

# Blind Nonlinearity Equalization by Machine Learning based Clustering for Single- and Multi-Channel Coherent Optical OFDM

Elias Giacomidis, *Member IEEE, Member OSA*, Amir Matin, Jinlong Wei, *Member IEEE, Member OSA*, Nick J. Doran, *Senior Member IEEE, Senior Member OSA*, Liam P. Barry, *Senior Member IEEE, Senior Member OSA* and Xu Wang, *Senior Member IEEE, Senior Member OSA*

**Abstract**—Fiber-induced intra- and inter-channel nonlinearities are experimentally tackled using blind nonlinear equalization (NLE) by unsupervised machine learning based clustering (MLC) in ~46-Gb/s single-channel and ~20-Gb/s (middle-channel) multi-channel coherent multi-carrier signals (OFDM-based). To that end we introduce, for the first time, Hierarchical and Fuzzy-Logic C-means (FLC) based clustering in optical communications. It is shown that among the two proposed MLC algorithms, FLC reveals the highest performance at optimum launched optical powers (LOPs), while at very high LOPs Hierarchical can compensate more effectively nonlinearities only for low-level modulation formats. FLC also outperforms K-means, fast-Newton support vector machines, supervised artificial neural networks and a NLE with deterministic Volterra analysis, when employing BPSK and QPSK. In particular, for the middle channel of a QPSK WDM coherent optical OFDM system at optimum -5 dBm of LOP and 3200 km of transmission, FLC outperforms Volterra-NLE by 2.5 dB in Q-factor. However, for a 16-quadrature amplitude modulated single-channel system at 2000 km, the performance benefit of FLC over IVSTF reduces to ~0.4 dB at a LOP of 2 dBm (optimum). Even when using novel sophisticated clustering designs in 16 clusters, no more than additional ~0.3 dB Q-factor enhancement is observed. Finally, in contrast to the deterministic Volterra-NLE, MLC algorithms can partially tackle the stochastic *parametric noise amplification*.

**Index Terms**—Machine learning, clustering, coherent detection, nonlinearity mitigation, coherent optical OFDM.

## I. INTRODUCTION

Current optical networks are based on standard single-mode fiber (SSMF) cables as opposed to more futuristic, high-

capacity multiple spatial modes (few-mode) fibers. Employing few-mode could form the most plausible alternative towards the desirable bandwidth capacity increase [1, 2]. The Kerr effect is a nonlinear phenomenon which causes distortion to the propagated optical signal and it is proportional to its power [1, 2], resulting in the deceleration of the data transmission. Attempts to combat fiber-induced nonlinearities have been performed by nonlinearity compensators [3–5] which tackle deterministic nonlinearities. These techniques however, result in modest improvements because the interaction between nonlinearity and random noises in a long-distance network such as from concatenated erbium doped fiber amplifiers (EDFAs) (i.e. the parametric noise amplification phenomenon [6]) add significant stochastic nonlinear distortion. Especially at low transmitted powers, the received data reveal more entropy meaning they have higher randomness due to EDFAs non-deterministic noise. Moreover, all proposed nonlinearity compensators present high complexity [3–5] being impractical for real-time communications. The aforementioned random noises of the network can be partially tackled by digital machine learning algorithms that perform nonlinear equalization (NLE), such as unsupervised and supervised algorithms: machine learning clustering (MLC) using K-means and Gaussian mixture [7–9], and classification machines [10], e.g. artificial neural networks (ANN) [11–13] and convolutional neural network-based deep learning [14, 15].

Multi-carrier technologies such as coherent optical OFDM (CO-OFDM) was proposed to enhance flexibility in the network and compensate both fiber chromatic dispersion and polarization mode dispersion (PMD) without adding sophisticated digital algorithms. Unfortunately, the unsortable high peak-to-average power ratio (PAPR) causes a number of problems such as high-power consumption, in-band distortion, and spectrum spreading. A high PAPR causes crosstalk effects in subcarrier-based OFDM appearing more random rather deterministic [13]. ANN has resulted in improved signal quality(Q)-factor in both single-channel [10, 11] and wavelength division multiplexing (WDM) CO-OFDM [12]. However, ANN is optimized using multiple convergence-steps and a high amount of training data ( $\geq 10\%$  [10–12]) is also required which adds complexity and limit signal capacity, respectively.

In this work we experimentally demonstrate, for the first time, MLC-based Hierarchical and Fuzzy-logic C-means (FLC) in single-channel and WDM CO-OFDM for up to 3200

This work was supported by the Partnership Resource Project of Quantum Communications Hub (EPSRC), the EU Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713567, the SFI CONNECT Research Centre and Sterlite Techn. Ltd. We thank S. T. Le, and M. E. McCarthy for their support.

E. Giacomidis, A. Matin and X. Wang are with Heriot-Watt University, School of Engineering & Physical Sciences, EH14 4AS, Edinburgh, UK(email:{e.giacomidis,a.amirali,x.wang}@hw.ac.uk).

E. Giacomidis (is now) and Liam P. Barry are with Dublin City University, Glasnevin 9, Dublin, Ireland (email: liam.barry@dcu.ie).

J.L. Wei is with Huawei Technologies Düsseldorf GmbH, European Research Center, Riesstrasse 25, 80992 München, Germany(email: jinlong.wei@huawei.com).

