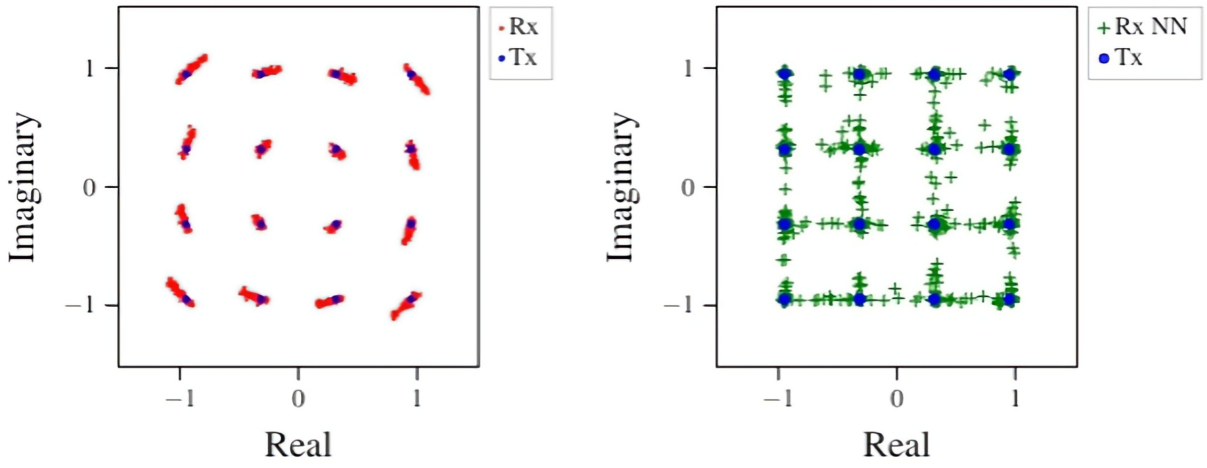
Research Thesis

Investigation of Neural Equalization-Enhanced Phase Noise Mitigation



Shalini Rai

Date of hand out:	April 09, 2024
Date of hand in:	March 3, 2025
Supervisor:	Sebastian Jung
	Tim Janz

submitted to

University of Stuttgart
Institute of Telecommunications

Abstract

This thesis explores the challenges of using Neural Networks (NN) for Equalization-Enhanced Phase Noise (EEPN) mitigation in communication systems, specifically focusing on the pitfalls associated with low Achievable Information Rate (AIR) and model generalization. The primary issue addressed is the phenomenon of the 'jail window', which occurs when a neural network fails to generalize across different operating conditions, such as varying linewidths and link distances. This problem is exacerbated by the use of Mean Squared Error (MSE) as a loss function, which tends to overemphasize large errors and can lead the model to memorize noise patterns rather than learn underlying signal characteristics. The thesis investigates the impact of MSE on the formation of 'jail window', supported by existing literature on its role in signal processing tasks. Through this exploration, the thesis aims to provide insights into the limitations of neural networks in communication systems, offering suggestions for improving model robustness and generalization. Ultimately, the work shifts the focus from EEPN mitigation to a broader discussion on the inherent challenges and potential solutions when applying neural networks to complex signal processing problems.

Title page image: EEPN and Jail Window Effect

Contents

Notations

x	scalar variables, especially in time domain: italic lower case letters
X	scalar variables, especially in frequency domain or random variables: italic upper case letters
$\mathbf{x}$	vector variables: boldface upright lower case letters
$\mathbf{X}$	matrix variables: boldface upright upper case letters
$x(t)$, $H(z)$	functions of continuous variables: argument is placed in round parentheses
$x[k]$	functions of discrete variables: argument is placed in brackets
x_k	scalar elements of a vector or element of a sequence: element index is a subscript
$X_{r,c}$	scalar elements of a matrix: row and column indices are subscripts
$a \cdot b$, $\mathbf{a} \cdot \mathbf{b}$, $\mathbf{H} \cdot \mathbf{x}$	scalar product of two scalars, vectors or matrices
$\mathbf{a} \times \mathbf{b}$, 5.1×10^{-9}	cross product of two vectors or scientific notation of numbers
$a * b$	linear convolution
$a \circledast b$	circular convolution
$P[X < 0]$	Probability of an event on a random variable X
$P[X = x \mid Y = y]$, $P[x \mid y]$	Probability of the event $X = x$ under the condition $Y = y$
$p_X(x)$, $p(x)$	continuous probability density function of a random variable X

$p_{X\|Y=y}(x)$, $p(x\|y)$	probability density function of a random variable X under the condition $Y = y$
$\mathrm{E}[X]$	expected value of a random variable X
$X(\omega) = \mathscr{F}\{x(t)\}$, $X(\omega) \circ\!\!-\!\!\bullet x(t)$	Fourier transform from time-domain signal $x(t)$ into frequency-domain spectrum $X(\omega)$
$A(z,t)$	complex envelope of the optical field
z	propagation distance
t	time in a moving frame
α	attenuation coefficient
β_2	Group Velocity Dispersion (GVD) parameter
γ	nonlinear Kerr coefficient
B_j	experimentally determined coefficients
ω_j	resonance frequencies of the material
ω	optical angular frequency
β_0	phase constant
v_g	group velocity
β_1	inverse of group velocity
β_3	third-order dispersion (TOD)
λ	wavelength of light
c	speed of light in vacuum
$\phi(t)$	accumulated phase noise
D	dispersion coefficient
L	fiber length
f_0	central optical frequency

$X_{LO}(f)$	local oscillator spectrum
δ	Total EEPN
$h_{ps}(t)$	pulse shaping filter response
$S(f)$	power spectral density of the LO laser
f_s	symbol rate
Δv	linewidth of the laser
$\|C_i\|^2$	power of the individual constellation point
$E[\|C\|^2]$	average power of the constellation
$w_{ji}^{(l)}$	weight connecting neuron i in layer l to neuron j in layer $l+1$
$a_i^{(l)}$	activation of neuron i in layer l
$b_j^{(l)}$	bias
η	Learning rate (step size)
$\nabla \mathscr{L}(w^{(t)}; x_i, y_i)$	Gradient of the loss function
$H(X)$	entropy of the transmitted signal
$H(X\|Y)$	conditional entropy
X	transmitted symbols
R	received signal
$f(R)$	equalized signal
$\mathbb{E}[\cdot]$	expectation over the signal distribution

1 Introduction

1.1 Introduction to Investigation of Neural EEPN Mitigation

In coherent optical communication systems, the interplay between chromatic dispersion (CD) compensation and laser phase noise gives rise to a phenomenon known as Equalization-Enhanced Phase Noise (EEPN) [1, 2, 3]. Unlike conventional phase noise, EEPN is caused by the Digital Signal Processing (DSP) algorithms, particularly dispersion compensation, that are integral to coherent receivers. This effect was quantified by Shieh and Ho in their work, which demonstrated that due to the non-commutability of the phase noise and channel dispersion, a new source of impairment is identified that is originated from the receive laser phase noise enhanced by the electronic equalization. This (electronic) EEPN imposes a tighter constraint on the receive laser phase noise for the systems with large chromatic dispersion and high symbol rate [1]. The EEPN is found to scale with accumulated CD, linewidth as well as symbol rates. Consequently, the linewidth requirements for Local Oscillator (LO) become more stringent for systems using electronic CD compensation. Simulations demonstrate that EEPN introduces significant optical signal-to-noise ratio (SNR) penalties in long-haul terrestrial links, particularly when using high spectral efficiency formats. Moreover, the phase noise component caused by EEPN is approximately twice as strong as the corresponding amplitude noise [4].

As the limits of model-based compensation become apparent under dynamic channel conditions and in long-haul systems, recent research has increasingly turned to data-driven approaches. Machine learning and, in particular, deep learning techniques have been applied to develop neural network (NN)–based [5] equalizers that can learn the nonlinear relationships governing impairments in coherent optical systems. Freire et al. discuss the broader challenges and opportunities of applying machine learning to optical communications, noting that these methods hold promise for mitigating complex impairments [6]. In parallel, deep learning frameworks have been demonstrated for nonlinear equalization; for example, Jarajreh et al. have shown that deep neural network architectures can be employed to enhance DSP performance in coherent systems [7].

Despite these promising results, neural network–based approaches to EEPN mitigation introduce new challenges, such as the need for extensive training data that capture the full variability of dispersion and phase noise interactions, sensitivity to hyperparameter settings, increased computational complexity compared to traditional methods, and usage of appropriate loss functions. Initially, the project aimed to develop a neural network-based solution for compensating EEPN. However, as the experiments progressed, results showed phenomena such as

the ‘jail window‘ [6], which is indicative of poor generalization. This observation prompted a reevaluation of the approach, leading to a deeper exploration of the pitfalls associated with neural network-based compensation techniques in communication systems. This thesis aims to systematically investigate these pitfalls and evaluate the robustness of neural equalization strategies in mitigating EEPN. Through a combination of theoretical analysis and experimental validation, the work seeks to chart a pathway toward more reliable DSP architectures for next-generation coherent optical systems.

2 Preliminaries

2.1 Optical Channel

An optical communication channel refers to the medium through which optical signals propagate, typically an optical fiber. The fiber-optic channel is subject to various impairments, including attenuation, dispersion, non-linearity, and noise, all of which influence signal transmission [8].

The fundamental equation governing optical signal propagation in a fiber is the Nonlinear Schrödinger Equation (NLSE), which accounts for dispersion, fiber loss, and nonlinear effects

$$\frac{\partial A(z,t)}{\partial z} + \underbrace{\frac{\alpha}{2}A(z,t)}_{\text{Attenuation Term}} - \underbrace{i\frac{\beta_2}{2}\frac{\partial^2 A(z,t)}{\partial t^2}}_{\text{Disersion Term}} + \underbrace{i\gamma|A(z,t)|^2 A(z,t)}_{\text{Non-linearity Term}} = 0 \tag{2.1}$$

This equation is fundamental in optical fiber communication, particularly for understanding the trade-offs between non-linearity and dispersion.

In long-haul optical fiber communication, dispersion and non-linearity interact in a complex way, often leading to signal distortions. The interaction of Kerr non-linearity and dispersion gives rise to effects such as self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM), all of which impact system performance.

The optical channel is also affected by polarization-mode dispersion (PMD), where different polarization components of the signal travel at slightly different speeds, leading to distortions. The combined effects of these impairments necessitate digital signal processing (DSP) and machine learning-based equalization in modern optical networks.

2.1.1 Chromatic Dispersion

Chromatic dispersion (CD) [8, 9] occurs due to the frequency dependence of the refractive index $n(\omega)$. This dependence causes different spectral components of an optical pulse to travel at different speeds, leading to pulse broadening. It is also referred to as intramodal dispersion, or GVD.

Neglecting both, nonlinear effects ($\gamma = 0$) and attenuation ($\alpha = 0$), the NLSE of Eq.(2.1) simplifies to,

$$\frac{\partial A(z,t)}{\partial z} - i\frac{\beta_2}{2}\frac{\partial^2 A(z,t)}{\partial t^2} = 0 \tag{2.2}$$

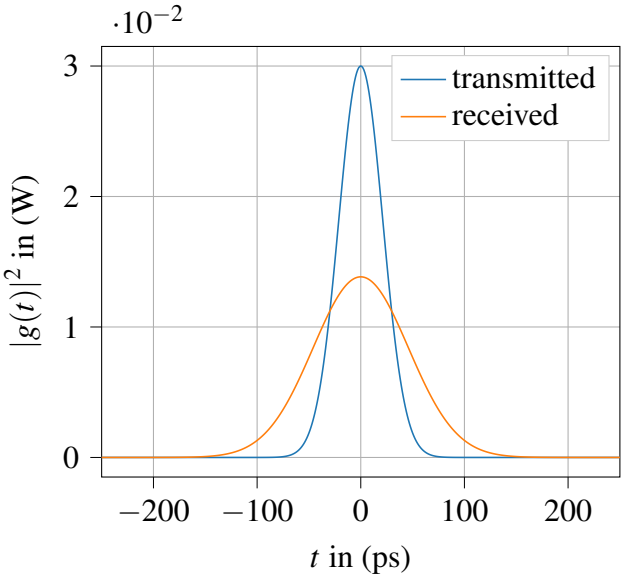


Figure 2.1: Chromatic Dispersion in Standard Single Mode Fiber with Gaussian Impulse and Fiber Length of 80km

Figure 2.1 shows that compared to the transmit impulse the received one has significantly broadened in time. The absolute value of the spectrum, nevertheless, stays the same.

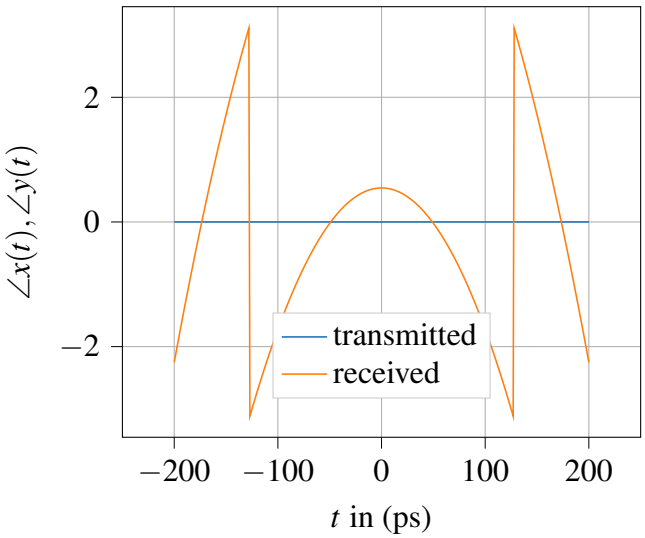


Figure 2.2: Chromatic Dispersion Phase Plot for Standard Single Mode Fiber with Gaussian Impulse and Fiber Length of 80km

Figure 2.2 shows plot of the phase of the received signal where the typical parabolic shift can be seen.

