

Temporal Energy Analysis of Symbol Sequences for Fiber Nonlinear Interference Modelling via Energy Dispersion Index

Kaiquan Wu, *Student Member, IEEE*, Gabriele Liga, *Member, IEEE*, Alireza Sheikh, *Member, IEEE*, Frans M. J. Willems, *Life Fellow, IEEE*, and Alex Alvarado, *Senior Member, IEEE*

Abstract—The stationary statistical properties of independent, identically distributed (i.i.d.) input symbols provide insights on the induced nonlinear interference (NLI) during fiber transmission. For example, *kurtosis* is known to predict the modulation format-dependent NLI. These statistical properties can be used in the design of probabilistic amplitude shaping (PAS), which is a popular scheme that relies on an amplitude shaper for increasing spectral efficiencies of fiber-optic systems. One property of certain shapers used in PAS—including constant-composition distribution matchers—that is often overlooked is that a time-dependency between amplitudes is introduced. This dependency results in symbols that are *non-i.i.d.*, which have time-varying statistical properties. Somewhat surprisingly, the effective signal-to-noise ratio (SNR) in PAS has been shown to increase when the shaping blocklength decreases. This blocklength dependency of SNR has been attributed to time-varying statistical properties of the symbol sequences, in particular, to variation of the symbol energies. In this paper, we investigate the temporal energy behavior of symbol sequences, and introduce a new metric called *energy dispersion index* (EDI). EDI captures the time-varying statistical properties of symbol energies. Numerical results show strong correlations between EDI and effective SNR, with absolute correlation coefficients above 99% for different transmission distances.

Index Terms—Constant-Composition Distribution Matching, Fiber Channel Models, Fiber Nonlinearities, Probabilistic Amplitude Shaping.

I. INTRODUCTION

CONSTELLATION shaping (CS) and forward error correction (FEC) are two crucial elements to realize near capacity-achieving transmission for the additive white Gaussian noise (AWGN) channel. In recent years, a novel coded modulation framework called probabilistic amplitude shaping (PAS) [1] has attracted wide attention for its capacity-achieving performance. PAS elegantly integrates FEC and probabilistic shaping (PS) [2], where the PS functionality

This work is supported by the Netherlands Organisation for Scientific Research via the VIDI Grant ICONIC (project number 15685). The work of Alex Alvarado is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 757791). The work of G. Liga is supported by the EuroTechPostdoc programme under the European Union’s Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No 754462).

The authors are with the Information and Communication Theory Lab, Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands (e-mails: {k.wu, g.liga, a.sheikh, f.m.j.willems, a.alvarado}@tue.nl).

A. Sheikh is with imec, Holst Centre, High Tech Campus 31, 5656 AE Eindhoven, The Netherlands (email: alireza.sheikh@imec.nl).

is enabled by an amplitude shaper. One popular amplitude shaper is constant-composition distribution matcher (CCDM) [3]. Other well-known shapers include multiset-partition distribution matcher (MPDM) [4], product distribution matcher (PDM) [5], enumerative sphere shaping (ESS) [6], [7], and shell mapping [8].

In the context of fiber optical communications, numerous studies about PAS have been carried out in simulation [9]–[11] and experiments [12], [13]. Record spectral efficiencies (SEs) have been achieved in field trials [14], [15]. However, unlike the AWGN channel, SEs in the nonlinear fiber channel are limited by the Kerr effect. This effect causes nonlinear interference (NLI), which becomes a substantial part of the total noise experienced by the transmitted signals [16].

Mitigation of NLI can be achieved by optimizing statistical properties of the transmitted symbol sequences. Given a number of constellation points, it is possible to change the constellation geometry [17], [18] or the probability mass function (PMF) of the constellation symbols [9], [19], [20]. These techniques are often referred to as geometric and probabilistic shaping, respectively. In both cases, by assuming symbols to be independent identically distributed (i.i.d.), the stationary statistical properties of symbols are optimized. In the former, the support of the PMF is optimized, while in the later, the probabilities are optimized.

An alternative approach is to manipulate the *temporal* structure of the symbol sequence, which is believed to exert great influence on the NLI [21], [22]. The NLI can be modeled as inter-symbol interference (ISI) [23], i.e., the NLI experienced by a transmitted symbol only depends on adjacent symbols. The number of interfering symbols is usually referred to as channel memory. Therefore, transmitted symbols that comply to certain temporal structures could be used to suppress NLI. One example of this approach is the so-called temporal shaping introduced in [22], which generates correlated symbols from a ball-shaped constellation. By taking the correlation between symbol energies into account, [22] showed that transmitting correlated symbols leads to a new correlated term compared to i.i.d. symbol sequences (see [22, Eqs. (5) and (9)]). Temporal shaping was later realized using a finite state machine [24]. Although improved tolerance to nonlinearities can be achieved by controlling the temporal structure of the symbol sequences, a guiding principle for the optimization of the temporal structure is still unknown. This could be due to the fact that the statistical analysis of non-

i.i.d. symbol sequences is in general more difficult than its i.i.d. counterpart.

PAS generates dependent symbols because the employed shaper imposes hidden temporal structure on the amplitude blocks. Recently, several studies based on the PAS architecture have observed the impact of the resulting temporal structure on the effective SNR. The effective SNR has been shown to depend on the shaping blocklength: using short shaping blocklengths has been shown to offer significant effective SNR gains due to a weaker presence of nonlinearities [11], [25]–[27]. In [28], the relationship between carrier phase recovery and shaping blocklength on the nonlinear shaping gain was studied. In addition, the symbol mapping strategy, i.e., the way the shaped amplitudes are mapped to multi-dimensional symbols, has also been shown to be a crucial factor in the effective SNR performance of the system. It was observed in [25], [29] that instead of using four amplitude shapers independently for I/Q and X/Y dimensions, using one amplitude shaper across four dimensions improves the effective SNR.

The assumption of i.i.d. input symbols can be justified in scenarios such as (i) uncoded transmission, and (ii) systems employing a random interleaver that breaks dependencies between symbols. Based on the i.i.d. assumption, state-of-the-art Gaussian noise (GN) and enhanced Gaussian noise (EGN) models provide insights about the connection between the stationary statistical properties of the transmitted signal and the NLI. The GN model concludes that the NLI power scales as the cube of the average symbol energy [30]–[32]. By relaxing the *Gaussianity* assumption of the GN model [33], the EGN model shows that the standardized fourth moment (also known as kurtosis¹) of the transmitted symbols is an important metric in predicting the modulation-dependent NLI [34]–[36]. To mitigate the NLI, an optimized PMF was designed in [20] with the help of kurtosis. In [37], the blockwise composition of the symbol sequences having low kurtosis was selected for transmission.

To address the scenario of non-i.i.d. input symbols, the heuristic finite-memory GN model was proposed in [21], which considers a time-windowed symbol energy within the channel memory for the computation of the NLI. Recently, to analyze the temporal structure of symbol sequences, it was shown in [25], [38] that that symbol sequences with fewer clusters of identical symbols offer reduced NLI. A heuristic metric called run ratio was also proposed in [25]. The perturbative time-domain model of [39] can in principle also be used to analyze non-i.i.d. input symbols.

The discussion above leaves multiple open questions about the effects of the temporal structure of the symbol sequences on the NLI. Perhaps the most important question is that a metric that accounts for time-varying statistical properties of symbol sequences and enables a precise NLI prediction is still unknown. This paper is devoted to presenting a detailed explanation for the blocklength dependency on effective SNR,

¹The sixth-order moment reflects self-channel interference, which is marginal in long distance transmission using wavelength-division multiplexing (WDM).

and introducing a metric that is able to capture the effect of time-varying statistics on the NLI.

The contributions of this paper are three. First, we show that the temporal structure of the symbol sequence shaped by CCDM yields correlated symbols. The second contribution is to introduce a novel metric called energy dispersion index (EDI) to evaluate the variance of the symbol energies within a window. This metric is inspired by the channel models in [21], [22] and is a function of the autocorrelation function of the symbol energy’s variance within a window. An almost perfect correlation between the EDI and the effective SNR for different transmission distances is observed in numerical analysis using CCDM. Lastly, we also give analytical expressions for the autocorrelation and the EDI of the QAM symbol sequences generated by CCDM.

The rest of the paper is organized as follows. A review of the channel memory estimation and relevant channel models for the optical fiber channel is presented in Sec. II. In Sec. III, the statistical properties of symbol sequences shaped by CCDM are studied. The EDI, which is the main contribution of the paper, is presented along with the simulation results in Sec. IV and Sec. V. Conclusions are drawn in Sec. VI.

II. MODELING NLI WITH WINDOWED ENERGY

In this section, the estimation of memory in the fiber-optic channel and the effective SNR considering the NLI are reviewed. We then discuss the effects of temporal energy behavior of symbol sequences on the NLI, from the perspectives of channel models with memory [21], [22], [39]. We end this section with a review of metrics available in the literature that predict the NLI induced by input symbols.

A. Channel Memory and Effective SNR

In a nonlinear fiber channel, amplifiers along with conventional linear digital signal processing (DSP) effectively compensate chromatic dispersion (CD) and attenuation. In this paper, we consider the channel output after DSP steps excluding nonlinearity compensation. Thus, the residual noise comes mainly from amplified spontaneous emission (ASE) and NLI. The ASE noise is a random process independent of the signal, whereas the NLI noise is dependent on the signal. The interplay between CD and fiber nonlinearity induces nonlinear interactions among a number of past and future symbols. The number of interfering symbols is often referred to as the channel memory.