N. J. Doran is with Aston University, B4 7ET, Birmingham, UK (e-mail: n.j.doran@aston.ac.uk).

km of SSMF for single-polarization. Hierarchical and FLC essentially are compared with the benchmark K-means clustering or simply K-nearest neighbors [9], the advanced classification supervised fast-Newton support vector machine (F-SVM) [13] and ANN-NLE [10–12] and the reduced complexity Volterra-based NLE using the 3rd order Kernel-based inverse Volterra-series transfer function (IVSTF)-NLE [4]. It is shown that FLC reveals the highest performance at optimum launched optical power (LOP), outperforming both K-means and ANN based NLE. FLC also compensates more effectively fiber-induced nonlinearities than the deterministic IVSTF, especially for low number of clusters, i.e. for binary phase-shift keying (BPSK) and quaternary PSK (QPSK). For a ~20-Gb/s QPSK middle ‘worst-case’ WDM channel (the channel suffering the most from inter-channel nonlinearities) at 3200 km and optimum -5 dBm of LOP, FLC outperforms IVSTF by 2.5 dB in Q-factor. However, for 16 quadrature amplitude modulation (16-QAM) in ~46-Gb/s single-channel CO-OFDM at 2000 km, the Q-factor improvement provided by FLC reduces to ~0.4 dB. Even when using alternative more sophisticated clustering designs in 16 clusters, no more than additional ~0.3 dB of Q-factor improvement is observed. Finally, in contrast to the deterministic IVSTF, MLC algorithms can partially tackle the stochastic nonlinearity of parametric noise amplification [6].

In Section II we analyze the principles of the proposed (Hierarchical & Fuzzy-logic C-means) MLC algorithms for optical communication systems with procedures similar to other scientific research areas (e.g. economics). Section III summarizes the experimental and simulated setups which are similar to our previous setups reported in Refs. [10–13] with the exception of implementing the proposed MLC algorithms as new NLEs at the receiver side. Section IV presents the experimental results using MLC (including the benchmark K-means [9]), the traditional machine learning based ANN [11], and the deterministic IVSTF [4] in single-channel QPSK CO-OFDM and WDM 16-QAM CO-OFDM at 3200 and 2000 km, respectively. In this Section, we also investigate the impact of alternative novel advanced clustering designs on 16-QAM CO-OFDM. Finally, in Section V we provide the conclusion to this work.

## II. PRINCIPLE OF MLC ALGORITHMS FOR CO-OFDM

The new adopted clustering algorithms of Hierarchical, Fuzzy logic C-means are discussed in this Section. The adopted MLC algorithms are performed in frequency domain in the CO-OFDM receiver just before decoding and after the fast-Fourier transform (FFT) processing, thus avoiding an additional time- to-frequency domain conversion block. Moreover, clustering processing is performed directly on complex data in contrast to ANN-NLE [11], thus further reducing the complexity of the system. It should be noted however, that when performed MLC separately on real and imaginary part the performance of the algorithms was degraded since they were unable to account for cross-information between the amplitude and phase of the signal.

### A. Hierarchical clustering

Clustering using a Hierarchical approach in OFDM is a multi-step process. It is essentially segmented into agglomerative techniques, being processed by a number of  $n$  symbols mixed into sub-groups, and divisive approaches, which isolate  $n$  OFDM symbols into multiple effective groups for each subcarrier [14–16]. Due to the statistical structure of agglomerative methods, they most commonly characterized by a two-dimensional (2-D) diagram. This diagram is widely identified as dendrological (from the Greek word ‘tree’). An illustration of such dendrogram is depicted in Fig. 1, demonstrating the divisions or fusions made at each successive stage of analysis. Hierarchical clustering harnessing agglomerative processing harvests a number of symbol partitions ( $P$ ):  $P_n, P_{n-1}, \dots, P_1$ . Where  $n$  corresponds to single symbol-based clusters and  $l$  to one group encompassing the total  $n$  cases. During each step, this technique merges the nearest two clusters. However, there are many agglomerative approaches to estimate the distance between clusters [14]. In this work, the least complex single-linkage (D) clustering is implemented in which as mentioned in Ref. [15]: “the distance between groups is defined as the distance between the closest pair of objects/symbols, where only pairs consisting of one object/symbol from each group are considered”. In single-linkage,  $D(r,s)$  is described by Eq. (1) in which a symbol  $i$  belongs to a cluster  $r$  and vice-versa. Afterwards, the shortest (min.) distance for each ‘symbol pair ( $i,j$ )’ is calculated among cluster  $s$  and  $r$ . At every step, these clusters are combined together so that the new developed cluster to reach the “minimum pairwise distances between the symbols” [15].

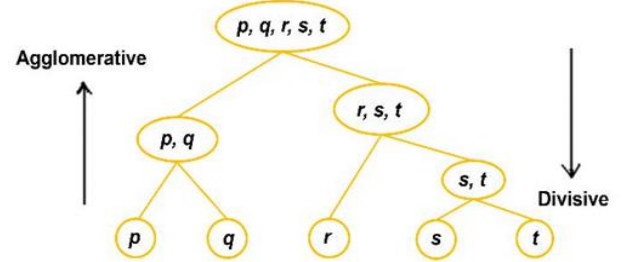


Fig. 1. Conceptual dendrogram for agglomerative and divisive Hierarchical based clustering [15].

Hierarchical clustering is comprised of six steps [16]:

1. Enter the number of targeted clusters, e.g. four for QPSK.
2. Initiate disjoint cluster having zero level ( $L(0) = 0$ ) and order ( $m = 0$ ).
3. Identify the least unrelated pair of clusters ( $r, s$ ) w.r.t.

