Nonlinar compensation algorithm for WDM system in optical fiber

Xinyu Xiao, Bin Dong, Zhennan Zhou

March 19, 2022

Contents

Abstract

In long-haul optical communication systems, compensating nonlinear effects through digital signal processing (DSP) is difficult due to intractable interactions between Kerr nonlinearity, chromatic dispersion (CD) and amplified spontaneous emission (ASE) noise from inline amplifiers. We constructed a pytorch-based optical fiber communication system including Tra nsmitter, fiber channel model and receiver. We design some digital back propagation models and test their performence. We get some improvements compared with some latest work.

1 Communication System

1.1 overview

There are the most important components in a fiber communication system (see Figure [1]):

1. Transmitter: turn a complex symbol sequence to a wave form signal and transmitted to a fiber channel.

$$\{X_n\}_{n=0}^{N} \rightarrow u(0,t), t \in R^+$$

   Where $X_n \in \mathbb{C}$ are the transmmited symbols, $u \in \mathbb{C}$ is the wave form signal. We assume total N symbols here.

2. Fiber channel: evolution the analog signal from $z = 0$ to $z = L$.

$$u(0,t) \rightarrow u(L,t)$$

3. Reciever: sampled data points from the received signal $u(L,t)$ and turn this data to a complex symbol sequence.

$$u(L,t) \rightarrow \{Y_n\}_{n=0}^{N}$$

   where $Y_n \in \mathbb{C}$ is output symbols.

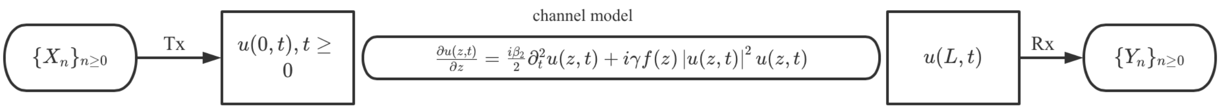


Figure 1: data flow

Because of the noise in fiber channel, we get a channel conditional probability:

$$P(Y|X)$$

where $X = (X_1, X_2, \ldots, X_N)$, $Y = (Y_1, Y_2, \ldots, Y_N)$.
Our target

1. In practice. We want to increase the bit error rate(BER) as much as we can.

$$BER = Pr(X_n = Y_n)$$

2. In theorem. We want to know the capacity limit in such a nonlinear communication system.

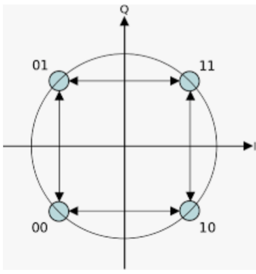
$$C = \sup_{P(X)} \frac{1}{N} I(X:Y)$$

   Here $I(X;Y)$ is the mutual information between two random vectors. This formula is the famous Seccond shannon theorem.

1.2 Transmitter Model

$$\begin{cases} u(0,t) = \sum_{i=M}^{-M} A_i(0,t) exp(i(\omega_i - \omega_0)t) \quad t \in [0, T_f] \\ A_i(0,t) = \sqrt{P_i} \sum_{k=0}^{N-1} x_k^i g(t - kT) \end{cases} \tag{1}$$

Where $g(t)$ is the Raise-Cosin signal pulse controled by a roll-off parameter β. Show as formula (2) and figure (2). T is symbol period and $T_f = N * T$ is the signal length. P_0 is the peak power. $x_k^i \in \mathbb{C}$ is complex symbol (See Remark (1.1)). ω_0 is the central carrier frenquency.



Remark 1.1. *[4QAM]*

$$x_j^k \in \mathcal{X} = \{\pm 1 \pm i\}$$

$$g(t) = \begin{cases} \frac{\pi}{4T}\operatorname{sinc}\left(\frac{1}{2\beta}\right), & t = \pm\frac{T}{2\beta} \\ \frac{1}{T}\operatorname{sinc}\left(\frac{t}{T}\right)\frac{\cos\left(\frac{\pi\beta t}{T}\right)}{1-\left(\frac{2\beta t}{T}\right)^2}, & \text{otherwise} \end{cases} \tag{2}$$