For a fiber-optic system with group velocity dispersion β_2 and optical bandwidth $\Delta\omega$, the dispersive length L_D is defined as [21, Sec. II-B]

$$L_D = 1 / (\Delta\omega^2 |\beta_2|), \quad (1)$$

where L_D indicates the distance scale over which pulse broadening effects become significant. Given the propagation distance L , the *one-sided* channel memory M is roughly approximated as [21, Sec. II-B]

$$M \approx L / L_D. \quad (2)$$

At time instant i , given the transmitted symbol X_i , a vector consisting of X_i along with its M neighbouring symbols is denoted as $\mathbf{X}_{i-M}^{i+M} = [X_{i-M}, \dots, X_i, X_{i+1}, \dots, X_{i+M}]^2$. By considering the NLI and the ASE as AWGNs, the received symbol Y_i can be expressed as

$$Y_i = X_i + Z_{\text{ASE},i} + Z_{\text{NLI},i}(\mathbf{X}_{i-M}^{i+M}). \quad (3)$$

Due to the nonlinear interference, the NLI term $Z_{\text{NLI},i}$ is expressed as a function of \mathbf{X}_{i-M}^{i+M} .

The SNR is usually evaluated under the implicit assumption of ergodic signal and noise processes. Therefore, the empirical estimations of the signal power and noise variance converge to the corresponding statistical values. Hence, the *effective* SNR including the NLI is defined as

$$\text{SNR}_{\text{eff}} \triangleq \frac{\mathbb{E}[|X|^2]}{\text{Var}[Z_{\text{ASE}}] + \text{Var}[Z_{\text{NLI}}]} \quad (4)$$

$$= \frac{\mathbb{E}[|X|^2]}{\mathbb{E}[|Y - X|^2]} \approx \frac{\mu_{|X|^2}}{\mu_{|Y-X|^2}}. \quad (5)$$

Often the r.h.s. of (5) is used to approximate the effective SNR in simulations and experiments.

B. Fiber Channel Models with Finite Memory

We review related channel models that consider the temporal effects of the transmitted symbols and also fit the general model in (3). A first-order perturbation model can be derived from the nonlinear Schrödinger equation in discrete-time domain [39]. Assuming a single channel at time instant $i = 0$, $Z_{\text{NLI},0}$ is modeled as [39, Eq. (57)]

$$Z_{\text{NLI},0}(\mathbf{X}_{-\infty}^{+\infty}) = \mathcal{J}\gamma \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} S_{h,j,l} X_h X_j X_l^*, \quad (6)$$

where γ is the nonlinear coefficient. The product of the symbol triplets are weighted by complex perturbation terms $S_{h,j,l}$ that quantify self channel interference³. In (6) we use the notation $\mathbf{X}_{-\infty}^{+\infty}$ to emphasize that (6) is in theory an infinite-memory channel.

The perturbation terms $S_{h,j,l}$ exhibit very different magnitudes depending on their indices (see for example [40, Fig. 5]). To capture the dominant terms contributing to the NLI, we consider the terms with $j = l$ as previously done in [22, Sec. II-B]. Furthermore, we then truncate the indices within the channel memory $2M$. Hence, (6) is approximated as

$$Z_{\text{NLI},0}(\mathbf{X}_{-M}^{+M}) \approx \mathcal{J}\gamma \sum_{h=-M}^M X_h \sum_{l=-M}^M S_{h,l,l} |X_l|^2, \quad (7)$$

²*Notation:* Throughout this paper, random variables are denoted by uppercase letters X and their (deterministic) outcomes by the same letter in lowercase x . Sequences are denoted by boldface letters \mathbf{X} (random) or \mathbf{x} (deterministic). We use subscript and superscript to denote the boundary of a random sequences, i.e., $\mathbf{X}_i^j = [X_i, X_{i+1}, \dots, X_j]$. Expectations, variances and autocovariances are denoted by $\mathbb{E}[\cdot]$, $\text{Var}[\cdot]$, and $\text{Cov}[\cdot]$, respectively. The probability of $X_i = x$ is denoted by $\mathbb{P}_{X_i}(x)$. Conditional probability is denoted as $\mathbb{P}_{Y_i|X_i}(y|x)$. The imaginary unit is $j \triangleq \sqrt{-1}$.

³The cross-channel interference can be expressed in the same form as the symbol triplet product times the complex perturbation terms (see [39, Eq. (60)]).

which shows that the NLI noise $Z_{\text{NLI},0}$ depends on the weighted symbol energies within the vector of \mathbf{X}_{-M}^{+M} .

For the transmission of *non*-i.i.d. symbol sequences, the sum of the ASE and the NLI in (3) is approximated in [21, Eq. (8)] as

$$Z_{\text{ASE},0} + Z_{\text{NLI},0}(\mathbf{X}_{-M}^{+M}) \approx \tilde{Z}_0 \sqrt{P_{\text{ASE}} + \eta \left(\frac{\sum_{k=-M}^M |X_k|^2}{2M+1} \right)^3}, \quad (8)$$

where \tilde{Z}_0 is a zero-mean unit-variance circularly symmetric complex Gaussian random variable, P_{ASE} is the ASE noise variance, and η is a real, non-negative constant quantifying the NLI. When compared to (7), (8) is a more radical approximation in the sense that the energies of all interfering symbols within a window of length $2M+1$ are assumed to make equal contribution to NLI.

In general, the expressions (7) and (8) show that the sum of (weighted) symbol energies within a window plays an important role in determining the NLI. Transmitting correlated symbols can therefore suppress the NLI. These observations motivate us to focus on the symbol energies and their temporal correlations.

C. Related Metrics for NLI

One well-known metric for showing variations of symbol energies is peak-to-average power ratio (PAPR). PAPR has been widely used for analysis of orthogonal frequency-division multiplexing (OFDM) fiber transmission [41]. PAPR partially reflects energy variation of symbols and thus can be regarded as a rough indication of nonlinearity tolerance of the transmitted signal [42, Sec. II-D], [43, Sec. II]. PAPR is defined as

$$\Theta \triangleq \frac{|X|_{\max}^2}{\mathbb{E}[|X|^2]}, \quad (9)$$

where $|X|_{\max}^2$ is the maximum symbol energy in the symbol sequence.

Another important metric (originated from the EGN model) is kurtosis. The EGN model predicts strong NLI if the transmitted symbols sequence has high kurtosis. For zero-mean constellations, kurtosis is defined as

$$\Phi \triangleq \frac{\mathbb{E}[|X|^4]}{\mathbb{E}[|X|^2]^2}. \quad (10)$$

To explain the blocklength dependent effective SNR, run ratio was proposed in [25, Sec. III-B]. Symbol sequences with high run ratio are considered to be more prone to the NLI. Run ratio is defined as

$$\text{R}_r \triangleq \frac{1}{T} \left(1 + \sum_{i=1}^{T-1} \bar{\delta}(x_{i-1}, x_i) \right), \quad (11)$$

where T is the number of transmitted symbols, and $\bar{\delta}(x_{i-1}, x_i)$ is a decision function which is equal to 1 when $x_{i-1} \neq x_i$ or 0 otherwise.

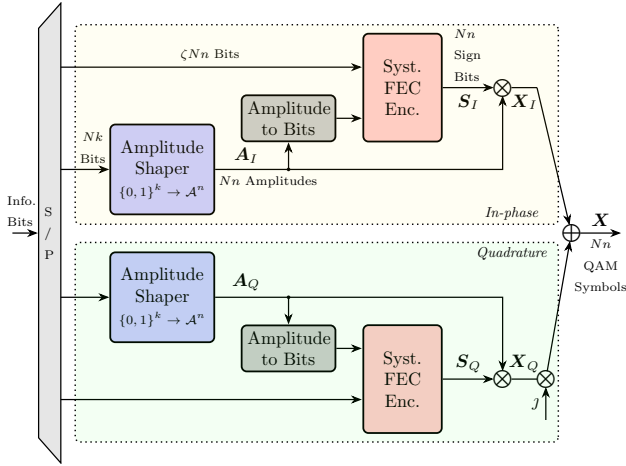


Fig. 1. Transmitter schematic diagram for a PAS architecture using 1D symbol mapping.

III. QAM SYMBOLS SHAPED BY CCDM

This section first reviews the generation of QAM symbol sequences in the PAS architecture. We then show that the symbols generated by CCDM are statistically dependent from each other. Motivated by results in Sec. II-B, which showed that the correlation among symbols could lead to a reduced NLI, we propose to analyze the autocorrelation of the symbol energies. This analysis will be later used in Sec. IV for the derivation of the EDI.

A block diagram of a PAS transmitter is illustrated in Fig. 1. To obtain a nonuniformly distributed QAM symbol sequence \mathbf{X} , two sequences of shaped pulse-amplitude modulation (PAM) symbols \mathbf{X}_I and \mathbf{X}_Q are independently generated in the in-phase and quadrature dimensions, respectively. The resulting QAM symbols are then $\mathbf{X} = \mathbf{X}_I + j\mathbf{X}_Q$. In this paper, similar to [29, Sec. II-C], we refer this mapping strategy as 1D symbol mapping⁴. When obtained from independent information bit sequences, \mathbf{X}_I and \mathbf{X}_Q are also independent of each other.

A. System Model

As shown in Fig. 1, the generation of PAM symbols relies on a systematic FEC engine and an amplitude shaper. A fixed-length amplitude shaper encodes k information bits to an *amplitude codeword* of blocklength n , and thus the shaping rate is k/n . The amplitude set is denoted as $\mathcal{A} = \{\Delta, 3\Delta, 5\Delta, \dots\}$ where Δ is a scaling factor.

For simplicity of exposition, let us consider the in-phase dimension and one FEC codeword. As shown in Fig. 1, a FEC codeword consists of N amplitude codewords. After a serial to parallel conversion, Nk information bits go into the amplitude shaper, and ζNn bits go into FEC, where ζ denotes the fraction of information in sign bits (see [1, Eq. (29)]). The amplitude shaper converts Nk bits into N amplitude codewords, which we call an *amplitude sequence* $\mathbf{A}_I = [A_0, A_1, \dots, A_{Nn-1}]$. The amplitudes in \mathbf{A}_I are then labeled into bits (see Fig. 1) and fed into the FEC along

with ζNn bits. The labeling used is the binary-reflected Gray code [44]. The FEC parity bits together with ζNn bits serve as sign bits \mathbf{S}_I to yield the PAM symbol sequence $\mathbf{X}_I = [X_{I,0}, X_{I,1}, \dots, X_{I,Nn-1}]$, where $X_{I,i} = (-1)^{S_{I,i}} A_{I,i}$ ($i = 0, 1, \dots, Nn-1$). The same procedure is conducted in the quadrature branch to generate \mathbf{X}_Q .

