 doi:10.20944/preprints201909.0018.v1

Article

A Blind Nonlinearity Compensator Using DBSCAN Clustering for Coherent Optical Transmission Systems

Elias Giacoumidis [1,*], Yi Lin [1], Mutsam Jarajreh [2], Sean O'Duill [1], Kevin McGuinness [3], Paul F. Whelan [4], and Liam P. Barry [1]

[1] Dublin City University, School of Electronic Engineering, Radio and Optical Laboratory, Glasnevin 9, Dublin, Ireland.
[2] Fahad Bin Sultan University, Tabuk, Saudi Arabia, Computer Engineering Department.
[3] Insight Centre for Data Analytics, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland.
[4] Vision Systems Group, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland.

* Correspondence: elias.giacoumidis@dcu.ie; Tel.: +353-1-700-8598

Abstract: We experimentally demonstrate, for the first time, blind nonlinearity compensation using the Density-Based Spatial-Clustering of Applications with Noise (DBSCAN) algorithm for a 40-Gb/s 16-quadrature amplitude modulated being transmitted at 50 km of standard single-mode fiber. At high launched optical powers, DBSCAN offers up to 0.83- and 8.84-dB enhancement in Q-factor when compared to conventional K-means clustering and linear equalization, respectively.

Keywords: fiber optics communications; coherent communications; machine learning; clustering; nonlinearity cancellation

---

1. Introduction

Coherent optical communications have been proposed as a viable solution for maximizing the signal capacity in both short-reach and long-haul communications. However, Kerr-induced fiber nonlinearity prevents channel capacity from approaching the Shannon limit, especially when the signal power is high. Endeavours to surpass the Kerr nonlinearity limit have been performed by techniques that in principle compensate deterministic nonlinearities. For example, nonlinearities can be combated by either inserting an optical phase conjugator (OPC) at the middle point of the link [1], or by inverting the fiber effects among multiple frequency stabilized optical signals [2]. However, OPC reduces the flexibility in an optically routed network, whereas in [2], a digital back propagation (DBP) [3] pre-compensator is used which is of excessive complexity. Other famous techniques include hybrid pre- and post-compensation [4], Volterra-based nonlinear equalization (NLE) [5], phase-conjugated twin-waves (PC-TW) [6], and the nonlinear Fourier transform (NFT) [7]. Unfortunately, pre-/post-compensation algorithms and Volterra-NLE present marginal performance enhancement, PC-TW sacrifices signal capacity and NFT is unpractical for real-time signal processing. Above all, the aforementioned deterministic methods are unable to tackle stochastic nonlinearities such as the amplified spontaneous emission (ASE) noise induced from optical amplifiers.

Unsupervised machine learning clustering has been recently introduced in optical communications for blind (training-data-free) nonlinear equalization (BNLE). Such unsupervised algorithms can tackle stochastic nonlinearities and include for example fuzzy-logic C-means [8], K-means [8, 9], hierarchical [8], affinity propagation [10] and Gaussian mixture [11] clustering. However, the received constellation diagrams in [8-11] involve Gaussian-circular clusters of symbols, and hence there is an uncertainty if machine learning clustering can be effective for non-circular rotated clusters, as a direct result from very strong nonlinear phase noise.

In this work, we address the aforementioned issue by experimentally demonstrating the first BNLE that harnesses the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [12] algorithm in 40-Gb/s 16-QAM coherent optical signals being transmitted at 50 km. As a proof-of-concept, DBSCAN is tested for very-high launched optical powers (LOPs), where the clusters of the received constellation diagrams are vastly rotated by means of self-phase modulation (SPM). Two novel modified DBSCAN methods are also proposed, in which the "un-clustered" noisy points are further processed using (1) K-means, and (2) the minimum distance between an unlabelled point and the clustered points. We show that DBSCAN offers up to 0.83-dB Q-factor improvement over K-means and 8.84-dB when compared to linear equalization at +16 dBm of LOP. This occurs because DBSCAN can effectively recover non-circularly-symmetric (elliptical form) noisy clusters by effectively combating SPM.

