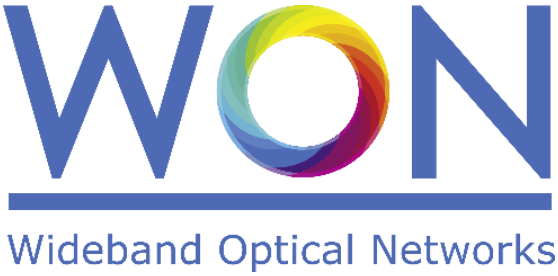
Marie Skłodowska-Curie (MSCA) – Innovative Training Networks (ITN)
H2020-MSCA-ITN European Training Networks



Wideband Optical Networks [WON]

Grant agreement ID: 814276

WP2 – Digital signal processing and system modelling

Deliverable D2.1 Digital pre-compensation techniques for wideband opto and/or electronic devices

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 814276.

Document Details

Work Package	WP2 – Digital signal processing and system modelling
Deliverable number	D2.1
Deliverable Title	Digital pre-compensation techniques for wideband opto and/or electronic devices
Lead Beneficiary:	Fraunhofer HHI
Deliverable due date:	31 July 2021
Actual delivery date:	10 November 2021
Dissemination level:	Public

Project Details

Project Acronym	WON
Project Title	Wideband Optical Networks
Call Identifier	H2020-MSCA-2018 Innovative Training Networks
Coordinated by	Aston University, UK
Start of the Project	1 January 2019
Project Duration	48 months
WON website:	https://won.astonphotonics.uk/
CORDIS Link	https://cordis.europa.eu/project/rcn/218205/en

WON Consortium and Acronyms

Consortium member	**Legal Entity Short Name**
Aston University	Aston
Danmarks Tekniske Universitet	DTU
VPIphotonics GmbH	VPI
Infinera Portugal	INF PT
Fraunhofer HHI	HHI
Politecnico di Torino	POLITO
Technische Universiteit Eidhoven	TUE
Universiteit Gent	UG
Keysight Technologies	Keysight
Finisar Germany GmnH	Finisar
Orange SA	Orange
Technische Universitaet Berlin	TUB
Instituto Superior Tecnico, University of Lisboa	IST

Abbreviations

BO	Bayesian optimization
DAC	Digital-to-analog converter
DAs	Driver Amplifiers
DP	Dual-polarization
DPD	Digital pre-distortion
DSP	Digital Signal Processing
FOM	Figure of merit
ILA	Indirect learning architecture
ML	Machine learning
OMFT	Optical multi-format transmitter
RRC	Root-raised cosine pulse
SI	System Identification
SUT	System under test
WB	Wideband

CONTENTS

EXECUTIVE SUMMARY

The present scientific deliverable is a part of the Work Package 2 “Digital signal processing and system modelling” of the ETN project WON “Wideband Optical Network”, funded under the Horizon 2020 Marie Sklodowska-Curie scheme Grant Agreement 814276.

The deliverable D2.1 describes the concepts of DSP algorithms to estimate and compensate for the impairments of prototyped components related to fibre propagation when transmitting over S- to L-band. The discussion carried out in this report is described as follows. First, the concepts of digital pre-distortion of opto electronic devices are explained and the requirements for its use in wideband operation are consequently reviewed. Then, this report introduces a novel nonlinear system identification scheme based on Bayesian optimization, which is used to characterize a standard C-band optical transmitter. It is shown that this scheme can reduce the convergence time by 46% to characterize black-box systems based on Volterra series with respect to other traditional approaches. Besides, when applied to different setup configuration schemes (e.g., different symbol rate, different amplifier gains) this new method is able to derive adaptable filter designs that can account for the strength of the imposed nonlinear distortions. Finally, the results of an experimental evaluation are exposed where this proposed nonlinear system identification technique is utilized to generate Volterra and memory polynomial pre-distortion filters in the S-, C- and L-band (approximately 120 nm). It is shown that the proposed solution outperforms standard pre-distortion methods such as linear pre-distortion and nonlinear pre-distortion based on a grid-search heuristic over 120 nm.

1. OVERVIEW ON DIGITAL PRE-DISTORTION FOR OPTO AND/OR ELECTRONIC DEVICES

Wideband (WB) optical systems have shown to be a promising solution to cope with the ever-growing increase in network traffic by efficiently utilizing the spectral resources of deployed optical fiber infrastructure. In WB systems, the main objective is to extend the wavelength division multiplexing (WDM) transmission window beyond the typical C+L-band (~5-10 THz) through the utilization of other optical amplification bands (e.g., O-, E- and S-band), which leads to an additional 44 THz of available transmission bandwidth in single-mode fibers. The realization of WB-based systems requires new devices, such as modulators and amplifiers, which cover broader optical bandwidths in comparison to the conventional C+L-band technology. However, despite recent advances, the technological immaturity and the associated infrastructure upgrade costs make this approach unfit for their short-term commercial deployment. Another alternative is to optimize standard C-band technologies for out-of-band use via the utilization of sophisticated digital signal processing (DSP) tools permitting the efficient correction of distortions that arise from transmitters in S/L-band operation [3]. In this regard, one promising solution is the use of digital pre-distortion (DPD) schemes [4], which is thoroughly discussed in the course of this report.