The amplitude shaper imposes constraints on the amplitudes of each amplitude codeword. In this paper, the amplitude shaper we consider is CCDM. Due to the constant-composition (CC) constraint, given blocklength n , the number of amplitudes $a \in \mathcal{A}$ in every codeword is a constant denoted by n_a , and thus $n = \sum_{a \in \mathcal{A}} n_a$. For any realization of the amplitude sequence, the amplitude frequency distribution of a is n_a/n . Therefore, the amplitude codewords generated by CCDM are simply permuted versions of each other.

In this paper, we make two assumptions about the CCDM codebook. In the following section, these assumptions will allow us to consider the CCDM codewords as a genuine process of drawing amplitudes without replacement [3, Sec. IV]. The first assumption is that the codewords are equally likely, which is justified in the scenario where the information bits are independent and uniformly distributed. The second assumption is that the CCDM codebook contains all possible amplitude permutations. The total number of permutations is given by the multinomial coefficient of the composition (see [45, Eq. (5)]), which is denoted as N_C . For k input bits, CCDM selects 2^k permutations as codewords. In general, $2^k < N_C$ because (i) k is chosen to be variable to for example achieve rate adaptivity, or (ii) N_C usually not be exactly a power of two. In this paper, we always use the largest possible value of k , and thus, (i) above can be ignored. On the other hand, due to (ii) above, CCDM does not always generate all N_C permutations. In this paper, we use *emulated* CCDM codewords that include all amplitude permutations. We will show later in the paper that this approximation is very good when compared to “exact” CCDM, where not all permutations are used.

B. Statistical Dependency Among CC Amplitudes

The dependency of amplitudes in a FEC frame is illustrated in Fig. 2. For notation simplicity, in what follows we drop the subscripts of \mathbf{A} and \mathbf{S} that indicate in-phase or quadrature. Furthermore, we pay little attention to sign bits since they only determine the polarity of PAM symbols. In terms of the amplitude sequence \mathbf{A} , Fig. 2 schematically shows that the amplitudes from different amplitude codewords of length n are independent of each other. Fig. 2 also shows that amplitudes *within* an amplitude codeword are mutually *dependent*.

The amplitude dependency within amplitude codewords comes from the CC constraint. During the arithmetic encoding of a CCDM that can be modeled as drawing without replacement process, the probability of drawing an amplitude is always updated by subtracting the previous number of amplitudes from the composition. The CC constraint further implicitly ensures that the sum of amplitudes in a CCDM codeword is a constant, i.e.,

$$\sum_{j=0}^{n-1} A_j = \sum_{a \in \mathcal{A}} a n_a. \quad (12)$$

⁴This mapping is also called *inter-pairing* in [25, Sec. II-B].

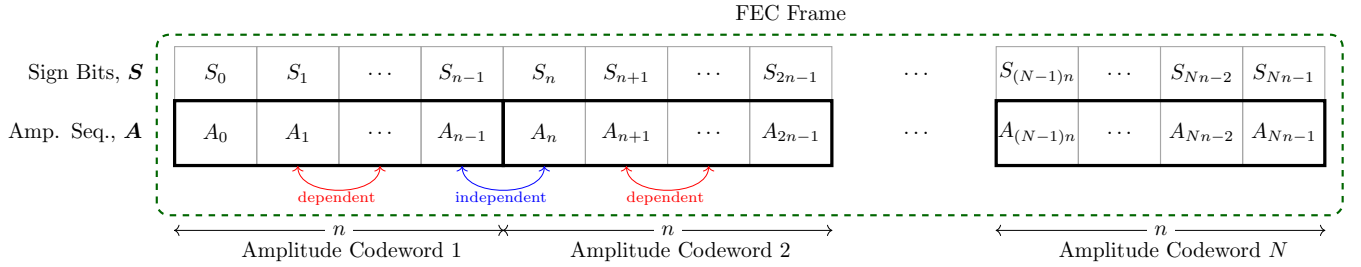


Fig. 2. Illustration of FEC frame and amplitude dependency. One FEC frame is formed by nN amplitudes and nN sign bits. Amplitudes within the same codeword are dependent, while amplitudes from different codewords are independent.

For any two different time instants i and $i + \tau$, we then have

$$A_i = -A_{i+\tau} - \sum_{\substack{j=0 \\ j \neq i, i+\tau}}^{n-1} A_j + \sum_{a \in \mathcal{A}} an_a. \quad (13)$$

Expression (13) shows that these amplitudes A_i and $A_{i+\tau}$ within CCDM codewords exhibit a *linear relationship*. This observation shows that amplitudes within an amplitude codeword are indeed linearly dependent.

In the next example, we use a CCDM shaping trellis to explain the statistical dependency between amplitudes. The trellis shows all possible amplitude sequences and their accumulated energies. We construct the trellis based on the following rules:

- Each path in the trellis represents a unique amplitude sequence.
- At time instant i , the vertical position of the nodes represents the accumulated energy $E_i^{\text{acc}} \in \mathcal{E}$, where

$$E_i^{\text{acc}} \triangleq \sum_{t=0}^{i-1} |A_t|^2, \quad (14)$$

$E_0^{\text{acc}} = 0$, and \mathcal{E} is the set of possible accumulated energy levels.

- The numbers labeling the trellis states represent the value of $\mathbb{P}_{A_i|E_i^{\text{acc}}}(a|e)$, i.e., the conditional probability of amplitude $A_i = a$ given the accumulated energy $E_i^{\text{acc}} = e$ for $i = 0, 1, \dots$, where $a \in \mathcal{A}$ and $e \in \mathcal{E}$.
- At time instant i , the edges indicate the amplitude A_i . The steeper the slope of the edge has, the larger the amplitude. The number next to the edges is $\mathbb{P}_{A_i, E_i^{\text{acc}}}(a, e)$, i.e., the probability of the paths starting from accumulated energy e using an amplitude a .

Example 1 (CCDM Trellis): Assume the use of CCDM with blocklength $n = 4$ and a composition of three amplitudes 1 and one amplitude 3 (i.e., $n_1 = 3$ and $n_3 = 1$), and thus, the amplitude PMF is $\mathbb{P}_A = [\frac{3}{4}, \frac{1}{4}]$. Fig. 3 shows the trellis of the first amplitude codeword, whose behavior will repeat for the subsequent codewords. The set of accumulated energy states is $\mathcal{E} = \{0, 1, 2, 3, 9, 10, 11, 12\}$. To explain the trellis, consider state $E_2^{\text{acc}} = 10$ and the three connected edges (shown with boldface text in Fig. 3). Since two paths reach this state, $\mathbb{P}_{E_2^{\text{acc}}}(10) = \mathbb{P}_{A_1, E_1^{\text{acc}}}(3, 1) + \mathbb{P}_{A_1, E_1^{\text{acc}}}(1, 9) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. At this node, we can only choose amplitude 1 for a_2 , and thus $\mathbb{P}_{A_2|E_2^{\text{acc}}}(3|10) = 0$ and $\mathbb{P}_{A_2|E_2^{\text{acc}}}(1|10) = 1$. The joint probability of amplitude $a_2 = 1$ starting from $e_2 = 10$ is thus $\mathbb{P}_{A_2, E_2^{\text{acc}}}(1, 10) = \mathbb{P}_{A_2|E_2^{\text{acc}}}(1|10) \mathbb{P}_{E_2^{\text{acc}}}(10) = 1 \times \frac{1}{2} = \frac{1}{2}$.

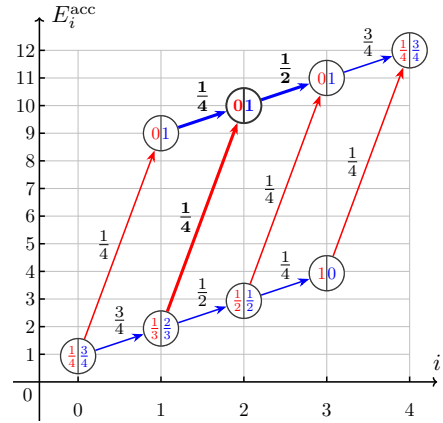


Fig. 3. CCDM shaping trellis with blocklength $n = 4$ for amplitude set $\mathcal{A} = \{1, 3\}$ with $\mathbb{P}_A = [\frac{3}{4}, \frac{1}{4}]$. The amplitude edges are labeled with $\mathbb{P}_{A_i, E_i^{\text{acc}}}(a, e)$. Each node contains $\mathbb{P}_{A_i|E_i^{\text{acc}}}(3|e)$ on the left (in red) and $\mathbb{P}_{A_i|E_i^{\text{acc}}}(1|e)$ on the right (in blue).

In Example 1, the amplitudes within a CCDM codeword are statistically dependent among each other. This can be seen from Fig. 3, where $\mathbb{P}_{A_i|E_i^{\text{acc}}}(a|e) = \mathbb{P}_{A_i}(a)$ is not always satisfied. The reason for this dependency is that $A_0 + A_1 + A_2 + A_3 = 6$ (see (12)), and any pair of amplitudes exhibits a linear relationship. On the other hand, for i.i.d. amplitude sequences, $\mathbb{P}_{A_i|E_i^{\text{acc}}}(a|e) = \mathbb{P}_{A_i}(a)$ is satisfied for $\forall a \in \mathcal{A}$.