2. DBSCAN description

In density-based clustering we make an assumption that clusters are densed regions in space, separated by regions of lower density [12]. A dense cluster is a region which is "density connected", i.e. the density of points in that region is greater than a minimum [13]. DBSCAN is an example that searches for dense areas and expands these recursively to find arbitrarily densed-shaped clusters. The two main parameters of DBSCAN are the ε ('Epsilon') and the 'minimum points'. The ε defines the radius of the "neighbourhood region" while the 'minimum points' define the minimum number of constellation points (i.e. symbols) that should be contained within that neighbourhood. DBSCAN arbitrarily pics-up a point until all of them have been visited. If the predefined number of 'minimum points' is within the radius-ε, then we consider all these points to be part of the same cluster. The clusters are then expanded by recursively repeating the neighbourhood calculation for each neighbouring point. However, for the unallocated points, if the number of points within the ε-neighbourhood is less than a predefined threshold, they designated to be "noisy" and not assigned to a particular cluster. Noisy data are not further processed in conventional DBSCAN. Here, we propose to apply a 2nd loop clustering only for these noisy data using: (1) K-means [8], or (2) the minimum distance between an unlabelled point and the clustered points. A schematic diagram for conventional DBSCAN is depicted in Fig. 1 when the minimum points is 4. In Fig. 1 we assume the following assumptions [14]:

a. **Epsilon neighbourhood (Nε):** A set of all constellation points within a distance 'ε'.
b. **Core point:** A constellation point whose Nε contains at least a 'minimum point' (including itself).
c. **Direct Density Reachable:** A point q is directly density reachable from a point p, if p is core point and $q \in$ Nε.
d. **Density Reachable:** Two constellation points are density reachable if there is a chain of 'direct density reachable' points that link these two points.
e. **Border Point:** A constellation point that is 'direct density reachable' but not a core point.

Noise: Constellation points not belonging to any point's Nε.

The steps related to the conventional and modified DBSCAN are listed below, where the algorithm converges until all constellation points have been allocated to a cluster or labelled as 'noisy' only if conventional DBSCAN is considered (step 5 below – 1st loop) [14, 15]:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 September 2019 doi:10.20944/preprints201909.0018.v1

1. Randomly select a point *p* (referred in Fig. 1) in the constellation map.
2. Retrieve all constellation points directly density-reachable from *p* that satisfy the condition of the radius ε limits.
3. If the constellation point *p* is a core point, a cluster is formed. Search recursively and find all of its density connected points and assign them to the same cluster as *p*.
4. If *p* is not a core point, the DBSCAN algorithm "scans" for the rest unvisited constellation points.
5. **DBSCAN 1st loop:** Points that are un-clustered are labelled as zero points ("noisy points") where linear equalization is performed only on these points; and then the conventional DBSCAN algorithm stops.
6. **DBSCAN 2nd loop (extra novel step):**

i. ***Method-1:*** K-means clustering is activated for the "noisy points" using the Lloyd's algorithm [8, 9]:

a. Assignment: Allocate each observation to the cluster whose mean has the least squared Euclidean distance ("nearest" mean) [8, 9].

b. Update: Calculate the new means to be the centroids of the observations in the new clusters [8, 9]. K-means converges when assignments do not change.

ii. ***Method-2:*** Calculation of minimum distance between the unlabelled "noisy points" and the clustered points.

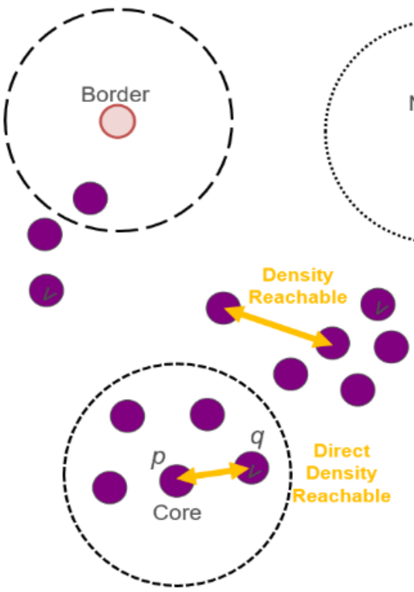


Figure 1. DBSCAN example for Min. Points = 4.