$$D(r,s) = \min\{d[i,j]\} \quad (1)$$

4. Increase the order by  $m=m+1$  and the clusters  $r$  and  $s$  into one cluster, creating a new cluster  $m$ . The level of such cluster is formed by

$$L(m) = d[r,s] \quad (2)$$

5. Upgrade the D “proximity matrix”, thus erasing the corresponding rows/columns from clusters  $r$  and  $s$ , while adding a single row/column related to the new created cluster. The proximity between the old,  $k$ , and new cluster, ( $r,s$ ), is given from Eq. (3)

$$d\{k, (r,s)\} = \min\{d[(k),(r)], d[(k,s)]\} \quad (3)$$

6. Stop process if total OFDM symbols are located in a single cluster, otherwise, return to second step.

### B. FLC: Fuzzy-logic C-means clustering

FLC permits OFDM symbols to fluctuate their membership degree (MD) while being allocated into many clusters [17–21]. FLC minimizes an objective function of the form:

$$F_m = \sum_{i,j} \sum_{l=1}^R \sum_{j=1}^L \mu_{ij}^m \|t_i - c_j\|^2 \quad (4)$$

where  $N$ ,  $R$ ,  $L$ , and  $m$ , are the total number of subcarriers, symbols, clusters, and a “Fuzzy partition matrix (FPM) exponent”, respectively. FPM regulates the ‘degree’ of clusters overlapping, for  $m$  greater than unity. Such overlapping is related to a Fuzzy one, denoting the degree of boundaries’ fuzziness between clusters. Where  $t_i$  is referred the  $i$ -th symbol,  $c_j$  is the centre of a  $j$ -th cluster, and  $\mu_{ij}$  refers to the MD of  $t_i$  into  $j$ -th cluster. Given a random symbol  $t_i$ , the sum of MD for the total clusters is unity.

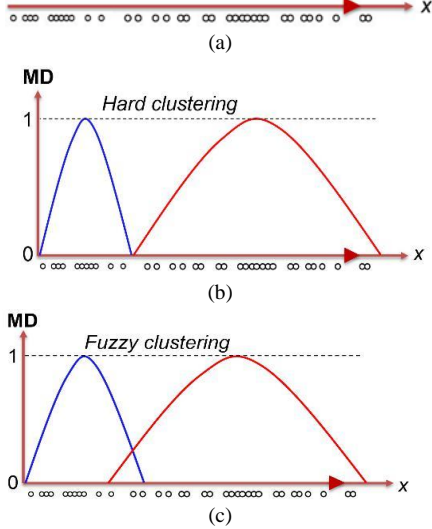


Fig. 2. Operation of FLC on OFDM symbols (circles): (a) Single-dimension data. (b) Hard clustering. (c) Fuzzy clustering. MD: membership degree.

FLC is comprised by the following six steps [21, 22]:

1. Enter the number of targeted clusters.
2. Initiate the cluster MD,  $\mu_{ij}$ .
3. Estimate the center of the cluster by the following expression

$$C_j = \sum_{i,j} \left( \frac{\sum_{l=1}^R \mu_{ij}^m t_i}{\sum_{l=1}^R \mu_{ij}^m} \right) \quad (5)$$

4. Update  $\mu_{ij}$  using Eq. (6)

$$C_j = 1 / \left( \sum_{i,j} \sum_{k=1}^L \|t_i - c_j\| / \|t_i - c_k\| \right)^{2/m-1} \quad (6)$$

5. Compute  $F_m$  using Eq. (4).
6. Return and perform second to fourth step until  $F_m$  is converged for a specified threshold.

A typical example of the FLC processing is illustrated in Fig. 2. In Fig. 2(a) single-dimension symbols are given with a MD

of either one or zero that can be formed in two distinct clusters. The MD is formed exclusively between one and zero and presented by the y-axis as shown in Figs. 2(b), 2(c), corresponding to Hard and Fuzzy clustering, respectively. In comparison to clustering with Hard decisions, the adopted FCL sets a new threshold between the two clusters with an updated MD for each symbol, that are based on the centroids of the clusters and the distance between the clusters’ centroids [17].

### III. EXPERIMENTAL AND SIMULATION SETUP

The experimental setup (identical to Ref. [12]) for multi-channel QPSK is shown in Fig. 3(a). It comprised a laser grid of 5 standard distributed feedback lasers (DFBs) on 100 GHz grid with the help of polarization maintaining multiplexer (PMM), which were substituted in turn by a 100 kHz linewidth laser. The 100 kHz linewidth DFBs are located between 193.5–193.9 THz. Additional loading channels (10 GHz of bandwidth) were generated using an amplified spontaneous emission (ASE) source that were spectrally shaped using a wavelength selective switch (WSS). The 20 loading channels were spread symmetrically around the test wavelengths so that the total bandwidth of the transmitted signal was 2.5 THz (see inset spectrum in Fig. 3(a)). A wideband filter was used to filter out-of-band ASE noise at the transmitter. The transmission path was an acousto-optic modulator (AOM) based re-circulating loop consisting of 4×100 km spans of Sterlite OH-LITE (E) SSMF, having 18.9–19.5 dB insertion loss. The loop switch was located in the mid-stage of the 1<sup>st</sup> Erbium-doped fiber amplifier (EDFA) and a gain flattening filter (GFF) was placed in the mid stage of the 3<sup>rd</sup> EDFA. After propagation, the signal was filtered using a 4.2 nm flat topped filter and coherently detected. Fig. 3(b) depicts the single-channel experimental setup (identical to Refs. [10], [13]) whereas 100 kHz linewidth DFB was modulated using a dual-parallel Mach-Zehnder modulator in IQ configuration fed by offline OFDM I-Q components. The transmission path at 1550.2 nm was a recirculating loop consisting of 20×100 km spans of E-SSMF controlled by AOM. The loop switch was located in the mid-stage of the 1<sup>st</sup>EDFA and a GFF was placed in the mid-stage of the 3<sup>rd</sup>EDFA. The optimum LOP was swept by controlling the output power of the EDFAs.