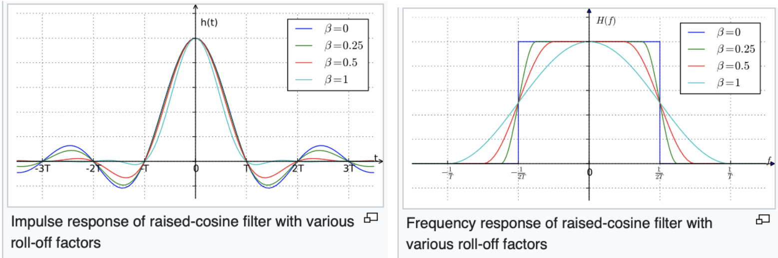


Figure 2: Raise-Cosin signal pulse

1.3 Fiber model

1.3.1 Nonlinear Schrodinger equation model (NLSE)

$$\begin{cases} \dfrac{\partial u(z,t)}{\partial z} = -\dfrac{\alpha}{2}u(z,t) + \dfrac{i\beta_2}{2}\dfrac{\partial^2 u(z,t)}{\partial t^2} + i\gamma\left|u(z,t)\right|^2 u(z,t) + n(z,t) \\ u(z,0) = u(z,T_f) \\ u(0,t) = \displaystyle\sum_{i=M}^{-M} A_i(0,t)exp(i(\omega_i - \omega_0)t) \\ A_i(0,t) = \sqrt{P_i}\displaystyle\sum_{k=1}^{N} x_k^i g(t-kT) \\ (z,t) \in [0,L] \times [0,T_f] \end{cases} \tag{3}$$

Where α,β_2,γ denote attenuation, group velocity dispersion, and fiber nonlinearity coefficient. $n(z,t)$ is a complex Gaussian noise process with autocorrelation (1.2)(see [3]).

Remark 1.2. *[noise term]*

$$\begin{cases} \left\langle n(z,\omega)\bar{n}\left(z',\omega'\right)\right\rangle_n = 2\pi\mathcal{Q}\delta\left(\omega-\omega'\right)\theta\left(\dfrac{W'}{2} - |\omega|\right)\delta\left(z-z'\right) \\ \left\langle n(z,t)\bar{n}\left(z',t'\right)\right\rangle_n = \dfrac{\mathcal{Q}}{\pi\left(t-t'\right)}\sin\left(\dfrac{W'\left(t-t'\right)}{2}\right)\delta\left(z-z'\right) \end{cases} \tag{4}$$

Where $\theta(x)$ is Heaviside theta-function.

Remark 1.3 (FT)**.**

$$n(z,\omega) = \int_{-\infty}^{\infty} dt e^{-i\omega t} n(z,t)$$

1.3.2 Couple Nonlinear Schrodinger Equation model (CNLSE)

Use the WDM initial expression (5), and neglect FWM terms(when can we neglect this term ?), we can get a simpler model (6).

$$u(z,t) = \sum_{i=M}^{-M} A_i(z,t)exp(i(\omega_i - \omega_0)t) \tag{5}$$

$$\frac{\partial A_i(z,t)}{\partial z} = -\frac{\alpha}{2}A_i(z,t) - \beta_{i1}\frac{\partial A_i(z,t)}{\partial t} + \frac{i\beta_{i2}}{2}\frac{\partial^2 A_i(z,t)}{\partial t^2} + i\gamma\left(|A_i|^2 + 2\sum_{q\neq i}|A_q|^2\right)A_i(z,t) + n(z,t) \tag{6}$$

1.3.3 Split Step Fourier Method

To solve the NLSE and CNLSE, we review the classical Split Step Fourier Method(SSFM) .

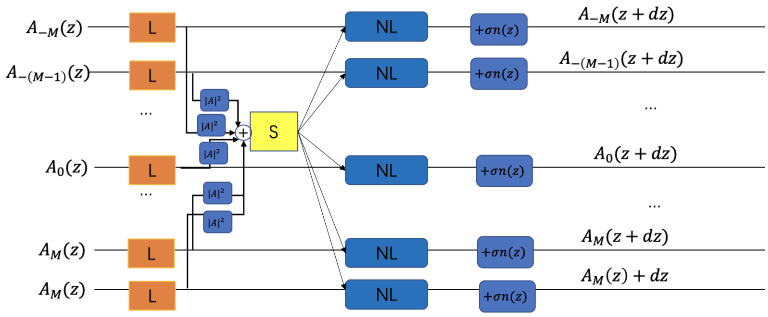


Figure 3: SSFM structure: one step

$$SSFM: \begin{pmatrix} A_{-M}(0,t) \\ A_{-(M-1)}(0,t) \\ \dots \\ A_0(0,t) \\ \dots \\ A_M(0,t) \end{pmatrix} \to \begin{pmatrix} A_{-M}(L,t) \\ A_{-(M-1)}(L,t) \\ \dots \\ A_0(L,t) \\ \dots \\ A_M(L,t) \end{pmatrix}$$

Using SSMF method: Linar step and Nonlinear step alternate. Assume each symbol take N_t points sampling at equal intervals. Define collocation points $t_j = j\frac{T}{N_t}(0 \leq j \leq N_f)$. Here, $N_f = N * N_t$.

$$A_i(z) := [A_i(z,t_0), A_i(z,t_1), \dots, A_i(z,t_{N_f})] \in \mathbb{C}^{N_f}$$

The transfer function:

$$\omega = \frac{2\pi}{T_f}[0, 1, \dots, \frac{N_f}{2}, -\frac{N_f}{2}+1, -\frac{N_f}{2}+2, \dots, -1] \in \mathbb{C}^{N_f}$$
$$H_i(dz) = \exp\left[\left(-\frac{\alpha}{2} - \mathrm{i}\beta_{i1}\omega - \mathrm{i}\beta_{i2}\frac{\omega^2}{2}\right)dz\right] \in \mathbb{C}^{N_f}$$