Sellmeier Equation for Material Dispersion

Fiber dispersion plays a critical role in propagation of short optical pulses because different spectral components associated with the pulse travel at different speeds given by $n(\omega)$. Even when the nonlinear effects are not important, dispersion-induced pulse broadening can be detrimental for optical communication systems. In the nonlinear regime, the combination of dispersion and non-linearity can result in a qualitatively different behavior. The frequency-dependent refractive index $n(\omega)$ of the fiber material is described by the Sellmeier equation,

$$n^2(\omega) = 1 + \sum_{j=1}^{m} \frac{B_j \omega_j^2}{\omega_j^2 - \omega^2} \tag{2.3}$$

This equation helps determine the wavelength dependence of the refractive index.

Propagation Constant and Dispersion Parameters

The propagation constant $\beta(\omega)$ is expanded as a Taylor series around a central frequency ω_0,

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{\beta_2}{2}(\omega - \omega_0)^2 + \frac{\beta_3}{6}(\omega - \omega_0)^3 + \dots \tag{2.4}$$

GVD

CD has two components:

- Material dispersion coefficient D_M: refractive index in core dependent on wavelength, n = n(λ), i.e., different wavelengths propagate at different phase velocities through the fiber.
- Waveguide dispersion coefficient D_W: account for geomtric constraints/interaction of core and cladding; a fraction of the wave propagates through the cladding which has a lower refractive index (also wavelength dependent).

Nonlinear effects in optical fibers can manifest qualitatively different behaviors depending on the sign of the GVD parameter. For wavelengths such that ($\lambda > \lambda_D$), the fiber is said to exhibit normal dispersion as ($\beta_2 > 0$).

The group velocity v_g is given by

$$v_g = \frac{1}{\beta_1} = \frac{c}{n_g}, \tag{2.5}$$

where the group index n_g is

$$n_g = n + \omega \frac{dn}{d\omega}. \tag{2.6}$$

The GVD parameter is

$$\beta_2 = \frac{d^2\beta}{d\omega^2} = \frac{1}{c}\left(2\frac{dn}{d\omega} + \omega\frac{d^2n}{d\omega^2}\right). \tag{2.7}$$

Dispersion Parameter

In optical fiber communications, dispersion is often described using the dispersion parameter D, which is related to β_2 by

$$D = -\frac{2\pi c}{\lambda^2}\beta_2 - \frac{\lambda}{c}\frac{d^2n}{d\lambda^2}. \tag{2.8}$$

Dispersion Compensation Techniques

To counteract dispersion effects, several compensation methods can be used such as,

- Dispersion - compensating fibers (DCFs) which are special fibers with high negative dispersion values to cancel standard fiber dispersion.
- Dispersion-shifted fibers (DSFs) which are designed to have zero dispersion at 1.55 m for minimal chromatic dispersion.
- EDC and Machine learning-based compensation where NNs dynamically adjust for dispersion distortions.

Chromatic dispersion is a key limitation in optical fiber communications, causing pulse broadening and inter-symbol interference. The group velocity dispersion β_2 and dispersion parameter D quantify dispersion effects, while compensation techniques like DCFs, DSFs, and digital equalization help mitigate its impact.

2.2 Laser Phase Noise

Phase noise arises due to random fluctuations in the phase of a signal, typically caused by imperfections in oscillators (such as laser sources in optical communication). These fluctuations degrade system performance by introducing phase errors in the received signal.

2.2.1 Phase Noise Model

Phase noise is modeled using random walk [10], a Wiener process, which is a discrete-time stochastic process that describes random motion. In the context of phase noise, the Wiener process represents the accumulation of random phase fluctuations over time. The model describe a process where the value at each step is the sum of random increments.

The Wiener process $W(t)$ is characterized by

$$W(t) - W(s) \sim \mathcal{N}(0, t-s) \tag{2.9}$$

for $t > s$, where $\mathcal{N}(0, t-s)$ is a normal distribution with mean 0 and variance $t-s$.

In discrete time, the phase noise $\phi(t)$ is modeled as a random walk

$$\phi(t) = \sum_{i=1}^{t} \Delta\phi_i \tag{2.10}$$

where $\Delta\phi_i$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance σ^2.

In optical communication systems, phase noise is applied to the transmitted signal by rotating its phase randomly. The resulting signal exhibits random phase fluctuations, which can distort the constellation diagram and decrease the Achievable Information Rate (AIR) in coherent detection systems.

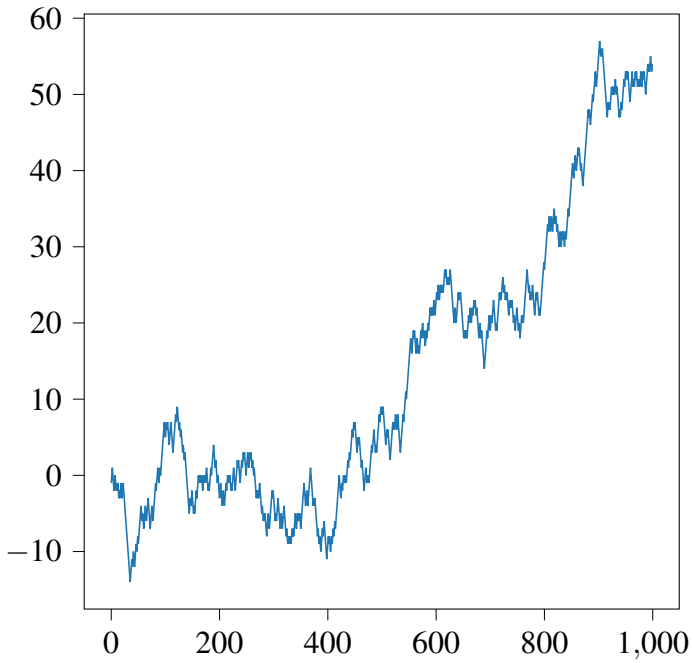


Figure 2.3: Random Walk

Figure 2.3 shows an example of a random walk process. A non-deterministic random generator is used to generate the noise bins, which are then accumulated. This gives a random walk-like noise process. The phase noise is applied to the signal by multiplying it with an exponential function of the complex noise.

Linewidth

Linewidth Δf in oscillators, such as local oscillators, is the full-width at half-maximum (FWHM) of the power spectral density of the oscillator output. It characterizes how stable the phase of an oscillator is over time. A narrow linewidth means that the oscillator produces a highly stable frequency, whereas a broad linewidth implies greater phase noise fluctuations [11].

The strength of phase noise is determined by the laser linewidth Δf, which characterizes the spectral purity of the laser source. The variance of the Gaussian increments $\Delta\phi_i$ is proportional to the laser linewidth and the time step Δt

$$\sigma^2 = 2\pi\Delta f\,\Delta t \tag{2.11}$$

This relationship captures how the laser linewidth influences the phase noise.

The impact of EEPN on the system performance can be described in terms of optical signal to noise ratio (OSNR) penalty. For a given OSNR penalty, the required linewidth decreases with an increase in fiber length, baud rate or required OSNR for a given constellation. This restricts the scalability of systems employing EDC [12].

The phase noise is applied to the transmitted signal $y(t)$ by multiplying it with a complex exponential term

$$y_{\text{noisy}}(t) = y(t) \cdot e^{j\phi(t)} \tag{2.12}$$

where $\phi(t)$ is the accumulated phase noise.

2.3 EEPN

Equalization-Enhanced Phase Noise (EEPN) is a phenomenon in coherent optical communication systems where the interplay between laser phase noise and electronic dispersion compensation (EDC) leads to enhanced phase noise in the received signal. This issue arises because EDC can interact with the phase noise of the transmitter laser or the local oscillator, creating an additional noise component that degrades system performance.

It is particularly significant in systems with high symbol rates and large amounts of chromatic dispersion. It manifests as phase and amplitude distortions in the received signal. The extent of EEPN depends on the system design, including the digital signal processing (DSP) configurations and the laser linewidths.

It becomes particularly significant in systems where tight laser linewidths and advanced compensation techniques are required. Wider linewidths exacerbate EEPN due to increased phase fluctuations.

It is directly proportional to the dispersion in the fiber link. Longer transmission distances or higher dispersion values increase the impact of EEPN. During transmission over optical fiber, the signal experiences chromatic dispersion, which spreads the optical pulses temporally.

After dispersion compensation (often performed electronically), the phase noise interacts with the dispersion compensation process.

EEPN causes phase distortion in the received signal constellation, leading to degraded symbol detection. It also lowers the system's AIR by introducing additional noise, reducing the effective capacity of the channel. For coherent optical systems operating at high data rates, EEPN becomes a limiting factor for system scalability and performance.

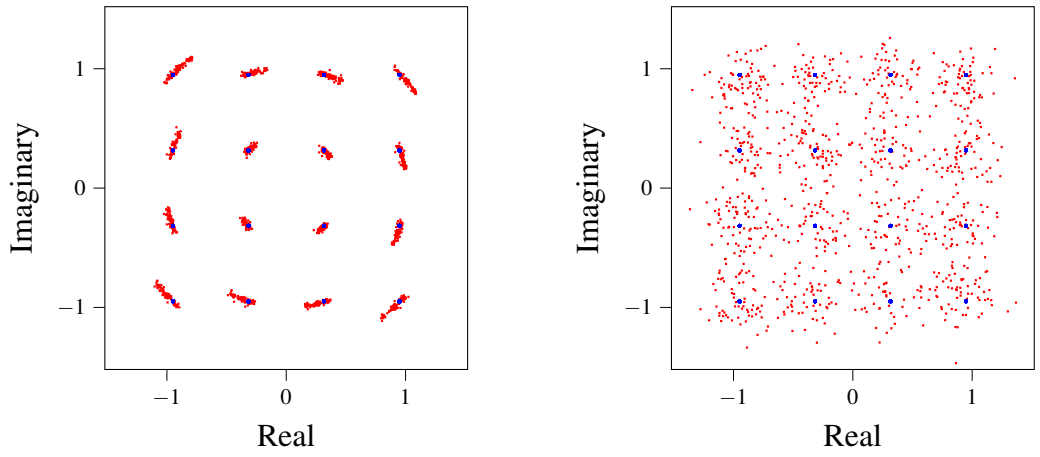


Figure 2.4: EEPN Effect

Figure 2.4 shows the effect of EEPN in system without and with gaussian noise. In the system without gaussian noise, a distinct ”blurring” or ”spinning” effect is visible, which increases with linewidth and chromatic dispersion. The spread is often elliptical or skewed due to the interaction of dispersion and phase noise.

Conventional strategies to mitigate EEPN may include optimizing DSP algorithms, using lasers with narrower linewidths, and reducing the low-frequency noise of the local oscillator.

The received signal $R(f)$ after EDC is given by

$$R(f) = \int_{-\infty}^{\infty} X(f - f_1) \cdot e^{jk(f_1^2 - 2ff_1)} \cdot X_{LO}(f_1)\, df_1, \tag{2.13}$$

The time-domain expression for the received signal $r'(t)$ after EDC is

$$r(t) = \int_{-\infty}^{\infty} X_{LO}(f_1) e^{-jkf_1^2} \left[x(t - kf_1/\pi) e^{j2\pi f_1 t} \right] df_1 \tag{2.14}$$

The variance of the received signal is given by

$$\sigma_{r'}^2(t) = E\left[|c_n|^2\right] \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left| h_{ps}\left(t - nT_s - \frac{kf}{\pi}\right) \right|^2 S(f)\, df \tag{2.15}$$

The total EEPN, denoted as δ, depends on the linewidth Δv of the local oscillator (LO) and is given by

$$\delta = \pi^2 |\beta_2| L f_s \Delta v \tag{2.16}$$

2.3.1 Modulus-Dependence of EEPN-Induced Phase Noise

EEPN is not uniformly distributed across constellation points, instead, it depends on the modulus (amplitude) of each point. For higher-order QAM, which has multiple modulus rings, EEPN is stronger in the outer rings due to its higher power levels [13].

For a given constellation point C_i, the total EEPN variance is

$$D_E(C_i) = \frac{\delta}{3}|C_i|^2 + \frac{2\delta}{3}E[|C|^2] \tag{2.17}$$

The EEPN-induced phase noise variance for each constellation point is

$$D(r\varphi, C_i) = \frac{\delta}{3}|C_i|^2 + \frac{\delta}{3}E[|C|^2] \tag{2.18}$$

The amplitude noise variance is

$$D(r, C) = \frac{\delta}{3}E[|C|^2] \tag{2.19}$$

The ovality $O(C_i)$, which measures the dominance of phase noise over amplitude noise, is given by

$$O(C_i) = \frac{D(r\varphi, C_i)}{D(r, C)} = \frac{|C_i|^2 + E[|C|^2]}{E[|C|^2]} \tag{2.20}$$

For higher-order QAM, ovality varies across different constellation rings.

2.4 Deep Learning and Neural Networks

Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts [14]. NNs are a fundamental concept in machine learning, inspired by the structure and function of the human brain. They consist of layers of interconnected neurons (nodes) that process data and learn patterns from input to output. Neural networks are widely used in image recognition, natural language processing, medical diagnosis, and autonomous systems.

2.4.1 Structure of a Neural Network

A basic artificial neural network (ANN) consists of:

1. Input Layer: Receives raw data (features).
2. Hidden Layers: Perform computations and transformations using activation functions.
3. Output Layer: Produces the final prediction or classification result.

Each neuron in a layer is connected to neurons in the next layer through weighted connections. The network learns by adjusting these weights during training.

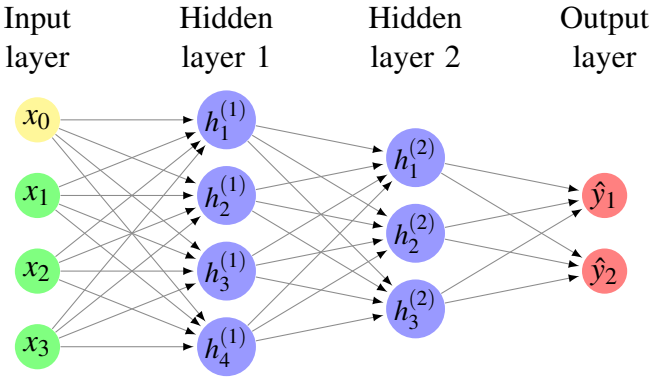


Figure 2.5: Basic Neural Network

Figure 2.5 shows a neural network with one input layer containing 4 neurons, two hidden layers containing four and three neurons respectively and an output layer containing two neurons.