1.1 Volterra-based DPD

In the most general sense, DPD consists of building a nonlinear model that can be used to synthesize a DPD filter, in turn, employed to compensate for the transmitter-induced distortions. Conventionally, an optical transmitter consists of digital-to-analog converters (DACs) followed by a quad set of driver amplifiers (DAs) and a dual-polarization (DP) in-phase/quadrature (IQ) modulator. Its output is a continuous-time waveform that when represented in its discrete form, i.e., $r[n]$, can be modelled with respect to the discrete input sequence ($s[n]$) via a truncated, time-invariant Volterra series [5]. Generally, Volterra-based DPD of optical transmitters is defined in two major stages:

1) System Identification (SI) – A Volterra series is used to build a model R of the system under test (SUT) based on measurements of the input waveform s and output waveform r. The mathematical relation between the n-th sample of the output ($r[n]$) and input ($s[n]$) waveform is described by Equation 1, where τ_p is an arbitrary delay used for non-causal filtering realizations, $h_p[c_1, \dots, c_p]$ is the p-th order Volterra kernel coefficients and m_p is the corresponding p-th order memory length. When performing the SI, h_p can be estimated via adaptive algorithms (e.g., least-squares estimation [6]), where the kernel coefficients are obtained such to optimally fit the relation between r and s (Equation 1). Once the estimation of the kernel coefficients (h_p, $\forall\, p \in \{1,..,P\}$) is complete, an emulated output (r') can be generated by applying the exciting signal (s) to the obtained Volterra model R. In ideal conditions, $r = r'$. Figure 1(a) summarizes with a block diagram the logic implemented in the SI.

$$r[n] = h_0 + \sum_{p=1}^{P} \sum_{c_1=0}^{m_1-1} \dots \sum_{c_p=c_{p-1}}^{m_p-1} h_p[c_1, \dots, c_p] \times \prod_{i=1}^{p} s[n - c_i - \tau_p] \quad (1)$$

2) Signal Pre-Distortion – An inverse model S of the SUT is synthesized and operates on s to generate a pre-distorted signal $\hat{s}$. When $\hat{s}$ is applied to the SUT, the output waveform is s', which in ideal conditions, $s' = s$ (Figure 1(b)). To generate the inverse model S, the indirect

learning architecture (ILA) is often utilized [7]. The two major benefits of the ILA are the fact that the derived pre-distortion filter S is applicable to arbitrary input signals and that the ILA does not rely on a pre-defined architecture of the distorting system R. In most practical cases, the inverse model S can be expressed as a truncated, time-invariant Volterra series. However, the design of Volterra-based DPD filters suffers from scalability issues, since the number of computed kernel coefficients grows exponentially with the length of the memory effects. One possible solution is to employ the so-called memory polynomial (MP) architecture for the inverse model S. MPs represent a very compact subset of the Volterra series and mathematically correspond to rewriting Equation 1, such that, $h_p[c_1, \dots, c_p] = 0, \forall\, c_1 \neq \cdots \neq c_p$, which in simple words is equivalent to only considering the main diagonal of the p-th order Volterra operator h_p.

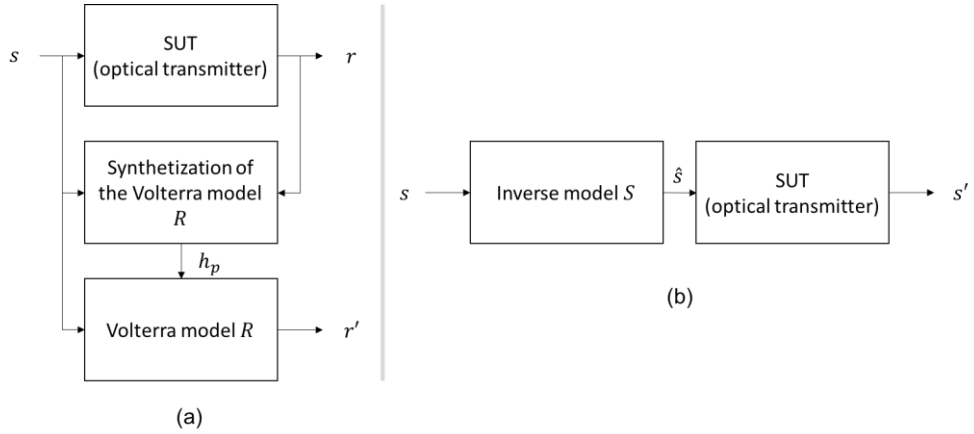


Figure 1: Block diagram of the: (a) Volterra-based SI for an optical transmitter, (b) signal pre-distortion.

After this brief introduction to the concepts concerning DPD of optical transmitters, we highlight in the following section the necessary requirements to adapt traditional DPD approaches for WB systems.

2. REQUIREMENTS FOR DPD IN WB SYSTEMS

One of the key facilitators to the design of WB systems is the capability to generate mechanisms that, by incorporating cognitive actions, can perceive current conditions, plan, decide, and act to optimize system performance. In simple words, with the expanded utilization of the telecom spectrum (e.g., with the addition of the S- and L-band), communication systems need to adapt their operation to the physical specificities imposed by each optical band/wavelength in an autonomous fashion, i.e., without (or with minimal) human intervention. In Volterra-based DPD, the memory vector ($\mathbf{m} = [m_1, \dots, m_p, \dots, m_P]$) that defines the architecture of the model S (Figure 2 shows h_1, h_2 and h_3 for $\mathbf{m} = [39, 5, 7, 3, 1]$) is an example of a parameter that can adapt the operation of the DPD scheme to the specific transmission conditions and functions as a good indicator of the strength of distortions of the modelled system. Traditionally, the tuning of $\mathbf{m}$ is empirically performed via manual configuration, or heuristic procedures inspired by grid-search [8], which are computationally expensive and, hence, inappropriate for cognitive applications.