Table I. Single- and multi-channel experimental OFDM parameters

Parameter	Value
Net bit-rate	18.2 Gb/s(WDM), 40 Gb/s(1-ch.)
Net bit-rate for ANN	16.84 Gb/s(WDM), 38 Gb/s(1-c.)
Raw bit-rate	20 Gb/s(WDM), 46 Gb/s(1-ch.)
Format of modulation	QPSK(WDM), 16-QAM(1-ch.)
Number of symbols	400
Symbol time duration	20.48 ns
Generated subcarriers	210
CP	2 %
Size of FFT & inverse(I)FFT	512
ANN Training overhead	10 %
ANN Train. symbol length	40 symbols
DFB linewidth	100 KHz
OH-LITE (E) fiber attenuation	18.9–19.5 dB/100 km
Number of spans	30(WDM), 20(1-chan.)
Length-per-span	100 km
Center wavelength	1550.2 nm

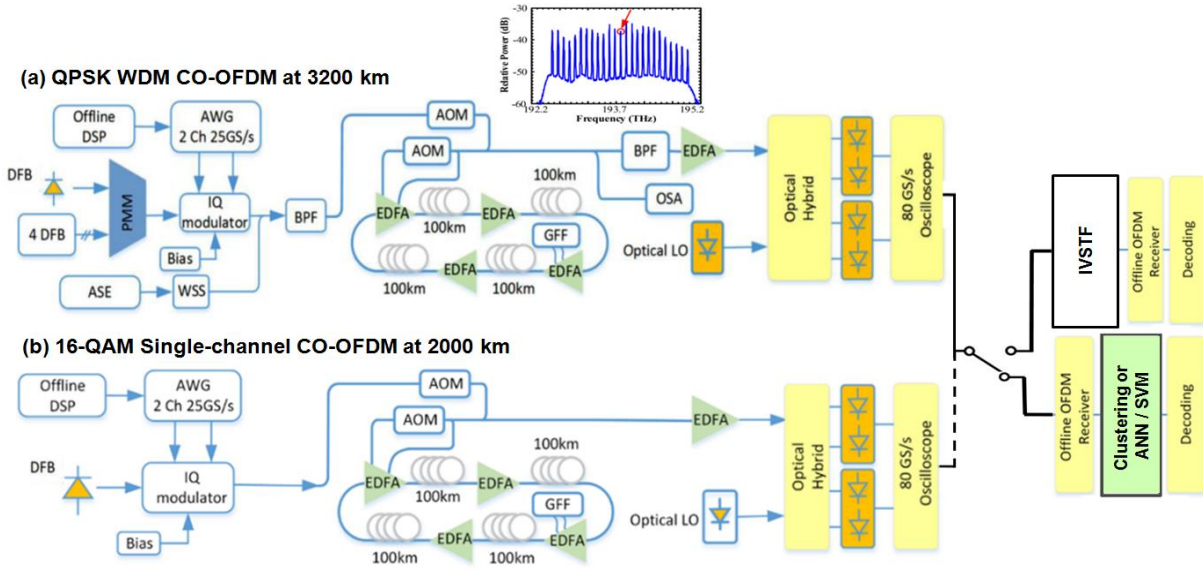


Fig. 3. Experimental setup of CO-OFDM equipped with clustering, ANN, SVM or IVSTF based NLE for: (a) multi-channel 20 Gb/s QPSK (middle channel) at 3200 km; (b) single-channel 40 Gb/s 16-QAM at 2000 km. AWG: arbitrary waveform generator, PMM: polarization maintaining multiplexer, WSS: wavelength selective switch, DFB: distributed feedback laser, AOM: acousto-optic modulator, GFF: gain-flatten filter, BPF: band-pass filter.

For both cases at the receiver, the incoming channel was combined with another 100 kHz linewidth DFB acting as local oscillator (LO). After down-conversion, the baseband signal was sampled using a real-time oscilloscope operating at 80 GS/s and processed offline in Matlab<sup>®</sup>. 400 OFDM symbols (20.48 ns length) were generated using a 512-point IFFT in which 210 subcarriers were modulated using QPSK (WDM) and 16-QAM (single-channel). To eliminate inter-symbol-interference from linear effects, a CP of 2% was included. The ANN and F-SVM training overheads were set at 10% (optimum value for linear equalization, LE, similarly to Refs. [10], [12], [13]) resulting in a training length of 40 symbols. For LE, IVSTF, ANN, and F-SVM the net bit-rate for the WDM system was fixed at 18.2 Gb/s after CP and removed, and 16.84 Gb/s after 10% of ANN and F-SVM training overhead is removed, while the raw bit-rate was 20 Gb/s. For the single-channel system the net and raw bit-rates were ~40 Gb/s and ~46 Gb/s, respectively. The offline OFDM demodulator included timing synchronization, frequency offset compensation, channel estimation and equalization with the assistance of an initial training sequence, as well as I-Q imbalance and CD compensation using an overlapped frequency domain equalizer employing the overlap-and-save method. The CO-OFDM transceiver and transmission parameters are depicted on Table I. The NLEs performances were assessed by Q-factor measurements averaging over 10 recorded traces (~10<sup>6</sup> bits), which was estimated from the bit-error-rate (BER) obtained by error counting after hard-decision decoding. The Q-factor is related to BER by  $Q = 20 \log_{10} [\sqrt{2} \operatorname{erfc}^{-1}(2BER)]$ .

For the simulated analysis, we used a co-simulated environment with VPI-transmission-maker<sup>TM</sup> and Matlab<sup>®</sup>. The former simulated the optical components including the SSF by the well-known split-step Fourier method via the nonlinear Schrödinger equation, and the latter simulated the digital signal processing units including OFDM modulation and demodulation. For the numerical analysis, a 9.1 Gb/s BPSK

single-polarization and single-channel CO-OFDM system was considered and transmitted at 500 km with 100 km span-length. The reason for not increasing the bit-rate of such low modulation format order was to relax the digital-to-analogue/analogue-to-digital converter bandwidth (sampling rate/analogue bandwidth) to a more realistic value.