2.4.2 Forward Propagation

Neural networks process data through forward propagation [15, 16], where information moves from input to output. Each neuron applies a weighted sum of its inputs followed by an activation function. The output of a neuron j in layer $l+1$ is

$$z_j^{(l+1)} = \sum_{i=1}^{n_l} w_{ji}^{(l)} a_i^{(l)} + b_j^{(l)}. \tag{2.21}$$

2.4.3 Activation Functions

Training also requires a loss function to measure the error and an optimization algorithm, such as stochastic gradient descent, to adjust the weights. Nonlinear activation functions like Rectified Linear Unit (ReLU) [17] or sigmoid [18] introduce complexity, enabling neural networks to learn intricate patterns in data. The activation function f is applied to get the neuron's output

$$a_j^{(l+1)} = f(z_j^{(l+1)}) \tag{2.22}$$

Some common activation functions are:

Sigmoid

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.23}$$

ReLU

$$f(x) = \max(0,x) \tag{2.24}$$

2.4.4 Backpropagation

Backpropagation [19] is a gradient estimation method commonly used for training a neural network to compute its parameter updates. Neural networks learn by minimizing an error function (loss) using gradient descent.

The loss function for classification is the cross-entropy loss

$$L = -\sum_i y_i \log \hat{y}_i \tag{2.25}$$

For regression tasks, Mean Squared Error (MSE) is used

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.26}$$

Optimization Algorithms

Neural networks are trained using optimization algorithms, such as stochastic gradient descent (SGD), that minimize a loss function. The process involves forward propagation of input data through the network to compute predictions, followed by backward propagation to update the weights based on the error.

SGD

$$w^{(t+1)} = w^{(t)} - \eta \nabla \mathscr{L}(w^{(t)}; x_i, y_i) \tag{2.27}$$

Regularization techniques, like dropout or L2 regularization, are often used to prevent overfitting and improve generalization.

2.4.5 Deep Neural Networks (DNNs)

The primary strength of neural networks lies in their ability to generalize from data, making them powerful tools for solving problems in classification, regression, and function approximation. A Deep Neural Network (DNN) has multiple hidden layers, making it capable of learning hierarchical features. Common architectures include feedforward neural networks (FNNs), convolutional neural networks (CNNs) [20] for image processing, and recurrent neural networks (RNNs) [21] for sequential data. Deep learning, a subset of machine learning, refers to neural networks with many layers, which are particularly effective for complex tasks.

Universal Approximation Theorem

A feedforward neural network with at least one hidden layer can approximate any continuous function

$$f(x) \approx \sum_i \alpha_i \sigma(w_i^T x + b_i) \tag{2.28}$$

where σ is a nonlinear activation function (like ReLU or Sigmoid).

2.4.6 Convolutional Neural Networks (CNNs)

CNNs are a class of deep learning models designed for processing grid-like data structures, such as images and time-series data. CNNs are particularly effective in tasks such as image classification, object detection, speech recognition, and medical imaging, due to their ability to capture spatial and hierarchical features.

Convolution Operation

The core operation in CNNs is the convolution, which applies a filter (or kernel) to the input data to extract features. The convolution operation is defined as

$$Z(i,j) = \sum_m \sum_n X(i-m, j-n)K(m,n) \tag{2.29}$$

Stride and Padding

In a convolutional neural network, the output size of a convolutional layer depends on the input size, kernel size, stride, and padding.

Stride controls how much the filter moves at each step. Larger strides reduce the output size.

Padding adds extra zeros around the input to control the output size and preserve spatial dimensions.

For an input of size $H \times W$, kernel size K, stride s, and padding p, the output size is

$$H_{\text{out}} = \frac{H-K+2p}{s}+1, \quad W_{\text{out}} = \frac{W-K+2p}{s}+1. \tag{2.30}$$

Pooling Layers: Reducing Dimensionality

Pooling helps in reducing computation, improving robustness to small translations, and preventing overfitting.

Max pooling takes the maximum value in a region

$$Z(i,j) = \max_{m,n} X(i+m, j+n) \tag{2.31}$$

Average pooling Computes the mean value

$$Z(i,j) = \frac{1}{K^2} \sum_{m,n} X(i+m, j+n) \tag{2.32}$$

Fully Connected Layers

After convolution and pooling, the high-level features are flattened and passed through fully connected (dense) layers

$$Z = WX + b \tag{2.33}$$

Fully connected layers enable classification or regression tasks based on learned features.

Backpropagation and Training in CNNs

CNNs are trained using gradient descent and backpropagation. The loss function measures the difference between predicted and true labels.

The gradients are computed using the chain rule and weights are updated using Stochastic Gradient Descent (SGD).

Regularization

Regularization techniques prevent overfitting by adding constraints to the model. Common techniques include:

L2 Regularization

$$L_{\text{reg}} = L + \lambda \sum_i \theta_i^2 \tag{2.34}$$

Dropout randomly drops units during training with probability p.

Batch Normalization

Batch normalization normalizes the activations of each layer to have zero mean and unit variance, stabilizing training and accelerating convergence.

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} y = \gamma \hat{x} + \beta \tag{2.35}$$

Residual Connections

Residual connections add the input of a block to its output, enabling the network to learn residual mappings.

$$y = F(x, \{W_i\}) + x \tag{2.36}$$

2.4.7 Performance Metrics

Mutual Information (MI) [22] quantifies the shared information between transmitted and received signals. It is defined as

$$I(X;Y) = H(X) - H(X|Y) \tag{2.37}$$

Using probability distributions, MI is

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)} \tag{2.38}$$

MI represents the maximum achievable data rate per symbol in ideal conditions, assuming an optimal decoder [22].

AIR is the maximum data rate that can be reliably transmitted over a real channel, accounting for practical decoding limitations [23],

$$\text{AIR} = I(X;Y) - \Delta_{\text{penalty}} \tag{2.39}$$

For a Gaussian channel

$$\text{AIR} = \log_2(1 + \text{SNR}) \tag{2.40}$$

where:

$$\text{SNR} = \frac{P}{N_0} \tag{2.41}$$

While MI represents the theoretical maximum data rate, assuming optimal modulation and decoding, AIR is a practical measure, which is always lower than MI due to imperfections in the receiver. AIR is computed using MI but adjusted for coding and system constraints. For Gaussian channels, MI and AIR converge when using optimal coding schemes.

2.4.8 Neural Network Equalization

In communication systems, equalization is essential for mitigating the effects of channel impairments such as inter-symbol interference (ISI), noise, and phase noise. Conventional equalization methods, like linear equalizers or decision feedback equalizers (DFEs), are often limited in their ability to adapt to non-linear or time-varying channel conditions. Neural networks, with their powerful ability to model non-linear relationships, have emerged as promising tools for equalization. [24]

One of the primary advantages of neural networks in this context is their ability to learn complex channel characteristics directly from data without relying on explicit channel models. This makes them especially effective for non-linear equalization, where traditional linear methods fail. Neural networks are also highly adaptable and can be trained online to respond to dynamic changes in channel conditions, making them suitable for real-time applications. Furthermore, in the context of equalization enhanced phase noise (EEPN), neural networks can effectively mitigate both amplitude and phase distortions, improving overall system performance.

To implement neural networks for equalization, the process typically begins with data preprocessing, where transmitted and received symbols are paired to create labeled training datasets. The choice of network architecture depends on the channel characteristics; for example, feedforward networks are often used for static channels, while recurrent or convolutional networks are preferred for dynamic or frequency-selective channels. The training process involves minimizing a loss function, such as MSE or maximizing metrics like AIR.

3 System Setup

3.1 System Overview

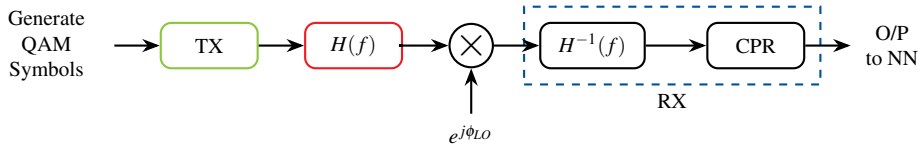


Figure 3.1: Block Diagram of Optical Communication System

The system represents a typical digital communication chain for optical transmission, consisting of three main components: a transmitter, a channel, and a receiver. The system is designed to simulate the transmission of 16-QAM symbols over an optical fiber channel, incorporating various impairments such as dispersion, additive noise, and phase noise. Additionally, the system includes digital signal processing (DSP) blocks to mitigate these impairments and recover the transmitted symbols.

It is a communication technique in which a signal is transmitted as light, with optical fiber serving as the medium for moving those light signals from one location to another. A signal transmitted through an optical fiber is transformed from an electrical signal into light, which is then converted back into an electrical signal at the receiving end. Optical fiber communication has been applied for a variety of telecommunication needs because it performs well in long-distance and high-speed data transfer.

3.1.1 Transmitter Model

Random symbols are generated from a 16-QAM constellation, representing the transmitted data. Each symbol is a complex value, where the real and imaginary parts correspond to in-phase (I) and quadrature (Q) components.

The QAM symbols are upsampled to increase the sampling rate, preparing the signal for pulse shaping. This step inserts zeros between symbol samples to increase the signal bandwidth and ensures that the signal has sufficient samples to represent the continuous waveform accurately.

A root-raised cosine (RRC) filter is applied to shape the transmitted pulses. Pulse shaping is essential to limit the bandwidth of the signal and minimize inter-symbol interference (ISI) at the receiver.

The signal is padded to ensure its length is compatible with subsequent processing steps, such as the fast Fourier transform (FFT) used in the channel model. It is used to avoid edge effects during filtering and match system constraints.

3.1.2 Channel Model

The signal is transmitted through a channel that introduces phase noise. The optical channel in the system is modeled as a dispersive medium, incorporating chromatic dispersion and potential phase noise, using the random walk model.

The signal is propagated through the optical fiber using the Split-Step Fourier Method (SSFM), which models the effects of chromatic dispersion. Dispersion causes the signal to spread out in time, leading to ISI.

Complex Gaussian noise is added to the signal to simulate the effects of thermal noise and other impairments in the optical system. It simulates the combined effects of optical amplifier noise and receiver noise, reducing the overall (SNR).

Additionally, phase noise is introduced to model laser imperfections, which cause random fluctuations in the signal's phase, leading to potential errors in demodulation.

A second SSFM block is used to compensate for dispersion by applying an inverse dispersion profile. This is possible because chromatic dispersion is a linear effect. This step helps to mitigate the effects of dispersion and recover the original signal.

3.1.3 Receiver Model

The receiver processes the received signal to recover the transmitted QAM symbols. It includes several DSP blocks to mitigate impairments and synchronize the signal.

A matched filter, implemented using the same RRC filter as in the transmitter, is applied to maximize the signa SNR and reduce ISI.

The signal is downsampled to reduce the sampling rate back to the symbol rate, simplifying further processing.

A timing recovery block with Gardner algorithm [25] is used to correct any timing errors in the received signal. Although no explicit timing error is introduced in the system, real-world hardware imperfections, such as clock drift and jitter, create small misalignments between the transmitted and received symbols. Even in simulation, phase noise and dispersion distortions introduce slight variations that affect symbol timing. Timing recovery aligns the sampling instants with the symbol boundaries, ensuring that each symbol is detected accurately. This is

particularly important in systems with pulse shaping, where the optimal sampling instants are not always obvious.

Phase noise, caused by laser linewidth and other factors, is compensated using a phase noise estimation and correction block. This step is crucial for coherent detection systems, where phase noise can significantly degrade performance.

The carrier phase recovery (CPR) [26] block estimates and corrects any residual phase errors in the received signal. This step ensures that the constellation diagram is properly aligned, enabling accurate symbol detection. This step refines the phase estimation by comparing received symbols to a reference constellation and adjusting for any residual phase offsets.

Finally, the recovered symbols are normalized to ensure they have the correct power level, simplifying further processing and analysis.

3.2 Simulation Setup

The parameters used in this simulation define the characteristics of the optical communication system, including signal shaping, fiber impairments, and receiver processing. Each parameter plays a crucial role in ensuring realistic modeling of the transmission and reception process.

The setup models a high-speed optical communication system with a 16-QAM signal of 1 million symbols that undergoes chromatic dispersion, phase noise, and attenuation over a 5000 km fiber link. The RRC pulse shaping filter ensures spectral control, while oversampling at 10 samples per symbol allows for precise DSP operations. Given the 200 kHz laser linewidth, phase noise will be a major impairment, requiring carrier phase recovery at the receiver. Additionally, timing recovery is necessary due to potential distortions introduced by dispersion and phase noise.

3.2.1 Pulse Shaping and Sampling Parameters

The roll-off factor determines the excess bandwidth in the RRC filter used for pulse shaping. A lower β (close to 0) results in a sharper filter with minimal excess bandwidth, reducing ISI. However, this also makes the signal more susceptible to timing jitter, requiring precise timing recovery at the receiver. A smaller roll-off factor reduces the occupied bandwidth but may increase sensitivity to timing errors. A larger roll-off factor increases robustness to timing errors but uses more bandwidth.

Filter span in symbols defines the length of the RRC filter window in terms of symbols. A larger span improves frequency domain localization, reducing ISI at the cost of increased computational complexity.

Each symbol is oversampled by a factor of 10 to allow for accurate pulse shaping and digital signal processing (DSP). This ensures smooth transitions between symbols and aids in

timing recovery, as interpolated samples provide finer granularity for symbol detection. A higher number of samples per symbol improves the accuracy of the simulation but increases computational load.

3.2.2 Fiber Optic Channel Parameters

Chromatic dispersion coefficient (in $\frac{ps^2}{km}$) causes different frequency components to travel at different speeds through the fiber. The negative value indicates anomalous dispersion, meaning higher frequencies travel faster than lower frequencies. This effect leads to pulse broadening and ISI, which must be compensated at the receiver.

Fiber Attenuation represents fiber loss, typically measured in $\frac{dB}{km}$. Optical signals attenuate as they propagate through the fiber due to absorption and scattering, requiring optical amplification in real-world systems. However, in this simulation, no amplification is used, meaning the signal strength will naturally decay over the fiber length. Higher attenuation increases the need for amplification or signal regeneration over long distances.

The total fiber length makes chromatic dispersion a significant impairment. Long-distance transmission also means the effects of phase noise, fiber non-linearity, and attenuation become critical in determining system performance. Longer fiber lengths increase the effects of dispersion and attenuation, requiring more robust compensation techniques.

