# Highly Linear IoT-over-Plastic Optical Fiber Transmission with low complexity Cluster-based Predistortion.

M.A. Losada, P. L. Carro, A. López, P. García-Dúcar, J. de Mingo, A. Valdovinos,

***Abstract*— In the context of ultra-low end applications such as smart homes, smart vehicles and Industrial Internet of Things (IIoT), the use of Plastic Optical Fibers (POFs) as communications backbone to enable the convergence of wireless and optical systems has emerged as a cost-effective and ruggedized solution. The transmission of standalone Narrowband IoT (NB-IoT) signals through large-core Step-Index POF (SI-POF) can be achieved by directly modulating a laser diode at the cost of introducing non-linear effects. In this work, the performance of a Radio-over-POF (RoPOF) link is enhanced by Digital Predistortion (DPD) using a memoryless polynomial approach whose complexity is reduced by scaling down the number of samples using a novel technique that preserves the data statistics by organizing the data into a number of clusters.**

***Index Terms*—Radio over Plastic Optical Fiber (RoPOF), non-linearities, predistorsion techniques.**

## I. Introduction

THE implementation of new paradigms such as smart homes, autonomous vehicles and Industry 4.0 demands a large number of connections among many low-power heterogeneous devices, which can be realized through the narrowband wide-area cellular transmission standard termed Narrowband Internet of Things (NB-IoT). This standard was developed by 3GPP [1] to support ultra-low end applications with extended coverage in combination with existing cellular technologies as Long Term Evolution (LTE) [2]. Currently, 5G architectures encompass mobile broadband, massive machine-type and critical communications, forming the backbone of intelligent IoT systems. As we transition toward 6G, these systems are expected to incorporate Artificial Intelligence (AI) to enable real-time decision making, requiring efficient IoT devices [3]. Given that AI algorithms rely on vast datasets and computational intensive training, both of which demand high transmission speed and low latency, future smart systems will increasingly require robust backhaul infrastructures. To support these evolving necessities, different optical cooperative technologies, such as fiber-wireless (Fi-Wi) networks, free space optical (FSO) systems, and Radio-over-Fiber (RoF) solutions have been proposed to deal with cellular or unlicensed IoT ecosystems [3]–[5]. In particular, RoF emerges as a key enabler by exploiting the high spectral efficiency, broad bandwidth, and low attenuation of fiber-optic infrastructure to act as the backbone for next-generation networks. This architecture facilitates the transport of radio signals and AI-related data between distributed edge nodes and centralized processing units, supporting the stringent latency and throughput requirements of future IoT applications.

In long-haul scenarios, singlemode or multimode glass fibers are the preferred option to constitute the wired backbone. However, in very-short-area environments, where it is imperative to curtail installation and maintenance costs, the fast deployment of NB-IoT services can be achieved by using Plastic Optical Fibers (POFs). In fact, POFs made of perfluorinated polymers with Graded Index (GI) profiles offer transmission performance comparable to that of multimode glass fibers, and have proven excellent performance in different Radio over POF (RoPOF) systems [6]–[8]. On the other hand, Poly(methyl methacrylate) or PMMA-based POFs, remain widely adopted in automotive and domestic environments [9]. Their larger core diameters (typically 980 $\mu$m) and higher numerical apertures (up to 0.6) make them easy to be handled, which simplifies installation and reduce operational costs. In addition, PMMA-based POFs display high resistance to impacts and harsh temperatures, and have low curvature and misalignment losses compared to those of glass and fluoropolymer-based fibers. These features make them well-suited to rugged and hostile environments, such as factories or aircraft [10]. Moreover, in terms of power efficiency and energy conservation, a notable advantage of IoT-over-POF systems is their capacity for reliable remote power delivery [11], [12]. In spite of these advantages, the use of large-core PMMA-based POFs has been mainly restricted to transmit baseband signals, using advanced modulations formats to enhance their performance [13]. Although 5G radio signals have been successfully transmitted through 50 and 75 meters of large-core PMMA GI-POFs [14], [15], very few attempts to use Step Index POFs (SI-POFs) in this context have been reported [16], [17]. However, even the scarce bandwidth of large-core SI-POFs can potentially cope with the transmission of narrowband signals provided the carrier frequency is below 2 GHz [18].

Considering the transmission limitations of SI-POFs, the NB-IoT standard presents a suitable framework for evaluating their performance in RoPOF links. NB-IoT is based on Orthogonal Frequency Division Multiplexing (OFDM) modulation for downlink transmission in the three possible operational

This research was funded by the Spanish Ministry of Science and Innovation (MICINN/FEDER), under grant number PID2021-122505OB-C33, by the Aragón Government (DGA), under grant numbers T20-23R and T31-23R and by European Union – Next GenerationEU under grant TSI-064100-2023-27. (Corresponding author: M.A. Losada). M.A. Losada (e-mail: alosada@unizar.es), and A. López (aliclope@unizar.es) are with the Photonics Technology Group (GTF), Aragón Institute of Engineering Research (i3A), University of Zaragoza, Zaragoza, Spain. P. L. Carro (plcarro@unizar.es), P. García-Dúcar (paloma@unizar.es), J. de Mingo (mingo@unizar.es), and A. Valdovinos (toni@unizar.es) are with the Communications Networks and Information Tecnhologies group (CeNIT), Aragón Institute of Engineering Research (i3A), University of Zaragoza, Zaragoza, Spain. All authors are with the Electronics Engineering and Communications Dept., EINA, University of Zaragoza, Zaragoza, Spain.

Manuscript received April 19, 2021; revised August 16, 2021.

modes: inband, guardband and standalone, each with a signal bandwidth of 180 kHz. In the standalone mode, the carrier frequencies range from 925 MHz to 960 MHz, spaced at 200 kHz [2]. These characteristics make standalone NB-IoT a promising candidate for demonstrating the feasibility of SI-POFs in short-range RoPOF applications. Large-core POFs offer the advantage of easy coupling with cost effective light sources, such as visible-light laser diodes, which support direct modulation and eliminate the need for expensive external modulators. On the other hand, OFDM-like signals are especially susceptible to nonlinear distortions which can readily occur in RoF transmission, not only at the power amplifiers, but mainly across different components of the optical link [19], [20]. Therefore, the application of specific techniques such as Digital Predistortion (DPD), is necessary to mitigate the system non-linearities while using high RF powers at the transmission end without detriment to the link power budget.

The memory polynomial model approximation to the Volterra series can describe most non-linear systems by varying its memory depth and polynomial order [21]. This model has been customarily used to improve the transmission performance in RoF links [22] and for NB-IoT applications [23]. However, the implementation of DPD in NB-IoT transmission represents a significant source of power consumption, posing a critical challenge for energy-efficient system design. Therefore, feature selection and feature extraction techniques have been proposed to reduce the complexity of obtaining the model and DPD parameters from experimental data [24], [25]. Undersampling strategies produce shorter data sequences, but they do not represent well the statistical properties of the system leading to less accurate models [26]. To overcome this problem, histogram-based methods have been proposed to reduce the number of data samples taking into account their statistical distribution [27], [28]. Following the later strategy, this work proposes a novel technique based on clustering to alleviate the computational requirements of DPD based on the polynomial model. The data is organized into a number of clusters before applying a sample reduction technique that preserves a higher proportion of data in the clusters which are more significant to represent the non-linear system behavior. To test the proposal, we implemented a Radio-over-POF (RoPOF) link designed to transmit downlink standalone NB-IoT signals through SI-POFs of different characteristics and lengths. Results demonstrate that the good performance of SI-POF for the transmission of NB-IoT can be enhanced by using the proposed Cluster-Based Sample-Selection (C-SS) method to extend the linear operative range of the link.

The main contributions of this paper are:

- The demonstration of the viability of single channel standalone NB-IoT over SI-POFs longer than 50 meters and the identification of the optical transmitter as the cause of non-linear distortion at high RF input powers.
- The proposal and validation of a sample selection method based on clustering to reduce complexity in DPD extending the range of usable input RF powers.

The remainder of this paper is structured as follows:

- Section II describes the experimental testbed and results of the transmission of NB-IoT signals through SI-POFs.
- Section III introduces our proposed technique of sample selection based on data classification in clusters to identify the DPD model with reduced complexity.
- Section IV evaluates the transmission of the DPD signals over the implemented link to assess the effectiveness of the proposed method.
- Section V discusses the results and points to potential directions for future research.
- Section VI summarizes our findings and conclusions.