Another property of i.i.d. amplitude sequences is the time-independent marginal probability of an amplitude. As can be seen in Example 1, the marginal probability of an amplitude within a CCDM codeword is also constant, i.e., for $\forall i \in \mathbb{Z}$,

$$\sum_{e \in \mathcal{E}} \mathbb{P}_{A_i, E_i^{\text{acc}}}(a, e) = \mathbb{P}_{A_i}(a) = \mathbb{P}_A(a) = \frac{n_a}{n}. \quad (15)$$

The property in (15) is due to the set of codewords being permutation invariant, as shown in [46, Lemma 1]. For example, $\mathbb{P}_{A_2}(1) = \mathbb{P}_{A_2, E_2^{\text{acc}}}(1, 10) + \mathbb{P}_{A_2, E_2^{\text{acc}}}(1, 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. The property in (15) holds for all CC amplitude sequences and will be used in the statistical analysis later in the paper.

A closer look at Example 1 also reveals that when multiple CCDM codewords are transmitted, the trellis extends in the time domain, and the probabilistic model of the first amplitude codeword repeats with period n . This repetition is studied in the next example.

Example 2 (Trellis Repetition and Blocklength): Fig. 4 shows a comparison of the CCDM shaping trellises for block-

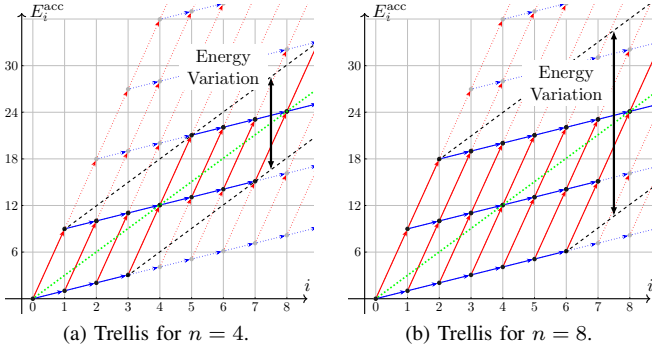


Fig. 4. CDDM shaping trellis comparison for blocklengths $n = 4$ and $n = 8$, with $\mathbb{P}_A = [\frac{3}{4}, \frac{1}{4}]$. The trellis marked by gray dotted lines show amplitude sequences never generated by CDDM.

lengths $n = 4$ and $n = 8$. Since both cases have the same \mathbb{P}_A , it can be seen that $E_4^{\text{acc}}/4 = 12/4 = 3$ for $n = 4$, which is equal to $E_8^{\text{acc}}/8 = 24/8 = 3$ for $n = 8$ (see Fig. 4). As a result, the green dotted lines in the middle of the trellises show that their accumulated energies grow at the same average rate. We also note that for blocklength $n = 8$, the trellis broadens (in the energy direction) and includes the trellis for $n = 4$. The vertical arrows between the black dashed lines shows that the variations of the accumulated energy growth for $n = 8$ is larger when compared to $n = 4$.

An important implication from the two examples above is that by introducing dependencies between amplitudes, CDDM artificially constrains the *dispersion* of the accumulated energy of the amplitude sequences (see black arrows in Fig. 4). For a fixed PMF \mathbb{P}_A , the accumulated energy dispersion is more constrained with shorter blocklength.

C. Statistics of Symbol Energies with CC Amplitudes

In this section, we will study the statistical properties of the QAM symbol sequences defined by CC amplitudes from the perspective of stochastic processes. In particular, we will show that the process of *symbol energies* is wide-sense cyclostationary (WSCS). The analysis in this section will lay the ground for the definition and formulation of the EDI metric in Sec. IV. We start by briefly reviewing some of the key properties of stochastic processes.

Definition 1 (First-order stationarity [47, Ch. 9-1]): A stochastic process \mathbf{X} is called *first-order stationary* if $\mathbb{P}_{X_i} = \mathbb{P}_{X_{i+\tau}}$ holds for any delay τ .

Definition 2 (Autocorrelation [48, Def. 13.13]): The autocorrelation of a stochastic process \mathbf{X} as a function of time slot i and delay τ is defined as

$$R_X(i, \tau) \triangleq \mathbb{E}[X_i X_{i+\tau}]. \quad (16)$$

Definition 3 (Autocovariance [48, Def. 13.12]): The autocovariance of a stochastic process \mathbf{X} is defined as

$$\text{Cov}[X_i, X_{i+\tau}] \triangleq \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_{i+\tau} - \mathbb{E}[X_{i+\tau}])]. \quad (17)$$

Moreover, for first-order stationary processes,

$$\text{Cov}[X_i, X_{i+\tau}] = R_X(i, \tau) - \mathbb{E}[X]^2. \quad (18)$$

Definition 4 (Wide-sense Cyclostationarity [47, Ch. 10-4]): A stochastic process \mathbf{X} is called *wide-sense cyclostationary* with period n if $\mathbb{E}[X_i] = \mathbb{E}[X_{i+mn}]$ and $R_X(i, \tau) = R_X(i+mn, \tau)$ hold for every $m \in \mathbb{Z}$.

In the definitions above, we used \mathbf{X} to denote a generic stochastic process. In what follows, we focus on the process of symbol energies defined as

$$\mathbf{E} \triangleq [\dots, E_{i-1}, E_i, E_{i+1}, \dots], \quad (19)$$

where

$$E_i \triangleq |X_i|^2 = A_{I,i}^2 + A_{Q,i}^2. \quad (20)$$

For a symbol energy process \mathbf{E} , we will refer to every n symbol energies as a *block*⁵, which are constructed based on two amplitude codewords (see Fig. 1). In addition, the autocorrelation of \mathbf{E} is

$$R_E(i, \tau) \triangleq \mathbb{E}[E_i E_{i+\tau}]. \quad (21)$$

Lemma 1: The QAM symbol energies \mathbf{E} defined by CC amplitudes is a first-order stationary process.

Proof: As shown in (15), the CC amplitude sequences \mathbf{A} are first-order stationary processes. Since a QAM symbol can be decomposed into two independent dimensions, the process of the energies \mathbf{E} in (19) is also first-order stationary. ■

Lemma 1 shows that the probability distribution of symbol energy \mathbb{P}_{E_i} is constant for any $i \in \mathbb{Z}$. From this first-order stationarity property in Lemma 1, it follows that

$$\mathbb{E}[E_i] = \mathbb{E}[|X_i|^2] = \mathbb{E}[|X|^2], \quad (22)$$

$$\text{Var}[E_i] = \text{Var}[|X_i|^2] = \text{Var}[|X|^2], \quad (23)$$

where

$$\mathbb{E}[|X|^2] = 2\mathbb{E}[A^2], \quad (24)$$

$$\text{Var}[|X|^2] = 2\mathbb{E}[A^4] - 2\mathbb{E}[A^2]^2, \quad (25)$$

and (25) was obtained using

$$\mathbb{E}[|X|^4] = 2\mathbb{E}[A^4] + 2\mathbb{E}[A^2]^2. \quad (26)$$

The second and fourth order moments of A in (24)–(26) can be computed by using the amplitude PMF $\mathbb{P}_A(a) = n_a/n$, i.e.,

$$\mathbb{E}[A^2] = \sum_{a \in \mathcal{A}} a^2 \frac{n_a}{n}, \quad (27)$$

$$\mathbb{E}[A^4] = \sum_{a \in \mathcal{A}} a^4 \frac{n_a}{n}. \quad (28)$$

Theorem 2: For the sequence of QAM symbol energies \mathbf{E} defined by CC amplitudes, if E_i and $E_{i+\tau}$ ($\tau \neq 0$) belong to the same block, we have

$$R_E(i, \tau) = \rho < \mathbb{E}[|X|^2]^2, \quad (29)$$

where

$$\rho \triangleq \frac{n\mathbb{E}[|X|^2]^2 - \mathbb{E}[|X|^4]}{n-1}, \quad (30)$$

⁵We avoid using “symbol energy codeword”, since the mapping between k input bits and n symbol energies is not one-to-one.

and

$$\text{Cov}[E_i, E_{i+\tau}] = -\frac{\text{Var}[|X|^2]}{n-1} < 0. \quad (31)$$

Proof: See Appendix A. ■

In Theorem 2, the negativity of autocovariance in (31) shows that the dependency between amplitudes is due to the fact that these amplitudes are inversely correlated. This negativity will also be used to prove Corollary 7 in Sec. IV. In addition, from (31) we obtain $\lim_{n \rightarrow \infty} \text{Cov}[E_i, E_{i+\tau}] = 0$, indicating that using long blocklengths weakens the linear dependency between symbol energies.

Lemma 3: The QAM symbol energies \mathbf{E} defined by CC amplitudes is a WSCS process.

Proof: The first condition of WSCS processes $\mathbb{E}[E_i] = \mathbb{E}[E_{i+mn}]$ is satisfied because of (22). The second condition of WSCS processes $R_E(i, \tau) = R_E(i + mn, \tau)$ follows from the fact the same probabilistic model repeats for every block of n symbols (see Fig. 2). ■

In practice, we are interested in the average autocorrelation of WSCS processes over the cyclostationarity period. For the process under investigation, such an average autocorrelation $\bar{R}_E(\tau)$ is defined as

$$\bar{R}_E(\tau) \triangleq \frac{1}{n} \sum_{i=0}^{n-1} R_E(i, \tau) \quad (32)$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[E_i E_{i+\tau}], \quad (33)$$

where (33) follows from (21). The average autocorrelation $\bar{R}_E(\tau)$ in (32) can be interpreted as a quantity that indicates the average linear dependency of all possible pairs of symbol energies separated by τ symbols.

The following theorem gives an analytical expression for the average autocorrelation in (33) for the considered QAM symbol sequences.

Theorem 4: The average autocorrelation $\bar{R}_E(\tau)$, $\tau \in \mathbb{Z}$ for CCDDM QAM symbol sequences with blocklength n generated using a composition such that $\mathbb{P}_A(a) = n_a/n, a \in \mathcal{A}$ can be expressed as

$$\bar{R}_E(\tau) = \begin{cases} \mathbb{E}[|X|^4], & \text{if } \tau = 0 \\ \frac{|\tau| \mathbb{E}[|X|^2]^2 + (n - |\tau|)\rho}{n}, & \text{if } 1 \leq |\tau| < n \\ \mathbb{E}[|X|^2]^2, & \text{if } |\tau| \geq n \end{cases} \quad (34)$$

where ρ is given in (30).