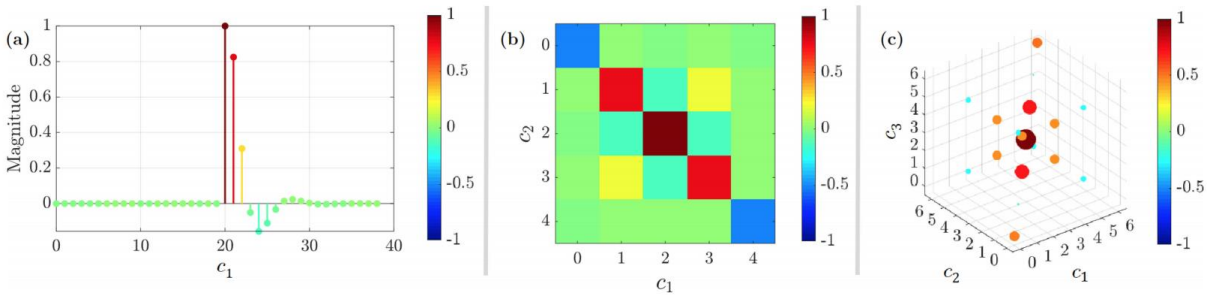


*Figure 2:*Normalized (a) first (h_1), (b) second (h_2), and (c) third (h_3) order Volterra kernel coefficients of a 5th-order synthetic filter for $\mathbf{m} = [39, 5, 7, 3, 1]$.

Given that the use of standard C-band technologies for out-of-band operation is expected to be a short-term alternative for the deployment of WB systems, one needs to think of efficient DPD strategies to adapt $\mathbf{m}$ to the specific impairments beyond the C-band (e.g., S- and L-band). This leads to three fundamental requirements for the development of DPD schemes for WB systems.

1) Time-efficiency – Given the dynamism of optical network operations and heterogeneity of the transmission parameters (e.g., modulation formats, symbol rates) deployed over multiple wavelengths, there is need for time-efficient tools that can quickly adapt $\mathbf{m}$ according to transmission conditions. Therefore, efficient SI must be carried out to accelerate the adaptation of the DPD scheme, thus avoiding latency in the network.
2) Adaptability – The architecture of the modelled system, embodied in $\mathbf{m}$, must take into account the wavelength dependency of the transmitter distortions. This means that for each wavelength, a specific model R must be obtained and consequently used to derive a specific inverse model S (Figure1(b)) that mitigates the nonlinear distortions.
3) Performance – The adaptation of $\mathbf{m}$ in the DPD scheme deployed in WB systems must demonstrate a reliable performance improvement in comparison to traditional approaches, in order to justify its use. Additionally, it is also desired that a reduced complexity is achieved, which also relaxes the criteria in terms of computational processing.

Given the aforementioned requirements for the design of DPD schemes for WB systems, we now introduce how the use of a machine learning (ML) approach can help us fulfil these important criteria.

3. BAYESIAN OPTIMIZATION FOR DPD

Machine learning (ML) has been claimed as a fundamental building block for the future of optical networks because its algorithms can learn from data, identify patterns and make decisions with minimal human intervention. More importantly, ML-based algorithms have shown great compatibility to solve standard optical communication problems, while reducing complexity of traditional approaches [9]. In this regard, we envision strong similarities between the optimization of the memory vector of a Volterra filter for DPD of optical transmitters and a design problem often dealt within ML applications, so-called hyperparameter tuning (HT) [10].

3.1 The HT problem

For decades, black-box models have attracted much attention in both academia and industry due to their efficiency on characterizing complex nonlinear systems. Nevertheless, an important challenge limiting the performance of these models is their intrinsic dependence on the selection of hyperparameters, i.e., any model parameter that can be set beforehand to control learning algorithms. Hyperparameters affect the speed and accuracy of the learning process of black-box models and, in contrast with conventional model parameters, hyperparameters cannot be easily estimated from the dataset.

As discussed in sub-section 1.1, in the SI, which is a black-box process, the least-squares algorithm is used to estimate a Volterra filter that optimally fits the input signal (s) to the SUT measured output (r). After that, an emulated SUT response (r') can be obtained by applying the resulting Volterra filter to s. The similarity between r' and r heavily relies on $\mathbf{m}$, which is considered a hyperparameter because it needs to be set beforehand to model the Volterra filter in which the p-th order kernel coefficients h_p are estimated. However, the lack of analytical formulas to calculate an optimal value for $\mathbf{m}$ and the restricted options of methods (e.g., manual or exhaustive search) opens the opportunity for the application of more sophisticated optimization techniques. One of these techniques is the Bayesian optimization (BO), which has its role in the context of DPD described in the following sub-sections

3.2 BO: a brief description

Given a black-box model that characterizes a SUT, in which a given arbitrary input s yields a response r, then the model accuracy can be evaluated through an objective function f. A hyperparameter entry, represented by a scalar (or vector) input θ, determines this evaluation, such that $f = f(\theta, s, r)$, which for simplicity can be written as $f = f(\theta)$, $f: \Theta \rightarrow \mathbb{R}$. In order to estimate the optimal model accuracy, f must be subject to an optimization process with respect to θ. However, in most cases, this optimization of f is bounded by two important restrictions, they are:

1) Computational complexity – The number of evaluations performed on f is limited, typically in the range of a few hundreds. This condition frequently arises because each evaluation takes a substantial amount of time.
2) Non-differentiability – Typically, first- and second-order derivatives of f with respect to θ, i.e., $f'(\theta)$ and $f''(\theta)$, are not obtainable, thus, preventing the application of methods like gradient descent, Newton's method, or quasi-Newton methods.