Carrier frequency is the central optical frequency of the transmitted signal, neasured in Hz. It corresponds to a wavelength of approximately 1550 nm, which is commonly used in optical communication systems due to low fiber attenuation. The choice of carrier frequency affects the system's compatibility with optical components and the fiber's attenuation characteristics.

3.2.3 Time and Distance Normalization Parameters

Time Normalization Factor scales time-related quantities to ensure numerical stability in simulations. In optical systems, working with raw time values (in the order of femtoseconds or picoseconds) can lead to computational issues, so a normalization factor is applied. A smaller value increases the precision of time-related computations but may require finer discretization.

Distance Normalization Factor is used for normalizing distances in the simulation. It ensures that fiber propagation equations remain numerically stable by rescaling the values to a computationally manageable range.

3.2.4 Laser and Receiver Parameters

The laser linewidth represents the spectral width of the transmitter's optical source. A greater linewidth introduces phase noise, which can degrade signal quality, especially in higher-order modulation formats (16-QAM). To counteract this impairment, the receiver applies CPR.

3.2.5 Modulation and Transmission Parameters

Each transmitted symbol carries 4 bits (ie. Modulation Order of 4), indicating that the system uses 16-QAM. Higher modulation orders increase spectral efficiency but also make the signal more sensitive to noise, phase errors, and nonlinearities.

The batch size defines how many independent signal sequences are processed in parallel. A batch size of 1 suggests a single independent transmission, rather than a mini-batch approach commonly used in deep learning. Larger batch sizes improve computational efficiency but require more memory.

Number of symbols transmitted in each simulation run defines the total number of symbols transmitted per batch. A larger value means longer transmission sequences, allowing for a more realistic simulation of optical channel impairments over extended distances which improves the statistical reliability of the simulation results but increase computation time.

3.3 Neural Network Architecture

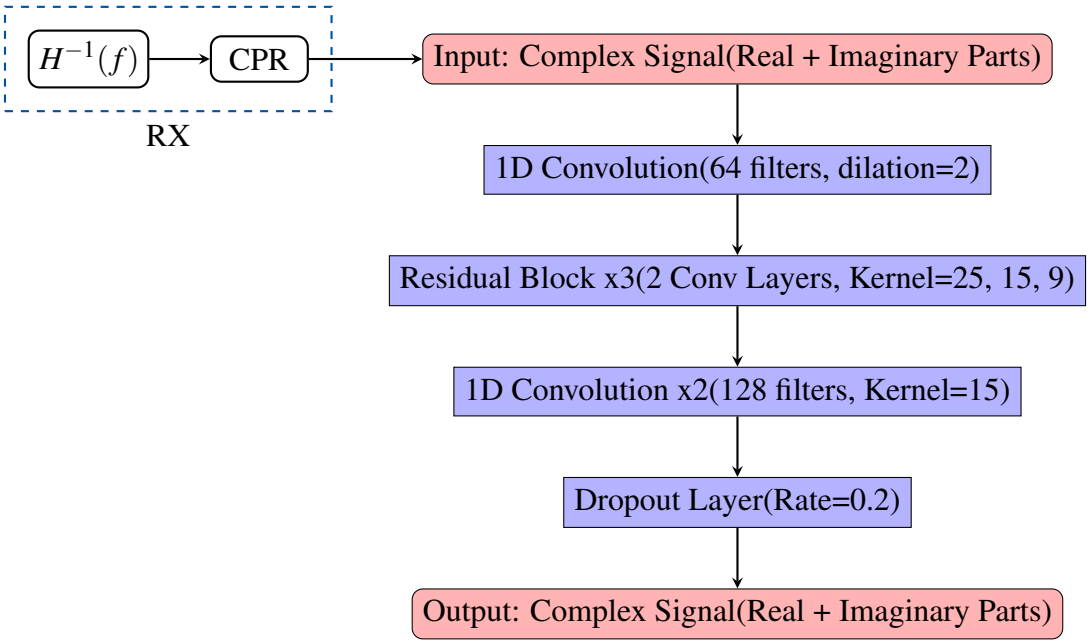


Figure 3.2: Flowchart of Neural Network Architecture

Figure 3.2 shows a custom-built CNN architecture that combines residual learning, dilated convolutions, and dropout for efficient and robust noise mitigation.

The neural network architecture is a deep convolutional model designed to process complex-valued signals. The architecture consists of convolutional layers, residual blocks, batch normalization, dropout, and an output layer, all structured to handle real and imaginary components of complex data separately.

The network is implemented using TensorFlow and Keras, and it includes several key components to enhance its performance and robustness.

Graph Mode Optimization is used to optimize the network for execution in TensorFlow's graph mode, which improves performance by compiling the model into a static computation graph, reducing overhead and enabling faster execution.

3.3.1 Input Representation and Transformation

The network is fed with output from CPR for training with 1 million complex-valued symbols. Since the symbols are random and have no sequence or limit, each trial of training and testing utilizes freshly transmitted symbols. The network is structured to process complex-valued

inputs, which are separated into their real and imaginary components. This is necessary because conventional deep learning frameworks, such as TensorFlow, operate on real-valued tensors. By treating the real and imaginary parts as two separate feature channels, the network maintains the structure of the input while allowing for effective learning of amplitude and phase variations. This stage converts the complex-valued signal into a format compatible with real-valued convolutional layers. It also preserves the phase and amplitude information of the signal, which is critical for tasks such as phase recovery or channel equalization.

3.3.2 Initial Convolutional Layer: Feature Extraction and Dilation

The first convolutional layer applies one-dimensional convolution with a kernel size of 50 and a dilation rate of 2. The significance of these values lies in their ability to capture information over a wider receptive field while maintaining computational efficiency. A larger kernel size enables the network to capture long-range dependencies, which is crucial in handling dispersed signals in communication systems. The dilation rate increases the spacing between the kernel elements, exponentially expanding the receptive field without increasing the number of parameters.

This layer also applies ReLU activation, which introduces non-linearity to model complex relationships in the data. The inclusion of L2 regularization penalizes large weight values, ensuring that the network remains generalizable and prevents overfitting.

3.3.3 Residual Blocks: Hierarchical Feature Learning

The model employs three stacked residual blocks, each consisting of two convolutional layers, batch normalization, and a residual connection. Residual connections are fundamental in deep architectures as they address the vanishing gradient problem by allowing gradients to flow through the network during backpropagation. Residual learning allows the network to learn residual mappings, which are easier to optimize than direct mappings.

The kernel size decreases across blocks enables the network to capture fine-grained features in later layers. Each residual block consists of convolutional layers with progressively smaller kernel sizes:

- Kernels of length 25 in the first residual block capture broad frequency components and coarse structure.
- Kernels of length 15 in the second residual block refine intermediate-level features.
- Kernels of length 9 in the third residual block extract fine-grained details, helping to resolve subtle distortions in the signal.

Batch normalization is applied after each convolutional layer to stabilize activations and ensure faster convergence during training. This normalization mitigates internal covariate shift, making the network less sensitive to initialization and improving generalization.

3.3.4 Additional Convolutional Layers: Feature Expansion

Beyond residual learning, additional convolutional layers are employed with 128 filters and a kernel size of 15. This expansion phase increases the network's representational capacity, allowing it to model more complex transformations. By doubling the number of filters, the network learns a richer set of features, enhancing its ability to distinguish between noise and true signal components.

3.3.5 Dropout for Regularization

To prevent overfitting, the network integrates dropout with a rate of 0.2, which results in 20% of neurons being randomly deactivated during training. Dropout forces the network to learn redundant representations, ensuring that no single neuron becomes too dominant. This is particularly beneficial in high-dimensional signal processing tasks, where overfitting to specific frequency components or distortions can degrade real-world performance.

3.3.6 Output Layer

The final layer of the network consists of two filters, each corresponding to one of the real and imaginary components of the output. The absence of an activation function in this layer ensures that the network outputs a linear transformation of the learned features, preserving the continuous nature of the signal. The real and imaginary parts are then recombined into a complex-valued output, reconstructing the processed signal for further downstream tasks.

4 Results

This section presents a detailed investigation into the effectiveness and limitations of using DNNs for mitigating EEPN in optical communication systems. The experiments are aimed to evaluate the performance of the NN-based equalizer in compensating for EEPN and other impairments. However, the results revealed that the NN-based equalizer introduced a phenomenon known as the ‘jail window‘, which is characterized by the clustering of received symbols in specific regions of the complex plane. This section provides a detailed analysis of the results, including the ‘jail window‘ phenomenon, and discusses the implications for EEPN mitigation.

4.1 MSE Performance

The SNR vs Linewidth plot illustrates the SNR in dB versus Linewidth in Hz for different link distances. Outputs are shown as line charts with NN-based equalizer out and without NN-based equalizer, ie. outut from CPR block in the Receiver.

Each condition is measured across different link distances of 1000 km, 2000 km, 4000 km, and 5000 km and several linewidths of 100kHz, 200kHz, 300kHz, 400kHz, 500kHz, 750kHz and 1000kHz to capture performance of equalization over a variety of conditions.

In optical fiber systems, increasing signal power beyond a certain point leads to nonlinear impairments, which do not scale favorably with power. This creates a nonlinear Shannon limit, beyond which increasing power does not increase capacity and may even degrade it. This limit is typically in the range of 13–14 dB for modern fiber-optic systems.

A recent study published in Journal of Lightwave Technology (Guo et al., 2024) discusses the generalized signal-to-noise ratio (GSNR) and its impact on system capacity, highlighting that power optimization is crucial in maximizing transmission efficiency in fiber-optic networks. The study suggests that beyond a certain SNR, the gains in capacity diminish due to these nonlinear effects [27]. Therefore, a baseline of 14dB is set for SNR results.

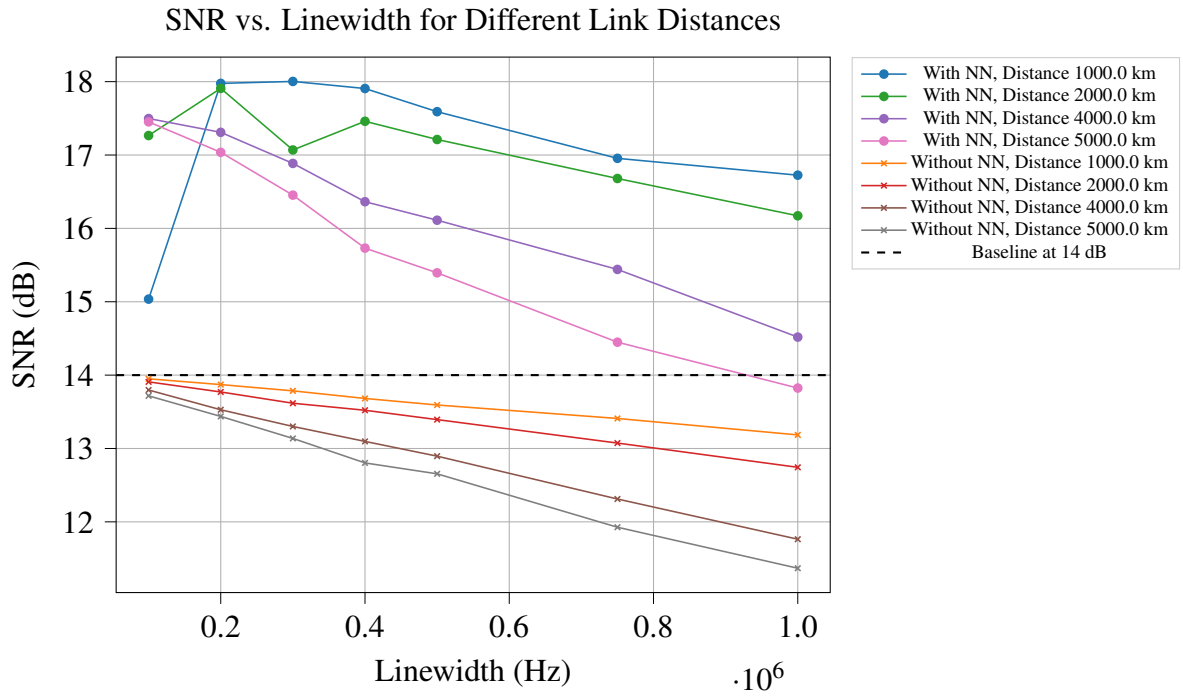


Figure 4.1: SNR vs Linewidth for Different Link Distances with MSE loss function

Figure 4.1 shows SNR vs Linewidth plot for various Link Distances when NN-based equalizer performance is measured on MSE loss function. It is evident from the plot that SNR for received symbols with NN is much higher than the baseline SNR of 14dB.

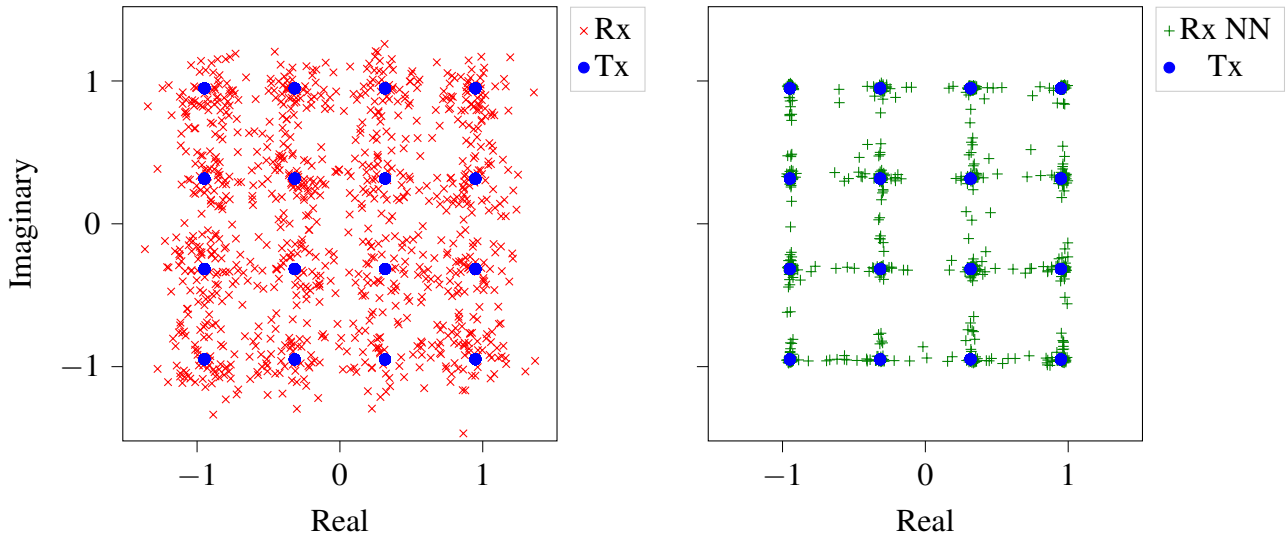


Figure 4.2: Scatter Plot for Tx vs Rx with and without NN for linewidth 750kHz and link distance 2000km with MSE Loss Function

4.1.1 Pitfall from MSE Loss Function: Jail Window

The ‘jail window‘ effect arises from a mismatch between the true transmission performance metric, such as the AIR, and the metric minimized by the NN, typically the MSE loss. This mismatch leads to a disconnect between the objective function optimized by the NN and the desired outcome of improving AIR. However, since AIR is non-differentiable, it cannot be directly used as a loss function for NN training [6].

