New Perspectives on Perturbation-based Pre-Distortion and Post-Compensation for Nonlinear Optical Transmission

Chuang Xu[1]* and Alan Pak Tao Lau[1]
[1]Photonics Research Institute, Department of Electrical and Electronic Engineering,
The Hong Kong Polytechnic University, Hong Kong SAR, China
*chuang.xu@connect.polyu.hk

Abstract: *We show that addition-multiplication (AM)-model is closest to nonlinear fiber channel among the 1st-order perturbation models and the reverse MA-model is the best pre-distortion and post-compensation method because of equivalence to propagating an inverse fiber.*
Keywords: *Fiber Nonlinearity, Perturbation, Pre-distortion, Compensation*

I. INTRODUCTION

Perturbation (PB) modeling solves the nonlinear Schrödinger equation (NLSE) by perturbative expansion. It provides analytical expressions for nonlinear (NL) distortion and hence valuable insights into different nonlinear effects and their dependence on system parameters, facilitating the design of transmitter (Tx) pre-distortion (PD) or receiver (Rx) post-compensation (PC) techniques to mitigate NL distortion. In PB analysis, the NL distortion is modeled as an additive term that contains a certain combination of information symbol triplets over the transmitted signal. However, the original additive model quickly loses its accuracy as launch power (LP) increases, thus improved models such as additive-multiplicative (AM)-model [1], second-order PB model [2] were studied. Machine learning [3] is also introduced to optimize the PB coefficient for better approximating NLSE as well as multiplicative-additive (MA)-models [4]. Intuitively, once a (forward) PB model is chosen, the PD/ PC algorithm is readily determined as they attempt to reverse the forward model. However, various versions of PD/PC have been shown in the literature usually without explaining why a specific version is chosen, and sometimes the adopted PD/PC method is not identical to (the reverse of) the chosen forward model [5, 6] despite their good transmission performance.

In this paper, we propose using gradient descent (GD) method on the information symbols to iteratively obtain the PD signal, which is possibly the optimal solution of the PD problem. It then serves as a benchmark for comparing different versions of PD/PC and reveals that AM-model is a better approximation to the NLSE channel, while surprisingly PD and PC based on the MA-model also give optimal transmission performance. We explain the phenomenon by showing PD and PC can be seen as backpropagation under the framework of the 1st-order PB of NLSE.

II. PRINCIPLE

The NLSE is given by $jq_z + \beta_2/2 q_{tt} + \gamma|q|^2 q = 0$, where $q = q(z,t)$ is the complex envelope of the signal, β_2 and γ are the dispersion and nonlinearity coefficient. The Tx signal $q(0,t) = \sqrt{P}\sum_k a_k u_k(0,t)$, is a pulse train of basic pulses $u_k(0,t)$ modulated by unit-averaged-power symbols a_k, where $u_k(0,t) = u(0, t-kT)$, T is the symbol duration and P is the averaged LP. By 1st-order PB, the matched-filtered Rx symbol b_k is approximated by a_k plus an additive NL distortion Δa_k, $b_k \approx \sqrt{P}a_k + P\sqrt{P}\Delta a_k$ with $\Delta a_k = \sum_{m,n,p} a_{m+k}a_{n+k}a^*_{p+k}C_{m,n,p}$, where $C_{m,n,p}$ is the PB coefficient given by $C_{m,n,p} = i\gamma\mathcal{E}\int_0^L f(z)\int_{-\infty}^{+\infty} u_0^*(z,t)u_m(z,t)u_n(z,t)u_p^*(z,t)dt$ with signal power profile $f(z)$ and $\mathcal{E}$ being the energy of $u(0,t)$ [7]. To accurately estimate Δa_k, the range of m,n,p should be large enough to include all neighboring symbols that interfere during propagation. For a sinc-shaped basic pulse, its envelope gradually approaches a rectangle with a width of $\pi|\beta_2|z/T$ during propagation [8], so the minimum range for m,n,p is $M = \pi|\beta_2|L/T^2$. It is found that the dominant terms in $C_{m,n,p}$ are those with $p = m+n$, and other terms are usually neglected in practice.