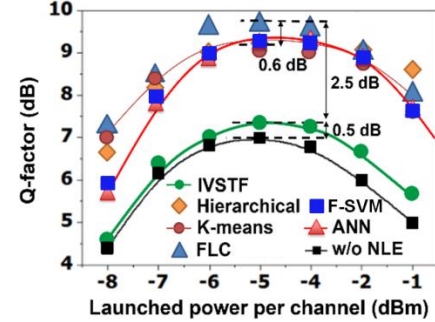


Fig. 4. Transmission performance comparison for all adopted MLC algorithms, ANN, F-SVM, IVSTF, and without (w/o) using NLE (i.e. LE) in terms of launched optical power (LOP) per channel for QPSK WDM CO-OFDM at 3200 km of fiber propagation.

#### IV. RESULTS AND DISCUSSION

In Fig. 4 results for QPSK WDM CO-OFDM at 3200 km are depicted, related to the Q-factor against the LOP per channel. From Fig. 4 it is evident that MLC can effectively tackle inter-channel (cross-phase modulation, XPM, four-wave mixing, FWM) and intra-channel nonlinearities (self-phase modulation and inter-subcarrier XPM and FWM). In comparison to IVSTF, MLC can enhance the Q-factor up to ~2.5 dB at optimum LOPs. Among FLC, Hierarchical, and K-means clustering, FLC outperforms within the range of optimum and very low LOPs, the latter which is partially contributed to the stochastic *parametric noise amplification*. However, at very high LOPs per channel (i.e. -1 dBm) Hierarchical clustering can tackle more effectively the strong nonlinear crosstalk effects. An example of the FLC improved performance at a very low power is depicted in Fig. 5, where the received



constellation diagrams for FLC and K-means are compared at a LOP per channel of  $-8$  dBm. Fig. 5 shows that FLC improves the Q-factor by 0.3 dB by making some versatile nonlinear decisions on “rotated” (distorted) OFDM symbols in contrast to K-means which is limited to linear decision boundaries. Essentially, FLC allocates the distorted symbols more effectively on the valid clusters using the minimum distance from the centroid values (white ‘x’ in Fig. 5). In comparison to the advanced supervised *classification-based* F-SVM and ANN-NLE, FLC can combat inter-channel nonlinearities more effectively leading to a Q-factor enhancement of  $\sim 0.6$  dB at the optimum LOP per channel of  $-5$  dBm as depicted in Fig. 4. Moreover, FLC and the rest of adopted MLC algorithms can tackle *parametric noise amplification* better than F-SVM and ANN, without the need of capacity-consuming training data, showing the great potential of clustering algorithms for QPSK signals.

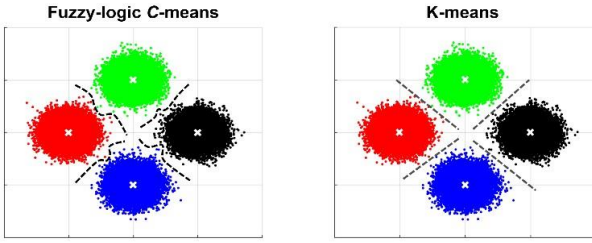


Fig. 5. Example of FLC and K-means boundaries on received constellation diagrams at  $-8$  dBm of LOP per channel for middle-channel QPSK WDM CO-OFDM (white ‘x’ denote the centroid per cluster).

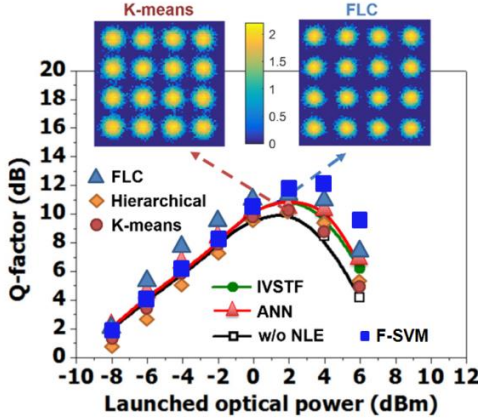


Fig. 6. Q-factor vs. LOP of 16-QAM CO-OFDM at 2000 km for adopted MLC algorithms, F-SVM, ANN, IVSTF and w/o NLE.

In Fig. 6, results from the 16-QAM single-channel CO-OFDM are illustrated at 2000 km of fiber transmission. For 16 clusters, MLC is not as effective as previously in 4 clusters (QPSK). This can be explained from Fig. 7, where the Q-factor distribution for the middle subcarriers is plotted using the best MLC algorithm, i.e. the FLC, and IVSTF for QPSK WDM CO-OFDM and 16-QAM CO-OFDM at optimum LOPs of  $-5$  and 2 dBm, respectively. Essentially, this comparison is contributed to the ‘stochastic vs. deterministic’ nonlinear effects on centre subcarriers which suffer the most from inter-subcarrier XPM and FWM. As shown from Fig. 7(a), FLC in 4 clusters can improve the Q-factor up to 3.8 dB on middle subcarriers compared to IVSTF. This performance improvement is mainly due to the partial compensation of ASE noise and inter-subcarrier intermixing deterministic

nonlinear effects that appear random due to the high PAPR [10]. In Fig. 7(b), the stochastic nonlinear effects on middle subcarriers cannot be effectively compensated because the accumulated random phase noise for 16 clusters is much higher than 4 clusters. Nevertheless, from Fig. 6, even marginally it is evident that FLC outperforms Hierarchical clustering, K-means, ANN and IVSTF, while reaching the transmission performance of an F-SVM at optimum 2 dBm of LOP. An example of the FLC performance improvement is depicted in inset of Fig. 6, where the received 16-QAM constellation diagrams for FLC and K-means are shown at optimum 2 dBm of LOP. In the linear regime of Fig. 6, FLC improves the Q-factor which is connected to the tolerance improvement of *parametric noise amplification*; where in contrast, Hierarchical clustering and K-means are not effective having worst performance than linear equalization.