One of the fundamental ideas of BO is the capability to iteratively create a surrogate model $f^* = p(f|D)$ that estimates the value of the objective function f for an arbitrary input θ, i.e., $f(\theta)$, conditioned on a limited sub-set of n-observed data points ($\mathcal{D} = \{f(\theta_1), f(\theta_2), \ldots, f(\theta_n)\}$). To build f^*, the BO algorithm models $p(f|D)$ as a Gaussian Processes (GP) which permits to represent the posterior distribution $p(f|D)$ by the normal distribution $\mathcal{N}(\mu, \sigma^2)$, where [11]:

$$\mu(\theta) = \mathbf{k}^T(\theta)\mathbf{K}^{-1}\mathbf{z} \quad (2)$$

$$\sigma^2(\theta) = k(\theta, \theta) - \mathbf{k}^T(\theta)\mathbf{K}^{-1}\mathbf{k}(\theta) \quad (3)$$

Equations 2 and 3 are fully determined by the kernel covariance function $k: \Theta \times \Theta \rightarrow \mathbb{R}$, the n-by-1 vector $\mathbf{z}$ and the n-by-n Gram matrix $\mathbf{K}$. The kernel covariance function $k(\theta, \theta')$ is built by applying a covariance function between two arbitrary entries $\theta, \theta' \in \Theta$, i.e., $k(\theta, \theta') = Cov(\theta, \theta')$, consequently yielding the 1-by-n vector $\mathbf{k}^T(\theta)$, where $[\mathbf{k}(\theta)]_{u,1} = Cov(\theta, \theta_u)$ for $u \in \{1, \ldots, n\}$, and the scalar $k(\theta, \theta) = Cov(\theta, \theta)$. Finally, the vector $\mathbf{z}$ and the Gram matrix $\mathbf{K}$ are respectively defined as $[\mathbf{z}]_{u,1} = f(\theta_u)$ and $[\mathbf{K}]_{u,v} = k(\theta_u, \theta_v)$, where $u, v \in \{1, \ldots, n\}$.

Once the surrogate model $f^* = \mathcal{N}(\mu, \sigma^2)$ is updated, it can be directly used to estimate the global optimum of the function f, which is usually performed with the aid of an analytical acquisition function $a(\theta)$ (e.g., expected improvement [11]). As the number of observed data points in $\mathcal{D}$ grows, the resemblance between f^* and f increases and the algorithm rapidly learns the location of the global optimum of f without directly applying any algebraic operation on this function.

Now, we introduce how this concept can be integrated into SI of optical transmitters and used in the synthetization of a DPD filter.

3.3 BO for SI of optical transmitters

To incorporate BO into SI, we restructure the conventional SI block diagram in Figure 1(a) to search for the $\mathbf{m}$ that minimizes the identification error e_{SI} between r and r' (Figure 3), here quantified by the normalized mean squared error (NMSE) (Equation 4). This new scheme is hereafter referenced as Bayesian-based SI and was initially introduced in [12].

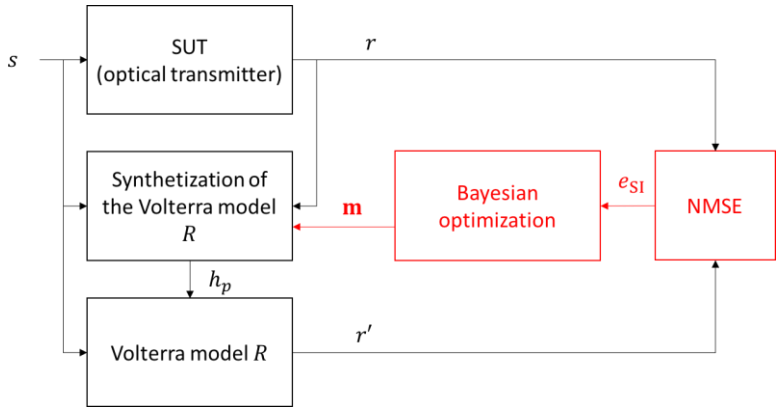


Figure 3: Bayesian-based SI for tuning of the SUT filter design ($\mathbf{m}$).

$$e_{\mathrm{SI}}(\mathbf{m}) = \frac{Var(r' - r)}{Var(r)} \tag{4}$$

Since $1 - e_{\mathrm{SI}}(\mathbf{m})$ is a proper figure of merit (FOM) of the SUT model accuracy, we mathematically write the HT problem as:

$$\max_{\mathbf{m} \in \mathcal{M}} [1 - e_{\mathrm{SI}}(\mathbf{m})] \tag{5}$$

One important boundary condition in this optimization problem is the complexity of the SUT model S, which can be quantified by the total amount of computed filter coefficients. Therefore, we set an upper bound to the amount of computed coefficients (M_C), thus limiting the space of possible solutions that can be searched in the BO. This boundary condition is expressed by the following inequality:

$$M_C \geq \sum_{p=1}^{P} \frac{(m_p + p - 1)!}{(m_p - 1)!\, p!} \tag{6}$$

3.4 Bayesian-based SI for time-efficiency

Before we investigate the application of the Bayesian-based SI in WB scenarios, it is important to demonstrate how this approach satisfies the three requirements presented in section 2.

The first requirement, time-efficiency, is assessed by comparing the proposed [13] approach to a heuristic memory tap optimization method introduced in [8]. In this scheme, a method was proposed to optimize the number of orders and memory taps of the Volterra series and follows the described method. First, the heuristic initializes the SI with a single-tap 1st-order Volterra filter, i.e., the least-squares estimation to obtain h_p is carried out for a memory vector $\mathbf{m} = [1]$. Then, m_1 is unitarily incremented until e_{SI} reaches an error floor. After that, m_1 is fixed and this procedure repeated for higher orders (e.g., m_2, m_3,…). The algorithm stops when an optimal e_{SI} is found after adding new orders. Despite providing an accurate estimation, this approximation gives rise to a difficult implementation constraint, i.e., the high number of evaluations of e_{SI} to find $\mathbf{m}_{opt}$, which arises from the grid-search-like nature of the assumed strategy.