Proof: The function $R_E(i, \tau)$ for different values of τ and i is illustrated in Table I, where for simplicity only $\tau > 0$ is shown. When $\tau = 0$, $R_E(i, 0) = \mathbb{E}[|X|^4]$ as shown in the first column in Table I. For $\tau \neq 0$, $E_{i+\tau}$ and E_i are either at the same block and thus correlated, or at two different blocks and thus independent of each other. For example, when $i = 0$, E_0 is only correlated with the following $n-1$ symbol energies, yielding $R_E(0, \tau) = \rho$ for $\tau = 1, 2, \dots, n-1$ (see Fig. 2 and the blue cells in the $i = 0$ row of Table I). When $\tau = n$, E_n is the energy of the first symbol in the second block, and thus E_0 and E_n are independent that gives $R_E(0, n) = \mathbb{E}[|X|^2]^2$.

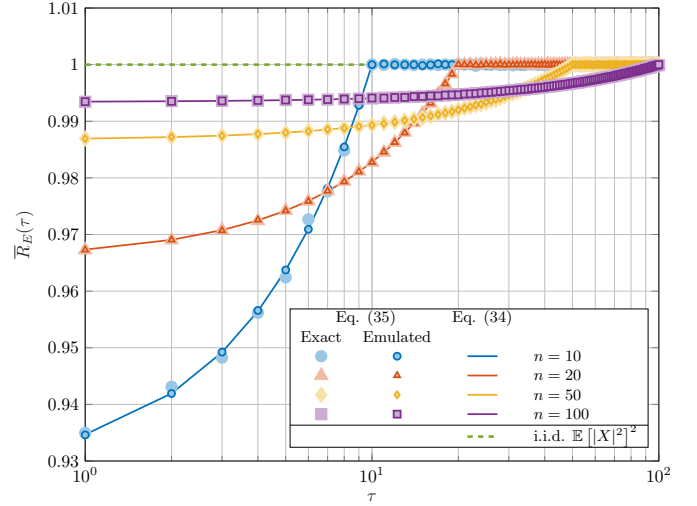


Fig. 5. Simulation results of $\hat{R}_E(\tau)$ in (35) (markers) by using exact or emulated CCDDM codewords, and analytical results of $\bar{R}_E(\tau)$ given by Theorem 4 (solid lines) for normalized ($\mathbb{E}[|X|^2] = 1$) 64QAM symbol sequences with $\mathbb{P}_A = [0.4, 0.3, 0.2, 0.1]$. For all cases $\bar{R}_E(0) = \mathbb{E}[|X|^4] = 1.65$.

This is shown by the red cells in the $i = 0$ row of Table I. The other rows in Table I can be calculated in a similar way. $\bar{R}_E(\tau)$ is the column-wise average of $R_E(i, \tau)$. ■

Note that $\bar{R}_E(\tau)$ is an even function and that $\bar{R}_E(\tau)$ in (34) is completely determined by the blocklength n , the second, and fourth order moments of $|X|$. The latter can be calculated via (24)–(25) and (27)–(28).

Furthermore, $\bar{R}_E(\tau)$ can be approximated by the sample autocorrelation function $\hat{R}_E(\tau)$ in a Monte Carlo simulation. With T samples, $\hat{R}_E(\tau)$ is defined as

$$\hat{R}_E(\tau) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} |x_t|^2 |x_{t+\tau}|^2. \quad (35)$$

When $T \rightarrow \infty$, $\hat{R}_E(\tau) \rightarrow \bar{R}_E(\tau)$. This ergodicity follows from Slutsky's theorem [47, Thm. 12-2], due to the fact that as $\tau' \rightarrow \infty$, the correlation between symbol energies vanishes and thus $\text{Cov}[E_i^2 E_{i+\tau}^2, E_{i+\tau'}^2 E_{i+\tau+\tau'}^2] \rightarrow 0$.

Example 3 (Average Autocorrelation): Fig. 5 shows the average autocorrelation $\bar{R}_E(\tau)$ in Theorem 4 and its estimation $\hat{R}_E(\tau)$ for 64QAM symbol sequences (normalized to $\mathbb{E}[|X|^2] = 1$). Four different shaping blocklengths are compared, which use the same distribution $\mathbb{P}_A = [0.4, 0.3, 0.2, 0.1]$, and have the same $\bar{R}_E(0) = \mathbb{E}[|X|^4] = 1.65$ (not shown in Fig. 5). I.i.d. sequences with the same distribution are also considered. Fig. 5 shows $\hat{R}_E(\tau)$ in (35) for sequences generated using exact CCDDM, emulated CCDDM, and for a system with a symbol-level interleaver to emulate the i.i.d. property. Fig. 5 shows that the emulated results differ from exact CCDDM only for $n = 10$, where only 65% of the total number of permutations N_C are used as codewords. Except this minor mismatch for $n = 10$, $\bar{R}_E(\tau)$ in Theorem 4 approximates well the true autocorrelation function $\bar{R}_E(\tau)$ in all other cases. It can also be seen in Fig. 5 that for i.i.d. symbol sequences or $\tau \geq n$, $\bar{R}_E(\tau) = \mathbb{E}[|X|^2]^2 = 1$. In the cases of CC symbol sequences with $0 < \tau < n$, due to

TABLE I

THE AUTOCORRELATION $R_E(i, \tau)$ AND ITS AVERAGE $\bar{R}_E(\tau)$ FOR $\tau > 0$. THE THREE CASES OF $R_E(i, \tau)$ IN (55) CORRESPOND TO THE FIRST COLUMN (IN GRAY), THE UPPER LEFT PART (IN BLUE) AND THE BOTTOM RIGHT (IN RED), RESPECTIVELY.

i	τ								
	0	1	2	...	$n-2$	$n-1$	n	$n+1$...
0	$\mathbb{E}[X ^4]$	ρ	ρ	...	ρ	ρ	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...
1	$\mathbb{E}[X ^4]$	ρ	ρ	...	ρ	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\vdots	\vdots
$n-2$	$\mathbb{E}[X ^4]$	ρ	$\mathbb{E}[X ^2]^2$...	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...
$n-1$	$\mathbb{E}[X ^4]$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...
$\bar{R}_E(\tau)$	$\mathbb{E}[X ^4]$	$(n-1)\rho/n + \mathbb{E}[X ^2]^2/n$	$(n-2)\rho/n + 2\mathbb{E}[X ^2]^2/n$...	$2\rho/n + (n-2)\mathbb{E}[X ^2]^2/n$	$\rho/n + (n-1)\mathbb{E}[X ^2]^2/n$	$\mathbb{E}[X ^2]^2$	$\mathbb{E}[X ^2]^2$...

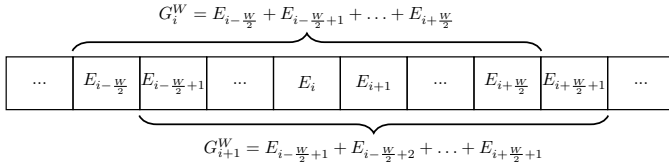


Fig. 6. Window energies G_i^W and G_{i+1}^W given symbol energy sequence \mathbf{E} .

$R_E(i, \tau)$ being smaller than $\mathbb{E}[|X|^2]^2$, the average symbol energy dependency manifest itself as a deviation between $\bar{R}_E(\tau)$ and $\mathbb{E}[|X|^2]^2$ (see (29) and the second case in (34)). For each blocklength, as τ increases, the symbol energy dependency gradually decreases and vanishes when $\tau = n$ such that E_i and $E_{i+\tau}$ always belong to two independent blocks. Moreover, as n increases, the curves approach $\mathbb{E}[|X|^2]^2$ (i.e., the deviation decreases), which implies a weaker dependency.

IV. ENERGY DISPERSION INDEX

In this section, EDI is introduced as a figure of merit to qualitatively predict the NLI power. In Sec. II-A, we showed that memory is one of the main phenomena affecting the NLI. In previous NLI metrics (see Sec. II-C), the effect of non-i.i.d. input sequences and their interaction with the channel memory was not considered. EDI brings together these two elements by capturing the statistical properties of the input sequence over a time window which is comparable to the channel memory.

A. Definition

EDI is designed to be a sliding window statistic with window length W . The windowed energy at time instant i , G_i^W , is defined as the total energy of $\mathbf{X}_{i-W/2}^{i+W/2}$, i.e.,

$$G_i^W \triangleq \sum_{j=i-W/2}^{i+W/2} E_j = \sum_{j=i-W/2}^{i+W/2} |X_j|^2. \quad (36)$$

Fig. 6 illustrates the windowed energies G_i^W and G_{i+1}^W and shows the sliding window effect.

The windowed energy is a random variable, and thus, we define the *windowed energy process* as

$$[\dots, G_{i-1}^W, G_i^W, G_{i+1}^W, \dots]. \quad (37)$$

For CCDM QAM symbol sequences, the windowed energy process is the sum of $W+1$ time-shifted WSCS processes

of symbol energies of period n , and thus the windowed energy process is also WSCS with the same period, i.e., n [49, Ch. 17.2-Prop. 1]. Therefore, $R_{G^W}(i, \tau)$ and $\mathbb{E}[G_i^W]$ are periodic with period n . Furthermore, the variance of the windowed energy G_i^W (see (18)) is given by

$$\text{Var}[G_i^W] = \text{Cov}[G_i^W, G_i^W] = R_{G^W}(i, 0) - \mathbb{E}[G_i^W]^2. \quad (38)$$

With periodic $R_{G^W}(i, 0)$ and $\mathbb{E}[G_i^W]$, $\text{Var}[G_i^W]$ in (37) also varies cyclically with period n .