Δa_k consists of phase noise ISPXM and circularly symmetric noise IFWM which we denote by A_k and B_k respectively, i.e., $\Delta a_k = A_k a_k + B_k$. When LP goes high, the accuracy of 1st-order PB quickly declines as it neglects higher-order NL distortion, and AM-model was proposed to take account of higher-order NL contribution. Originally, $b_k = \sqrt{P}a_k + \sqrt{P^3}\Delta a_k \approx \sqrt{P}a_k e^{PA_k} + \sqrt{P^3}B_k$, which we call MA-model hereafter (while it may also be confusingly referred to as AM-model in the literature, e.g. [4]). To possibly include higher-order IFWM, AM-model makes addition first then multiplication, arriving at $b_k \approx \left(\sqrt{P}a_k + \sqrt{P^3}B_k\right)e^{PA_k}$, which turns out to be a better approximation to the true NL distortion. Given an N-length symbols sequence $\sqrt{P}\hat{\boldsymbol{a}} = \sqrt{P}[\hat{a}_1, \hat{a}_2, \dots, \hat{a}_N]$ to be transmitted with average power P, according to AM-model, the NL-distorted sequence at Rx is $\left[\sqrt{P}\hat{\boldsymbol{a}} + \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})\right]e^{P\boldsymbol{A}(\hat{\boldsymbol{a}})}$, where $e^{\boldsymbol{A}(\hat{\boldsymbol{a}})}$ is shorthand for

This work was supported by the Hong Kong Government Research Grants Council General Research Fund (GRF) under Project PolyU 15220120 and PolyU 15225423.

taking exponential to each element of $\boldsymbol{A}(\hat{\boldsymbol{a}})$, $\boldsymbol{A}(\hat{\boldsymbol{a}})$ (pure imaginary) and $\boldsymbol{B}(\hat{\boldsymbol{a}})$ are vectors denoting the ISXPM and IFWM distortion generated by $\hat{\boldsymbol{a}}$, respectively. Intuitively, the PD sequence should be given by $\sqrt{P}\hat{\boldsymbol{a}}e^{-\boldsymbol{PA}(\hat{\boldsymbol{a}})} - \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})$, i.e., the reverse operation of AM-model layer by layer, denoted by PD-AM. However, we noticed that actually $\left[\sqrt{P}\hat{\boldsymbol{a}} - \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})\right]e^{-\boldsymbol{PA}(\hat{\boldsymbol{a}})}$ works much better, which is not obvious at first glance. We denote it by PD-MA as it is the reverse operation of MA-model. Different PD and PC methods corresponding to their PB models are summarized in Table.1 for convenient comparison hereafter.

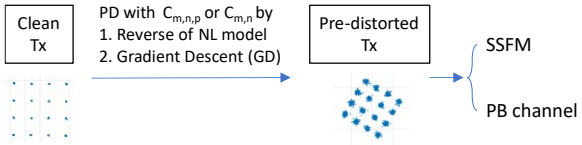

Table. 1. PB models and corresponding PD, PC. $\boldsymbol{b}$, $\boldsymbol{d}$ denotes Rx and its decision.