To benchmark Bayesian-based SI [13] with [8], we synthesized a 5th-order Volterra model that emulates the response of a SUT and is depicted by Figure 2, for which $\mathbf{m}$ is known and equal to $\mathbf{m} = [39, 5, 7, 3, 1]$. Then, a random eight-level pulse-amplitude modulation (PAM-8) training sequence (s) with 10^5 symbols at 1 Sa/symbol is fed to the emulated SUT model to obtain the corresponding output signal (r). This pair of waveforms, i.e., s and r, is hence provided to the proposed Bayesian-based SI and to the heuristic approach. Then, both techniques are used to blindly learn the optimal $\mathbf{m}$. The SUT model was emulated with *VPItoolkit™ DSP Library*. For the Bayesian-based SI, we set a maximum filter complexity of $M_C = 155$, which ensures that $\mathbf{m}_{opt} \in \mathcal{M}$.

The Wallclock time (elapsed processing time) after each iteration loop for both techniques was then used as FOM. According to the results shown in Figure 4, Bayesian-based SI is able to reach the minimum identification error e_{SI}, 46% faster compared to the benchmarked approach. This indicates that using BO to identify the optimal memory tap distribution of a Volterra filter brings the advantage of reducing the convergence time, contributing to the time-efficiency requirement.

3.5 Bayesian-based SI for adaptable transmission scenarios

The second requirement presented in section 0 concerns the adaptability of the DPD scheme when subject to multiple transmission configuration setups. In order to demonstrate how the proposed Bayesian-based SI stands out as a promising candidate for adapting the modelling of the SUT, we evaluate its performance for two different scenarios, namely, (1) different DA gains and (2) different symbol rates.

In the both scenarios, we set up an optical back-to-back testbed and carry out an experimental validation of the proposed approach. For this analysis, we divided the experimental methodology into two main parts, i.e., (1) SI and (2) performance evaluation. In the first part (i.e., SI), a 96 GBd DP-16QAM "probe" signal was generated using a 215 random bit sequence followed by a root-raised cosine (RRC) pulse shaping filter with roll-off factor 0. The 96-GBd symbol rate ensures that the identification of the SUT Volterra model covers sufficient frequency components to guarantee that the test signals used in the performance evaluation will not be cut off in the frequency domain when applied to the inverse model S.

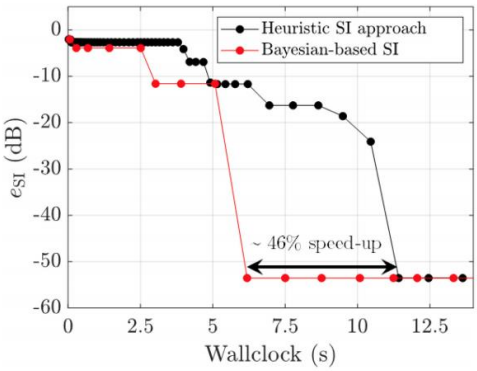


Figure 4: The proposed Bayesian-based SI is benchmarked against the heuristic approach introduced in [8], showing a 46% reduction in convergence time.

Then, the four sample sequences for the quadrature components (XI, XQ, YI and YQ) were uploaded to a 4-channel 120 GSa/s Keysight arbitrary waveform generator (AWG) with 3-dB bandwidth of 45 GHz (able to generate signals with frequency components above 45 GHz). The AWG was used to drive a 40-GHz optical multi-format transmitter (OMFT) from ID Photonics based on a high-bandwidth coherent driver module (HB-CDM). Together, the OMFT and the AWG comprise the SUT. An external cavity laser (ECL) at fixed wavelength (1550 nm) was used for the DP-IQ modulator and for the local oscillator (LO). Then, the optical signal was transmitted, received and digitized using an optical coherent receiver (70 GHz) followed by a 256 GSample/s real-time oscilloscope (RTO) with 110 GHz analog bandwidth. At the receiver DSP, Stokes space based polarization demultiplexing, clock recovery, resampling, frequency offset correction and carrier phase recovery were performed. The received and the transmitted samples of the probe signal quadrature components were finally provided to the Bayesian-based SI. At this stage, the Bayesian-based SI was processed for different filter complexities M_C [13].

In the second stage (performance evaluation), we tested the obtained solutions derived in the SI. This means that the kernel coefficients h_p obtained from the Bayesian-based SI were used in the synthetization of the model S (realized through ILA). Subsequently, the obtained DPD filter was used to pre-distort the test signal that unlike the probe signal was generated with a RRC pulse-shape (roll-off factor of 0.1), for which a single modulation formats (DP-64QAM) and two symbol rates (64 and 80 GBd) could be selected. After the transmission and reception of the pre-distorted test signal, the bit-error ratio (BER) was measured by counting the errors in 1 million bits per measurement point. Further details on the realization of the described experimental assessment can be found in [13].

3.5.1 Different amplifier gains

In order to verify that the proposed approach is able to adapt the DPD design according to the system configuration, we tested three gain configurations (g_1, g_2 and g_3) of the DAs to excite different degrees of transmitter nonlinearity. These gain configurations correspond to three levels of nonlinear system excitation: weak (g_1,), strong (g_2,) and highly nonlinear (g_3,), such that $g_1 < g_2 < g_3$.