1. Linear step

$$A_i(z+dz) = L_{H_i(dz)}(A_i(z)) := ifft(fft(A_i(z)).*H_i(dz)) \tag{7}$$

2. Nonlinear step

$$A_i(z+dz) = NL_{dz,S_i}(A_i(z)) := A_i(z).*\exp\left(\mathrm{i}\gamma dz*(S_i + |A_i|^2)\right) := A_i(z).*\exp\left(i\gamma dz*(2\sum_{q\neq i}|A_q(z)|^2 + |A_i(z)|^2)\right) \tag{8}$$

We use S_i to denotes the power sum of the other channels.

$$S_i = 2\sum_{q\neq i}|A_q(z)|^2 \tag{9}$$

The whole process show in Figure [3].

Algorithm 1 SSFM

Input: input transmmitted signals $A_i(0)(-M \leq i \leq M)$, step size dz, noise level σ,transmmitted length L
Output: output result

$K = \frac{L}{dz}$
for s = 1,...,K **do**
  Linear step:
$$A_i^1(z) = L_{H_i(dz)}(A_i(z))$$
  Calculate S_i:
$$S_i = 2\sum_{q\neq i} |A_i^1(z)|^2$$
  Nonlinear step:
$$A_i^2(z) = NL_{dz,S_i}(A_i^1(z))$$
  Add noise: sample a gaussian noise $n(z)$.
$$A_i(z+dz) \leftarrow A_i(z+dz) + \sigma * n(z)$$
  $z \leftarrow z + dz$
return $A_i(L)$, $-M \leq i \leq M$

1.4 Receiver Model

There are three steps in our receiver model.

1. Digital back propagation. This step is the most import part to compensate all kinds of interference. We will design some algorithm in Section 2.
$$\hat{A}_0(t) = DBP(A_0(L,t)) \tag{10}$$
2. Filter.
$$\tilde{x}_k^0 = \frac{1}{\sqrt{P_0}\|g\|_2^2}\int_{-\infty}^{+\infty} \hat{A}_0(t)g_k(t)dt \tag{11}$$
Here we define $g_k(t) = g(t-kT)$, then g_k is othorgonal with respect to diffrent k.
3. Classification.
$$\hat{x}_k^0 = \underset{x\in\mathcal{X}}{argmin}\, \|x - \tilde{x}_k^0\| \tag{12}$$

We can show the whole communication system as Figure 4.

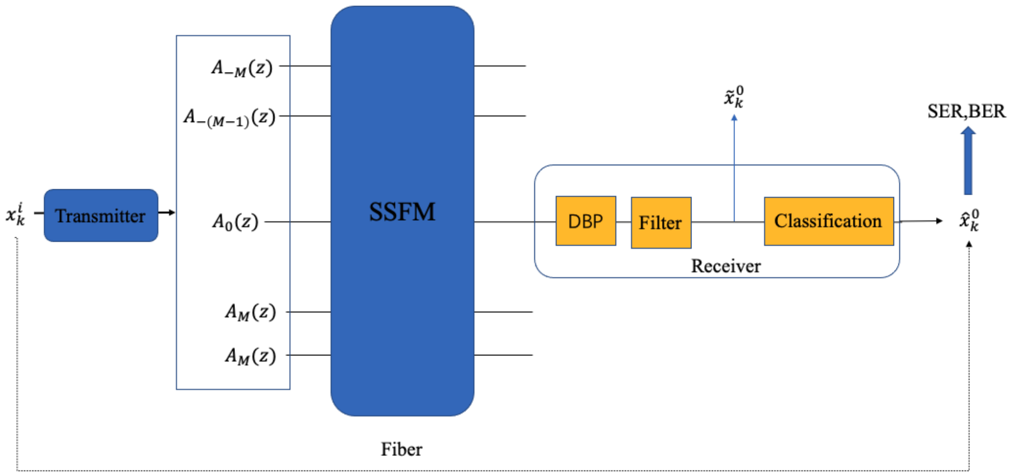


Figure 4: communication system

2 Method:Digital back propagation

This section aims to design the digital back propagation algorithm (See (10)). We concentrate on the central channel $A_0(z,t)$.

$$DBP : A_0(L,t) \rightarrow \hat{A}_0(t)$$

The main challege is that we can not get the information of other channels. So we can not implement full channel digital propagation (FDBP). We can construct several digital compensation method:

2.1 Dispersion only DBP (DO-DBP)

This method only compensate the linear distortion.