PB Model		Pre-Distortion			Post-Compensation (by $\boldsymbol{d}$ or $\boldsymbol{b}$)	
Additive	$\sqrt{P}\hat{\boldsymbol{a}} + \sqrt{P^3}\Delta\hat{\boldsymbol{a}}$	PD-A	$\sqrt{P}\hat{\boldsymbol{a}} - \sqrt{P^3}\Delta\hat{\boldsymbol{a}}$	GD-A	PC-A	$\sqrt{P}\boldsymbol{b} - \sqrt{P^3}\Delta\boldsymbol{d}$
MA	$\sqrt{P}\hat{\boldsymbol{a}}e^{\boldsymbol{PA}(\hat{\boldsymbol{a}})} + \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})$	PD-MA	$\left[\sqrt{P}\hat{\boldsymbol{a}} - \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})\right]e^{-\boldsymbol{PA}(\hat{\boldsymbol{a}})}$	GD-MA	PC-MA	$\left[\sqrt{P}\boldsymbol{b} - \sqrt{P^3}\boldsymbol{B}(\boldsymbol{d})\right]e^{-\boldsymbol{PA}(\boldsymbol{d})}$
AM	$\left[\sqrt{P}\hat{\boldsymbol{a}} + \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})\right]e^{\boldsymbol{PA}(\hat{\boldsymbol{a}})}$	PD-AM	$\sqrt{P}\hat{\boldsymbol{a}}e^{-\boldsymbol{PA}(\hat{\boldsymbol{a}})} - \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})$	GD-AM	PC-AM	$\sqrt{P}\boldsymbol{b}e^{-\boldsymbol{PA}(\boldsymbol{d})} - \sqrt{P^3}\boldsymbol{B}(\boldsymbol{d})$

Fig. 1. Different PD signals are generated and fed into SSFM or PB channel for performance comparison.

An ideal PD for AM-model can be interpreted as to find a sequence $\boldsymbol{a} = [a_1, a_2, \dots, a_N]$ that (approximately) solves the equation $\left[\sqrt{P}\boldsymbol{a} + \sqrt{P^3}\boldsymbol{B}(\boldsymbol{a})\right]e^{\boldsymbol{PA}(\boldsymbol{a})} = \sqrt{P_0}\hat{\boldsymbol{a}}$. We propose iteratively solving it by gradient descent method with a cost function of $J(\boldsymbol{a}) = \left\|\left[\sqrt{P}\boldsymbol{a} + \sqrt{P^3}\boldsymbol{B}(\boldsymbol{a})\right]e^{\boldsymbol{PA}(\boldsymbol{a})} - \sqrt{P_0}\hat{\boldsymbol{a}}\right\|^2$, where P is alterable during iteration, and P_0 is the fixed designed LP. The complex gradient [9] of $J(\boldsymbol{a})$ is derived to be $\nabla J(\boldsymbol{a}) = [\nabla_1, \nabla_2, \dots, \nabla_N]$, where $\nabla_k = \frac{\partial J(\boldsymbol{a})}{\partial a_k^*} = Pa_k + P^2\left(B_k + \frac{\partial B_k}{\partial a_k^*}a_k^* + \frac{\partial B_k^*}{\partial a_k^*}a_k\right) + P^3\left(\frac{\partial B_k}{\partial a_k^*}a_k^* + \frac{\partial B_k^*}{\partial a_k^*}a_k\right) - \sqrt{PP_0}\hat{a}_k e^{-PA_k}\left(1 - P\frac{\partial A_k}{\partial a_k^*}a_k^* + P\left(\frac{\partial B_k}{\partial a_k^*} - P\frac{\partial A_k}{\partial a_k^*}B_k^*\right)\right) - \sqrt{PP_0}\hat{a}_k^* e^{PA_k}\left(P\frac{\partial A_k}{\partial a_k^*}a_k + P\left(\frac{\partial B_k}{\partial a_k^*} + P\frac{\partial A_k}{\partial a_k^*}B_k\right)\right)$, where $A_k = 2\sum_n a_{n+k}a_{n+k}^* C_{0,n,n}$, $B_k = \sum_{(m,n,p)\notin S} a_{m+k}a_{n+k}a_{p+k}^* C_{m,n,p}$ with $S = \{m,n,p: (m=0, n=p)\ or\ (n=0, m=p)\}$, and $\frac{\partial A_k}{\partial a_k^*} = 2a_k C_{0,0,0}$, $\frac{\partial B_k}{\partial a_k^*} = \sum_{(m,n,0)\notin S} a_{m+k}a_{n+k}C_{m,n,0}$, $\frac{\partial B_k^*}{\partial a_k^*} = 2\sum_{(m,0,p)\notin S} a_{m+k}a_{p+k}^* C_{m,0,p}$. The GD algorithm updates a_k by $a_k(n+1) = a_k(n) - \mu\nabla_k(n)$, with learning step μ. A contra-rotated $\hat{\boldsymbol{a}}$ by amount of ISXPM may be used as an initial sequence to accelerate the convergence. After GD iteration, the Rx signal can be quite clean, the corresponding PD Tx signal is denoted by GD-AM. Similarly, gradient for the additive and MA-model can also be derived and GD-A, GD-MA signal can be obtained.