The performance evaluation was carried out with a test signal configured to DP-64QAM at 64 GBd at maximum OSNR (44.9 dB). As can be seen in Figure 5, the filter architecture that results in lowest BER for the gain g_1 (indicated by the pink star-like marker) is a 3rd-order Volterra filter ($\mathbf{m}_{opt} = [76, 3, 9]$, $M_c = 50$). When gain g_2 is tested, not only m_1 and m_2 incorporate 4 and 2 more taps, respectively, but there is also the inclusion of a 4th order ($\mathbf{m}_{opt} = [80, 3, 11, 3]$, $M_c = 400$). Finally, for the highly nonlinear regime, represented by the gain g_3, the design for lowest BER is a 5th-order filter ($\mathbf{m}_{opt} = [200, 11, 15, 1, 3]$, $M_c = 1000$).

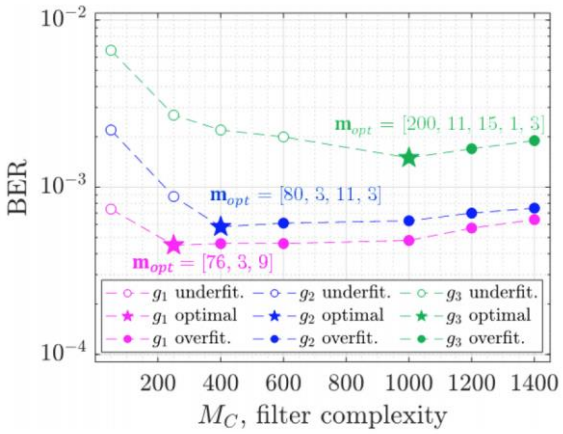


Figure 5: BER validation (at fixed OSNR = 44.9 dB) for different amplification gains $g_1 < g_2 < g_3$ as function of the filter complexity M_C. It is possible to distinguish an optimum among underfitted and overfitted filter designs.

3.5.2 Different symbol rates

As previously mentioned, during the SI a probe signal at 96 GBd was used to excite the SUT. This enables to synthesize a model R that covers sufficient frequency components to ensure that the test signals (at 64 and 80 GBd) will not be cut off in the frequency domain when applied to the inverse model S. The downside of such procedure is that the 96 GBd signal also excites transmitter nonlinearities at frequencies where the 64 or 80 GBd waveforms have no spectral support. Consequently, it is expected that at higher symbol rates the output signal of the SUT will manifest stronger distortions. The next logical step is to know whether our proposed scheme can adapt the filter design to reflect the necessity of additional or fewer memory taps to model the distortions induced in different symbol rate regimes.

For this test, the SI was performed in a similar way with respect to the analysis presented in sub-section 0, except for the fact that now a fixed DA gain (g_2) was set. In the performance evaluation, the test signal was configured to a fixed modulation format (DP-64QAM) while the chosen symbol rates were either set to 64 GBd or 80 GBd.

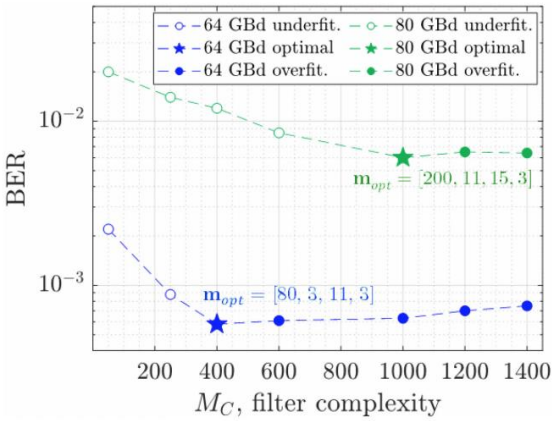


Figure 6: BER validation (at fixed OSNR = 44.9 dB) for different symbol rates (64 and 80 GBd) as function of the filter complexity M_C. It is possible to distinguish an optimum among underfitted and overfitted filter designs.

As depicted in Figure 6, the higher symbol rate regime (80 GBd) requires a filter complexity ($\mathbf{m}_{opt} = [200, 11, 15, 3]$, $M_C = 1000$) 2.5× higher than the 64 GBd case ($\mathbf{m}_{opt} = [80, 3, 11, 3]$, $M_C = 400$) to achieve the lowest BER. This demonstrates that the proposed approach can tailor specific filter designs for operation in different symbol rates.

4. DPD FOR WB SYSTEMS

At the beginning of this report, we mentioned that one alternative for short-term deployment of WB systems is to optimize standard C-band technologies for out-of-band use via the utilization of sophisticated digital signal processing (DSP) tools. In order to demonstrate that the proposed DPD approach based on BO is a promising enabler for the future of WB transmission, we experimentally demonstrate the application of this scheme to identify and mitigate distortions when using a C-band transmitter on a 120-nm S+C+L-band (1470 – 1590 nm) optical coherent setup. The experiment is described as follows and shown in Figure 7.