$$\hat{A}_0(0) = L_{H_0(-L)}(A_0(L))$$

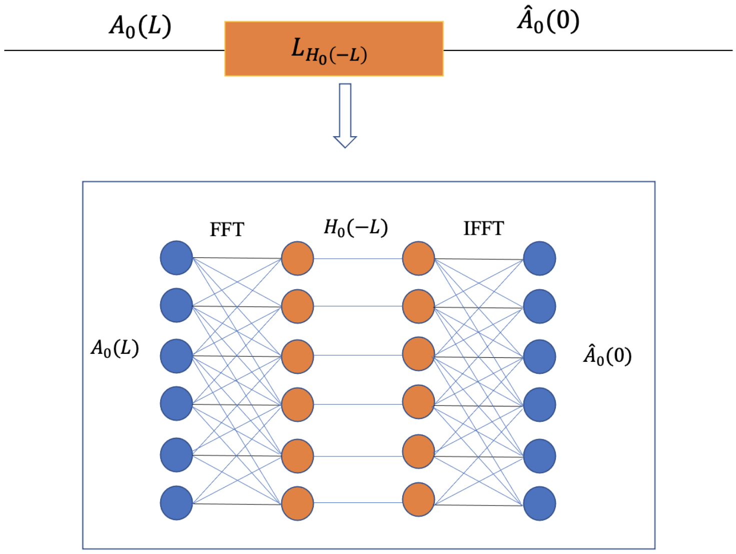


Figure 5: Dispersion Only DBP

2.2 Single channel DBP (SC-DBP)

$$\hat{A}_0(0) = \left(\prod_{i=1}^{K} NL_{-dz,S_0=0} \circ L_{H_0(-dz)}\right) A_0(L,t)$$

where K is L/dz the span numbers.

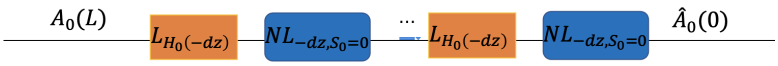


Figure 6: Single Channel DBP

2.3 Neural Network DBP (NN-DBP)

Notice that (7) is a linear transform and (8) is a point wise nonlinear transform. SSFM are composed with Linear transform and nonlinear activations just like deep neural network.[2] and [1] find this similiar structure between SSFM and DNN. Set some parameters in SSFM to free then get NN-DBP.

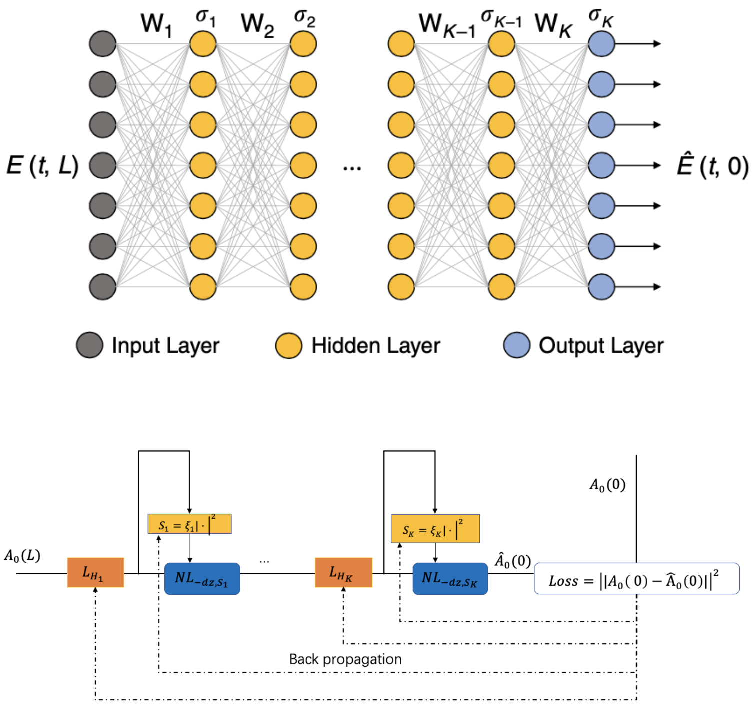


Figure 7: NN structure [1]

$$D_i(u) = NL_{-dz,S_0=\xi_i|u|^2} \circ L_{H_i}$$

$$\hat{A}(0) = \left(\prod_{i=1}^{K} D_i\right) A_0(L)$$

where $\{\xi_i, H_i\}$ is the trained parameters.

$$loss = \sum_{a_k^i \sim Uniform(\mathcal{X})} \|\hat{A}(0,t) - A(0,t)\|^2$$

2.4 Meta-1 DBP

Use two fully connected networks to estimate S_0 and H_0 in each step. We want to guess some information from $A_0(L)$. This is a meta neural network.

$$S_0 = \phi(u, \theta_i), H_0 = \psi(u, \mu_i) \tag{13}$$

$$D_i(u) = \left(NL_{-dz,S_0=\phi(u;\theta_i)} \circ L_{H_0=\psi(u;\mu_i)}\right)(u)$$

$$\hat{A}(0,t) = \left(\prod_{i=1}^{K} D_i\right) A_0(L,t)$$

where ϕ and ψ is two fully connected networks, $\{\theta_i, \mu_i\}$ is the trained parameter.