In Fig. 8, a simulated analysis is performed for lower level than QPSK format, i.e. BPSK, to evaluate the performance of the adopted and benchmark MLC algorithms in comparison to ANN and IVSTF. As depicted in Fig. 8, the performance improvement of the adopted MLC algorithms from benchmark NLEs increases for BPSK CO-OFDM. In particular, a Q-factor improvement of 3.2 dB is achieved at optimum LOP of  $-10$  dBm using the best MLC algorithm of FLC. This confirms the fact that less amount of constellation clusters contributes to enhanced transmission performance by MLC.

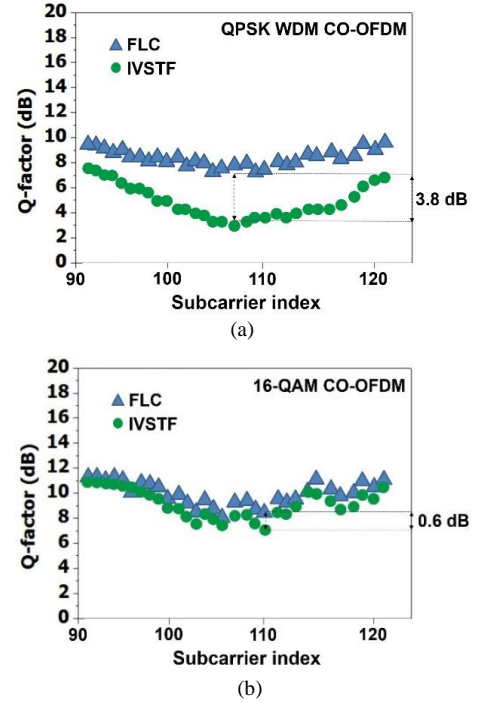


Fig. 7. Middle subcarrier index Q-factor distribution for FLC and IVSTF over (a) QPSK WDM at 3200 km of transmission and optimum LOP per channel of  $-5$  dBm; (b) 16-QAM single-channel at 2000 km and optimum LOP of 2 dBm.

Due to the modest improvement of MLC algorithms in 16 clusters for 16-QAM OFDM modulation, alternative novel clustering grouping designs have been tested. These designs are based on groups of two or four clusters due to the remarkable performance of MLC on BPSK and QPSK. The novel designs feature two extra cases: Case-1, where a single

Table II. Summary of key experimental results: Q-factor improvement (in dB) from linear equalization for each technique.

Technique	QPSK WDM LOP = -8 dBm	16-QAM Single-channel LOP = -6 dBm	QPSK WDM LOP = -5 dBm	16-QAM Single-channel LOP = 2 dBm
Fuzzy-logic Case-2	—	<b>1.8 dB</b>	—	<b>1.4 dB</b>
Fuzzy-logic	<b>2.9 dB</b>	<b>1.8 dB</b>	<b>2.8 dB</b>	<b>1.1 dB</b>
Hierarchical clustering	2.2 dB	1.2 dB	2.3 dB	0.2 dB
K-means	2.6 dB	-0.2 dB	2.2 dB	0.2 dB
ANN	1.2 dB	0.7 dB	2.3 dB	0.3 dB
F-SVM	1.4 dB	0.7 dB	2.3 dB	1.4 dB
IVSTF	0.2 dB	0.7 dB	0.5 dB	0.8 dB

“large” group of four clusters and subsequent four groups of four clusters are considered; and Case-2, in which a single group of four clusters and six groups of two clusters are performed. The latter case was inspired by the well-known accumulation of nonlinear phase noise on outer clusters in 16-QAM. Fig. 9 illustrates the adopted clustering designs in which the grouping centroids (denoted with ‘x’) are also depicted: light-blue on step 1 and white on step 2 for Case-1; black for single-step Case-2. Fig. 10 shows the performance of the best MLC algorithm, i.e. the FLC, for experimental 16-QAM CO-OFDM in single-channel transmission at 2000 km. It is shown that both clustering designs have almost identical performance with the conventional clustering approach, with the exception of Case-2 at very high LOPs where up to ~0.3 dB increase in Q-factor is observed, reaching the transmission performance of the F-SVM.

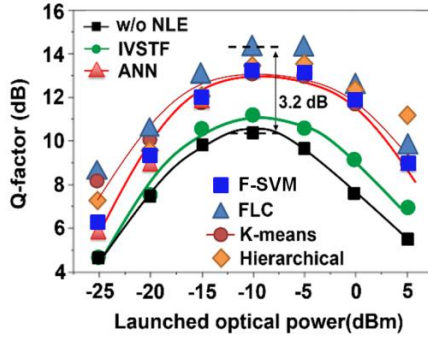


Fig. 8. Simulated CO-OFDM transmission performance for 9.1 Gb/s binary phase-shift keying in single-channel system at 500 km using FLC, Hierarchical clustering, K-means, F-SVM, ANN, IVSTF and w/o NLE (i.e. LE).

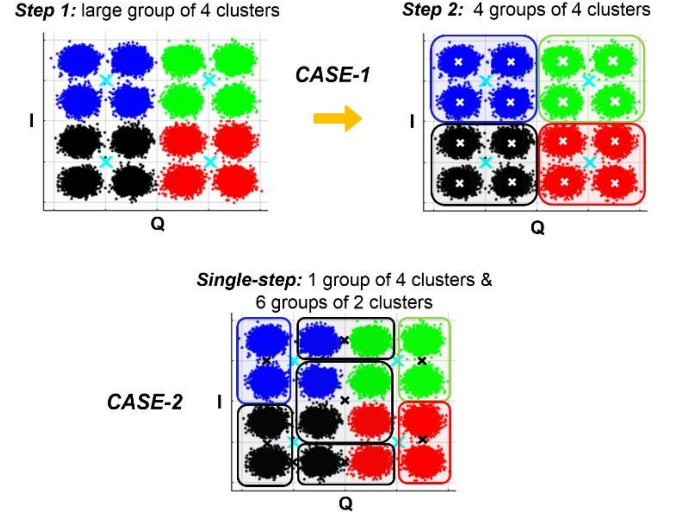


Fig. 9. Alternative clustering grouping in 16-QAM. In first case, a single group with four clusters is performed and afterwards four groups with four clusters. In second case, single-step is performed where a single group of four clusters and six groups of two clusters are performed.