III. Simulation Results

We conducted split-step-Fourier-method (SSFM) simulations on an 8×100 km EDFA system with fiber loss of 0.2dB/km, and EDFA NF of 5dB. 1024 30Gbaud 16QAM-modulated single-polarization symbols are pre-distorted by the GD method with $C_{m,n,p}$ and $C_{m,n}$, respectively, and zero roll-off sinc waveform was used. PD-A, PD-AM, and PD-MA signals are also generated for comparison. With those system configurations, the minimum range of m, n, p is evaluated to be 46.3, so we restricted $m, n, p \in [-50, 50]$. After ideal CD compensation and matched filtering, the SNR of the received signal is evaluated by $1/EVM_{RMS}^2$ where $\mathrm{EVM_{RMS}}$ stands for the root mean squared error vector magnitude. To verify the effectiveness of GD, we fed those pre-distorted signals into the PB channel again and compared the new Rx SNR. As shown in Fig. 2(a), the SNR of conventional PD schemes all significantly decreases as LP increases, while all GD signals return a flat SNR curve, indicating GD method solves the nonlinear PB equation quite well across different LP. Note that the power of GD-A is much lower than the designed ones at high LP because the additive model generates NL distortion with excess power, which is considered by the GD algorithm supporting alterable signal power. Distributions of the Rx signal of GD-AM vs PD-AM are shown in the insets.

Fig. 2(b) shows the results for different PD techniques in noiseless SSFM propagation, and their SNR all drops due to the inherent modeling error between PB and NLSE while the GD-AM drops the least. A successful PD method requires first a good enough model of NLSE and also a good, if not perfect, solution corresponding to the model. Thus, AM is the best approximation to NLSE among the three PB models, and GD-AM works well for such an approximation. It also gives the best SNR performance in SSMF with ASE loading as shown in Fig. 2(c). Compared with PD-AM, the peak SNR of GD-AM with $C_{m,n,p}$ is improved by 1.1dB, while that for $C_{m,n}$ case reduces to 0.4dB. Unexpectedly, PD-MA performs very closely to GD-AM that is supposedly the optimal solution, and even their waveforms are very close to each other (Fig. 2(d)), but PD-MA is not the reverse operation of AM-model. We try to explain why GD-AM and PD-MA essentially arrive at the same pre-distortion solution for AM channel through different approaches in the next section.

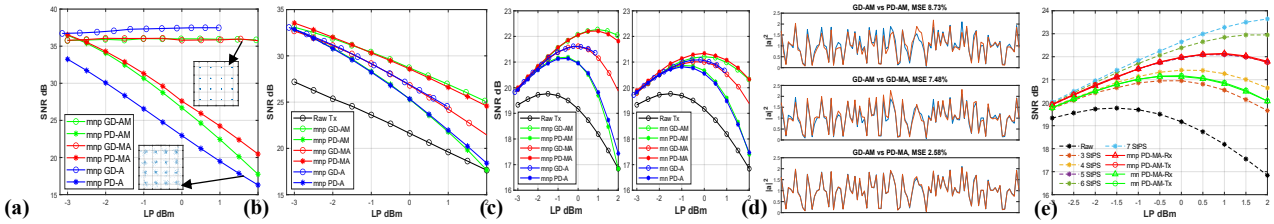

Fig. 2. Performance of raw Tx signal and different PD signals run through (a) PB channel; (b) SSFM ASE-free; (c) SSFM ASE-loading; (d) Waveform comparison between GD-AM and PD-MA; (e) Comparison between PC-MA-Rx and DBP of different steps per span (StSP).