$$loss = \sum_{a_k^i \sim Uniform(\mathcal{X})} \|\hat{A}(0,t) - A(0,t)\|^2$$

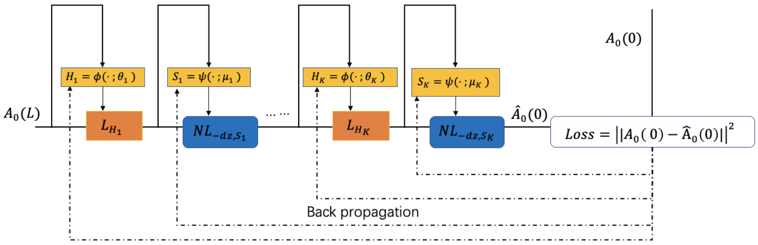


Figure 8: Meta-1 DBP structure

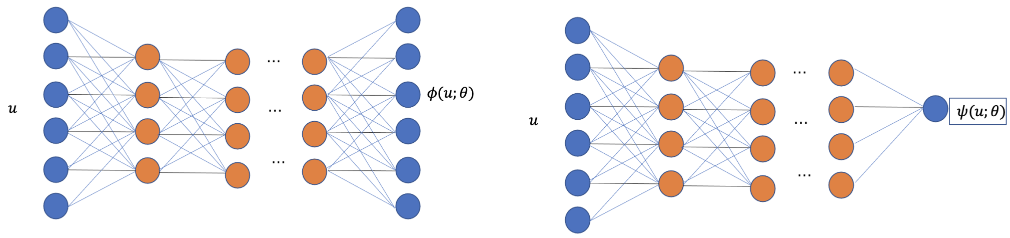


Figure 9: FNN structure

2.5 Meta-2 DBP

$$S_0 = \phi(u, \theta_i)|u|^2, H_0 = \psi(u, \mu_i) \tag{14}$$

Change the meta form.

$$D_i(u) = \left(NL_{-dz,S=\phi(u;\theta_i)|u|^2} \circ L_{-dz,H=\psi(u;\mu_i)}\right)(u)$$

$$\hat{A}(0,t) = \left(\prod_{i=1}^{K} D_i\right) A_0(L,t)$$

where ϕ and ψ is two fully connected networks, $\{\theta_i, \mu_i\}$ is the trained parameter.

$$loss = \sum_{a_k^i \sim Uniform(\mathcal{X})} \|\hat{A}(0,t) - A(0,t)\|^2$$

2.6 Meta-3 DBP

Share the meta-net weight in different steps.

$$D(u) = \left(NL_{-dz,S=\phi(u;\theta)} \circ L_{-dz,H=\psi(u;\mu)}\right)(u)$$

$$\hat{A}(0,t) = \left(\prod_{i=1}^{K} D_i\right) A_0(L,t)$$

where ϕ and ψ is two fully connected networks, $\{\theta, \mu\}$ is the trained parameter.

$$loss = \sum_{a_k^i \sim Uniform(\mathcal{X})} \|\hat{A}(0,t) - A(0,t)\|^2$$

3 Experiments result

The details of experiment are showed in Table 2. We construct two training data sets.

1. Data set A: $P_i = 50[mW]$.

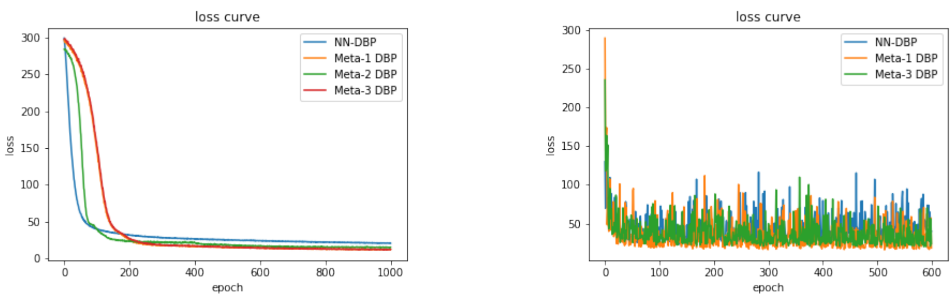
2. Date set B: P_i is i.i.d sampled from Uniform([50 mW,60 mW]) in each epoch.