3. Experimental Setup

Fig. 2 depicts the schematic diagram of the experimental setup of the 10 GBaud (40 Gb/s) 16-QAM coherent signal. In the transmitter-DSP, look-up-table-based pre-distortion was used to mitigate the opto-electronic components impairments similarly to [16]. A narrow linewidth (<100 kHz) external cavity laser (ECL) was tuned to 1549.5 nm and using an arbitrary waveform generator (AWG) operating at 20 GS/s, two uncorrelated pseudo-random level signals (2^{15}-1) were applied to the IQ modulator to generate the 16-QAM signal. After IQ modulation the optical signal was transmitted over 50 km of standard single-mode fiber (SSMF). At the receiver, noise loading was added using an optical amplifier to set different optical signal-to-noise ratio (OSNR) values and subsequently the optical signal was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 September 2019 doi:10.20944/preprints201909.0018.v1

converted to an electrical one using a homodyne coherent receiver. Afterwards, the signal was captured by a real-time oscilloscope sampled at 50 GS/s for offline receiver-DSP, in which the data was first resampled to 2 samples/point using a priori knowledge of the clock frequency. Then the constant modulus algorithm (CMA) combined with multi-modulus algorithm (MMA) was utilized for signal equalization. An Mt power frequency drifting compensation method was employed to compensate the frequency offset between the signal and the local oscillator in the coherent receiver. The decision-directed phase-locked loop (DDPLL) method was employed for the carrier phase recovery. Finally, machine learning was processed before hard decision and bit-error-rate (BER)/Q-factor (=$20\log_{10}[\sqrt{2}erfc^{-1}(2BER)]$) calculation similarly to other reported work with machine learning signal processing [17-22].

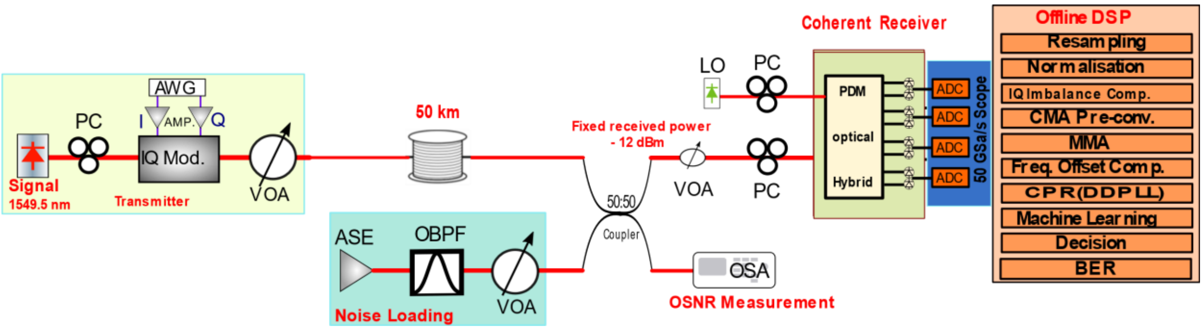


Figure 2. Experimental setup for a 40-Gb/s 16-QAM coherent optical signal transmitted at 50 km, incorporating machine learning clustering. PC: polarization controller, OBPF: optical band-pass filter, LO: local oscillator, CMA/MMA: constant/multi-modulus algorithm, CPR (DDPLL): carrier phase recovery (decision-directed phase-locked loop).

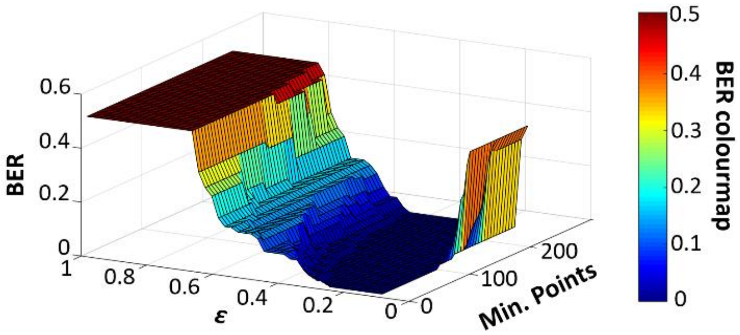


Figure 3. DBSCAN optimization for 16-QAM transmission over 50 km at +16 dBm of launch power: BER vs. ε, Min. Points.