## II. Standalone NB-IoT RoPOF transmission

This section is organized in two subsections. The first describes the experimental setup and equipment and details the characteristics of the different components. The second shows several examples of the measurements obtained for different conditions to illustrate the behavior of RoPOF transmission.

### A. Testbed for standalone NB-IoT RoPOF transmission

Fig. 1 is a schematic of the experimental setup developed to implement a RoPOF link based on a directly modulated laser diode (LD) to transmit downlink standalone NB-IoT signals through a SI-POF spool. The transmitter was based on one of the two LDs whose characteristics: threshold current, $I_{th}$, maximum recommended bias current, $I_{max}$, maximum emitted optical power, and peak wavelength $\lambda$, are specified in Table I. The LD was biased and modulated using an LDM9T mount, with DC and RF inputs. The range of bias currents, $I_b$, tested for each LD are also included in Table I. They were chosen to be as far as possible from the laser threshold current avoiding saturation of the optical receiver.

TABLE I
Laser Diodes' Characteristics

| LD model | $I_{th}$ (mA) | $I_{max}$ (mA) | $P$ (mW) | $\lambda$ (nm) | $I_b$ (mA) |
|---|---|---|---|---|---|
| L520P50 | 44 | 160 | 50 | 520 | 60-100 |
| L515A1 | 28 | 110 | 10 | 515 | 80-110 |

The modulated laser output signal was transmitted through a spool of one of the tested fibers whose characteristics are in Table II. At the receiver end, the optical signal was converted back to the electrical domain by an integrated receiver SPD-2_520 (SPD). This SPD has a nominal bandwidth of 1.2 GHz and its responsivity has been optimized for green light. We determined that the SPD entered saturation when the averaged received optical power was above 2.66 mW (4.25 dBm) at 520 nm.

A four-port Software Defined Radio (SDR) platform (Xilinx Zynq 7000) acted as digital/analog interface to generate the downlink NB-IoT signals. This platform can support analogue signals from 0.6 to 4 GHz with a bandwidth of 20 MHz. We chose the carrier frequencies at the lower and upper limits of range specified by the standard for downlink standalone NB-IoT: 925 MHz and 960 MHz [2]. An

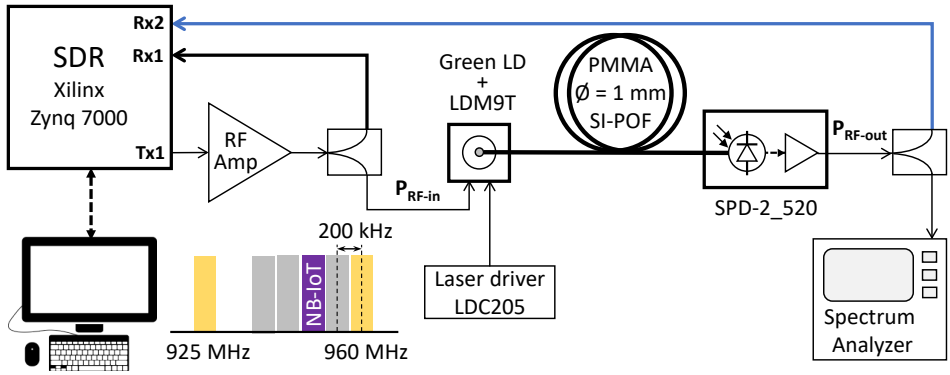


Fig. 1. Testbed for the RoPOF transmission of downlink standalone NB-IoT. The inset shows an schematic of the NB-IoT spectrum with the chosen carrier frequencies in yellow.

schematic of the NB-IoT spectrum is shown in the inset of Fig. 1, highlighting the two selected carrier frequencies. The spacing between carriers is 200 kHz and their bandwidth is 180 kHz. Standard-compliant downlink NB-IoT signals with QPSK modulation format were generated using a sampling frequency of 1.92 MHz and oversampled by a factor of 16 before transmission improving symbol and frequency offset estimation. The signal to be transmitted (output port Tx1) was then amplified using a RF amplifier (ZHL4240, with 39 dB minimum gain and 31 dBm saturation power). As our focus was to assess and reduce the non-linear effects of the opto-electronic system, we ensured that in all the experiments the amplifier was working inside its linear range. The amplified signal was divided to directly modulate the LD and to be fed-back to the SDR board (port Rx1) to serve as reference (input signal) and to monitor the behavior of the amplifier. At the receiver end, the signal from the SPD was also split and sent back to the SDR board through port Rx2 (output signal), and to a spectrum analyzer (Agilent 4404B) to measure received RF power.

The signals received at board ports Rx1 and Rx2 were downsampled to 1.92 MHz before being used in Section III to obtain the polynomial model and DPD signals. In the experimental validation described in Section IV, the DPD input signals were transmitted and the DPD output signals received at port Rx2 were analyzed to assess the performance of the DPD using our proposed method for sample selection. Signal generation, processing, and transmission performance analysis were carried out with a computer directly connected to the SDR platform using software based on the LTE toolbox of MATLAB.

We used three SI-POFs manufactured by Mitsubishi with a core diameter of 980 $\mu$m and different numerical apertures (NAs). The MH4001 (MH), also called ESKA-MEGA, is a low NA fiber specifically designed for bandwidth-demanding applications. The GH4002 (GH) is a more standard SI-POF with 0.5 NA, also termed ESKA-PREMIER. Finally, the BH4001 (BH) has a thermoset polymer jacket to make it resistant to high temperatures. Its higher NA implies that this fiber type presents lower bandwidth-length products. PMMA SI-POFs exhibit lower attenuation at green wavelengths compared to the more commonly used red, which typically shows attenuation values around 0.18–0.2 dB/m [29]. This motivated our use of green-emitting LDs as optical sources. The fiber characteristics are summarized in Table II. In the table, the range of spool lengths used for each fiber type is also specified.

TABLE II
SI-POFs CHARACTERISTICS

| Fiber | NA | Attenuation (dB/m) | Lengths (m) |
|---|---|---|---|
| MH4001 (MH) | 0.3 | 0.15 | 50-80 |
| GH4002 (GH) | 0.5 | 0.09 | 50-65 |
| BH4001 (BH) | 0.6 | 0.09 | 50 |

### B. Measurements of standalone NB-IoT over POF

The previously described testbed was used for evaluating the feasibility of large-core SI-POF based RoPOF links for standalone NB-IoT transmission. For this purpose, we measured the received signals for different lengths of the three SI-POFs (from 50 to 80 meters), using one of the two LDs in the transmitter biased with several currents. Not all fibers and lengths were tested with both LDs or for all bias currents. The more suitable were chosen for each experiment and will be specified in the text. Transmission performance was quantified in terms of Error Vector Magnitude (EVM) and Adjacent Channel Power Ratio (ACPR). In this subsection we present two representative analyses. On one hand, we show the performance for different fiber lengths of the increased-capacity MH fiber, which is the one that allowed for longer transmission lengths. On the other hand, we show the effect of LD bias current over the performance metrics in the case of the GH fiber, the most standard SI-POF. The temperature-resistant BH fiber was also tested although transmission performance was worse than with the other two fibers.

Figure 2 shows the EVM and ACPR (Fig. 2(a) and 2(b), respectively) as a function of the RF input power ($P_{RF-in}$ in Fig.1) for spools of different lengths of the MH fiber (60, 65, 70 and 80 meters). The LD employed was the L515A1, biased at a constant current of 110 mA, and the carrier frequency was 925 MHz. The gray lines show the values of EVM and ACPR of the reference signal in the corresponding graph. For all fiber lengths, both EVM and ACPR exhibit similar behavior:

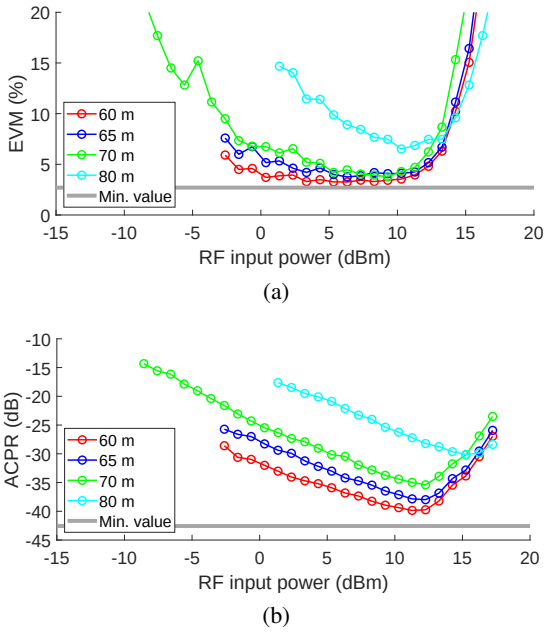


Fig. 2. EVM (a) and ACPR (b) for the RoPOF transmission of downlink standalone NB-IoT at 925 MHz. The signal was transmitted through various lengths of MH, using the L515A1 biased at 110 mA.