IV. DISCUSSION

Originally, we applied GD method on PD to seek a possible better-performing PD solution, but it turns out that the existing PD-MA is quite good enough though with huge simplicity but mismatched to the channel model, then the GD solution may be saved for an upper bound and the problem becomes why it is the case. We find it can be explained by interpreting PD as (pre-) backpropagation. Consider an inverse bulk CD compensation of $\beta_2 L$ followed by a virtual fiber link with opposite parameters to the real one as shown in Fig. 3(a). By PB analysis, the perturbative NL distortion of the reverse transmission Δa_k^r is derived (shown in Fig. 3) to be *negative* to that of a real transmission, where $u^0(z,t)$ and $\underline{u}^0(z,t)$ are the 0th-order solution for the real and reverse fiber link, respectively, $D_z^{\beta_2}$ is the dispersion operator that disperses a signal through a fiber for distance z with dispersion coefficient β_2, and $\tilde{f}(z)$ is the power profile for the reverse fiber link. Given that the forward transmission is well approximated by AM-model, a clean Tx symbol sequence $\sqrt{P}\hat{\boldsymbol{a}}$ will be distorted into $\left[\sqrt{P}\hat{\boldsymbol{a}} - \sqrt{P^3}\boldsymbol{B}(\hat{\boldsymbol{a}})\right]e^{-P\boldsymbol{A}(\hat{\boldsymbol{a}})}$ after a reverse transmission, which is exactly the PD-MA signal in form. Thus, PD-MA for the real transmission may be interpreted as a two-stage process: a clean signal goes through the reverse and real transmission in turn, which reasonably generates less NL distortion overall. In this two-stage process, PD-Tx signal effectively serves as a 'middle-stage' signal. To generalize this idea, alternatively, we can move the reverse transmission to the end of a real transmission, and then it corresponds to the PC case, where the *distorted* Rx signal should serve as the 'middle-stage' signal. Thus, counter-intuitively, the *distorted* Rx signal can be directly used for PC by reversely operating MA-model, denoted by PC-MA-Rx besides the convention of using decision of Rx to reconstruct NL distortion and eliminating the NL distortion from Rx signal in the PC-AM manner. This conjecture is verified by the result in Fig. 3(c) (d), with PB coefficient of $C_{m,n,p}$, the performance of PC-MA-Rx is nearly the same as PC-AM-Tx (which is the ideal PC) across the whole LP range, and this still holds in a more practical case when only $C_{m,n}$ is used, indicating a practical significance that the feedback structure of feeding Rx decision to PC may be removed.

Moreover, in this forward-backward configuration, the cascaded CD and −CD compensation cancel out each other, and then the reverse transmission is nothing but an ideal digital backpropagation (DBP) [10] in the sense of AM-model, which originates from 1st-order PB but includes somewhat higher-order effects by the exponential modification. Thus, it is expected that PC-MA-Rx can approach similar performance to DBP to a certain extent. Fig. 2 (d) shows that with $C_{m,n,p}$ and $C_{m,n}$ used, PD-MA-Rx performs closely to DBP of 5 StSP and 3 StSP, respectively.