4. Results

We transmit our 40-Gbit/s 16-QAM waveform with +16 dBm LOP over 50 km. Two parameters are needed to optimize the DBSCAN algorithm to produce the lowest BER, namely ε and the minimum number of points. The calculated BER while scanning for ε and the minimum number of points is shown in Fig. 3. The lowest BER can be found for 0.45 < ε < 0.1 and when the minimum points are less than 120. In Fig. 4, the performance of clustering algorithms is shown for different LOPs and two values of received OSNR: 30 and 15 dB. In Fig. 4, the performance benefit of machine learning clustering over linear equalization is significant for both OSNR values, especially when using DBSCAN method-2 resulting in up to 8.8-dB Q-factor improvement. This is attributed to the compensation of SPM since single-channel transmission is carried out. Results indicate that DBSCAN-based BNLE is a robust soft-clustering method when very strong nonlinear phase noise is present and where linear equalization fails completely. Moreover, DBSCAN method-2 has the highest Q-factor along the whole range of LOPs. Compared to DBSCAN method-1 and K-means, method-2 increases the Q-factor by up to about 0.7 and 0.83-dB, respectively, by better handling highly rotated clusters that become almost elliptically shaped. This is because the overlapping (soft) clustering ability of method-2 is more powerful than the common hard (exclusive) clustering of K-means and method-2 (which also includes K-means for the noisy constellation points). This is confirmed by the received constellation diagrams of Fig. 5(b) related to +16 dBm of LOP (OSNR=30 dB). On the other hand, the quite similar performances between DBSCAN algorithms and K-means at lower LOPs and OSNR is due to the existence of nearly circular-Gaussian clusters which are not rotated. This is corroborated in the received 16-QAM constellation diagrams of Fig. 5(a) at +10 dBm of LOP (OSNR=15 dB). In the left constellation diagram of Fig. 5(a), the DBSCAN "noisy" points are also presented in the 1st loop of the algorithm.