The experimental testbed used in this evaluation is shown in Figure 7. The SUT is composed by a 4-ch 120 GSa/s Keysight AWG, which was used to drive a commercially available C-band OMFT comprising a quad-set of DA and a LiNbO3 DP IQ-modulator. Three tunable ECL sources with linewidth < 100 kHz were utilized to cover the tested wavelengths in the S-band (1470 – 1520 nm), C-band (1530 – 1550 nm) and L-band (1560 – 1590 nm). To deliver a constant input optical power of 16 dBm to the transmitter, a thulium-doped fiber amplifier (TDFA) and an erbium-doped fiber amplifier (EDFA) were jointly used with the S- and L-band ECLs, respectively. Then a 96-GBd DP-16QAM signal was generated with a 215 random bit sequence and shaped with a root-raised cosine (RRC) pulse filter with roll-off factor 0. The four resulting sample sequences for the quadrature components, i.e., XI, XQ, YI and YQ, were uploaded to the AWG, which was used to drive the OMFT. In the following sequence, the signal is transmitted back-to-back, received and digitized using an optical

coherent receiver followed by a 200 GS/s real-time scope. For the measurements performed in C- and L-band, two individual ECLs were used as LOs for intradyne reception, whilst for the S-band, due to unavailability of a second S-band ECL, a self-homodyne reception was performed. At the receiver DSP, Stokes space based polarization demultiplexing, resampling, frequency offset correction and carrier phase recovery were respectively carried out. At last, the demultiplexed and the reference samples of the quadrature components were used as inputs for the optimization of $\mathbf{m}$ of the SUT Volterra model and provided to the ILA algorithm. Then, the synthetization of four DPD filter types is carried out. To benchmark our approach, we first synthesize a linear and a Volterra DPD filter, where the tap distributions are chosen in the following way. First, for the Volterra filter, we follow a similar approach to what is proposed in [8], i.e., the taps are incremented until the MSE curve reaches an error floor. However, unlike [8], we assumed increments of 100, 2, 2, 1 and 1 taps, for p = 1, 2, 3, 5 and 5, respectively, leading us to $\mathbf{m} = [500,10,10,5,5]$. For the linear case, we simply set $\mathbf{m} = [500]$. At last, we synthesize a Volterra and a MP DPD filter using the $\mathbf{m}$ obtained from the BO (hereafter referenced as autonomous tap optimization). With all these four DPD filters, we obtain the pre-distorted samples using a RRC pulse-shaped (roll-off factor of 0.1) DP-32QAM at the symbol rate of 64 GBd. Then the signal is transmitted over 40-km SSMF and a noise-loading stage is used to set a constant optical signal-to-noise ratio (OSNR). At the receiver, chromatic dispersion compensation was added to the DSP chain and BER calculated over 1 million bits.

The first logical step is to investigate how the SUT behaves in out-of-band operation. For that, we initially assess the performance of the transmitter when used in the S-band. Figure 8(a) shows the BER curves of the transmission experiment at a fixed OSNR of 32.5 dB when the linear ($\mathbf{m} = [500]$) and Volterra ($\mathbf{m} = [500,10,10,5,5]$) DPD filters without autonomous tap optimization are used. As can be seen, Volterra-based DPD significantly improves the system-level performance by reducing BER in comparison to linear DPD, revealing that nonlinear DPD is key to enable the utilization of standard C-band technology in MB regime. Then, the autonomous DPD is performed and the optimized DPD Volterra and MP filters are tested. In Figure 8(b) , we see that the optimized DPD filters not only improve the Q-factor for the S-band, but also over the entire C/L-band interval. This difference becomes even more significant when transmission is performed at lower wavelengths (e.g., 1470 nm), where the gain by autonomously optimizing the memory tap distribution nears 0.4 dB in Q-factor,

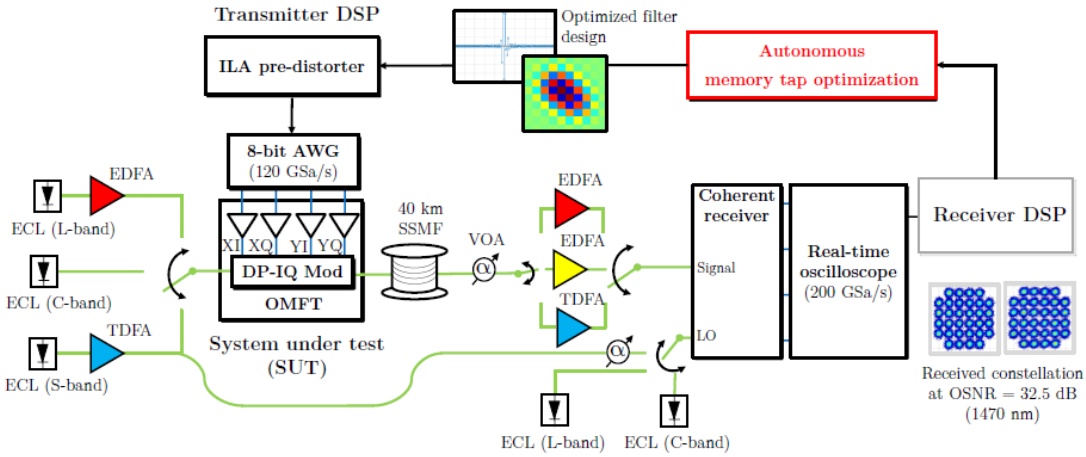


***Figure 7**: Experimental setup. VOA: variable optical attenuator.*

as indicated in Figure 8(b). When filter complexity is evaluated in terms of the number of computed coefficients (Figure 8(c)), it is also possible to infer that at 1470 nm the autonomous DPD provides more compact model designs by reducing the number of computed coefficients of the Volterra and MP DPD schemes by 48% and 65%, respectively, when compared to Volterra without autonomous tap optimization.

The results shown in Figure 8, thoroughy discussed in [14], demonstrate the impact of an autonomous DPD scheme based on BO and that performs adaptation of memory tap distribution for different wavelengths. It has been shown that over a total optical bandwidth of 120 nm, the autonomous DPD performed with Volterra and memory polynomial filters can improve system-level performance, in comparison to a scheme without the autonomous tap optimization. This improvement reaches 0.4 dB in Q-factor at 1470 nm at the gain of reducing filter complexity by 48% and 65%, for Volterra and memory polynomial DPD filters, respectively. The application of such scheme has revealed the importance that self-adaptive DSP algorithms have on improving the use of traditional C-band technologies in WB transmission.

