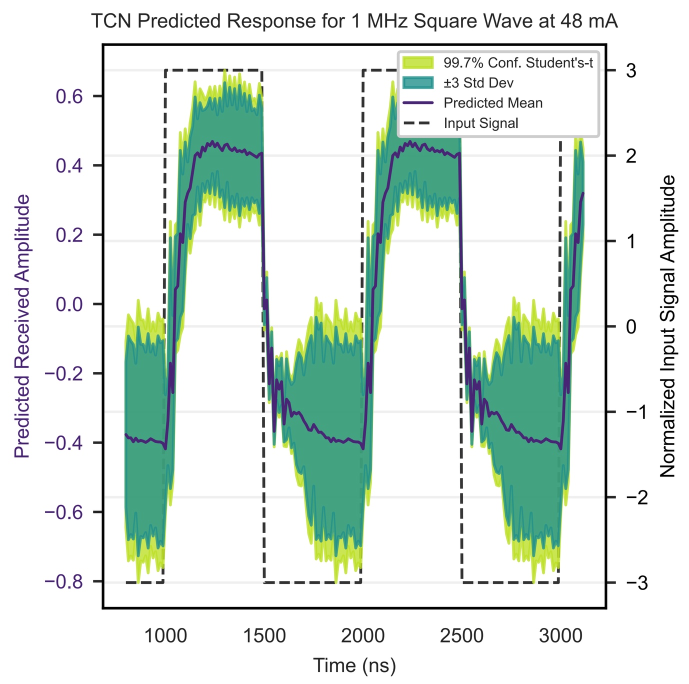
- **Predict the mean, standard deviation, and degrees of freedom of received signals over time** with a temporal convolutional network (TCN)
- This framework allows a heteroscedastic noise distribution with wide tails to be learned under a **maximum likelihood framework**
- Most current approaches predict the mean received signal under a mean squared error approach, which implicitly assumes additive white Gaussian noise, which is often not true with nonlinear systems with dynamics and memory

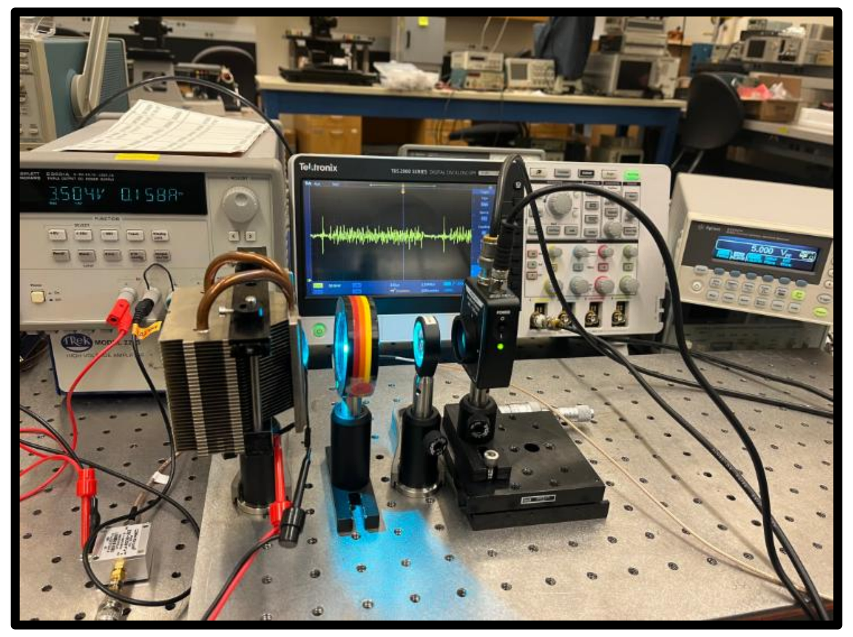
Comparison with Memory Polynomial

- Currently, many real-time systems use truncated forms of Volterra series, such as memory polynomials (MPs), because they are linear in the parameters, efficient to deploy at scale, and can be solved with least-squares procedures
- We compared TCN encoder/decoder EVM% loss trained on an MP channel model and a TCN channel model
- **Consistently, at lower DC bias (higher nonlinearity lower power consumption), the TCN channel model results in better equalization**

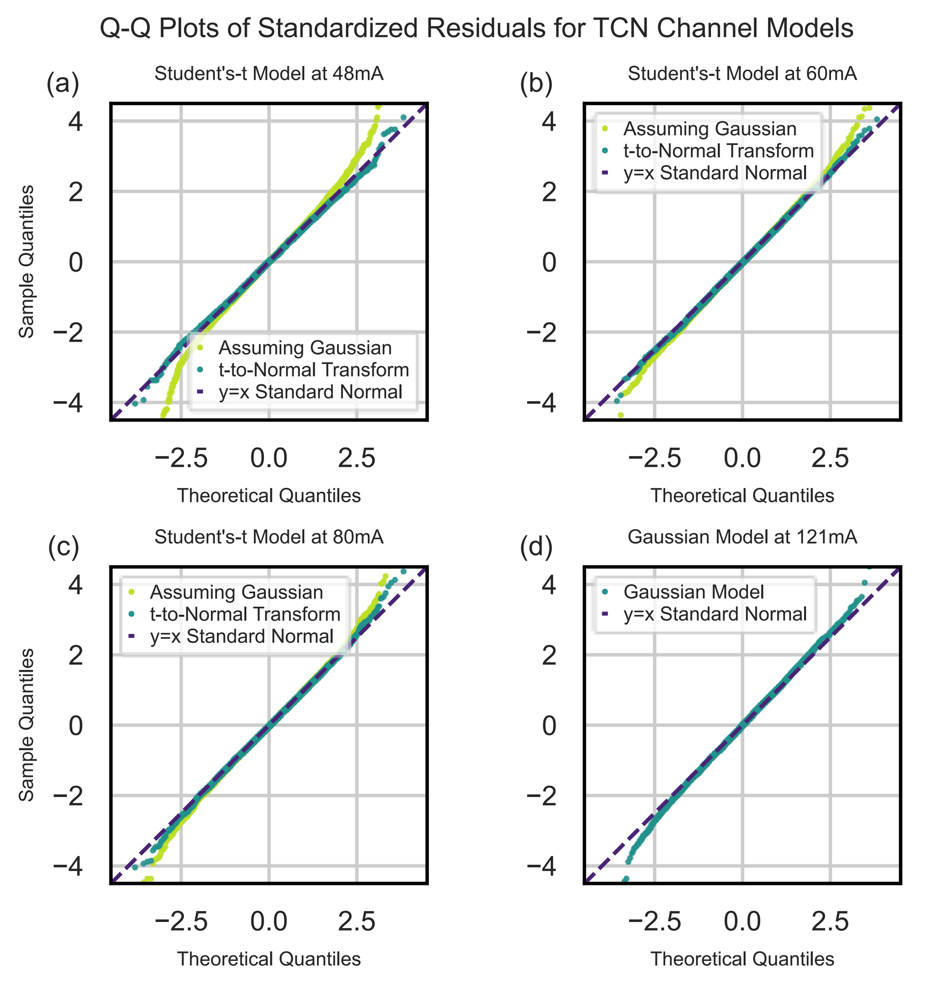


Example Probabilistic Output | Experimental Photo





Q-Q Plot Validation of Learned Channel Distribution



Acknowledgements

- Thank you to my advisors, Lih Lin and Banghua Zhu, for feedback and mentorship
- This work is supported by the National Science Foundation through Award No. CMMI-2227285 (for machine learning development) and ECCS-2430679 (for experimental implementation)

Advisors: Banghua Zhu and Lih Lin