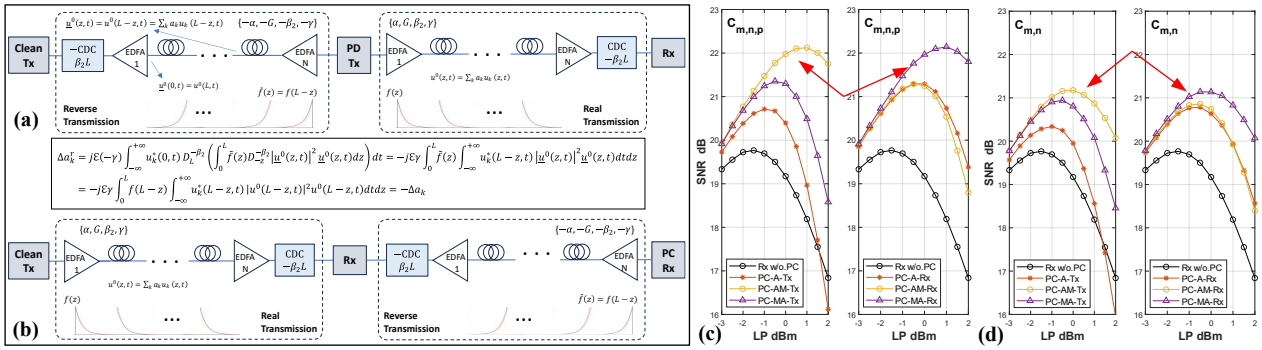

Fig. 3 (a) Schematic of interpreting PD and PC as backpropagation with inverse CD compensation; (b) Performance of different PC schemes.

V. CONCLUSIONS

We used GD method to iteratively obtain PD signal based on perturbation analysis of fiber NL distortion. Different PB models and corresponding PD schemes are compared. As a benchmark, GD-AM reveals that PD-MA is surprisingly also a very good but simple solution though mismatched with the AM-model. To explain this phenomenon, we propose new theoretical understanding of PD by treating it as backpropagation in the framework of perturbative NL analysis, which is then further generalized to the PC case and compared with DBP. The result justifies the AM-model and unexplained adoption of PD-MA in the literature.

REFERENCES

[1] Z. Tao, Y. Zhao, Y. Fan, L. Dou, T. Hoshida and J. C. Rasmussen, "Analytical Intrachannel Nonlinear Models to Predict the Nonlinear Noise Waveform," in Journal of Lightwave Technology, vol. 33, no. 10, pp. 2111-2119, 2015.

[2] S. K. Orappanpara Soman, A. Amari, O. A. Dobre and R. Venkatesan, "Second-Order Perturbation Theory-Based Digital Predistortion for Fiber Nonlinearity Compensation," in Journal of Lightwave Technology, vol. 39, no. 17, pp. 5474-5485, 2021.

[3] A. Barreiro, G. Liga and A. Alvarado, "Data-Driven Enhancement of the Time-Domain First-Order Regular Perturbation Model," in Journal of Lightwave Technology, vol. 41, no. 9, pp. 2691-2706, 2023.

[4] Liang, Xiaojun, and Shiva Kumar. "Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links," in Optics Express, vol.22, no. 24, pp. 29733-29745, 2014.

[5] S. Tharranetharan, S. K. O. Soman and L. Lampe, "Joint Fiber Nonlinearity Mitigation and Compensation for Digital Sub-Carrier Multiplexing System," in IEEE Photonics Journal, vol. 16, no. 4, pp. 1-17, 2024.
[6] Gao, Ying, et al. "Reducing the complexity of perturbation-based nonlinearity pre-compensation using symmetric EDC and pulse shaping," In Optics Express, vol. 22, no. 2, pp. 1209-1219, 2014.
[7] A. Ghazisaeidi, "A Theory of Nonlinear Interactions Between Signal and Amplified Spontaneous Emission Noise in Coherent Wavelength Division Multiplexed Systems," in Journal of Lightwave Technology, vol. 35, no. 23, pp. 5150-5175, 2017.
[8] J. Azana, N. K. Berger, B. Levit and B. Fischer, "Spectro-temporal imaging of optical pulses with a single time lens," in IEEE Photonics Technology Letters, vol. 16, no. 3, pp. 882-884, 2004
[9] S. S. Haykin, Adaptive filter theory, 5th edition. Pearson, 2014.
[10] E. Ip and J. M. Kahn, "Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation," in Journal of Lightwave Technology, vol. 26, no. 20, pp. 3416-3425, 2008.