In this scenario, the NN continues to minimize the Euclidean distance between predictions and labels, yet this does not result in a corresponding increase in AIR, diverging from the primary goal of equalization. The NN may instead exploit periodicities in the dataset, which deviates from its intended purpose as an equalizer.

Moreover, the ‘jail window‘ constellation patterns violate the Gaussian channel assumption typically used in the calculation of performance metrics. This violation reduces the accuracy of metrics based on Gaussian assumptions, such as the Effective Signal-to-Noise Ratio (ESNR), while metrics like the Q-factor derived directly from AIR remain accurate. Consequently, the use of Gaussian-based metrics can lead to misleading results, falsely suggesting that the NN performs well, when in reality, the benefits provided by the ‘jail window‘ constellation are significantly overestimated.

Notably, several studies [7, 28, 29] have reported ‘jail window‘ constellations following NN-based equalization. However, these works often fail to acknowledge the deeper implications and consequences of this phenomenon, highlighting the need for careful analysis of the metrics used to evaluate NN performance.

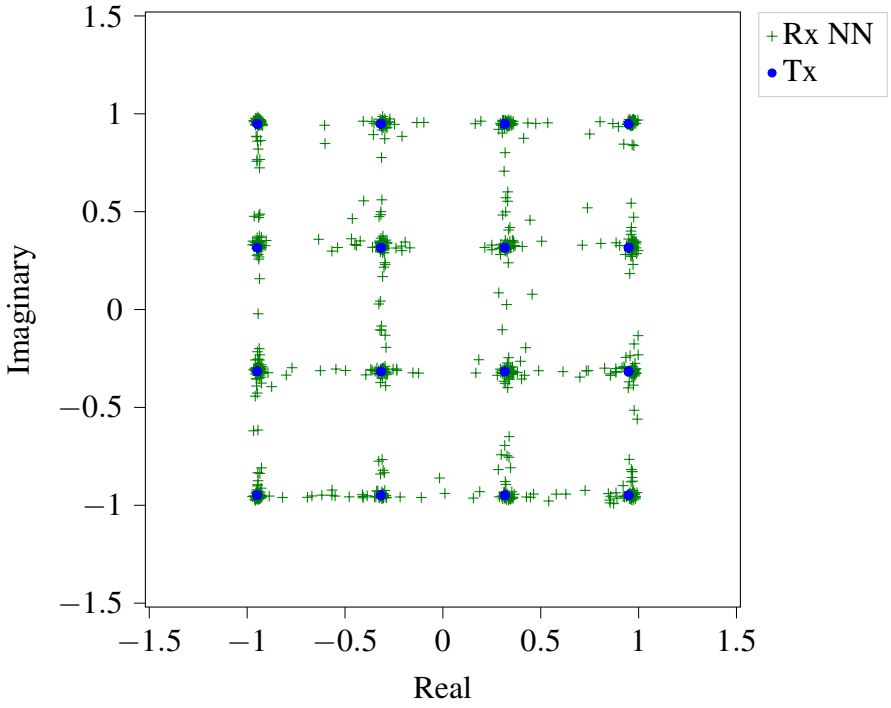


Figure 4.3: Jail Window Effect

4.2 Mitigation Techniques for Jail-Window Effect

4.2.1 Entropy-Regularized Mean Square Error (MSE-X)

MSE-X combines the traditional MSE loss with an entropy regularization term to handle uncertainty and noise in the signal. The entropy term helps improve generalization by encouraging the model to maintain a level of uncertainty in its predictions, which can be crucial for systems like phase noise compensation, where noise and impairments play a significant role [30].

The Mean Square Error (MSE) cost function is commonly used for nonlinear equalization, but it does not necessarily maximize the Achievable Information Rate (AIR). The Entropy-Regularized MSE (MSE-X) introduces an additional entropy term that improves AIR while preserving a high AIR.

The standard MSE cost function is defined as

$$\mathrm{MSE}(X, f(R)) = \mathbb{E}[|f(R) - X|^2] \tag{4.1}$$

While MSE ensures a low reconstruction error, it does not optimize for soft-decision Forward Error Correction (SD-FEC) because it does not explicitly account for probability distributions.

To improve performance, an entropy-based regularization term is introduced. The conditional probability of the transmitted symbol given the equalized output is

$$Q_{X|Y}(x|y) = \frac{P_X(x)Q_{Y|X}(y|x)}{Q_Y(y)} \tag{4.2}$$

Assuming a Gaussian channel model

$$Q_{Y|X}(y|x) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(y-x)^2}{2\sigma^2}\right) \tag{4.3}$$

The marginal distribution of Y is

$$Q_Y(y) = \sum_{x\in X} P_X(x)Q_{Y|X}(y|x). \tag{4.4}$$

To improve the achievable information rate,

$$\mathbb{E}[-\log Q_{X|Y}(X|Y)] \tag{4.5}$$

is minimized, which directly relates to maximizing mutual information.

By combining the MSE objective with the entropy regularization term, the following relation is obtained

$$\text{MSE-X}(X, f(R)) = \underbrace{\mathbb{E}[|f(R)-X|^2]}_{\text{MSE Term}} - 2\sigma^2 \underbrace{\mathbb{E}[-\log Q_Y(f(R))]}_{\text{Entropy Regularization Term}} \tag{4.6}$$

This function consists of:

- A standard MSE term that ensures signal reconstruction accuracy,
- A regularization term that preserves soft information by maximizing entropy.

When the noise variance σ^2 is small, MSE-X behaves similarly to standard MSE. However, as noise increases, entropy regularization plays a greater role, preventing excessive clustering of constellation points.

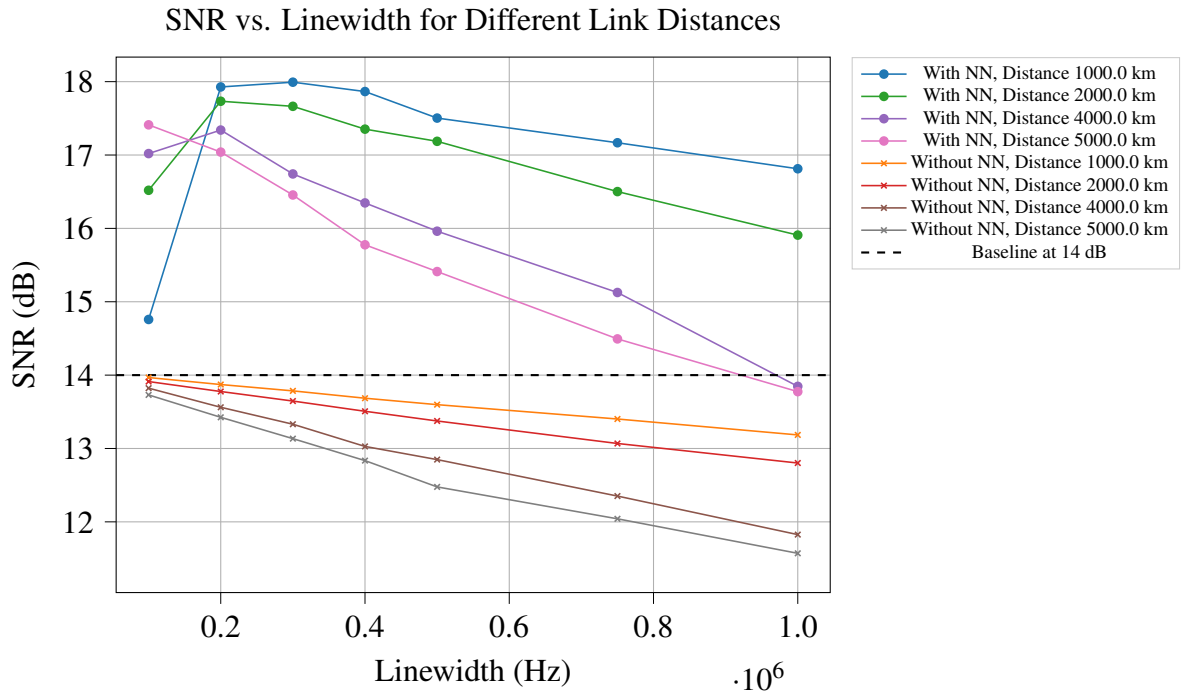


Figure 4.4: SNR vs Linewidth for Different Link Distances with MSE-X loss function

Figure 4.4 shows SNR vs Linewidth plot for various Link Distances when NN-based equalizer performance is measured on a novel MSE-X loss function. It is evident from the plot that SNR for received symbols with NN is much higher than the baseline SNR of 14dB.

4.2.2 AIR with Early Stopping : SNR Findings

AIR can be used as a criterion for early stopping, which is a technique to prevent overfitting and ensure that the model generalizes well to unseen data.

Early stopping is a regularization technique used to halt the training of a machine learning model before it starts to overfit the training data. The idea is to stop the training once the model reaches a point where it no longer improves on a validation set, thus preventing unnecessary overfitting to the noise in the data.

The challenge lies in determining when to stop training to avoid this overfitting while ensuring the model has learned enough from the data.

As training progresses, AIR can be used as a proxy for model performance in terms of the signal's ability to withstand noise and interference, especially in communication systems.

If the AIR plateaus or stops increasing over a set number of epochs, it indicates that the model has likely learned most of what it can, and further training might lead to overfitting. Although, it can also be caused due to hitting a local minima instead of global minima.

Ideally, If the model's AIR on a validation set starts to decrease or stabilize, it signals that the model is beginning to overfit the training data, and training should be stopped.

$$\text{Best Loss AIR} = \max(\text{AIR}_{\text{train}}, \text{Best Loss AIR}) \tag{4.7}$$

MSE Loss

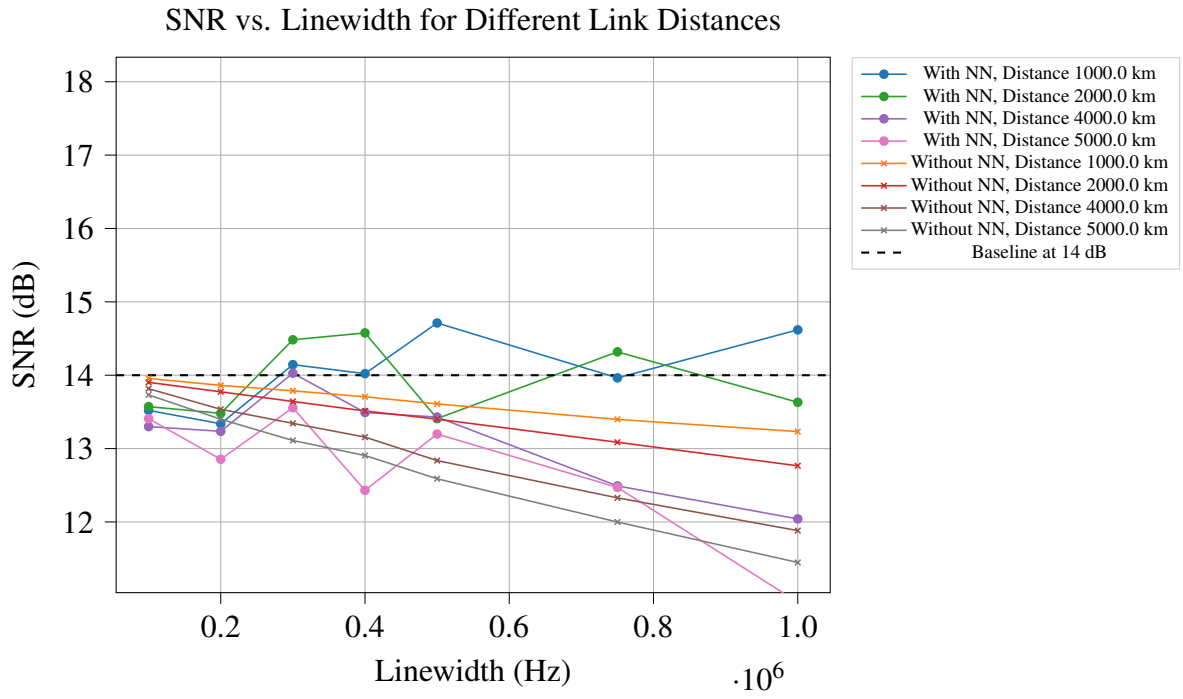


Figure 4.5: SNR vs Linewidth for Different Link Distances with MSE Loss Function

Figure 4.5 shows SNR performance for NN-based equalizer output with MSE loss function and a patience of 100 resulting in early stopping at around 300 epochs. The low SNR is evidence of insuffient training of the NN-based algorithm and thus, a poor performance.