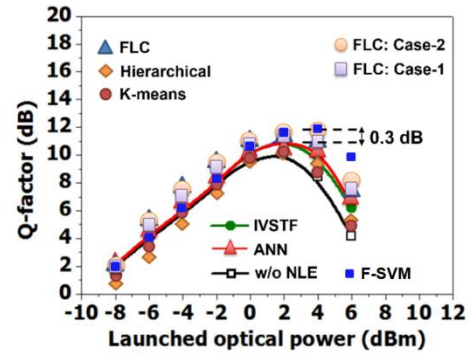


Fig. 10. Impact of alternative clustering designs (Case-1 vs. Case-2) on Q-factor vs. LOP for FLC using 16-QAM CO-OFDM at 2000 km of fiber propagation. A comparison with benchmark NLEs and MLC algorithms that employ the conventional clustering approach is also included.

## V. CONCLUSION

While Hierarchical and Fuzzy-logic C-means clustering (FLC) have been applied in medicine [19], economics [18, 21], wireless sensor networks [14] and other research areas, they have never been implemented in optical communications for signal quality improvement. In this work, for the first time, these two machine learning based clustering (MLC) algorithms were experimentally demonstrated for blind nonlinearity equalization in ~46-Gb/s single-channel and ~20-Gb/s (middle-channel) WDM CO-OFDM. MLC tackled more effectively intra- and inter-channel nonlinearities at 3200 km compared to the supervised F-SVM/ANN and the

deterministic IVSTF for low-level modulation formats (i.e. BPSK, QPSK). When more clusters were considered in single-channel 16-QAM at 2000 km, F-SVM outperformed to the other algorithms at high launched powers. Even when using sophisticated novel clustering designs in 16 clusters, no more than additional  $\sim 0.3$  dB Q-factor enhancement by FLC was observed. From both simulated analysis for BPSK modulation and experimental QPSK/16-QAM demonstrations, FLC had the highest performance among all adopted MLC algorithms at optimum LOPs. In contrast to IVSTF, FLC revealed better potential in tackling the stochastic *parametric noise amplification*. Finally, it should be noted that due to experimental restrictions and limitations on available resources, 16-QAM CO-OFDM was only demonstrated for single-channel and BPSK CO-OFDM for simulated analysis.

A summary of the key results for low and optimum LOPs are depicted in Table II, where the Q-factor improvement (in dB) from linear equalization (i.e. w/o NLE) is shown for each adopted algorithm. From Table II, it is clearly identified (in bold blue) that FLC, 'FLC with clustering design of Case-2' for 16-QAM have always the best performance at low powers. Considering the fact that MLC algorithms can potentially have lower complexity than IVSTF [4], F-SVM [9] and ANN [10], this work is very useful not only for next-generation high-capacity core networks but also for real-time optical communications.