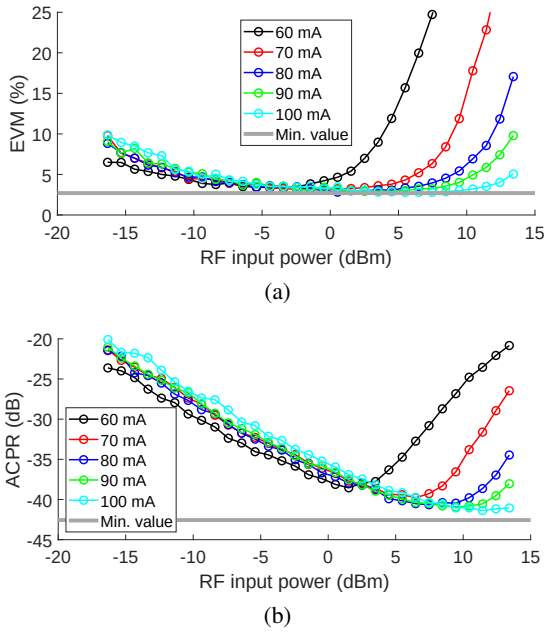


Fig. 3. EVM (a) and ACPR (b) for the RoPOF transmission of downlink standalone NB-IoT at 925 MHz. The signal were transmitted through 50 m of GH, using the L520P50 source, biased at multiple current levels.

their values initially decrease with increasing RF input power, reaching a minimum and subsequently rise at higher power levels. The high ACPR at low input powers is attributed to a low Signal to Noise Ratio (SNR), while the increase at high powers is indicative of system non-linearities. Notice that, at higher input powers, the EVM increases at the same RF input power level for all tested lengths. The differences between fiber lengths emerge at lower RF input powers, where longer fibers have higher EVM and ACPR, as SNR decrease with fiber length due to attenuation and modal dispersion. In any case, for all fiber lengths, there is range of input RF powers were the EVM value is below the standard limit of 17.5 % [1].

To assess LD bias current impact over transmission performance, we transmitted downlink NB-IoT at a carrier frequency of 925 MHz through a spool of 50 meters of GH fiber using the L520P50 biased at 60, 70, 80, 90 and 100 mA. The individual curves of EVM and ACPR shown in Fig. 3(a) and Fig. 3(b), respectively, exhibit the same general behavior that was described above. However, higher LD bias currents shift the onset of non-linear degradation in EVM and ACPR to higher input powers. We argue that, at lower bias currents, close to the LD threshold, the high Peak-to-Average Power Ratio (PAPR) of the downlink signal (9 to 12 dB) leads to clipping of its lowest amplitude values, resulting in signal degradation and spectral regrowth. Thus, the observed non-linear effects at high RF input powers can be attributed to the direct modulation of the LD, particularly when the bias is closer to threshold currents. Therefore, to save energy operating at lower bias currents, a technique to compensate non-linear effects must be applied.

To illustrate the effects of these non-linear effects over the NB-IoT standalone signals, we consider as an example the transmission through a 50 meters GH fiber at 960 MHz for a

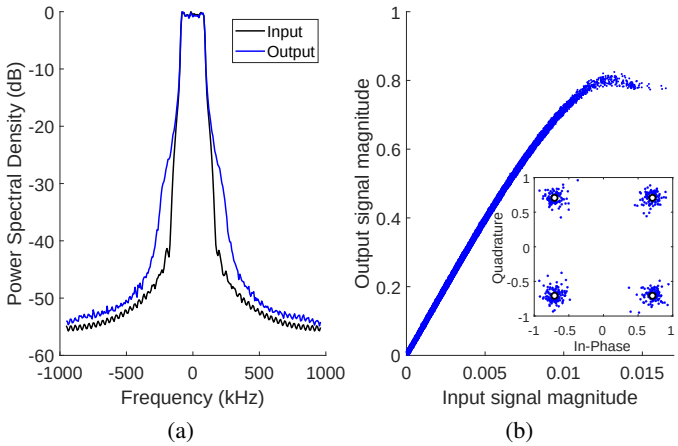


Fig. 4. Experimental results of NB-IoT transmission at 960 MHz, 80 mA bias current and 12.4 dBm RF input power: (a) Baseband PSD of input and output signals. (b) AM/AM characteristic. Inset: QPSK reference (black) and received (blue) symbols.

LD bias current of 80 mA and a $P_{RF-in}$ of 12.4 dBm. Fig. 4(a) shows the Normalized Power Spectral Density (PSD) of the input (black line), and output (blue line) signals. The system attenuation, $P_{RF-in} - P_{RF-out}$ was measured to be 37 dB in this case. This power difference is not visible in the graph because the PSDs have been normalized to reveal the spectral regrowth at both sides of the output signal spectra. The representation of the AM/AM characteristic shown in Fig. 4(b), confirms the non-linear system behavior, and the QPSK constellations in the inset illustrate its effects over the received data (blue). In this case, the EVM and ACPR of the input signal were 3.6 % and -46.6 dB, respectively. The EVM of the received waveform was 10.76 % that is still below the standard specified limit for QPSK modulation (17.5 %), but the ACPR was -38.33 dB,

slightly above the limit of -40 dB.

## III. DPD Cluster-based Sample Selection (C-SS) Method

In this section, we describe the proposed technique and use experimental signals to illustrate its operation, to choose its design parameters and to compare its performance to that of other sample selection techniques. It is structured in three subsections. The first explains the polynomial approximation of the Volterra series to obtain the coefficients of the system model and the DPD. The second describes the algorithm devised to obtain the data organized in clusters and the selection process (C-SS method). In the last section, the proposal is compared to similar approaches.

### A. Polynomial system and DPD models

To describe the non-linear behavior of the system and introduce the DPD technique, we used the classical approximation of the Volterra series by the memory polynomial model [21]. According to this model, the input-output characteristic of the system for discrete-time signals of $N$ samples can be described through the following expression:

$$y(n) = \sum_{m=0}^{M-1} x(n-m) \sum_{p=0}^{P-1} w_{mp} |x(n-m)|^p, \quad n = 1, \ldots, N \tag{1}$$

where $x(n)$ is the input signal, $y(n)$ is the output signal and $w_{mp}$ are the weights or coefficients of the model. In this equation, $P$ represents the order of the non-linearity, and $M$ the memory depth of the model.

To determine the values of memory depth and polynomial order that best suit the non-linear behavior of our system, we obtained the models ($w_{mp}$ coefficients) for different combinations of both design parameters: $P$ was varied from 1 to 11 and $M$ from 1 (no memory) to 3. Modeling performance for each case was obtained through the Normalized Mean Square Error (NMSE), calculated as the normalized mean square of the difference of the measured and modeled output signals:

$$\text{NMSE} = 10 \log_{10} \left( \frac{\sum_{n=1}^{N} |y(n) - \hat{y}(n)|^2}{\sum_{n=1}^{N} |y(n)|^2} \right), \tag{2}$$

where $y(n)$ is the output signal measured at Rx2, and $\hat{y}(n)$ is the output signal obtained with the model from the measured input signal (Rx1). Additionally, in order have into account not just modeling performance but also model complexity, we applied the Akaike Information Criterion (AIC) as an overall metric of the quality of each model that allowed us to select the $M$ and $P$ values best suited to our data. The AIC was computed using the following expression [30]:

$$\text{AIC} = 2N_p + 2\,(\text{NMSE}), \tag{3}$$

where $N_p$ is the number of parameters that in this case is $P \cdot M$, and NMSE is given in dB. The AIC and the model NMSE were estimated for 126 pairs of input and output signals measured under different conditions: fiber types, carrier frequencies, bias currents, and RF input powers. The values of the input RF power were chosen in the range where the non-linear effects were evident. Since no significant differences were observed among the results obtained under different conditions, we pooled the values from all signal pairs and calculated the average and standard deviation for each memory depth and polynomial order, which are shown in Fig. 5.