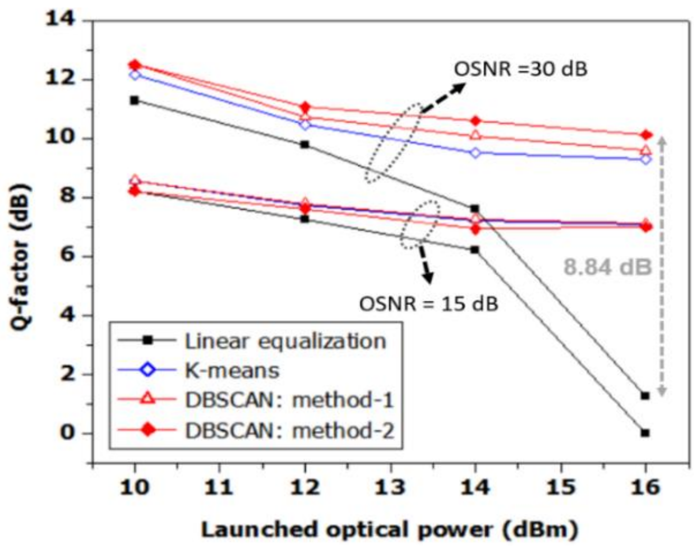
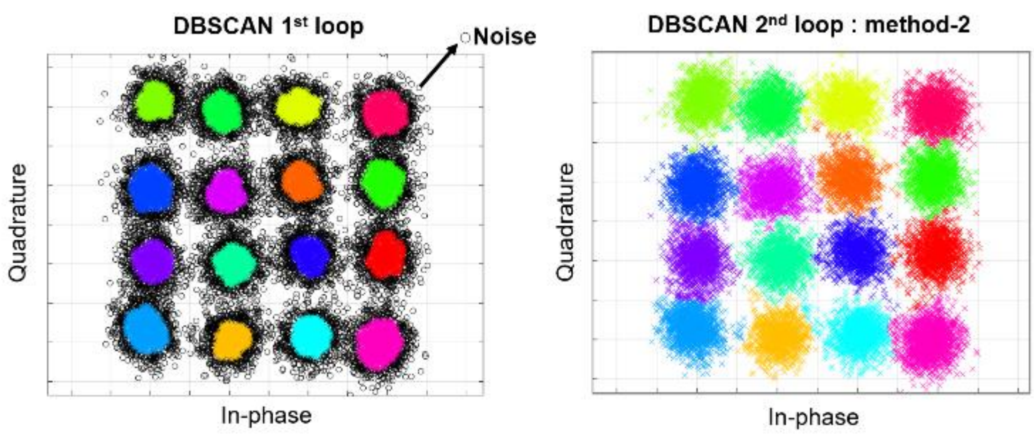
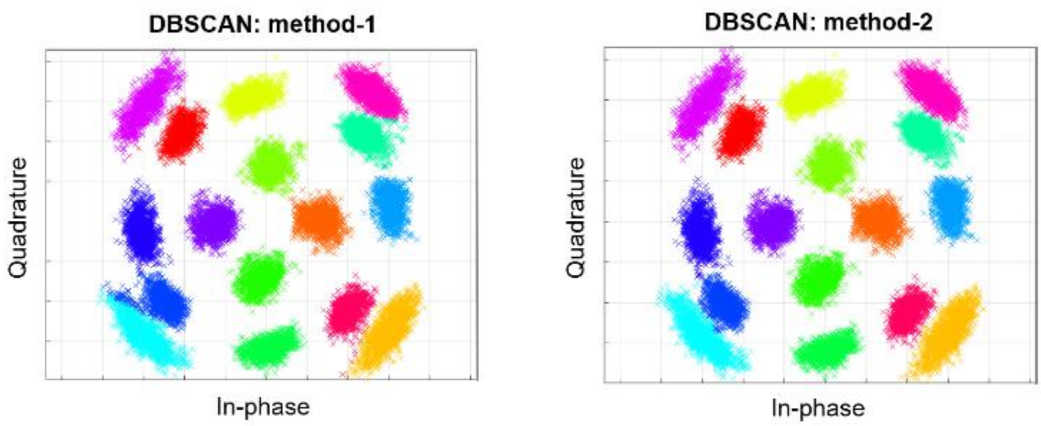


Figure 4. DBSCAN vs. K-means for 16-QAM transmission at 50 km for different launched optical powers (LOPs) when OSNR is 30, 15 dB.



(a)



(b)

Figure 5. Received constellation diagrams for (a) DBSCAN 1st loop (left), method-2/2nd loop (right) at +10 dBm of LOP [OSNR=15 dB]; and (b) DBSCAN method-1 (left), method-2 (right) at +16 dBm of LOP (OSNR=30 dB).

5. Conclusion

We experimentally demonstrated the first DBSCAN-BNLE for 16-QAM at 50 km. Two novel DBSCAN methods were proposed, in which the "un-clustered" noisy constellation points were processed using (1) K-means, and (2) the minimum distance between an unlabelled point and the clustered points. Compared to linear equalization, method-2 improved the Q-factor up to 8.8-dB by combating SPM. Method-2 ability for overlapping clustering resulted in Q-factor improvement over method-1 and K-means (exclusive clustering), when vastly rotated clusters of nearly elliptical form occur. Once optimized, DBSCAN proved to be a robust BNLE for very strong nonlinear phase noise.

The complexity of DBSCAN is $O(n^2)$, where n is the number of points. Detailed complexity analysis will be reported in future work.

Author Contributions: E.G, Y.L., M.J., S.O., K.M., P.F.W. and L.P.B. wrote the paper. E.G. and Y.L equally contributed to this work.