Definition 5 (Energy Dispersion Index): EDI is defined as

$$\Psi \triangleq \frac{\overline{\text{Var}[G^W]}}{\overline{\mathbb{E}[G^W]}}, \quad (39)$$

where

$$\overline{\mathbb{E}[G^W]} \triangleq \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[G_i^W], \quad (40)$$

$$\overline{\text{Var}[G^W]} \triangleq \frac{1}{n} \sum_{i=0}^{n-1} \text{Var}[G_i^W]. \quad (41)$$

EDI in Definition 5 measures the windowed energy variations. EDI in (39) is defined as the ratio of the average windowed energy variance ($\overline{\text{Var}[G^W]}$) to the average windowed energy mean ($\overline{\mathbb{E}[G^W]}$). As shown in (40) and (41), these two averages have the same form of the average autocorrelation $\bar{R}_E(\tau)$ in (32).

B. Alternative Formulations

EDI in Definition 5 can also be expressed in terms of the second-order statistics of the input symbols. In this section, we introduce such formulation of the EDI for both CCDM QAM symbol sequences as well as i.i.d. symbol sequences. Note that we view the process of i.i.d. symbol sequences as a special case of a WSCS process with period $n = 1$.

Theorem 5: The average windowed energy mean and average windowed energy variance for CCDM QAM symbol sequences can be expressed as

$$\overline{\mathbb{E}[G^W]} = (W+1)\mathbb{E}[|X|^2] \quad (42)$$

and

$$\overline{\text{Var}[G^W]} = (W+1)\text{Var}[|X|^2] - W(W+1)\mathbb{E}[|X|^2]^2 + 2 \sum_{\tau=1}^W (W-\tau+1)\bar{R}_E(\tau), \quad (43)$$

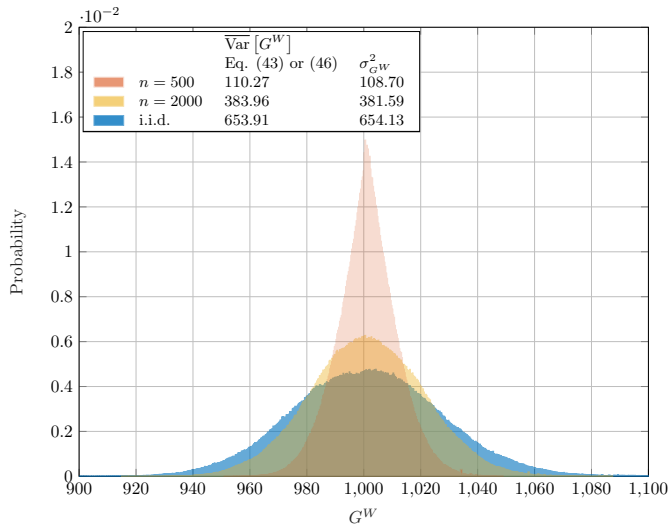


Fig. 7. Histogram of windowed energy G^W for $W = 1000$. Normalized ($\mathbb{E}[|X|^2] = 1$) 64QAM symbol sequences with $\mathbb{P}_A = [0.4, 0.3, 0.2, 0.1]$. The average windowed energy mean is $\mathbb{E}[G^W] = 1000$.

respectively, where $\overline{R}_E(\tau)$ is given by (34).

Proof: See Appendix B. ■

For i.i.d. symbol sequences, (42) also holds. On the other hand, $\text{Var}[G_i^W]$ for i.i.d. symbol sequences is the sum of $W + 1$ energy variances, therefore,

$$\overline{\text{Var}}[G^W] = \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=i-W/2}^{i+W/2} \text{Var}[E_j] \quad (44)$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} (W+1) \text{Var}[|X|^2] \quad (45)$$

$$= (W+1) \text{Var}[|X|^2], \quad (46)$$

where (45) follows from (23).

Example 4 (Windowed Energy Histograms): Fig. 7 illustrates how the probabilities of windowed energies depend on the blocklength n , based on a number of windowed energy samples. For simplicity, we only show the results of emulated CCDDM amplitude codewords, since almost the same result is obtained when using exact CCDDM. Fig. 7 shows that the mean value of G_i^W is independent of n as shown in (42). Fig. 7 also shows that an increase of n results in a heavier tail, i.e., a larger probability of observing a large windowed energy. The windowed energy variance increases as blocklength n increases, which is in good agreement with our observations in Example 2. Fig. 7 also shows that the estimated windowed energy variance $\sigma_{G^W}^2$ is well-approximated by $\overline{\text{Var}}[G^W]$ given in (43) and (46), where the discrepancies are caused by the limited number of samples.

Based on Theorem 5, the EDI of CCDDM QAM symbol sequences can be obtained directly by substituting (43) and (42) in (39), i.e.,

$$\Psi(n, W) = \mathbb{E}[|X|^2] [\Phi - (W+1)] + \frac{2 \sum_{\tau=1}^W (W-\tau+1) \overline{R}_E(\tau)}{(W+1) \mathbb{E}[|X|^2]}, \quad (47)$$

where $\overline{R}_E(\tau)$ and Φ are given by (34) and (10), respectively. Notation $\Psi(n, W)$ is used to emphasize the dependency of the EDI on the blocklength n and window length W .

Apart from n and W , in view of (10), (30) and (34), EDI in (47) is also determined by the kurtosis Φ as well as the second and fourth order moments of $|X|$. Furthermore, after dividing (46) by (42) and using (10), the EDI of i.i.d. is obtained, which is given by

$$\Psi = \mathbb{E}[|X|^2] (\Phi - 1). \quad (48)$$

It can be seen that (48) is independent of n and W , but is still a function of the kurtosis Φ . Recall that kurtosis from EGN model is derived based on i.i.d. symbols assumption, hence (48) means that EDI can give the same indication of the NLI as kurtosis does in the case of i.i.d. symbols.

We have derived a closed-form expression for the EDI in (47) (and for i.i.d. symbols in (48)). In the next section, we will investigate properties of the EDI, and compare them against the estimated EDI from Monte Carlo simulations.

C. Properties

In what follows, we first show that for certain values of blocklength n and window length W , EDI depends linearly on n . This corresponds to a regime where W (and thus, implicitly the channel memory) is larger than n . We then give bounds on EDI for arbitrary values of n and W .

Theorem 6: When $n \leq W + 2$, the EDI $\Psi(n, W)$ in (47) depends linearly on n via

$$\Psi_{\text{lin}}(n, W) = \frac{n+1}{3(W+1)} \mathbb{E}[|X|^2] (\Phi - 1). \quad (49)$$

Proof: See Appendix C. ■

Corollary 7: For any finite blocklength n , the EDI in (47) is upper- and lower-bounded as

$$0 \leq \Psi(n, W) \leq \mathbb{E}[|X|^2] (\Phi - 1). \quad (50)$$

The bounds are achieved for asymptotic values of W , i.e.,

$$\lim_{W \rightarrow 0} \Psi(n, W) = \mathbb{E}[|X|^2] (\Phi - 1). \quad (51)$$

$$\lim_{W \rightarrow \infty} \Psi(n, W) = 0, \quad (52)$$

Proof: See Appendix D. ■

The lower bound in (52) can be intuitively understood as follows. When W is much larger than n , the windows always include multiple complete blocks of QAM symbols, and the compositions of amplitudes within a large window “hardens”, yielding a reduced fluctuation of the window energies. The upper bound in (51) is identical as the EDI of i.i.d. symbol sequences in (48). This upper bound indicates that as window length decreases (less memory in the metric), the impact of symbol energy correlations on EDI decreases.

Note that the EDI for constant-modulus constellations (such as, e.g., phase-shift keying), is identically 0, due to the fact that the symbols have constant energy. The sliding window energy is thus constant. This reflects the fact that EDI is a metric specifically designed to capture NLI fluctuations in PAS systems with different shaping blocklengths, and such systems cannot be designed using constant modulus constellations.

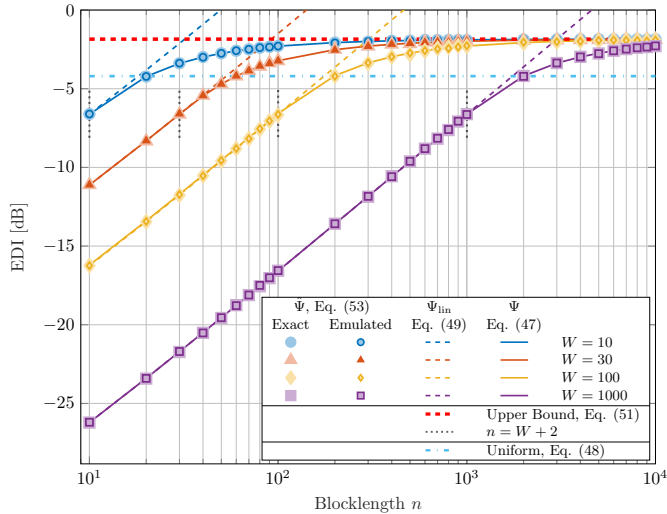


Fig. 8. Simulation results (markers) using (53), analytical results (solid lines) in (47) and (48), and linear EDI in (49) (in dB) vs. blocklength for normalized ($\mathbb{E}[|X|^2] = 1$) 64QAM symbol sequences with $\mathbb{P}_A = [0.4, 0.3, 0.2, 0.1]$.

Example 5 (EDI and Blocklength): Fig. 8 shows the EDI (in dB) of symbol sequences for different values of n and W , as well as the EDI for i.i.d. uniform 64QAM sequences. The analytical results of EDI are computed by using (47), (48) and (49) as well as the EDI estimated as

$$\hat{\Psi} \triangleq \frac{\sigma_{G^W}^2}{\mu_{G^W}}, \quad (53)$$

where $\sigma_{G^W}^2$ and μ_{G^W} are the estimated windowed energy variance and the estimated windowed energy mean, respectively. Fig. 8 shows that all the simulation results $\hat{\Psi}$ match the analytical results Ψ in (47) very well. This is the case even for short blocklength ($n = 10$), indicating that the CCDM emulation approach we took in this paper has little impact on the EDI in (47). Fig. 8 also shows the linear EDI expression Ψ_{lin} in (49). The EDI curves can be seen to be segmented into a linearly blocklength-dependent region by $n = W + 2$. As n increases above $W + 2$, Ψ begins to diverge from Ψ_{lin} in (49). For a fixed n , as W increases, EDI approaches the lower bound in (52) (i.e., $-\infty$ in dB). On the other hand, EDI gradually reaches the upper bound in (51) as $W \rightarrow 0$.