## REFERENCES

- [1] A. D. Ellis et al, "Performance limits in optical communications due to fiber nonlinearity," *Adv. in Opt. & Phot.*, vol. 9, no. 3, pp. 429–503, Sep. 2017.
- [2] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, pp. 1027–1030, April 2001.
- [3] D. Rafique, J. Zhao, and A. D. Ellis, "Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM m-ary QAM transmission," *Opt. Exp.*, vol. 19, no. 6, pp. 5219–5224, Mar. 2011.
- [4] G. Gao, X. Chen, and W. Shieh, "Limitation of fiber nonlinearity compensation using digital back propagation in the presence of PMD," in *Proc. OFC, CA, USA, Mar. 4–8, 2016*, paper OM3A.5.
- [5] E. Giacomidis et al, "Volterra-Based Reconfigurable Nonlinear Equalizer for Coherent OFDM," *IEEE Phot. Technol. Lett.*, vol. 26, no. 14, pp. 1383–1386, July 2014.
- [6] X. Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," *Nature Phot.*, vol. 7, no. 7, pp. 560–568, May 2013.
- [7] D. Zibar et al, "Nonlinear impairment compensation using expectation maximization for dispersion managed and unmanaged PDM 16-QAM transmission," *Opt. Exp.*, vol. 20, no. 26, pp. B181–196, Dec. 2012.
- [8] D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, "Machine Learning Techniques in Optical Communication (Invited)," *IEEE J. of Lightw. Techn.*, vol. 34, no. 6, pp. 1442–1452, Mar. 2016.
- [9] D. Wang et al, "Nonlinearity Mitigation Using a Machine Learning Detector Based on k-Nearest Neighbors," *IEEE Phot. Technol. Lett.*, vol. 28, no. 19, pp. 2102–2105, Oct. 2016.
- [10] D. Wang et al, "Combating nonlinear phase noise in coherent optical systems with an optimized decision processor based on machine learning," *Opt. Commun.*, vol. 35, pp. 199–208, Jun. 2016.
- [11] E. Giacomidis et al, "Fiber Nonlinearity-Induced Penalty Reduction in Coherent Optical OFDM by Artificial Neural Network based Nonlinear Equalization," *Opt. Lett.*, vol. 40, no. 21, pp. 5113–5116, Oct. 2015.
- [12] M.A. Jarajreh, E. Giacomidis, I. Aldaya, S.T. Le, A. Tsokanos, Z. Ghassemlooy, and N.J. Doran, "Artificial Neural Network Nonlinear Equalizer for Coherent Optical OFDM," *IEEE Phot. Technol. Lett.*, vol. 27, no. 4, pp. 387–390, Feb. 2015.
- [13] E. Giacomidis et al, "Intra and inter-channel nonlinearity compensation in WDM coherent optical OFDM using artificial neural network based nonlinear equalization," in *Proc. OFC, CA, USA, Mar. 19–23, 2017*, paper Th2A.62.
- [14] D. Wang et al, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Exp.*, vol. 25, no. 15, pp. 17150–17166, 2017.
- [15] D. Wang et al, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photon. Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, 2017.
- [16] A. D. Ellis, S. T. Le, M. E. McCarthy, and S.K. Turitsyn, "The impact of parametric noise amplification on long haul transmission throughput," in *Proc. IEEE ICTON, Budapest, Hungary, July 5–9, 2015*, paper We.D1.5.
- [17] E. Giacomidis et al, "Reduction of Nonlinear Inter-Subcarrier Intermodulation in Coherent Optical OFDM by a Fast Newton-based Support Vector Machine Nonlinear Equalizer," *IEEE J. of Lightw. Techn.*, vol. 35, no. 12, pp. 2391–2397, 2017.
- [18] R. Zhang, J. Pan, D. Xie, and F. Wang, "NDCMC: A Hybrid Data Collection Approach for Large-Scale WSNs Using Mobile Element and Hierarchical Clustering," *IEEE Internet of Things J.*, vol. 3, no. 4, pp. 533–543, Aug. 2016.
- [19] <https://www.solver.com/xlminer/help/hierarchical-clustering-intro>
- [20] S. P. Borgatti, "How to Explain Hierarchical Clustering," *INSNA Connections*, vol. 17, no. 2, pp. 78–80, 1994.
- [21] M. Elena, "Fuzzy C-means clustering in Matlab," in *Proc. of 7th Inter. Days of Statist & Econ.*, Prague, Sep. 19–21, 2013, pp. 905–914.
- [22] G. Gan, C. Ma, and J. Wu. (2007). *Data Clustering: Theory, Algorithms, and Applications*. (1<sup>st</sup> ed., pp. 18–25). London: Wiley.
- [23] T. Nguyen and S. Nahavandi, "Modified AHP for Gene Selection and Cancer Classification Using Type-2 Fuzzy Logic," *IEEE Trans. on Fuzzy Syst.*, vol. 24, no. 2, pp. 273–287, Apr. 2016.
- [24] N. Akhtar and M. V. Ahmad, "A Modified Fuzzy C Means Clustering Using Neutrosophic Logic," in *Proc. IEEE Int. Conf. on Commun. Syst. & Netw. Technol.*, Apr. 4–6, Gwalior, India, pp. 1124–1128, 2015.
- [25] J. C. Bezdec, "Pattern Recognition with Fuzzy Objective Function Algorithms," *Plenum Press*, New York, 1981.
- [26] <https://www.mathworks.com/help/fuzzy/fcm.html>
- [27] S. Panda, S. Sahu, P. Jena, and S. Chattopadhyay, "Comparing Fuzzy-C Means and K-Means Clustering Techniques: A Comprehensive Study," *Springer Advances in Computer Science, Eng. & Appl.*, vol. 166, pp. 451–460, 2012.

**Elias Giacomidis** is a Marie-Curie Fellow at Dublin City University & SFI CONNECT Research Centre of Ireland. His project tackles the "capacity crunch" in Energy-efficient and Phase-Insensitive Coherent Communications (EPIC). He has previously worked for various prestigious optical communications research groups: Heriot-Watt University, University of Sydney CUDOS (deputy project leader), Aston University, Bangor University (PhD scholarship), Athens Information Technology, and Telecom-ParisTech, where he was also teaching optical system modeling. He has authored/co-authored >100 papers (including 1 invited JLT and 2 prestigious Postdeadline papers at ACOFT and CLEO-Pacific Rim) that appeared in international peer-reviewed journals and top conferences being cited 954 times by Google-Scholar with H-factor of 18. His research involves balanced theoretical and experimental exploration in >40G transmission systems with specialization in key modern DSP modulation techniques (OFDM, CAP, PAM-4 etc.) and nonlinear photonics for local networks, passive optical networks and next-generation flexible optical communications. Dr. Giacomidis is the principal investigator of the world's-first direct-detected optical Fast-OFDM system. He was the first to implement digital-based machine learning in optical communications (coherent optical OFDM) for fiber nonlinearity compensation.

**Amir Matin** received his BEng degree in electrical and electronics engineering from Heriot-Watt University, Edinburgh, UK, in 2013 and has several years of experience in industry (2013-2016). Currently he is pursuing his Ph.D. in digital signal processing and fast optical imaging systems. His research interest includes image, video and signal processing, compressive sensing and fast optimization algorithms.

**Jinlong Wei** (S'09–M'11–SM'17) is currently a senior researcher in Huawei Technologies Düsseldorf GmbH, European Research Center, Germany. He received his PhD degree in Electronic Engineering from Bangor University, Bangor, UK in 2010 and worked there afterwards as a post-doc researcher before joining the Centre for Photonics Systems of Cambridge University, UK as a research associate in 2011. He was awarded a Marie Curie

fellowship and joined ADVA Optical Networking SE, Germany as a senior engineer in Sept. 2014. He has participated in about 10 European national and international projects as well as projects with industry and realized a number of world-first system demonstrations with his colleagues. He has been contributing to the next generation 100 Gigabit Ethernet study within IEEE802.3. He has authored and coauthored over 120 journal and conference publications including over 10 invited and 5 public patents.

**Nick J. Doran, Liam P. Barry and Xu Wang** biographies not available at the time of submission.