We use different widths and depths for A and B. See Table 1. The loss curves are shown in Figure 10. Under setting A, we can see Meta-1 to Meta-3 can achive better minimizers than NN-DBP. These methods converge steadily. Under Setting B, training becomes untractable. The costellations are shown in Figure 11 and 12. The BER results are shown in Table 3 and Table 5. We mainly consider comparation of 4 methods: NN-DBP，Meta-1,Meta-2,Meta-3.We can get conclusions as follows:

1. Under setting A:
   - From the loss curve 10, the four methods all converge smoothly, and the convergence point from Meta-1 to Meta-3 is better than NN-DBP. The loss curves of Meta-1 and Meta-2 almost overlap, which also shows a certain equivalence of the two methods. Meta-3 converges faster than Meta-1 and Meta-2, but the minimum point reached in the end is not as good as the first two.
   - From the costellation diagram 11,when the test power is equal to the training power, the four methods all show the best results.When the test power is not equal to the training power, the non-linear phase shift is not perfectly compensated. Compared with the ideal result, there is a phase deflection overall
   - From table 3,when the test power is equal to 50mW, Meta-1 to Meta-3 are better than NN-DBP by about 2%. Other test power performances are not good, and there is no comparative value. 没有比较价值。
2. Under setting B:(Since Meta-1 and Meta-2 behave very similarly in A, the Meta-2 experiment was omitted here.)
   - From the loss curve 10, the convergence of the four methods is very difficult. Meta-1 is better than Meta-3. Meta-3 is better than NN-DBP
   - From the constellation diagram, when test power = 55mW, the four methods perform best. Other power values behave similarly to A, with a phase rotation difference overall. This means neural network Learned an average power result in the training set. This also shows that our method is invalid without knowing the power of the interfering channel
   - From table 3,when the test power is equal to 55 mW, Meta-1 and Meta-3 are better than NN-DBP by about 2%. Other test power performances are not good, and there is no comparative value.

setting	width	depth
A	60	2
B	120	3

Table 1: experiments details



(a) Loss curve for Data set A

(b) Loss curve for Data set B

Figure 10: Loss curve for Data set A and B

parameter	variable	quantity	unit
attenuation	α	0.2	[dB/km]
dispersion	β_2	-21.68	$[ps^2km^{-1}]$
nonlinear	γ	1.37	$[W^{-1}km^{-1}]$
peak power	P_0	50	[mW]
number of WDM channel	$2M+1$	3	1
length of fiber	L	100	[km]
symbol rate	$R_s = \frac{1}{T_f}$	10	[GHz]
central wavelength	λ_0	1550	[nm]
channel space	$\Delta\lambda$	0.4	[nm]
roll	β	0.2	1
noise level	σ	0.002	1
number of symbols	N	128	1
number of samples/symbol	N_t	8	1

Table 2: experiments details

Test power: 50[mW]		Test power: 52[mW]	Test power: 45[mW]
Method	Acc(100 times mean)	Acc(100 times mean)	Acc(100 times mean)
DO-DBP	0.586914	0.421367	0.9275
SC-DBP	0.0989844	0.216875	0.12375
NN-DBP	0.963242	0.886992	0.573984
Meta-1	0.989297	0.910117	0.559688
Meta-2	0.984766	0.895469	0.556758
Meta-3	0.975234	0.893203	0.589961

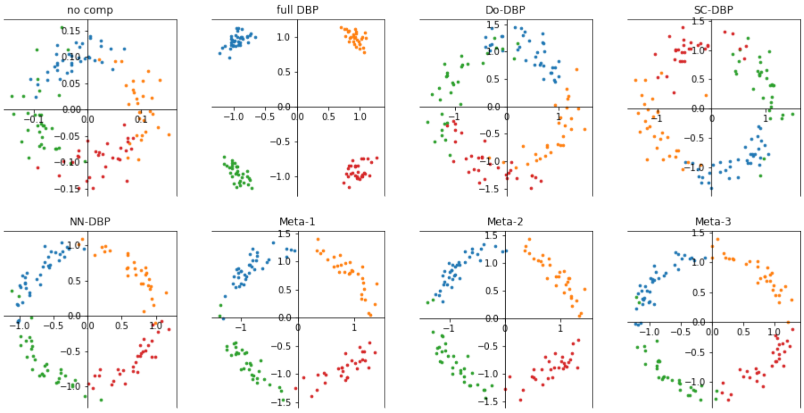
Table 3: Test result. Training data: A

Test power: 50[mW]	
Method	BER
no comp	0.518125
full DBP	0
DO-DBP	0.334219
SC-DBP	0.381836
NN-DBP	3.08594e-05
Meta-1	2.73437e-06
Meta-2	1.5625e-06
Meta-3	1.01563e-05

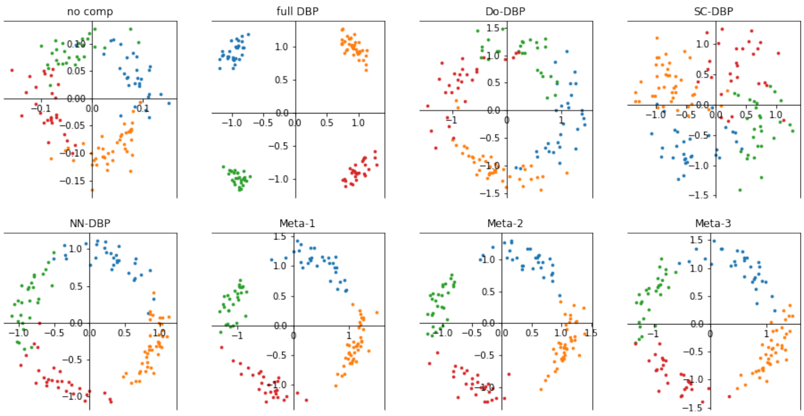
Table 4: TEST

Test power: 50[mW]		Test power: 55[mW]	Test power: 60[mW]
Method	Acc(100 times mean)	Acc(100 times mean)	Acc(100 times mean)
DO-DBP	0.587461	0.170586	0.318281
SC-DBP	0.0934766	0.410273	0.652031
NN-DBP	0.594141	0.948281	0.635977
Meta-1	0.600195	0.974766	0.606836
Meta-3	0.603828	0.964063	0.601992