It can be concluded from Example 5 that for CCDM QAM symbol sequences, if a symbol-level interleaver, or a very long blocklength is used, the EDI will approach the upper bound in (51) that is determined by the kurtosis. Hence, EDI can be viewed as a windowed version of kurtosis. For the purpose of NLI prediction, window length W should be carefully chosen such that the channel memory effect is properly reflected, as will be shown in the following section.

We conclude this section by emphasizing that Example 5 (and also Example 3) showed that the CCDM emulation approach of considering all N_C permutations as codewords gives very precise results. This assumption allowed us to find closed-form expressions for the average autocorrelation (see (34)) and the EDI (see (47)). In the next section, we therefore show results using the analytical expressions we have developed above, and thus, only consider emulated CCDM.

TABLE II
SIMULATION PARAMETERS.

Parameter	Value
Modulation	64 QAM
Pol.-mux.	Single
Wavelength (λ)	1550 nm
Symbol rate	32 GBd
WDM spacing (Δf)	50 GHz
WDM channels (N_{ch})	5
Pulse shape	root-raised cosine
Roll-off	10%
Fiber length	80 km
Fiber loss	0.2 dB/km
Dispersion parameter (D)	17 ps/nm/km
Nonlinear parameter (γ)	1.37 1/W/km
EDFA noise figure	6 dB
Oversampling factor	2 \times
QAM symbols per run	2 ¹⁸
No. of simulation runs	10

V. NUMERICAL RESULTS

In this section, we show that EDI can qualitatively predict the NLI magnitude when different blocklengths are used. To this end, we study the effective SNR in (4), where NLI is a substantial part of the total noise at relatively high power, and where NLI changes produce a change in effective SNR.

A. Simulation Setup

We consider an ideal single-polarization multi-span WDM fiber system with $N_{\text{ch}} = 5$ channels. Nonlinear noise caused by the Kerr effect, as well as ASE and CD are taken into consideration. The fiber propagation is simulated using the split-step Fourier method with a step size of 50 m. Other key simulation parameters are displayed in Table II. The channel of interest is located at the center of the WDM spectrum, where the channel spacing is $\Delta f = 50$ GHz. The signal is generated with root-raised cosine pulse shaping. After propagation over each span of standard single-mode fiber with span length 80 km, the attenuation is ideally compensated by an Erbium-doped fiber amplifier (EDFA). At the receiver, the channel of interest is filtered with a matched filter, followed by CD compensation and sampling.

The *one-sided* channel memory M is an important reference for the window length W in the EDI. M can be estimated using (1) and (2). In (2), the optical bandwidth is $\Delta\omega = N_{\text{ch}}\Delta f$, while group velocity dispersion β_2 is related to dispersion parameter D as shown in [16, Eq. (1.2.11)].

For shaped 64QAM transmission, the amplitude PMF $\mathbb{P}_A = [0.4, 0.3, 0.2, 0.1]$ is used. Note that EDI can be used for arbitrary PMF in principle. Quadrature phase-shift keying (QPSK) and uniform 64QAM are presented as baselines. QPSK is anticipated to provide optimal effective SNR performance, since constant modulus constellations completely remove the modulation-dependent NLI [22]. We also consider shaped 64QAM symbol sequences using a randomly generated symbol-level interleaver to emulate i.i.d. symbol sequences. The symbol sequences are always normalized to $\mathbb{E}[|X|^2] = 1$.

An overview of the NLI-related metrics is given in Table III. PAPR and kurtosis do not depend on the blocklength, and thus,

TABLE III
NLI METRICS OF THE EVALUATED SYMBOL SEQUENCES.
THE RESULTS FOR R_r ARE TAKEN FROM [25, FIG. 8].

Metrics	CC 64QAM	PS 64QAM	i.i.d. PS 64QAM	Uniform 64QAM	QPSK
Θ , PAPR	3.769	3.769	2.336	1	
Φ , Kurtosis	1.653	1.653	1.381	1	
R_r , Run Ratio	≈ 0.98	0.978	0.984	0.750	

these two metrics cannot predict the NLI differences caused by blocklength differences. Although run ratio can, to some extent, capture the blocklength dependency of the NLI, it does not take channel memory into account, and thus, cannot adapt to different distances. By contrast, EDI uses a window length depending on the distance.

With various blocklengths n , many pairs of effective SNR and EDI can be obtained. To quantify the correlation between effective SNR and EDI, Pearson's correlation coefficient [50, Ch. 11.1] is used, i.e.,

$$r_p \triangleq \frac{\text{Cov}[\text{SNR}_{\text{eff}}, \hat{\Psi}]}{\sqrt{\text{Var}[\text{SNR}_{\text{eff}}] \text{Var}[\hat{\Psi}]}}. \quad (54)$$

Coefficient values +1 or -1 indicates perfect correlation, while 0 indicates no correlation.

B. EDI and Effective SNR

We first investigate how EDI is correlated to effective SNR. Fig. 9 displays effective SNR and EDI vs. blocklength. At each transmission distance, the optimal launch power found for $n = 10$ is used (see Fig. 12 ahead). The effective SNRs achieved for $n = 10$ and $n = 10000$ at these launch powers are shown with filled triangles in Fig. 9. EDI is calculated using the optimal window lengths W^* , which will be discussed in Fig. 10. Fig. 9 shows that the estimated EDIs (circles) are in good agreement with the analytical EDI in (47) (solid lines), despite of slight fluctuations due to the limited number of transmitted symbols. Effective SNR almost follows the same trend as that of EDI, and their Pearson's linear correlation coefficients at three distances are at least 99.5%. Based on Theorem 6, the x-axes in Fig. 9 can be divided into two regions: $n < W^* + 2$ and $n > W^* + 2$. The blocklength-dependent region on the left shows that the effective SNR varies linearly with blocklength n . As n increases entering the region on the right, SNR begins to decrease slowly until exhibiting marginal differences, since for long blocklengths the NLI reduction brought by weakly-correlated symbol energies becomes insignificant. In this region, EDI is determined by kurtosis, and the EGN model is able to give accurate predictions of effective SNR, as demonstrated in [25, Fig. 7].

The optimal window length W^* used in Fig. 9 was chosen such that EDI yields the highest correlation with effective SNR. To this end, W^* is obtained by analyzing various window length W at each transmission distance, whose absolute value of correlation coefficient $|r_p|$ is shown in Fig. 10. It can be seen that $|r_p|$ reaches its peak for values W^* , which is much smaller than the estimated channel memory $2M$ given

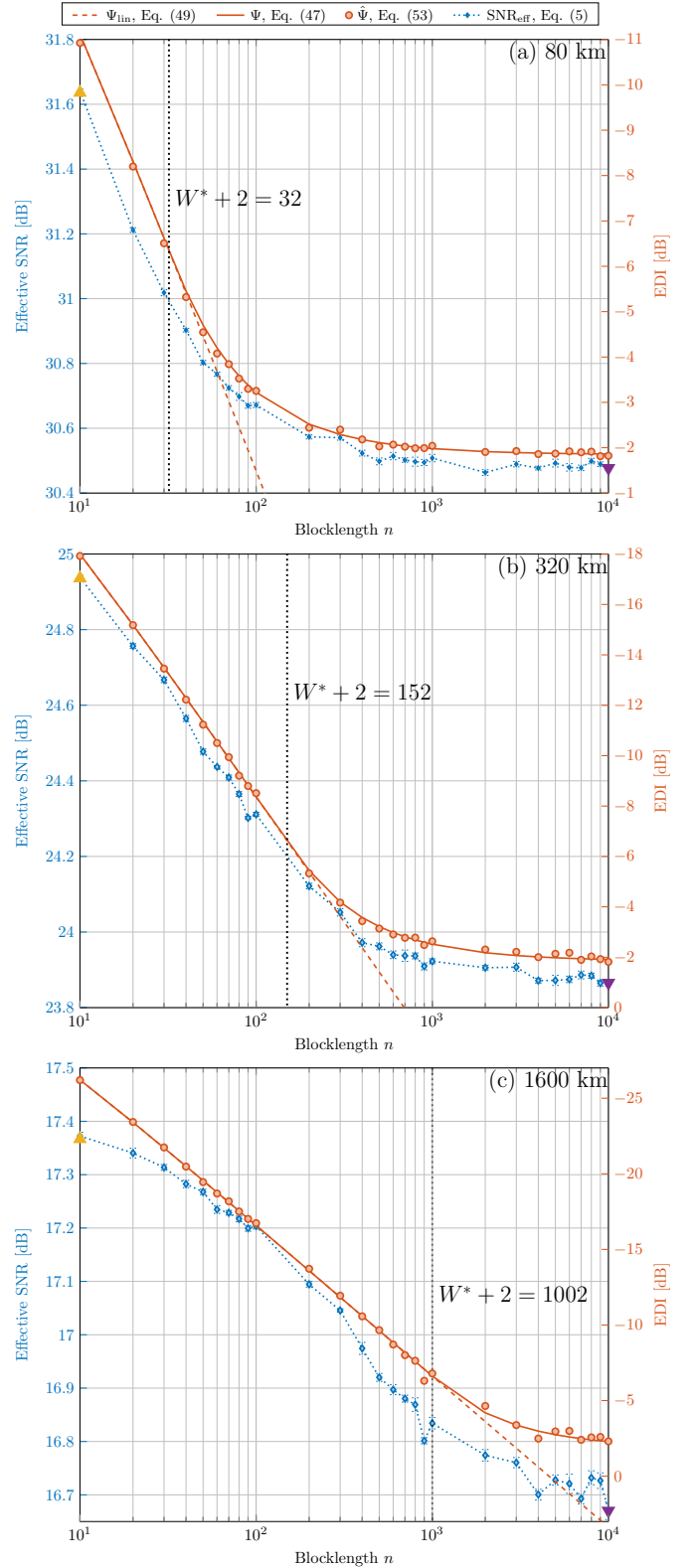


Fig. 9. Effective SNR (left axis) and EDI (right axis) vs. blocklength after transmission distances of (a) 80 km, (b) 320 km and (c) 1600 km. The launch powers are (a) -1.5 dBm, (b) -2.0 dBm and (c) -3.0 dBm. Error bars for effective SNRs represent 95% confidence interval. The EDI is shown in dB and inverted for convenience of comparison. The optimal window lengths W^* shown in the figure are used for the EDI calculation. The correlation coefficients r_p in (54) are (a) -0.998, (b) -0.999 and (c) -0.995.