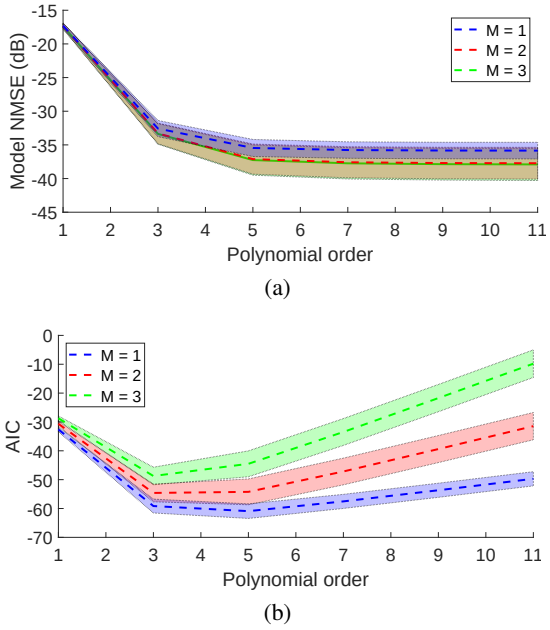


Fig. 5. Model NMSE error in dB (a), and Akaike information criterion (b), as a function or the polynomial order $P$ for memoryless models (blue), $M = 2$ (red) and $M = 3$ (green).

Fig. 5(a) illustrates that, although a memory depth of 2 yields a lower NMSE compared to the memoryless case, further increase in memory depth does not lead to substantial error reduction. Additionally, the average standard deviation for $M = 1$ is lower at 1.09, compared to 1.87 for memory depths of 2 and 3, respectively. The AIC values in Fig. 5(b) further support that, for standalone NB-IoT transmission over SI-POFs, memory depth is unnecessary which is consistent with expectations given the narrow 180 kHz bandwidth of a single channel. Although the average AIC curve for a one-tap memory reaches its minimum at a 5th-order polynomial, some individual cases show the minimum at order 7. Since the AIC increase between orders 5 and 7 is marginal, we adopt a 7th-order polynomial model.

Once the design parameters of the RoPOF polynomial model have been specified ($P = 7$, $M = 1$), we perform the correction by introducing the C-SS DPD technique using an indirect learning architecture (see Fig. 6). According to this approach, in a first training stage the coefficients in the feedback path are estimated using a Least Square (LS) criterion, while in a second stage the predistorter is set using a copy of the optimized feedback path coefficients.

Following the notation of Fig. 6 the estimated coefficients of the feedback path $\hat{\alpha}_{mp}$ are those that, for a given pair of system input and output signals $x(n)$ and $y(n)$, minimize the

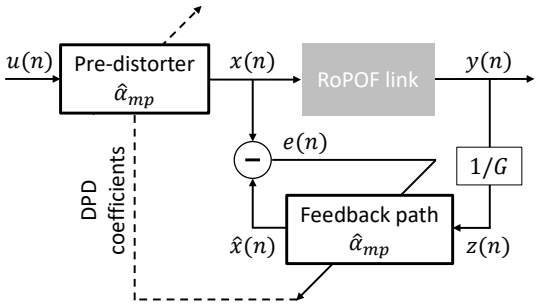


Fig. 6. Scheme of an indirect learning architecture for digital predistortion.

squared error:

$$\sum_{n=1}^{N} |e(n)|^2 = \sum_{n=1}^{N} |x(n) - \hat{x}(n)|^2 \tag{4}$$

In fact, $\hat{x}\,(n)$ can be described as a function of the polynomial coefficients and $y\,(n)$ as stated by the input-output characteristic of the feedback path, that in the memoryless case ($M = 1$) simplifies to:

$$\hat{x}\,(n) = \sum_{p=0}^{P-1} \hat{\alpha}_p z\,(n)\,|z(n)|^p, \tag{5}$$

where $\hat{\alpha}_{mp}$ reduces to $\hat{\alpha}_p$, as only $P$ coefficients are necessary, $z\,(n) = y\,(n)\,/G$ is the input to the feedback path and $G$ is the system gain obtained as:

$$G = \frac{1}{N} \sum_{n=1}^{N} \frac{|y\,(n)|}{|x\,(n)|} \tag{6}$$

In order to obtain the coefficients, Eq. (5) can be written in vectorial form as:

$$\hat{\mathbf{x}} = \mathbf{Z} \cdot \hat{\boldsymbol{\alpha}}, \tag{7}$$

where $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\alpha}}$ are the estimated signal at the output of the feedback path and the estimated coefficients in vector form that, for discrete-time signals with $N$ samples, can be written as:

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{x}\,(1) & \hat{x}\,(2) & \dots & \hat{x}\,(N) \end{bmatrix}^T$$

$$\hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat{\alpha}_0 & \hat{\alpha}_1 & \dots & \hat{\alpha}_{P-1} \end{bmatrix}^T \tag{8}$$

while matrix $\mathbf{Z}$ has the following structure:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_N \end{bmatrix}^T$$

$$\mathbf{z}_n = \begin{bmatrix} z\,(n)\,|z\,(n)|^0 & z\,(n)\,|z\,(n)|^1 & \dots & z\,(n)\,|z\,(n)|^{P-1} \end{bmatrix} \tag{9}$$

Consequently, the error can be expressed in vectorial form as:

$$\mathbf{e}\,(n) = \mathbf{x} - \mathbf{Z} \cdot \hat{\boldsymbol{\alpha}} \tag{10}$$

Finally, the least squared solution of this expression can be obtained by using the pseudoinverse of matrix $\mathbf{Z}$, so that the optimized coefficients can be calculated as:

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{Z}^H \cdot \mathbf{Z}\right)^{-1} \cdot \mathbf{Z}^H \cdot \mathbf{x} \tag{11}$$

### *B. Cluster-based Sample Selection (C-SS) algorithm*

As reducing the number of signal samples, $N$, to obtain the DPD coefficients has a favorable impact on the computational complexity independently on the applied method, it is desirable to consider some reduction technique to the collected data. Sample reduction can be directly performed by choosing a reduced number of the signals' samples, but the selected sample set would not convey the statistical distribution of the original signals. Therefore, organizing the data prior to sample reduction has been proposed to guide the selection process of the most significant input-output data samples. Previous proposals use the traditional histogram calculated dividing the magnitude of the input and output signals into uniform segments and counting the number of sample pairs in each bin [27], [28]. Here, we propose a classification method to organize the sample pairs into several clusters that tends to generate a more uniform mesh than the histogram and confers more flexibility to the selection process.

We begin forming pairs with the samples of the input and output signal magnitudes (or AM/AM characteristic). The classification of these sample pairs is performed by applying the k-means or Lloyd's algorithm [31], that organizes the sample pairs into $K$ clusters using the squared Euclidean distance. The algorithm proceeds as follows:

1) Choose $K$ initial cluster centers (centroids) using the k-means ++ algorithm [32]:
   - Select an observation randomly with uniform probability from the sample pairs as the first centroid.
   - Compute distances to this centroid and select next centroid randomly with a probability proportional to the distance to the first centroid.
   - Select each subsequent centroid with a probability proportional to its distance to the closest centroid already chosen.
   - Repeat until $K$ centroids are chosen.
2) Compute distances of all sample pairs to each centroid.
3) Assign each observation to the cluster with the closest centroid.
4) Recalculate $K$ new centroids as the average of the sample pairs in each cluster.
5) Repeat steps 2 through 4 until centroids do not change, or the maximum number of iterations is reached.