Funding: This work was supported by the Science Foundation Ireland through grant numbers 13/RC/2077, 12/RC/2276, 15/US-C2C/I3132, the HEA INSPIRE Programme, and the EU/EDGE Marie-Curie programme with grant number 713567.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Al-Khateeb, M. A. Z.; McCarthy, M. E.; Sánchez, C.; Ellis, A. D. Nonlinearity compensation using optical phase conjugation deployed in discretely amplified transmission systems. *Opt. Exp*. **2018**, *26*, 23945-23959.
2. Temprana, E. et al. Overcoming Kerr-induced capacity limit in optical fiber transmission. *Science* **2015**, *348*, 1445-1448.
3. Maher, R. et al. Linear and nonlinear impairment mitigation in a Nyquist spaced DP-16QAM WDM transmission system with full-field DBP. In Proc. ECOC, Cannes, France (IEEE, **2014**) P.5.10.
4. Lowery, A. J. Fiber nonlinearity pre- and post-compensation for long-haul optical links using OFDM. *Opt. Exp*. **2007**, *15*, 12965-12970.
5. Giacoumidis, E. et al. Volterra-based reconfigurable nonlinear equalizer for dual-polarization multiband coherent OFDM. *IEEE Photon. Technol. Lett*. **2014**, *26*, 1383-1386.

 doi:10.20944/preprints201909.0018.v1

6. Liu, X. Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit. *Nature Photon*. **2013**, *7*, 560-568.
7. Le, S. T.; Aref, V.; Buelow, H. Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit. *Nature Photon*. **2017**, *11*, 570-576.
8. Giacoumidis, E. Blind Nonlinearity Equalization by Machine Learning based Clustering for Single- and Multi-Channel Coherent Optical OFDM. *IEEE J. Lightw. Techn*. **2018**, *36*, 721-727.
9. Zhang, J.; Chen, W.; Gao, M.; Shen, G. K-means-clustering-based fiber nonlinearity equalization techniques for 64-QAM coherent optical communication system. *Opt. Exp*. **2017**, *25*, 27570-27580.
10. Giacoumidis, E. Affinity propagation clustering for blind nonlinearity compensation in coherent optical OFDM. In Proc. CLEO, San Jose, CA, USA (OSA, **2018**) p. STh1C.5.
11. Zibar, D.; Piels, M.; Jones, R.; Schäeffer, C. G. Machine Learning Techniques in Optical Communication. *IEEE J. Lightw. Techn*. **2016**, *34*, 1442-1452.
12. Ester, M.; Kriegel, H. P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proc. International Conf. on Knowl. Discov. and Data Mining, Portland, Oregon, USA, 1996, 226-231.
13. Shen, J. Real-Time Superpixel Segmentation by DBSCAN Clustering Algorithm', *IEEE Trans. on Image Proc*., **2016**, *25*, 5933-5942.
14. Boada, R.; Borkowski, R.; Monroy, I. T. Clustering algorithms for Stokes space modulation format recognition. *Opt. Exp*. **2015**, *23*, 15521-15531.
15. Lu, X. An I-Q-Time 3-dimensional post-equalization algorithm based on DBSCAN of machine learning in CAP VLC system. *Opt. Comms.* **2019**, *430*, 299-303.
16. Zhang, J.; Yu, J.; Chien, H.-C. Advanced linear and nonlinear compensations for 16QAM SC-400G unrepeatered transmission system. *Opt. Comms.* **2018**, *409*, 34-38.
17. Nguyen, T. et al. Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM. *Photon. J.* **2016**, *8*.
18. Giacoumidis, E. et al. Comparison of DSP-based nonlinear equalizers for intra-channel nonlinearity compensation in coherent optical OFDM. *Opt. Lett.* **2016**, *41*, 2509-2512.
19. Giacoumidis E. et al. Reduction of Nonlinear Inter-Subcarrier Intermixing in Coherent Optical OFDM by a Fast Newton-based Support Vector Machine Nonlinear Equalizer. *J. of Lightw. Techn.* **2017**, *35*, 2391-2397.
20. Giacoumidis, E. et al. Nonlinear Blind Equalization for 16-QAM Coherent Optical OFDM using Support Vector Machines. In Proc. ECOC, Düsseldorf, Germany, (IEEE, **2016**) p. Th.2.P2.
21. Mhatli, S. et al. A novel SVM robust model Based Electrical Equalizer for CO-OFDM Systems. *IET Commun*. **2017**, *11*, 1091-1096.
22. Giacoumidis E.; Tsokanos, A.; Ghanbarisabagh, M.; Mhatli, S.; Barry, L. P. Unsupervised Support Vector Machines for Nonlinear Blind Equalization in CO-OFDM. *Phot. Tech. Lett.*, **2018**, *30*, 1091-1094.