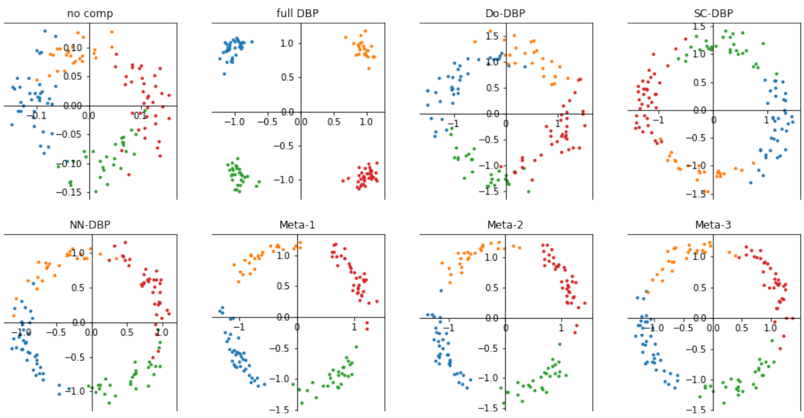
Table 5: Test result. Training data: B



(a) test power: 50mW

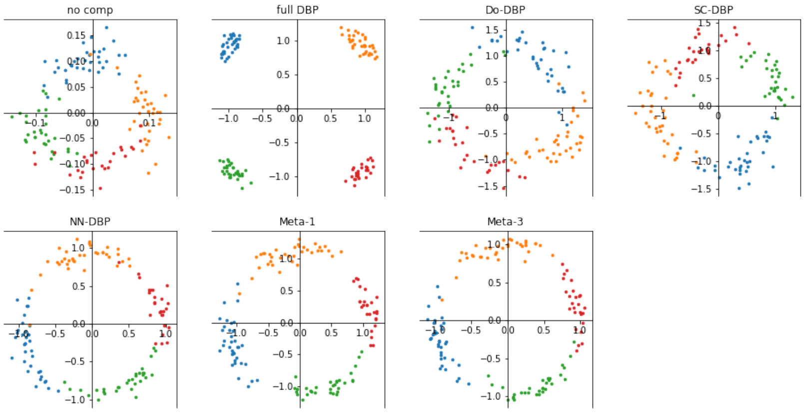


(b) test power: 55 mW

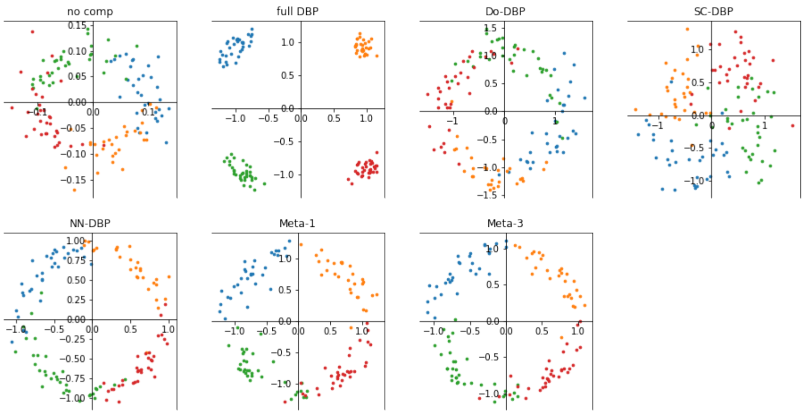


(c) test power: 45 mW

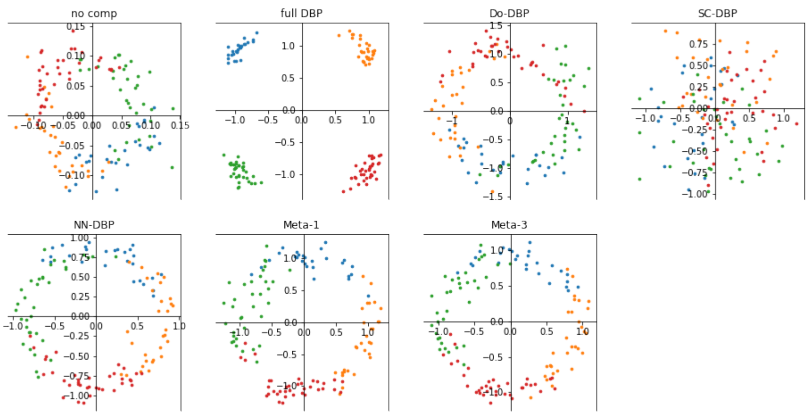
Figure 11: Plot the $\tilde{x}_k^0$. Training with Data Set A. Test with power: 50[mW],52[mW], 45[mW]