As a guide to assess a suitable number of clusters for our measurements, we have used the Silhouette Criterion [33]. The Silhouette coefficient is defined for a sample pair as an indicator of how closely it is matched to data within its cluster. A value near +1 indicates that the sample pair is far away from the neighboring clusters, while a value of 0 indicates that it is very close to the decision boundary between two neighboring clusters. For each of the 126 measured signals, we calculated the Silhouette coefficients with different number of clusters (from 4 to 64), and their average was obtained and presented in Fig. 7 in a bar diagram including the standard deviation. The results show that the maximum parameter value, 0.7560, is found for 4 clusters, but further research reveals that this value is too small to achieve a good sample selection. Thus, we chose $K = 16$, whose Silhouette value is only slightly

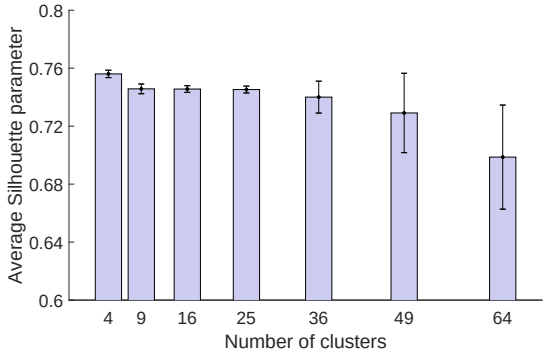


Fig. 7. Silhouette parameter averaged over all sample pairs of a set of measurements.

lower than the maximum, 0.7456, and also has the minimum standard deviation.

After the cluster classification has been completed, the reduction of the number of sample pairs was performed by dividing the number of samples in each cluster by a factor $R$ except when the cluster centroid was located at a significant distance from the linear behavior of the system. The idea is to keep the samples more representative of the non-linear characteristic that we aim to model and to correct with the DPD technique. Considering that $(c_{k_x}, c_{k_y})$ are the coordinates of the centroid of the $k$-th cluster in the plane defined by the AM/AM characteristic, the reduction indicator, $ri_k$, is calculated by comparing the difference of the centroid position to the linear behavior with its root mean squared (rms) value averaged over all centroids:

$$\left|c_{k_y} - G \cdot c_{k_x}\right| \underset{<}{\overset{ri_k \to F}{\geq}} \frac{1}{K}\sum_{i=1}^{K}\sqrt{\left|c_{i_y} - G \cdot c_{i_x}\right|^2} \quad (12)$$
$$ri_k \to T$$

This way, if this difference is lower than its rms value, $ri_k$ is false and the number of samples within the $k$-th cluster will be reduced by factor $R$. Otherwise, $ri_k$ is true, and the reduction will be $R' < R$. Fig. 8 illustrates the data classification and selection processes for the same case considered in Fig. 4. The sample pairs were organized in $K = 16$ clusters, and data selection was achieved by using $R = 16$ and $R' = 4$.

In Fig. 8(a), the AM/AM characteristic is shown representing the data within each cluster with a different color. This data classification renders a relatively uniform partition of the AM/AM characteristic. The cluster centroids are shown as black circles, and the black dashed line extrapolates the linear behavior shown at the lower input signal amplitudes, exposing that the three clusters with the highest centroid magnitudes deviate from the linear behavior. Fig. 8(b) shows the number of sample pairs found within each cluster (light-colored bars), and the selected number of pairs (dark-colored bars). According to the proposed method, more data pairs in the clusters whose centroids deviate from the fitting line are maintained in the selection process. This strategy tends to equalize the number of samples in the linear and non-linear portions of the AM/AM characteristics leading to a better estimate of the model and DPD coefficients. In this example, the number of selected samples relative to the initial total number that from this point

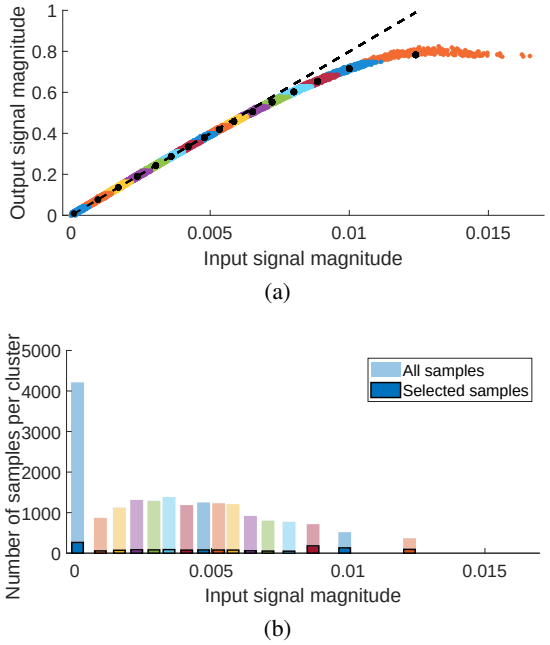


Fig. 8. System AM/AM characteristic after cluster classification (a); Number of sample pairs and of selected pairs within each cluster (b). In both graphs, different clusters are shown in different colors.

will be referred to as the Sample Selection Factor (SSF), was of $6.5 \times 10^{-2}$. In Section IV, we will demonstrate that lower SSFs can be achieved with good results.

### *C. Comparison of C-SS to undersampling and histogram-based classification*

In this subsection, we compare the performance of our proposal to other approaches of data reduction using the experimental signals obtained with the setup in 1. This time we chose as an example the results obtained for a spool of 65 meters of the MH fiber type. In these experiments, the laser was the L520P50 and the bias current 100 mA. The carrier frequency was 960 MHz and the RF input power was 12 dBm. We calculated the system and DPD coefficients reducing the number of sample pairs by three different methods: taking a reduced number of subsequent sample pairs (Undersampling Sample Selection, U-SS), obtaining the two-dimensional histogram and dividing the number of samples in each bin by a factor $R$ (Histogram Sample Selection, H-SS), and finally using the C-SS described above. The predistorted signal calculated with the DPD coefficients was used to compute the simulated DPD output signal. We varied the value of $R$ from 1 (no reduction) to 1500. We obtained the NMSE using Eq. (2), where $\hat{y}(n)$ is the simulated DPD output signal, and $y(n)$ is the target signal (that is, the input signal multiplied by the system gain, $G$). The results are shown in Fig.9, as the NMSE reduction (difference between the NMSE with and without DPD) for the different methods. For the H-SS and the C-SS methods, the algorithms were carried out 50 times. Each run, the number of selected samples was randomly selected within the bin or cluster, respectively. Thus, the curves for H-SS and C-SS in the figure show the average values with a continuous line and the standard deviations as error bars. On the right hand

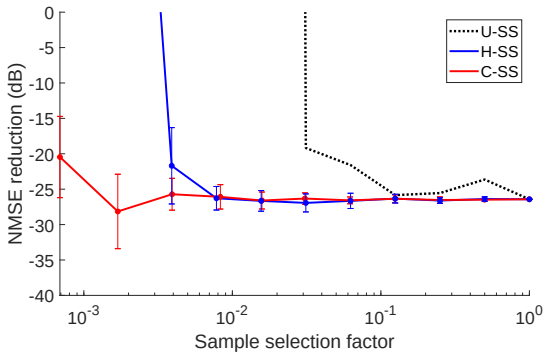


Fig. 9. Comparison of NMSE reduction for the DPD output signal obtained with U-SS, H-SS and C-SS relative to the uncorrected output signal

side of the graph, the value for an SSF of 1 is the DPD with no sample reduction ($R = 1$) that is the same for all methods. The U-SS method does not converge from an undersampling of $3 \times 10^{-2}$. The H-SS and the C-SS give similar results up to an SSF of $4 \times 10^{-3}$ but the H-SS does not converge for lower values. In these simulations, the results of the C-SS were still valid for an SSF of $7 \times 10^{-4}$, although with a larger standard deviation than for higher factors. The EVM and the ACPR were also obtained for all cases but their behavior is similar to the NMSE, and they are not shown here.