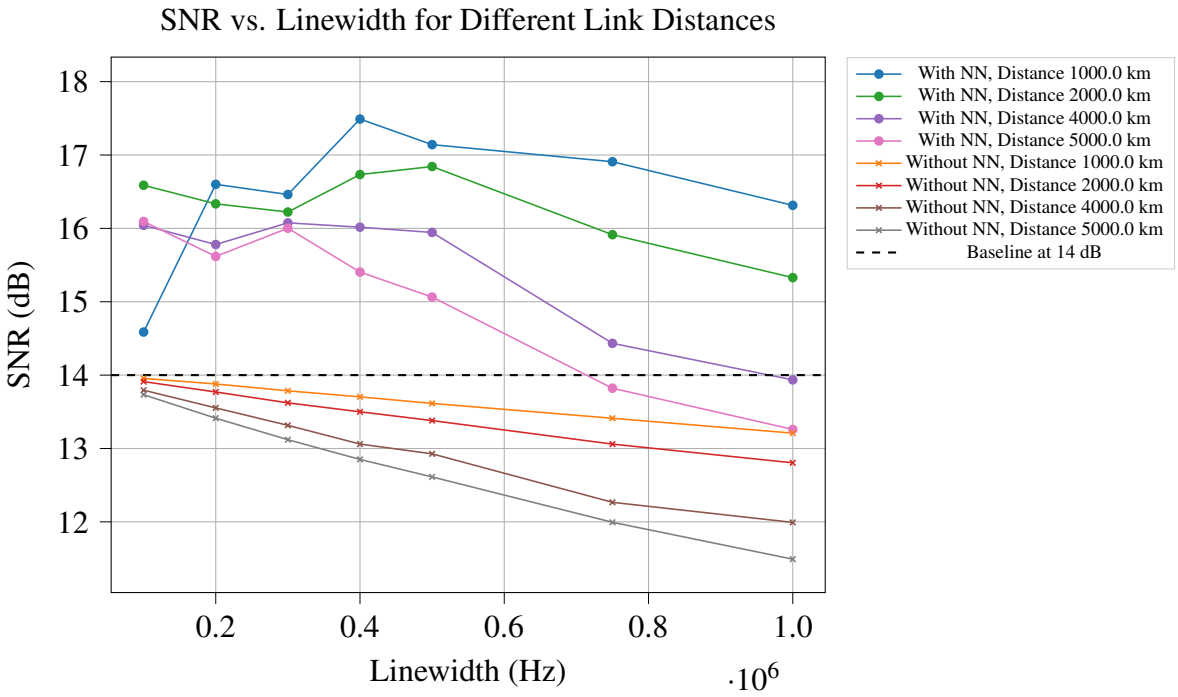


Figure 4.6: SNR vs Linewidth for Different Link Distances with MSE Loss Function

Figure 4.6 shows SNR performance for NN-based equalizer output with MSE loss function and a patience of 500 resulting in early stopping at around 1200 epochs. The SNR is almost the same as that without any early stopping but slightly lower.

MSE-X Loss

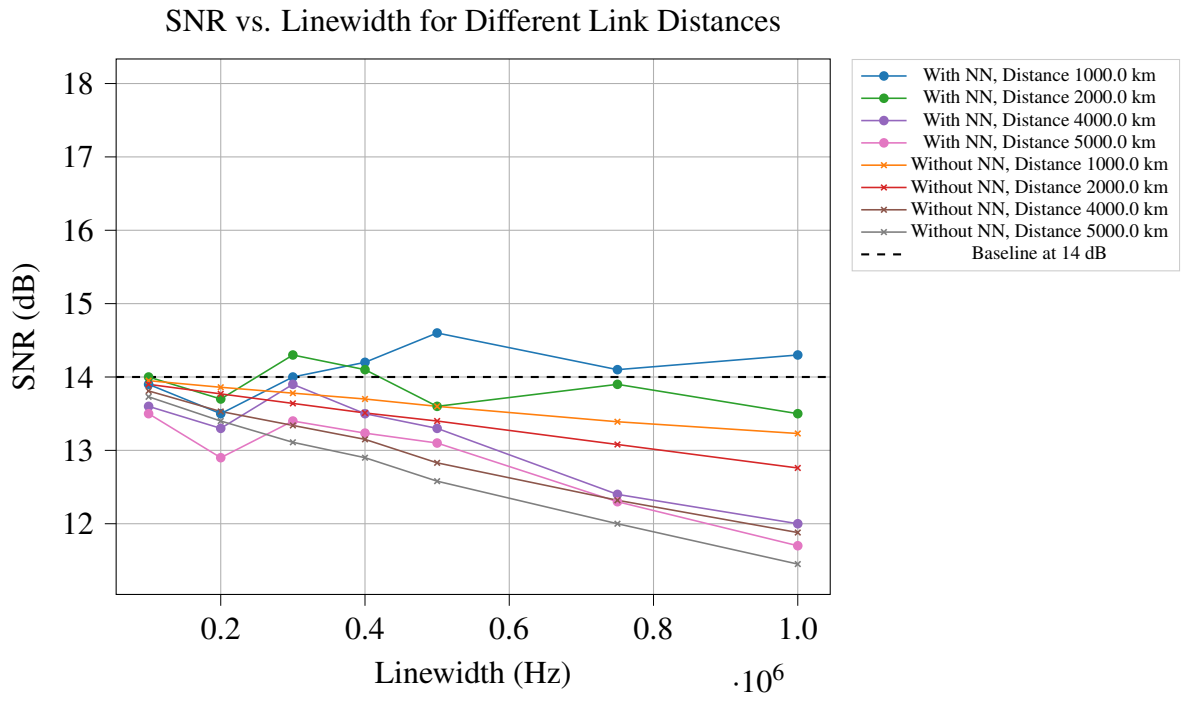


Figure 4.7: SNR vs Linewidth for Different Link Distances with MSE-X Loss Function

Figure 4.7 shows SNR performance for NN-based equalizer output with MSE-X loss function and a patience of 100 resulting in early stopping at around 300 epochs. The low SNR is evidence of insuffient training of the NN-based algorithm and thus, a poor performance.

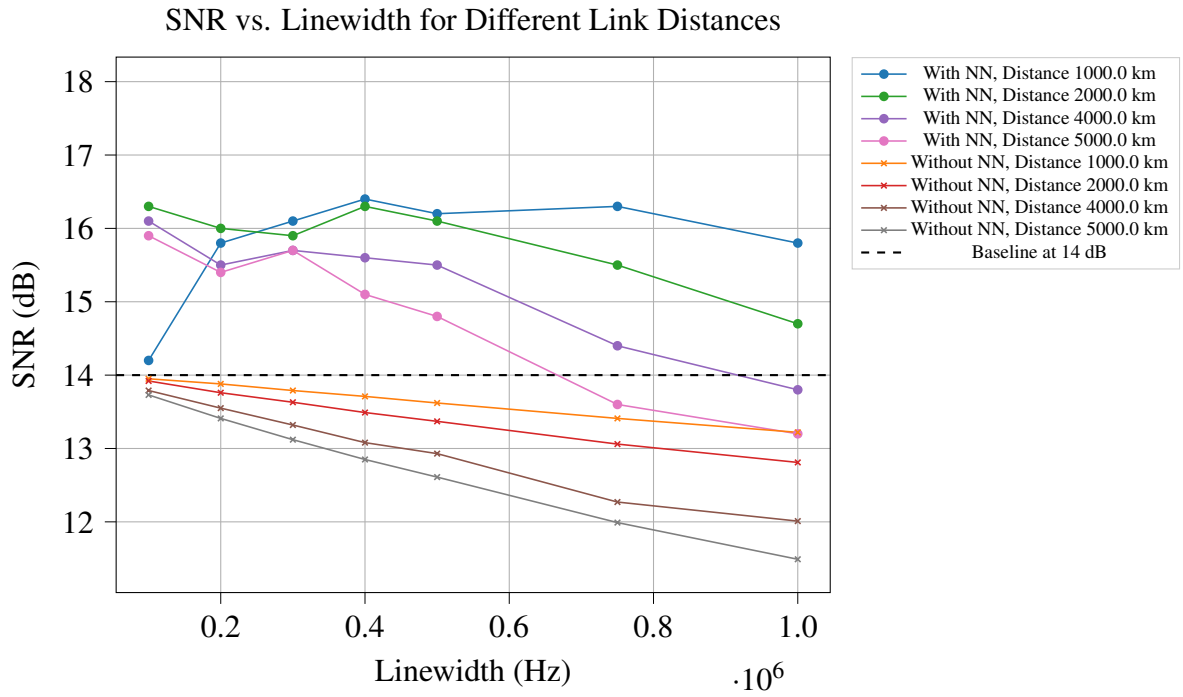


Figure 4.8: SNR vs Linewidth for Different Link Distances with MSE-X Loss Function

Figure 4.8 shows SNR performance for NN-based equalizer output with MSE loss function and a patience of 500 resulting in early stopping at around 1200 epochs. The SNR is almost the same as that without any early stopping but slightly lower and closer to baseline SNR as compared to that of without early stopping.

4.2.3 Key Observations

General Trends

As linewidth increases, SNR decreases for all distances. The SNR values without NN (solid lines with ”x” markers) are lower and decrease more gradually. The SNR values with NN-based equalizer (solid lines with circle markers) initially rise and then decline.

NN-based equalization significantly improves SNR, especially for shorter link distances (1000 km and 2000 km). However, at lower linewidths (200 kHz), SNR exceeds 18 dB, which seems unrealistically high due to residual EEPN.

For link distances of 4000 km and 5000 km, NN still improves SNR, but its advantage diminishes. This aligns with expectations because EEPN accumulates over long distances, limiting SNR improvement.

Key Differences

While the SNR values for outputs without NN remain consistent across different loss functions and performance metric combinations, the NN-based equalizer using the MSE loss function shows slightly higher SNR values compared to the MSE-X-based loss function. However, the difference is minor, and both exceed the 14 dB baseline.

SNR values across all early stopping results with the same patience levels are comparable. When the patience level is set to 100, the NN does not undergo sufficient training, as the original 2000 epochs are never reached, with training stopping before 300 epochs. This leads to degraded SNR performance, similar to that of the system without NN-based equalization.

For a patience level of 500 with AIR-based early stopping, training extends to approximately 900-1200 epochs, yielding results comparable to those without early stopping. While there is no significant performance improvement, the reduced number of epochs lowers memory and computational demands.

4.3 Constellation Diagram: Tx vs Rx Symbols

Constellation diagrams are a fundamental tool for visualizing the performance of digital communication systems, particularly in the presence of noise, distortion, and phase impairments. These scatter plots plot received symbols in the complex plane, allowing for an intuitive assessment of how well a signal has been transmitted and received.

In an ideal scenario, received symbols should cluster tightly around their intended positions, forming distinct, well-separated points. However, real-world impairments such as phase noise, amplitude distortions, and channel effects cause symbols to spread, leading to inter-symbol interference and degradation in signal quality.

Scatter plots provide a clear visualization of the impact of EEPN and how the neural network modifies the received symbols. EEPN causes symbol spreading and rotation, particularly at higher linewidths. The 16-QAM constellation becomes distorted, with clusters spreading outward. For long distances, the received symbols deviate significantly from their original positions.

4.3.1 MSE-X

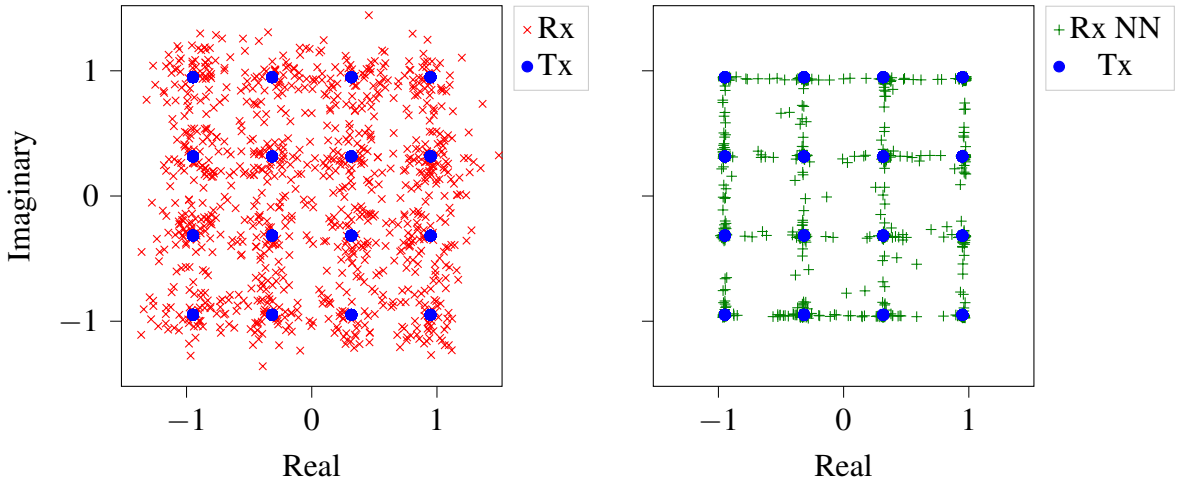


Figure 4.9: Scatter Plot for Tx vs Rx with and without NN for linewidth 750kHz and link distance 2000km with MSE-X Loss Function

Figure 4.9 shows much cleaner constellation for NN-based equalizer output with MSE-X loss function. The predicted symbols look much closer to transmitted symbols and jail window is much thinner as compared to with MSE loss function.

4.3.2 AIR with Early Stopping: Constellation Findings

MSE

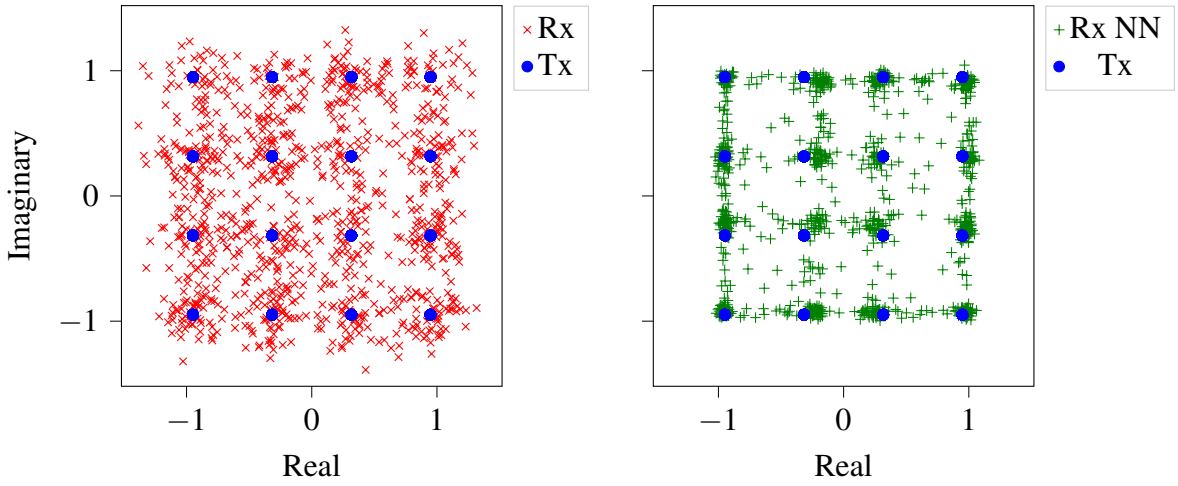


Figure 4.10: Scatter Plot for Tx vs Rx with and without NN for linewidth 750KHz and link distance 2000km with MSE Loss Function

Figure 4.10 shows presence of a dense ‘jail window‘ due to insufficient model training because of a very low patience level of 100. This can possibly be improved by setting a higher patience level so that AIR reaches a global minima ie. training does not stop prematurely.