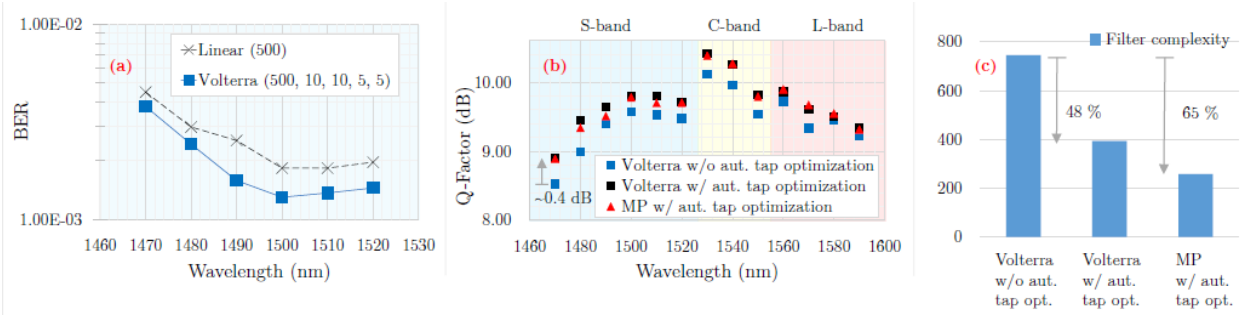


Figure 8: a) BER curves for fixed OSNR = 32.5 dB when linear (m = [500]) and Volterra (m = [500,10,10,5,5]) DPD are used in the S-band. (b) Q-factor curves for OSNR = 32.5 dB when Volterra (m = [500,10,10,5,5]) is compared with autonomous DPD using Volterra and MP filters. (c) Filter complexity measured in number of computed coefficients for DPD schemes of Figure 8(b) at 1470 nm.

5. REFERENCES

[1] A. Ferrari et al., "Assessment on the Achievable Throughput of Multi-Band ITU-T G.652.D Fiber Transmission Systems," in Journal of Lightwave Technology, vol. 38, no. 16, pp. 4279-4291, 15 Aug.15, 2020, doi: 10.1109/JLT.2020.2989620.

[2] L. Galdino et al., "Optical Fibre Capacity Optimisation via Continuous Bandwidth Amplification and Geometric Shaping," in IEEE Photonics Technology Letters, vol. 32, no. 17, pp. 1021-1024, 1 Sept.1, 2020, doi: 10.1109/LPT.2020.3007591.

[3] M. Sena et al., "Performance Evaluation of InP-Based DP-IQ Modulators for Multiband Transmission Systems," 2020 22nd International Conference on Transparent Optical Networks (ICTON), 2020, pp. 1-4, doi: 10.1109/ICTON51198.2020.9203554.

[4] P. W. Berenguer, T. Rahman, A. Napoli, M. Nölle, C. Schubert, and J. K. Fischer, “Nonlinear digital pre-distortion of transmitter components,” IEEE/OSA J. Lightw. Technol., vol. 34, no. 8, pp. 1739–1745, Apr. 2016.

[5] M. Schetzen, The Volterra and Wiener Theories of Nonlinear Systems. Malabar, FL, USA: John Wiley Sons, 1980.

[6] R. D. Nowak, “Penalized least squares estimation of volterra filters and higher order statistics,” IEEE Trans. Signal Process., vol. 46, no. 2, pp. 419–428, Feb. 1998.

[7] C. Eun and E. J. Powers, “A new volterra predistorter based on the indirect learning architecture,” IEEE Trans. Signal Process., vol. 45, no. 1, pp. 223–227, Jan. 1997.

[8] A. Richter, S. Dris, and N. André, “On the analysis and emulation of nonlinear component characteristics,” in Proc. Opt. Fiber Commun. Conf. Exhib., San Diego, CA, USA, 2019, Paper Th.1.D.1.

[9] F. Musumeci et al., “An overview on application of machine learning techniques in optical networks,” IEEE Commun. Surv. Tut., vol. 21, no. 2, pp. 1383–1408, Nov. 2018.

[10] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, “Hyperparameter tuning for big data using bayesian optimisation,” in Proc. Int. Conf. Pattern Recognit., Cancun, Mexico, Dec. 2016, pp. 2574–2579.

[11] J. Mockus, V. Tiesis, and A. Zilinskas, The Application of Bayesian Methods For Seeking the Extremum. New York, NY, USA: North Holland, 1978.

[12] M. Sena, M. Sezer Erkilinc, T. Dippon, B. Shariati, R. Emmerich, J. K. Fischer, and R. Freund, “An Autonomous Identification and Pre-distortion Scheme for Cognitive Transceivers using Bayesian Optimization,” presented at the European Conference Exhibition Optical Communication, Brussels, Belgium, Dec. 2020, Paper Tu.1.D.7.

[13] M. Sena *et al*., "Bayesian Optimization for Nonlinear System Identification and Pre-Distortion in Cognitive Transmitters," in *Journal of Lightwave Technology*, vol. 39, no. 15, pp. 5008-5020, Aug.1, 2021, doi: 10.1109/JLT.2021.3083676.

[14] M. Sena, R. Emmerich, B. Shariati, J. K. Fischer, and R. Freund, "Evaluation of an Autonomous Digital Pre-distortion Scheme for Optical Multiband Systems," in in Proc. Opt. Fiber Commun. Conf. Exhib., San Francisco, CA, USA, 2021, Paper F4D.5.