As far as the complexity is concerned, it is important to pay attention to the required number of operations in the proposed method to compare it to the other approaches. Before starting the training stage, the squared magnitudes of the sample pairs are obtained by $2 \cdot 2 \cdot N$ products and $2 \cdot N$ additions, resulting in a total of $6 \cdot N$ operations. These operations are necessary for both the H-SS and the C-SS, but not for the U-SS that feeds the input and output signals' complex values directly to the calculation of the DPD coefficients. The number of operations required in the proposed clustering method to obtain the cluster centroids (training stage) requires computing the square of Euclidean distances of all two-dimensional elements to the $K$ centroids i.e. $2 \cdot K \cdot N$ subtractions. Several iterations ($I$) can be necessary to obtain convergence, resulting in $I \cdot 2 \cdot K \cdot N$. As the convergence criteria imposed for the iterations to stop is very stringent, we studied the convergence of the algorithm and found that with a maximum of 100 iterations it succeeded 96 % of the times. When executed with a reduced number of iterations, the algorithm converges to a sub-optimal solution. We computed the NMSE error changing the number of iterations from 1 to 100 and did not find any significant variation of its average. We believe that the reason is that the initialization was performed using the k-means++ algorithm which provides a good seeding and that the improvement given in subsequent iterations does not have an impact in the extraction of the DPD coefficients [32]. The clustering overhead depends on the number of initial samples $N$, but it remains modest compared with the complexity of the polynomial model when the number of parameters $N_p$ is large. Using the complexity expressions reported in [27] for DPD coefficient estimation, illustrative numerical examples can be obtained. For instance, with 19200 samples, the clustering cost is $6.1 \times 10^5$, whereas an 11th-order depth-5 polynomial model ($N_p = 55$) requires $2.8 \times 10^8$ computations. Even when the number of samples is reduced to 300 (i.e., a sample selection factor of $1.5 \times 10^{-2}$), the DPD computational cost remains one order of magnitude higher than the clustering overhead ($5 \times 10^6$). For a very low number of parameters, such as in our case ($N_p = 7$), the clustering overhead becomes comparable to the DPD coefficient estimation complexity when using reduced data sets (e.g., $10^5$ operations for 300 samples). However, Fig.9 shows that convergence is only achieved at this low sample count when using a data classification method such as H-SS or C-SS. This highlights that the primary benefit of clustering is not only computational savings but also improved sample selection, which directly enhances DPD linearization performance. In fact, Kral et al. applied a method based on the histogram obtaining similar results to those obtained here with the C-SS when they optimized their selection process using a genetic algorithm [27].

## IV. Experimental Validation of the DPD Signal Transmission Performance

In this section, we present the experimental results obtained when the predistorted signals were transmitted over the RoPOF link. As a representative example, we chose the 50 meters spool of GH, as this fiber type is the most standard SI-POF with a NA of 0.5. We transmitted signals at carrier frequencies of 925 MHz and 960 MHz, for the L520P50 operated at three bias currents: 70, 80 and 90 mA. The RF input power values were set to produce an output signal whose initial ACPR was above the standard specified limit of -40 dB. These values were in the range from 7 dBm to 14 dBm, depending on the LD bias current and the carrier frequency, and are shown in Table III, along with the received optical power for each bias current.

TABLE III
Standalone NB-IoT Transmission Parameters

| LD bias current (mA) | 70 | 80 | 90 |
|---|---|---|---|
| $P_{RF-in}$ @925 MHz (dBm) | 8.96 | 12.13 | 13.44 |
| $P_{RF-in}$ @960 MHz (dBm) | 7.30 | 10.23 | 12.22 |
| Received optical power (dBm) | -0.46 | 0.41 | 1.24 |

For each case, the NB-IoT standalone downlink signals received at board port Rx1 (input signals) and Rx2 (output signals) were used to obtain the DPD coefficients for different SSFs using the C-SS algorithm. The predistorted input signals were calculated with these DPD coefficients and transmitted through the board to obtain the corresponding DPD output signal. This procedure was repeated several times for each SSF, calculating the averaged EVM and the ACPR of all signals involved to assess the performance of the C-SS method. In addition, the power loss between the uncorrected and the DPD output signals was estimated for each case. In addition, linearization performance was assessed in terms of NMSE, which was calculated using all samples of the experimental input and DPD output signals with Eq. (2).

To visualize the effect of the DPD with C-SS over the transmitted signals, Fig.10 compares the experimental results obtained with and without DPD. The selected example corresponds to a bias current of 80 mA and a carrier frequency of 960 MHz, the same used in Fig.4 and Fig.8, but with a SSF of $6.7 \times 10^{-3}$ in this case. Similar behavior was observed under other operating conditions and SSFs. Fig. 10(a), 10(b), 10(c), and 10(d) present the normalized PSDs, the AM/AM curves and the constellations for the experimental output signals without and with DPD, respectively. A comparison of the PSDs reveals substantial suppression of spectral regrowth in the DPD output signal (green) relative to the uncorrected output signal (blue), resulting in a PSD profile that closely matches the input signal (black) within the NB-IoT bandwidth (notice the gray boxes in the graph marking the positions of the nearest adjacent channels). Residual off-channel values in the DPD output signal are attributed to system noise that is beyond the modeling capabilities of the polynomial-based approach. This improvement is corroborated by the ACPR that, for this particular example, is -41.03 dB, 3.7 dB below the ACPR obtained after transmitting the signal without DPD, (-38.33 dB). Fig.10(b) displays the AM/AM representation of the original sample pairs (blue) along with the polynomial system model estimated with the selected dataset (cyan). The predistorted input signal calculated with the DPD coefficients to compensate the system non-linear behavior is also shown (red). The linearization performance is validated by the close alignment of the DPD output (green) with the ideal linear reference (black line). Although all samples are shown in the AM/AM plots, the DPD coefficients were calculated using only the reduced sample set. QPSK constellation diagrams in Fig.10(c) and Fig.10(d), further confirm the efficacy of the DPD technique. The EVM of the DPD output signal is 3.74 %, representing a 7-point reduction compared to the EVM of the uncorrected output signal that was (10.76 %). Additionally, the NMSE for the DPD output is -20.51 dB, indicating a 6 dB improvement over the initial value of -14.45 dB.

The results obtained for the 50-meter GH are shown in Fig. 11 for all tested conditions. For the carrier frequency of 925 MHz, the EVM and the ACPR are displayed as a function of the SSF on Fig.11(a) and Fig.11(b), respectively, and for 960 MHz, on Fig.11(c) and Fig.11(d). The values for the three bias currents are shown with different colors (blue for 70 mA, red for 80 mA and yellow for 90 mA). Individual data for the different measurements at each SSF are represented as colored dots, and their averages as solid lines of the same colors. For comparison purposes, the average EVM and ACPR of the output signals without DPD are included in the plots as dotted straight lines in the corresponding colors, while the thick gray lines show the average values for the original input signals (EVM of 3.5% and ACPR of -50.5 dB).

The plots show that both the EVM and the ACPR decrease when the DPD correction is applied, demonstrating its effectiveness for all bias currents and carrier frequencies. In fact, the EVM values for the DPD output signals remain close to 3.5 %, which is the average value for the original input signals. In contrast, ACPR values for the DPD output signals can be as low as -47 dB, although their average remains approximately

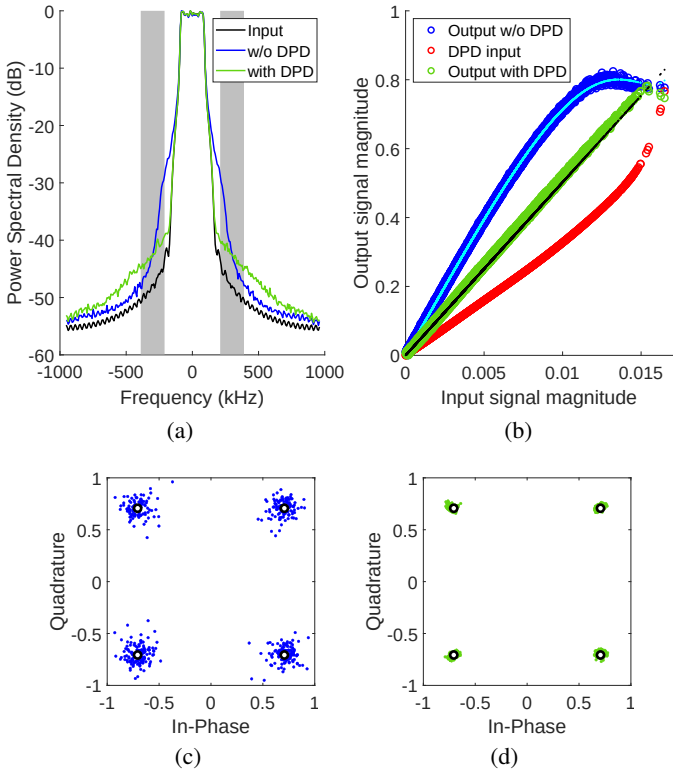


Fig. 10. Experimental results for a 50-m GH at 960 MHz with 80 mA bias current. (a) Baseband PSD of input signal and output signals with and without DPD. (b) AM/AM characteristics: output signal without DPD (blue), with DPD (green), DPD input signal (red) and polynomial model estimation (cyan). Black line shows lineal behavior. (c) QPSK received symbols without DPD. (d) QPSK received symbols with DPD.