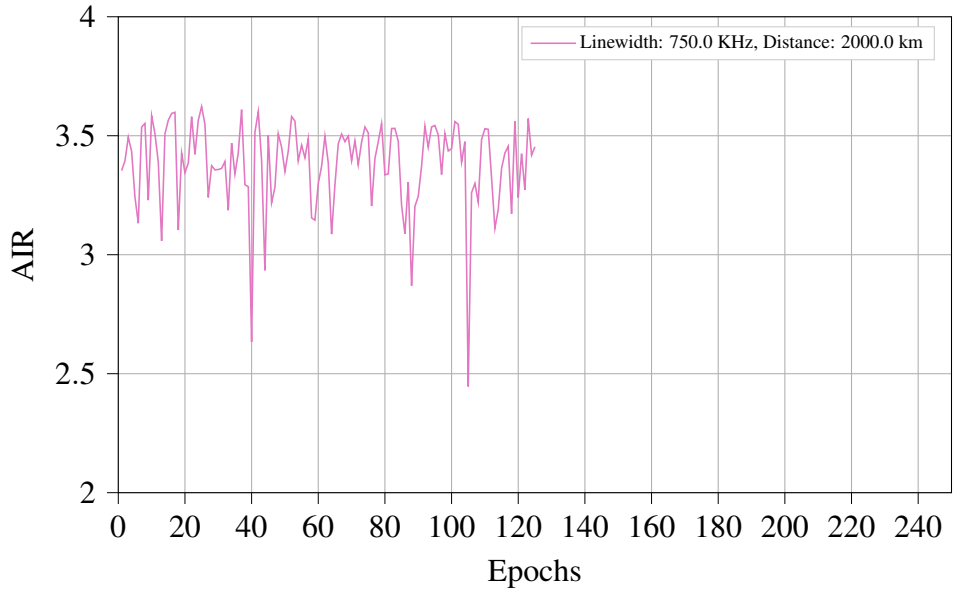


Figure 4.11: AIR vs Epochs for AIR with Early Stopping, MSE Loss Function, Patience = 100

Figure 4.11 depicts AIR values against epochs for linewidth of 750kHz and link distance of 2000km for MSE loss function. It can be seen that the model stops training when AIR does not exceed the maxima for 100 epochs, ie. around 125 epochs.

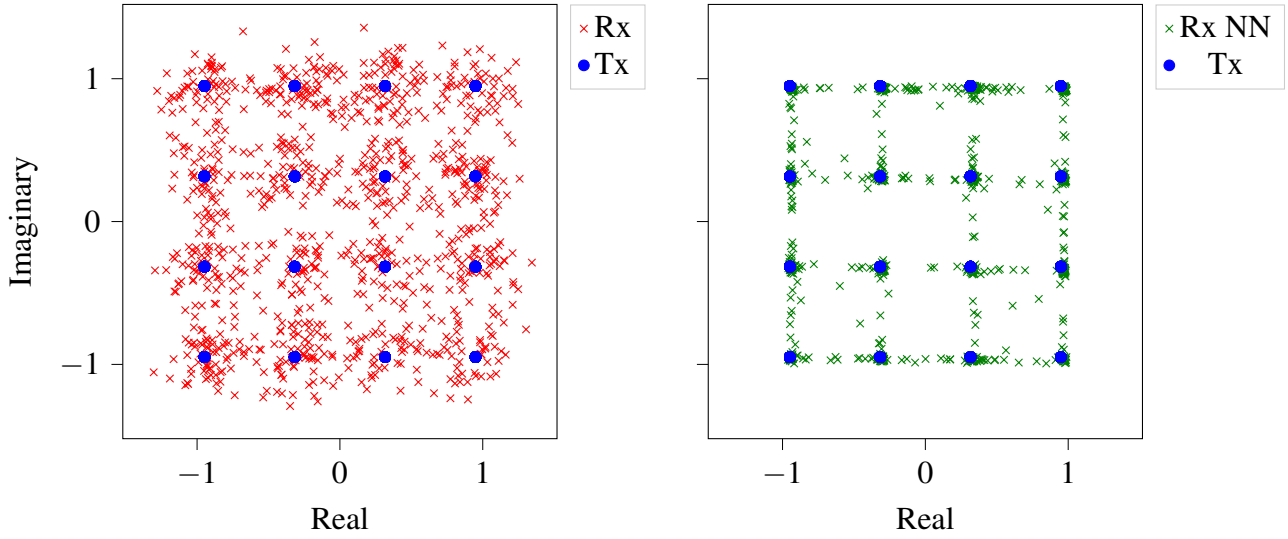


Figure 4.12: Scatter Plot for Tx vs Rx with and without NN for linewidth 750kHz and link distance 2000km with MSE Loss Function

Figure 4.12 shows presence of a significantly thin ‘jail window‘ due to a higher value of patience which is set at 500. The training seems to have run till around 770, so it did not stop too prematurely.

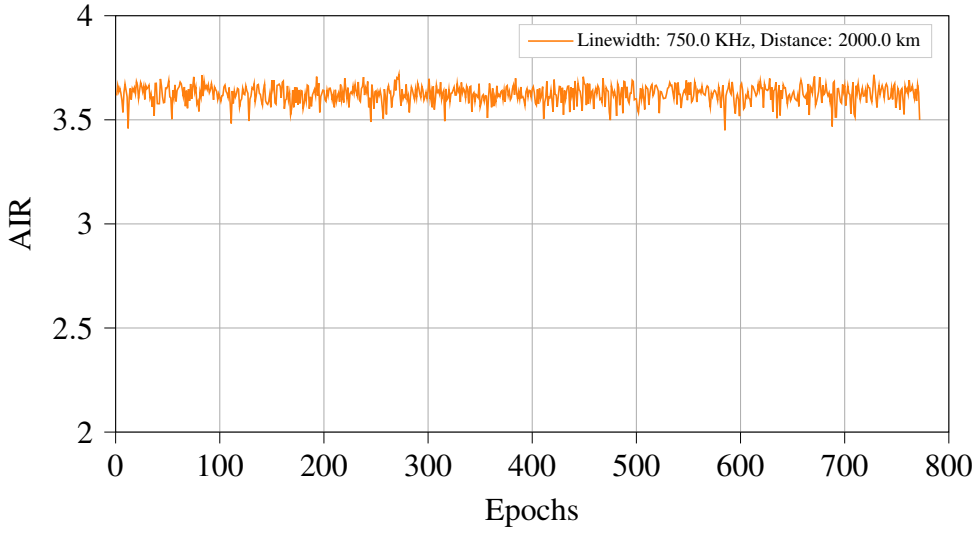


Figure 4.13: AIR vs Epochs for AIR with Early Stopping, MSE Loss Function, Patience = 500

Figure 4.13 depicts AIR values against epochs for linewidth of 750kHz and link distance of 2000km for MSE loss function. It can be seen that the model stops training when AIR does not exceed the maxima for 500 epochs, ie. around 770 epochs.

MSE-X

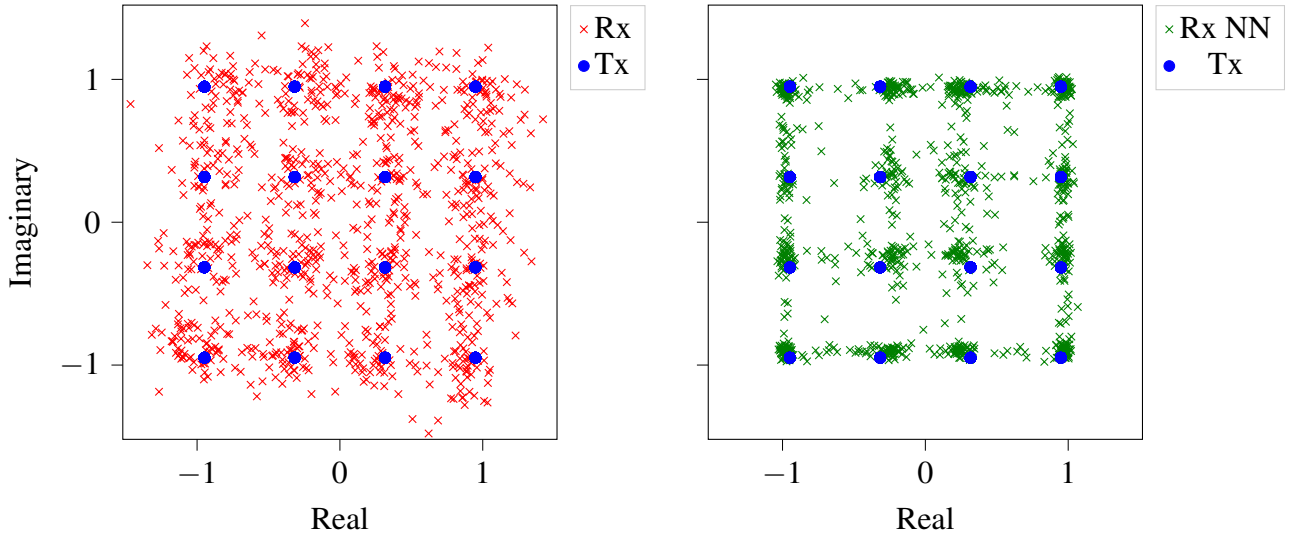


Figure 4.14: Scatter Plot for Tx vs Rx with and without NN for linewidth 750KHz and link distance 2000km with MSE-X Loss Function

Figure 4.14 shows presence of a dense ‘jail window‘ due to insufficient model training, due to a patience level of 100. This can likely be improved by setting a higher patience level so that AIR reaches a global minima ie. training does not stop prematurely.

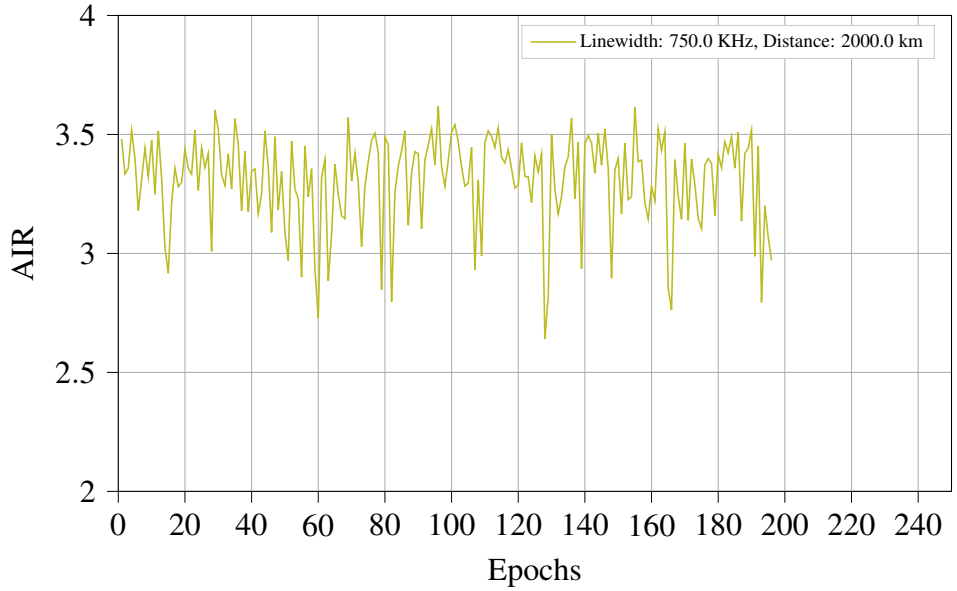


Figure 4.15: AIR vs Epochs for AIR with Early Stopping, MSE-X Loss Function, Patience = 100

Figure 4.15 depicts AIR values against epochs for linewidth of 750kHz and link distance of 2000km for MSE-X loss function. It can be seen that the model stops training when AIR does not exceed the maxima for 100 epochs, ie. around 190 epochs.

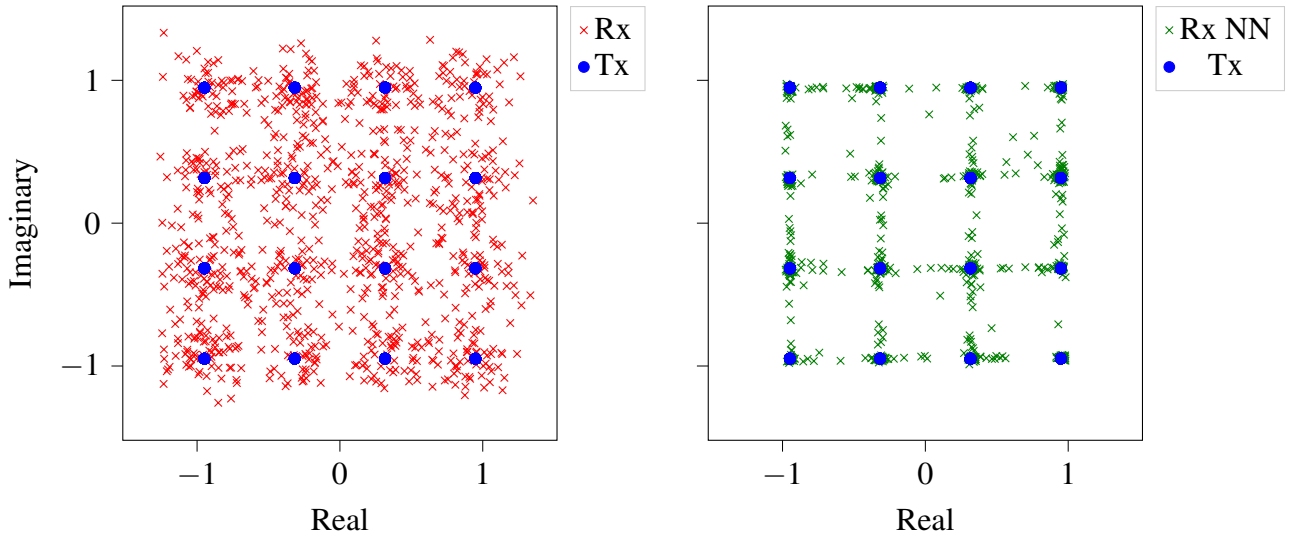


Figure 4.16: Scatter Plot for Tx vs Rx with and without NN for linewidth 750kHz and link distance 2000km with MSE-X Loss Function

Figure 4.16 shows presence of a sparse ‘jail window‘ due to a high patience level of 500. The

training seems to have stopped slightly prematurely and run for around 620 epochs.