(a) test power: 50mW



(b) test power: 55 mW



(c) test power: 60 mW

Figure 12: Plot the $\tilde{x}_k^0$. Training with data set B, Test with power: 50[mW],55[mW],60[mW]

4 Future work

1. Improve the Meta-NN structure. Use CNN,RNN etc. Test the boundary of our method.
2. Use the data from physical experiment.(Thank Qiri Fan(author of [1]) for provding the data).
3. Consider simulating the optical pulse propagation with Vector Manakov–PMD equation.

$$\frac{\partial \Psi}{\partial z} = \left(-\frac{1}{2}\alpha - j\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 \frac{\partial}{\partial t} - j\beta_2 \frac{1}{2}\frac{\partial^2}{\partial t^2}\right)\Psi \quad + j\frac{8}{9}\gamma|\boldsymbol{\Psi}|^2\boldsymbol{\Psi} + \begin{bmatrix} n_x(z,t) \\ n_y(z,t) \end{bmatrix}$$

4. Change the system parameter to approach real optical fiber system. In particular, increase symbol rate to 60 [GHz]. Increase number of spans to about 10. Change the distributed noise to EDFA noise. Reduce peak power to about 1[mW]. Increase number channels. The most important thing is to think about how to pose the problem. Under what circumstances can it be resolved, and under what circumstances will it reach the unsolvable boundary。
5. Improve the structure of the transmitter and receiver, set reasonable down sampling rules, and add some classic adaptive filters and adaptive phase noise estimators (FOE, CPE) to improve System performance.
6. Considering the calculation limitations of practical applications, balance calculation performance and accuracy improvement.

5 Digital Signal Processing Basics

5.1 Polarization model dispersion

Basic relationships in vacuum:

$$\lambda f = c, \omega = 2\pi f, k = \frac{2\pi}{\lambda} = \frac{\omega}{c} \tag{15}$$

When a beam of light goes into fiber, its frequency remains the same, wavelength changes, then its velocity

physic	variable	unit
wave length	λ	[m]
frequency	f	Hz
angular frequency	ω	[rads/s]
wave number	k	[rads/m]

Table 6: physical variable

changes.

$$v(\omega) = \frac{c}{n(\omega)}$$

$$\lambda f = v(\omega), \omega = 2\pi f, \beta = \frac{2\pi}{\lambda} = \frac{\omega}{v(\omega)} = \frac{n(\omega)\omega}{c} \tag{16}$$

In real fibers, small departures from cylindrical symmetry, occurring because of random variations in the core shape along the fiber length, result in a mixing of the two polarization states by breaking the mode degeneracy. Mathematically, the mode-propagation constant β becomes slightly different for the modes polarized in the x and y directions. This property is referred to as modal birefringence. The strength of modal birefringence is defined by a dimensionless parameter.

$$B_m = \frac{|\beta_x - \beta_\gamma|}{k_0} = |n_x - n_\gamma|$$

beat length:

$$L_B = \frac{2\pi}{|\beta_x - \beta_y|} = \frac{\lambda}{B_m}$$

5.2 Coherent detection

We use a electronic field to describe light:

$$\mathbf{E}(z,t) = \mathbf{E}_0 e^{i(kz-2\pi ft)} = (\hat{x}E_{0x} + \hat{y}E_{0u})\, e^{i(kz-2\pi ft)}$$

We use a Jones vector to describe a polarization state of light.

We consider Homodyne Receiver

$$E_s(t) = A_s(t) exp(j\omega_s t)$$

5.3 Adaptive filter theory

Some DSP techniques need to learn:

- Coherent receiver
- FDBP (filter DBP)
- FOE (Frequency offset estimater)
- BPM (Batch power normalization)
- MIMO (Multi Input Multi Output)
- CPE (carrier phase estimator)

Problems to derive:

- NLSE – CNLSE

6 Jax Programming

References

[1] Qirui Fan et al. "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning". In: *Nature Communications* 11.1 (2020), pp. 1–11.

[2] Christian Häger and Henry D Pfister. "Physics-based deep learning for fiber-optic communication systems". In: *IEEE Journal on Selected Areas in Communications* 39.1 (2020), pp. 280–294.

[3] Aleksei V. Reznichenko and Ivan S. Terekhov. "Path Integral Approach to Nondispersive Optical Fiber Communication Channel". In: *Entropy* 22.6 (2020). ISSN: 1099-4300. DOI: 10.3390/e22060607. URL: https://www.mdpi.com/1099-4300/22/6/607.