5 dB above the minimum ACPR. Sample selection has very little impact on DPD performance, as both EVM and ACPR values are only marginally higher at lower SSFs. In any case, both metrics are below the standard defined values even for the lowest SSF which was $8.3 \times 10^{-4}$. The variability observed in ACPR results is primarily attributed to system noise introduced by components such as the amplifier, laser, and receiver, which DPD correction cannot mitigate, as behavioral models only address nonlinearities. This explain why simulated NMSE results (Fig.9) show lower variability compared to experimental ACPR results (Fig.11), where noise is introduced when the calculated DPD input is transmitted through the real system. In contrast, EVM plots show less variability because signal processing applied prior to their calculation, whereas ACPR is calculated without filtering.

As the general behavior is practically independent on the laser LD bias current and on the carrier frequency, we have pulled in all data to obtain the average power loss and NMSE reduction. The power loss was calculated as the difference between the uncorrected and the DPD output signal powers, and the NMSE reduction as was described before in subsection III-C. They are both shown together in Fig.12, where the power loss is represented on the left axis (blue) and the NMSE reduction on the right axis (red).

The plotted curves do not exhibit a discernible trend with respect to variations in the SSF. The average power loss values

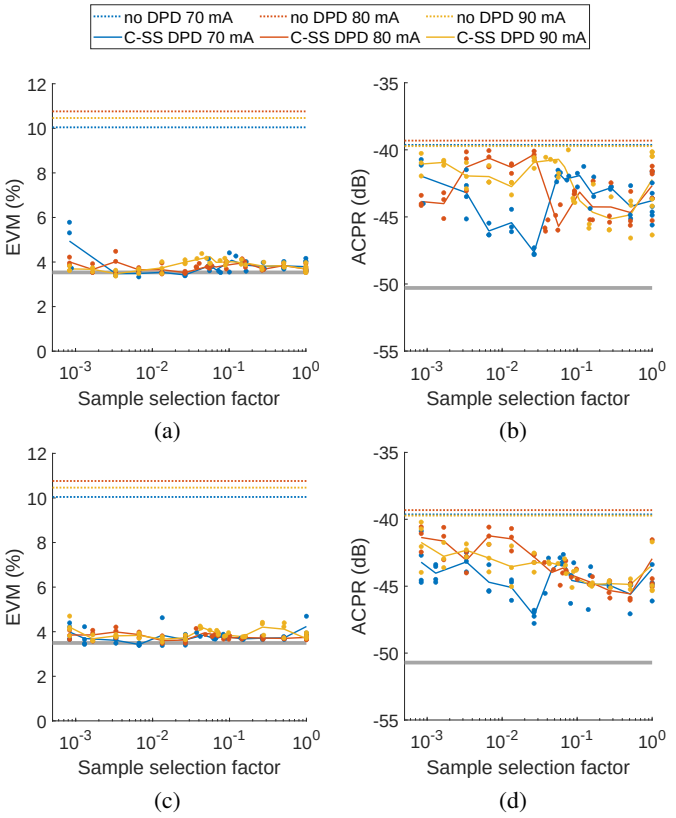


Fig. 11. EVM and ACPR versus SSF for a 50-m GH, at 925 MHz (graphs (a) and (b), respectively) and at 960 MHz (graphs (c) and (d), respectively). Different colors represent different bias. Dotted lines are values for the output signal without DPD. Continuous gray lines are the values for the input signal.

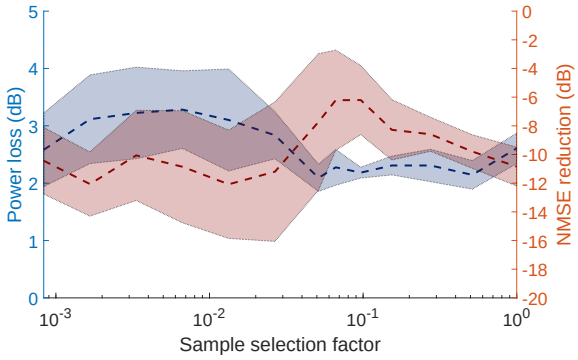


Fig. 12. Averages over carrier frequency and LD bias currents for a 50-m GH: Power loss between original and DPD output signals (blue) and NMSE reduction in the DPD relative to the uncorrected output signals (red).

lie within the range from 2 to 3 dB, while the NMSE reduction spans from 12 to 6 dB. Both parameters display substantial standard deviations, consistent with the high ACPR variability observed in Fig.11(b) and Fig.11(d). A significant conclusion derived from this analysis is the correlation between NMSE reduction and power loss: greater improvements in NMSE are generally associated with increased power loss. This trade-off suggests that enhanced linearization performance is achievable, albeit at the expense of reduced power efficiency.

## V. Discussion

We demonstrated effective transmission of narrowband RF signals over SI-POFs across distances ranging from 50 to 80 meters achieving excellent performance in terms of EVM. This supports their suitability for cost-effective deployments in short-area environments. Although this study solely focuses on single-channel downlink transmission for NB-IoT, our findings suggest the feasibility of transmitting more complex modulation formats and several channels. Specifically, we observed very low EVM values across a broad range of RF input powers for most fiber types and lengths that enable the transmission not only of QPSK, but also of higher-order schemes such as 16-QAM and even 64-QAM, in accordance with the 3GPP-defined performance thresholds: 12.5 % and 8 %, respectively [34].

A comparative analysis of EVM and ACPR as functions of input RF power across various fiber lengths reveals that attenuation and dispersion degrade the signal-to-noise ratio (SNR), thereby increasing the minimum achievable EVM and shifting its optimal point toward higher input power levels. These factors limit the maximum link distance achievable for a given fiber type. However, the observed non-linear effects at high RF input powers can be attributed to the transmitter's optical source when the bias-to-threshold current ratio is low but they are independent on fiber length. While directly modulated LDs offer simplicity and cost advantages, their performance is compromised by non-linearities near the threshold bias region. Increasing the bias current can alleviate these effects but at the expense of higher energy consumption and of introducing further non-linearities when eventually reaching the saturation segment of the laser characteristic curve. To address this trade-off, we propose the application of an effective digital predistortion technique to mitigate the non-linear effects, enabling energy-efficient operation by allowing the use of reduced laser bias currents without compromising performance.

In this work, we employ the polynomial model to characterize and compensate for system non-linearities. To reduce the volume of data required for estimating the model coefficients, we introduce a Cluster-based Sample Selection (C-SS) method, which organizes data into clusters that effectively capture their statistical properties. This clustering approach enables more efficient sample selection, minimizing the number of required data points. Although the training phase for determining cluster centroids introduces some computational overhead, our results show that even a single iteration yields a satisfactory classification, with negligible improvement in NMSE when increasing the number of iterations. Moreover, we assume that given a specific LD device and bias current, the training process can be performed offline and the clusters' centroids reused for the online DPD. Anyway, the data classification process scales linearly with the initial number of samples but, depending on the signal, preprocessing steps such as subsampling and zero-removal can be applied before the classification to reduce the initial dataset prior the cluster training stage. We have shown that narrowband signal transmission is adequately modeled using a 7th-order polynomial

without memory. However, for signals with wider bandwidths that require higher-order polynomial models with memory, the complexity reduction achieved by C-SS becomes increasingly significant. In these cases, the data reduction enabled by the C-SS justifies the classification overhead. Additionally, a clustering framework offers flexibility and simplicity for handling n-dimensional data, such as combining real and imaginary components of input and output signals, and provides a natural mechanism for incorporating memory depth.

Although we have proposed and validated a joined C-SS method with the polynomial model, C-SS can be integrated with other techniques to enhance sample selection efficiency. For example, it can support data reduction in feature extraction algorithms as suggested by Wang et al. [26], and significantly lower the computational burden of matrix factorization (QR decomposition or similar strategies). Also, in neural networks-based approaches like the Multi-Layer Perceptron (MLP) proposed by Pereira et al. whose training complexity scales with the number of neurons, memory and layers, an efficient data selection becomes critical [35]. C-SS prior to training can reduce the volume of data while preserving statistical relevance, thereby alleviating their computational load.