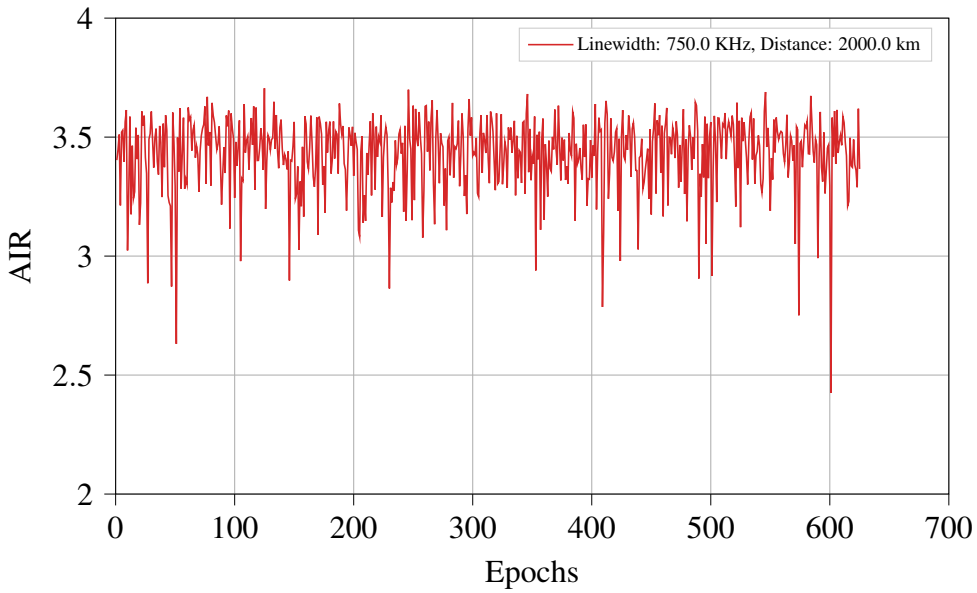


Figure 4.17: AIR vs Epochs for AIR with Early Stopping, MSE-X Loss Function, Patience = 500

Figure 4.17 shows AIR values against epochs for linewidth of 750kHz and link distance of 2000km for MSE-X loss function. It can be seen that the model stops training when AIR does not exceed the maxima for 500 epochs, ie. around 620 epochs.

4.3.3 Key Observations

General Trends

The scatter plots compare the Tx and Rx symbols for a link distance of 2000 km and a laser linewidth of 750 kHz, both without and with the NN-based equalizer. The plots reveal the jail window phenomenon and provide insights into the performance of the NN-based equalizer in mitigating EEPN.

Without the NN-based equalizer, the system suffers from severe phase noise, which degrades the quality of the received symbols. The spread in the scatter plot is consistent with the expected behavior of a system affected by EEPN.

The received symbols are clustered in specific regions of the complex plane, forming a jail window pattern. The clusters are neater and more compact compared to the case without the NN, indicating that the NN is reducing some of the noise.

However, the symbols are still not aligned with the transmitted symbols, and the clusters are offset from their ideal positions.

Key Differences

The scatter plot of transmitted versus received symbols with NN using the MSE-X loss function appears cleaner and more precise compared to that with the MSE loss function. This indicates that while MSE-X helps reduce the jail window effect, it does not completely eliminate it.

With AIR-based early stopping, the results for MSE and MSE-X loss functions are similar when using the same patience levels. Overall, symbols generated with a patience level of 500 are more precise than those obtained with a patience level of 100. This aligns with expectations, as models that undergo insufficient training and stop prematurely struggle to equalize effectively compared to those trained closer to the maximum epoch limit.

The scatter plots further demonstrate that the NN-based equalizer forms more distinct clusters of received symbols but fails to fully mitigate EEPN. The jail window phenomenon suggests that the NN introduces a systematic error, causing received symbols to cluster in specific regions instead of aligning accurately with transmitted symbols. This highlights that while the NN provides some improvement, it does not fully compensate for EEPN, indicating the need for further research into more effective mitigation strategies.

5 Conclusion and Outlook

This thesis has delved into the challenges encountered when applying neural networks to mitigate EEPN in communication systems, specifically focusing on the issue of loss function, exemplified by the 'jail window' effect. Through analysis, it was found that the use of the Mean Squared Error (MSE) loss function exacerbates jail window, leading to poor generalization. The work highlights the pitfalls of relying on MSE in neural network training, which tends to favor memorizing noise patterns instead of learning meaningful signal features. The shift in focus from compensating EEPN to documenting these pitfalls has provided valuable insights into the complexities of using neural networks for signal processing tasks. While neural networks show promise for various applications, their inherent limitations need to be addressed to improve their reliability and robustness in real-world communication systems.

Looking forward, future research should explore alternative loss functions and regularization techniques to mitigate the pitfalls and improve generalization. Incorporating more sophisticated loss functions, such as those based on perceptual or domain-specific metrics, could reduce the tendency of models to memorize noise patterns. Another potential avenue is the development of hybrid models that combine the strengths of traditional signal processing techniques with neural network-based approaches. This could help bridge the gap between theoretical models and practical implementations. Finally, advancing the interpretability of neural networks in communication systems could aid in understanding the decision-making process of these models and offer deeper insights into their behavior. By addressing these challenges, the application of neural networks in mitigation of EEPN and other signal processing tasks can be significantly improved, offering more reliable and efficient solutions for next-generation communication systems.

Bibliography

[1] W. Shieh and K.-P. Ho, "Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing," *Optics Express*, vol. 16, no. 20, pp. 15 718–15 727, 2008. (p. 1)

[2] S. Oda, C. Ohshima, T. Tanaka, T. Tanimura, H. Nakashima, N. Koizumi, T. Hoshida, H. Zhang, Z. Tao, and J. C. Rasmussen, "Interplay between local oscillator phase noise and electrical chromatic dispersion compensation in digital coherent transmission system," in *36th European Conference and Exhibition on Optical Communication*. IEEE, 2010, pp. 1–3. (p. 1)

[3] C. Xie, "Local oscillator phase noise induced penalties in optical coherent detection systems using electronic chromatic dispersion compensation," in *2009 Conference on Optical Fiber Communication*. IEEE, 2009, pp. 1–3. (p. 1)

[4] A. P. T. Lau, T. S. R. Shen, W. Shieh, and K.-P. Ho, "Equalization-enhanced phase noise for 100gb/s transmission and beyond with coherent detection," *Optics Express*, vol. 18, no. 16, pp. 17 239–17 251, 2010. (p. 1)

[5] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, no. 6, p. 386, 1958. (p. 1)

[6] P. J. Freire, A. Napoli, B. Spinnler, N. Costa, S. K. Turitsyn, and J. E. Prilepsky, "Neural networks-based equalizers for coherent optical transmission: Caveats and pitfalls," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 4: Mach. Learn. in Photon. Commun. and Meas. Syst., pp. 1–23, 2022. (p. 1), (p. 2), (p. 29)

[7] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical ofdm," *IEEE Photonics Technology Letters*, vol. 27, no. 4, pp. 387–390, 2014. (p. 1), (p. 29)

[8] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*. Springer, 2000, pp. 195–211. (p. 3)

[9] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013. (p. 3)

[10] K. Pearson, "The problem of the random walk," *Nature*, vol. 72, no. 1867, pp. 342–342, 1905. (p. 7)

[11] G. P. Agrawal, *Fiber-optic communication systems*. John Wiley & Sons, 2012. (p. 8)

[12] A. Kakkar, J. R. Navarro, R. Schatz, H. Louchet, X. Pang, O. Ozolins, G. Jacobsen, and S. Popov, "Comprehensive study of equalization-enhanced phase noise in coherent optical systems," *Journal of lightwave technology*, vol. 33, no. 23, pp. 4834–4841, 2015. (p. 9)

[13] H. Wang, X. Yi, J. Zhang, and F. Li, "Extended study on equalization-enhanced phase noise for high-order modulation formats," *Journal of Lightwave Technology*, vol. 40, no. 24, pp. 7808–7813, 2022. (p. 11)

[14] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2. (p. 11)

[15] A. Zell, "Simulation neuronaler netze," in *Simulation neuronaler Netze*. Oldenbourg Wissenschaftsverlag, 1994. (p. 12)

[16] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991. (p. 12)

[17] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323. (p. 13)

[18] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International workshop on artificial neural networks*. Springer, 1995, pp. 195–201. (p. 13)

[19] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, "Backpropagation and the brain," *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 335–346, 2020. (p. 13)

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. (p. 14)

[21] D. E. Rumelhart, G. E. Hinton, R. J. Williams *et al.*, "Learning internal representations by error propagation," 1985. (p. 14)

[22] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948. (p. 17)

[23] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999. (p. 17)

[24] C. Lee, J. Go, B. Baek, and H. Choi, "Neural network equalizer," in *Intelligent Computing: International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006. Proceedings, Part I 2*. Springer, 2006, pp. 204–215. (p. 17)

[25] F. Gardner, "A bpsk/qpsk timing-error detector for sampled receivers," *IEEE Transactions on communications*, vol. 34, no. 5, pp. 423–429, 1986. (p. 20)

[26] R. Farhoudi, A. Ghazisaeidi, and L. A. Rusch, "Performance of carrier phase recovery for electronically dispersion compensated coherent systems," *Optics express*, vol. 20, no. 24, pp. 26 568–26 582, 2012. (p. 21)

[27] N. Guo, G. Shen, N. Deng, and B. Mukherjee, "Can channel power optimization with gsnr flatness maximize capacities of c+ l-band optical systems and networks?" *Journal of Lightwave Technology*, 2024. (p. 27)

[28] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, and S. Pachnicke, "Deep neural network equalization for optical short reach communication," *Applied Sciences*, vol. 9, no. 21, p. 4675, 2019. (p. 29)

[29] S. Liu, P.-C. Peng, C.-W. Hsu, S. Chen, H. Tian, and G.-K. Chang, "An effective artificial neural network equalizer with s-shape activation function for high-speed 16-qam transmissions using low-cost directly modulated laser," in *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*. IEEE, 2018, pp. 269–273. (p. 29)

[30] F. Diedolo, G. Böcherer, M. Schädler, and S. Calabró, "Nonlinear equalization for optical communications based on entropy-regularized mean square error," in *European Conference and Exhibition on Optical Communication*. Optica Publishing Group, 2022, pp. We2C–2. (p. 30)

Declaration

I hereby declare that I have written this thesis independently and have not used any sources or aids other than those specified and that all statements taken verbatim or in spirit from other works have been marked as such.

The thesis has not yet been submitted to any other examination authority and has not yet been published. The electronic copy corresponds to the printed copies.

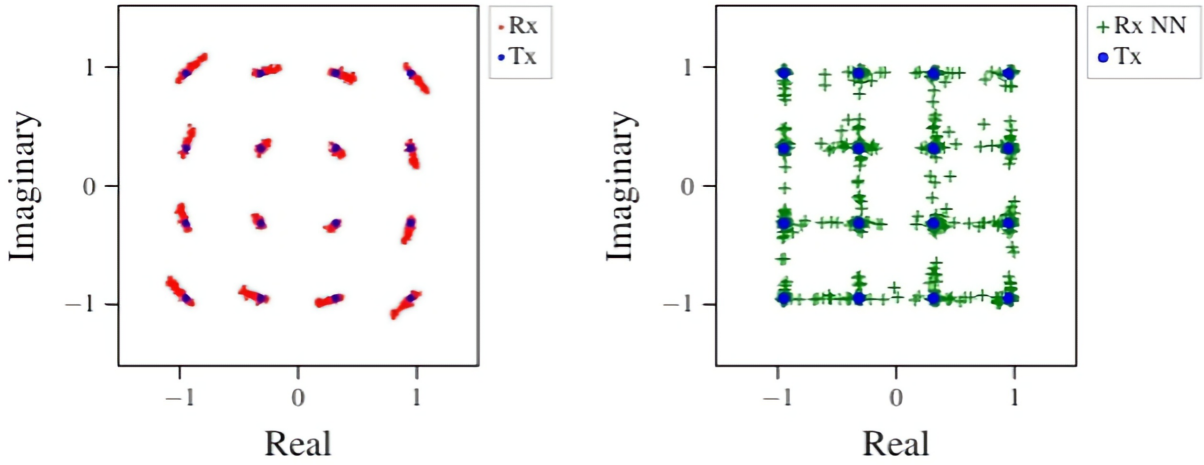
Stuttgart, March 03, 2025

Shalini Rai

Research Thesis

Investigation of Neural Equalization-Enhanced Phase Noise Mitigation

Shalini Rai



Date of hand in: March 3, 2025
Supervisor: Sebastian Jung
Tim Janz

Abstract

This thesis investigates Neural Networks (NN) for Equalization-Enhanced Phase Noise (EEPN) mitigation, highlighting the ‘jail window‘ effect, where models fail to generalize across linewidths and distances. It explores how Mean Squared Error (MSE) loss contributes to the pitfall, limiting Achievable Information Rate (AIR), and discusses strategies for improving NN robustness.

submitted to

University of Stuttgart
Institute of Telecommunications