## VI. Conclusion

We have demonstrated experimentally the viability of a RoPOF link to transmit downlink standalone NB-IoT over large-core SI-POFs. The performance in terms of EVM was excellent for fiber lengths up to 80 meters for some fiber types. These results were obtained by directly modulating a low-cost laser diode, simplifying the hybrid wireless-optical link design. To compensate the non-linear effects of the opto-electronic components, we proposed a DPD technique, whose complexity was reduced by classifying the data using a cluster partition method (C-SS). This will permit to bias the LD at lower currents resulting in energy savings. Moreover, our approach can be easily extended to domains with a higher dimension, and can be combined with other techniques.

## References


[1] "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," 3GPP, Technical Specification TS-36.211 V17.3.0, Mar. 2023, release Mar. 17.

[2] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, 2022, second quarter.

[3] N. Chen and M. Okada, "Toward 6G internet of things and the convergence with RoF system," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8719–8733, Jun. 2021.

[4] Y. C. Manie, C.-K. Yao, T.-Y. Yeh, Y.-C. Teng, and P.-C. Peng, "Laser-based optical wireless communications for internet of things (IoT) application," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 24 466–24 476, Dec. 2022.

[5] Y. Wen, F. Yang, J. Song, and Z. Han, "Free-space optical integrated sensing and communication based on DCO-OFDM: Performance metrics and resource allocation," *IEEE Internet of Things Journal*, vol. 12, no. 2, pp. 2158–2173, Jan. 2025.

[6] C. Lethien, C. Loyez, J.-P. Vilcot, R. Kassi, N. Rolland, C. Sion, and P.-A. Rolland, "Review of glass and polymer multimode fibers used in a wimedia ultrawideband MB-OFDM radio over fiber system," *Journal of Lightwave Technology*, vol. 27, no. 10, pp. 1320–1331, May 2009.

[7] K. Muramoto, A. Inoue, and Y. Koike, "Noise and distortion reduction in ofdm radio-over-fiber link by graded-index plastic optical fiber," *IEEE Photonics Technology Letters*, vol. 32, no. 13, pp. 835–838, Jul. 2020.

[8] U. Adnan, Z. Abbas, A. Gulistan, and S. Ghafoor, "Radio over plastic optical fibers – a tutorial and review," *Journal of Optical Communications*, vol. 45, no. s1, pp. s929–s940, 2024.

[9] O. Ziemann, J. Krauser, P. E. Zamzowr, and W. Daum, "Application of polymer optical and glass fibers," in *POF Handbook: Optical Short Range Transmission Systems*, 2nd ed. Springer, 2008, pp. 593–664.

[10] E. Arrospide, I. Bikandi, I. García, G. Durana, G. Aldabaldetreku, and J. Zubia, "Mechanical properties of polymer-optical fibres," in *Polymer Optical Fibres: Fibre Types, Materials, Fabrication, Characterisation and Applications*, ser. Series in Electronic and Optical Materials, C.-A. Bunge, T. Gries, and M. Beckers, Eds. Woodhead Publishing, 2017, vol. 89, pp. 201–2016.

[11] F. M. A. Al-Zubaidi, D. S. Montero, and C. Vázquez, "SI-POF supporting power-over-fiber in multi-gbit/s transmission for in-home networks," *Journal of Lightwave Technology*, vol. 39, no. 1, pp. 112–121, Jan. 2021.

[12] O. López-Lapeña and J. Polo-Cantero, "Time-division multiplexing for power and data transmission on optical fibers," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 785–17 792, 2022.

[13] G. Stepniak, M. Schüppert, and C.-A. Bunge, "Polymer-optical fibres for data transmission," in *Polymer Optical Fibres: Fibre Types, Materials, Fabrication, Characterisation and Applications*, ser. Series in Electronic and Optical Materials, C.-A. Bunge, T. Gries, and M. Beckers, Eds. Woodhead Publishing, 2017, vol. 89, pp. 217–310.

[14] F. Forni, Y. Shi, N. C. Tran, H. P. A. van den Boom, E. Tangdiongga, and A. M. J. Koonen, "Multiformat wired and wireless signals over large-core plastic fibers for in-home network," *Journal of Lightwave Technology*, vol. 36, no. 16, pp. 3444–3452, Aug. 2018.

[15] M. Waseem, A. López, P. L. Carro, and M. A. Losada, "Cellular guardband NB-IoT performance over PMMA plastic optical fibers," *Journal of Lightwave Technology*, vol. 41, no. 23, pp. 7302–7308, Dec. 2023.

[16] V. Thomas, D. Liu, M. El-Hajjar, and L. Hanzo, "Experimental demonstration of plastic optical fibre-based digitised radio over fibre downlink," *Electronics Letters*, vol. 51, no. 21, pp. 1679–1681, 2015.

[17] J. A. Apolo, B. Ortega, and V. Almenar, "Hybrid POF/VLC links based on a single LED for indoor communications," *Photonics*, vol. 8, no. 7, 2021.

[18] M. Waseem, A. López, P. L. Carro, and M. A. Losada, "Performance of large-core SI-POF for the transmission of standalone NB-IoT," in *2025 25th Anniversary International Conference on Transparent Optical Networks (ICTON)*, 2025, pp. 1–4.

[19] Z. Bouhamri, Y. Le Guennec, J.-M. Duchamp, G. Maury, A. Schimpf, V. Dobremez, L. Bidaux, and B. Cabon, "Multistandard transmission over plastic optical fiber," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 11, pp. 3109–3116, Nov. 2010.

[20] A. Giovannini, M. U. Hadi, L. I. Prat, N. Neji, Z. G. Tegegne, C. Viana, A.-L. Billabert, J.-M. Laheurte, J. Nanni, D. Masotti, G. Tartarini, and J.-L. Polleux, "Improved nonlinear model implementation for VCSEL behavioral modeling in radio-over-fiber links," *Journal of Lightwave Technology*, vol. 40, no. 20, pp. 6778–6784, Oct. 2022.

[21] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[22] L. C. Vieira, N. J. Gomes, A. Nkansah, and F. van Dijk, "Behavioral modeling of radio-over-fiber links using memory polynomials," in *2010 IEEE International Topical Meeting on Microwave Photonics*, 2010, pp. 85–88.

[23] K. S. Yuva, K. B. K. Seetharam, and S. Ramanath, "Digital predistortion for narrowband IoT applications," in *2018 International Conference on Signal Processing and Communications (SPCOM)*, 2018, pp. 302–306.

[24] P. L. Carro, C. Mateo, P. García-Dúcar, J. de Mingo, and A. Valdovinos, "DPD linearization complexity reduction of remote radio heads in C-RAN with radio over fiber fronthaul," in *2018 48th European Microwave Conference (EuMC)*, 2018, pp. 843–846.

[25] Q. A. Pham, G. Montoro, D. López-Bueno, and P. L. Gilabert, "Dynamic selection and estimation of the digital predistorter parameters for power amplifier linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 10, pp. 3996–4004, Oct. 2019.

[26] Y. Li, X. Wang, and A. Zhu, "Sampling rate reduction for digital predistortion of broadband RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 3, pp. 1054–1064, Mar. 2020.

[27] J. Kral, T. Gotthans, R. Marsalek, M. Harvanek, and M. Rupp, "On feedback sample selection methods allowing lightweight digital predistorter adaptation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 1976–1988, Jun. 2020.

[28] T. Wang and P. L. Gilabert, "Mesh-selecting for computational efficient PA behavioral modeling and DPD linearization," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 1, pp. 37–40, Jan. 2021.

[29] M. A. Losada, M. Mazo, A. López, C. Muzás, and J. Mateo, "Experimental assessment of the transmission performance of step index polymer optical fibers using a green laser diode," *Polymers*, vol. 13, no. 19, 2021.

[30] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd International Symposium on Information Theory*, B. N. Petrov and F. Csáki, Eds. Tsahkadsor, Armenia, USSR: Akadémiai Kiadó, 1971, pp. 267–281.

[31] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[32] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007, pp. 1027–1035.

[33] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[34] "Evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception," 3GPP, Technical Specification TS-36.104 V10.14.0, Apr. 2021, release 10.

[35] L. A. M. Pereira, L. L. Mendes, C. J. A. Bastos-Filho, and A. C. S. Jr, "Machine learning-based linearization schemes for radio over fiber systems," *IEEE Photonics Journal*, vol. 14, no. 6, pp. 1–10, 2022.