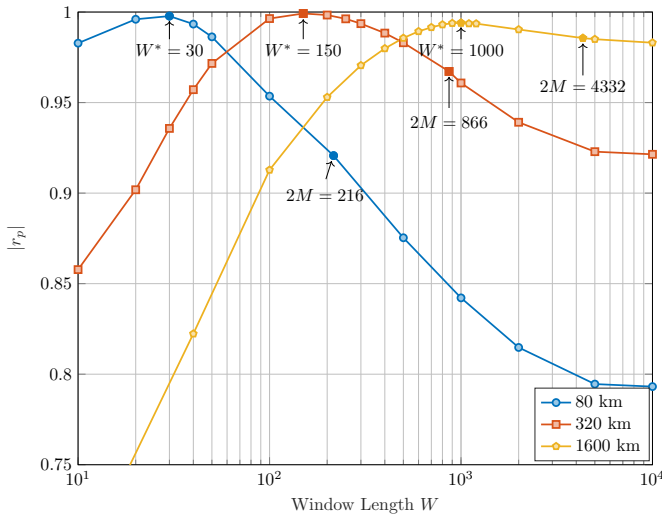


Fig. 10. Absolute value of Pearson's linear correlation coefficient $|r_p|$ in (54) between EDI and effective SNR vs. window length W for 80 km, 320 km and 1600 km, whose launch powers are -1.5 dBm, -2.0 dBm and -3.0 dBm, respectively. The optimal value of W is denoted by W^* . The channel memories $2M$ calculated using (1)–(2) are also indicated.

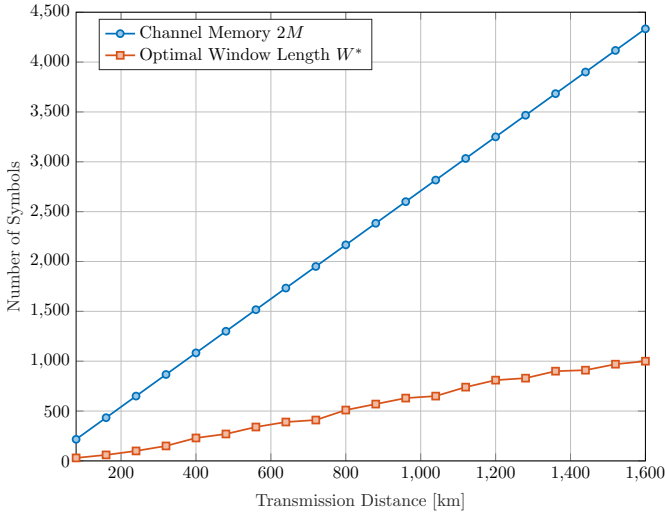


Fig. 11. The estimated channel memory $2M$ calculated using (1)–(2) and the optimal window length W^* ($|r_p| \geq 0.995$) at transmission distances from 80 km to 1600 km.

by (1)–(2). Incorrectly choosing W can lead to smaller values of $|r_p|$, and thus the SNR prediction by EDI is less accurate. However, Fig. 10 also shows that even if $2M$ is used as window length, EDI can still reflect the SNR variations with correlation coefficients above 90%. Thus, in practice one can directly use the estimated channel memory $2M$ as window length rather than finding the optimal one W^* . In general, Fig. 9 and Fig. 10 show that EDI and effective SNR are highly correlated with each other, indicated by a nearly perfect negative correlation. It can be concluded that EDI evaluated with the optimal window length is capable of reflecting the impact of blocklength and distance on the NLI.

Fig. 11 shows the relationship between W^* and estimated channel memory $2M$ at various transmission distances. All

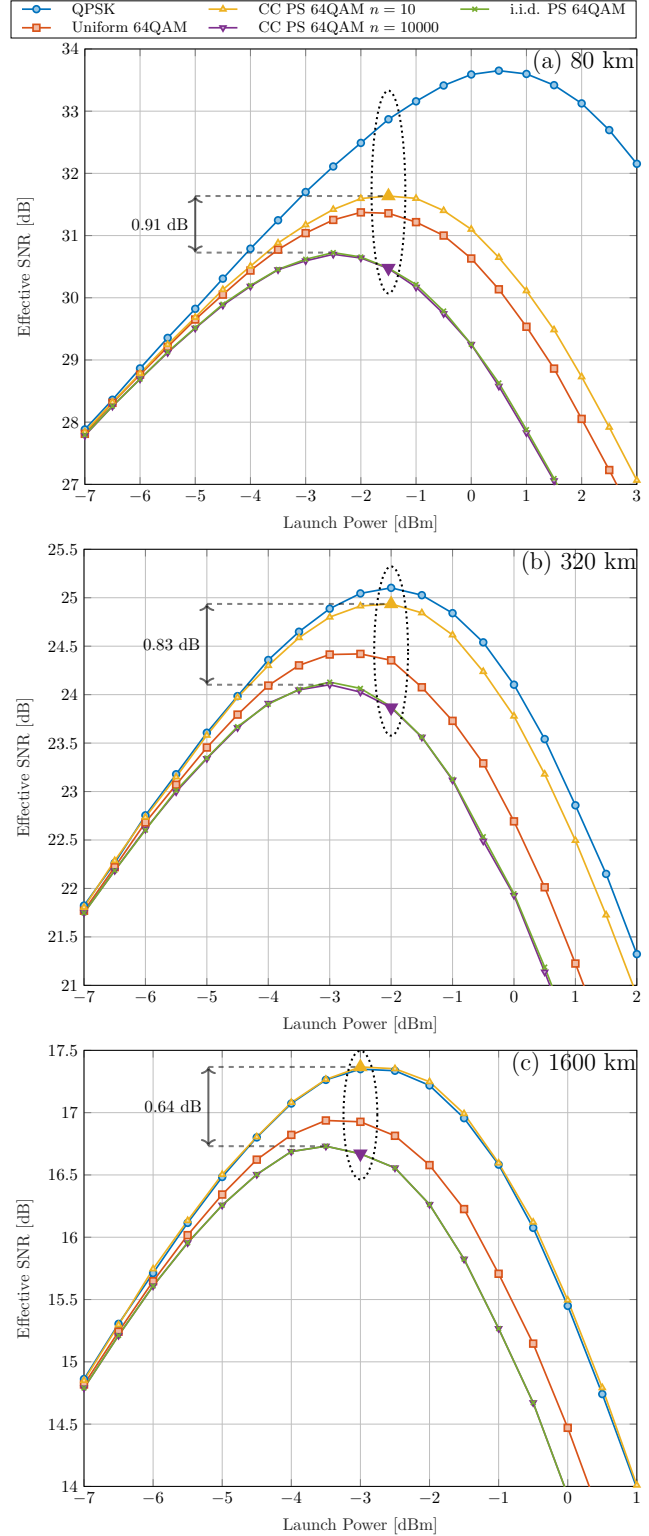


Fig. 12. Effective SNR vs. launch power after transmission distances of (a) 80 km, (b) 320 km and (c) 1600 km. PAS 64QAM with blocklength $n = 10$ and $n = 10000$ are displayed. Results of uniform 64QAM, i.i.d. PS 64QAM symbols, and QPSK are also included as references. The circled SNR performance correspond to the launch powers used in Fig. 9 (a)–(c) respectively.

W^* are obtained with at least 99.4% absolute correlation coefficients $|r_p|$. It can be seen in Fig. 11 that M scales linearly

TABLE IV
EDI Ψ (IN DB) OF THE EVALUATED SYMBOL SEQUENCES

Distance	CC PS 64QAM		i.i.d. PS	Uniform	QPSK
	$n = 10$	$n = 10000$	64QAM	64QAM	
80 km	-11.12	-1.86	-1.85	-4.20	$-\infty$
320 km	-18.00	-1.91	-1.85	-4.20	$-\infty$
1600 km	-26.21	-2.29	-1.85	-4.20	$-\infty$

with distance, as M is computed by using (2). Likewise, W^* also increases approximately linearly but at a slower rate. The optimal window length W^* is smaller than the estimated number of interfering symbols $2M$, which can be explained by two facts. First, (2) is a rough estimation of the channel memory. Second, the symbol energies within EDI window are assumed to have equal contributions to the NLI (as shown in (36) that all symbol energies are weighted by 1). Therefore, W^* generally represents an effective number of dominant interfering symbols that are involved in the NLI generation.

To conclude, we show effective SNR vs. launch power in Fig. 12, which shows that the impact of the shaping blocklength on effective SNR can be as large as that of modulation format. For simplicity, CCDDM 64QAM symbol sequences are only shown with ultra short blocklength $n = 10$ and long blocklength $n = 10000$, in general representing “good” and “bad” NLI-tolerant blocklengths. The EDI of different symbol sequences are given in Table IV. The filled triangles in Fig. 12 for $n = 10$ and $n = 10000$ correspond to the same effective SNR markers as shown in Fig. 9. The first observation from Fig. 12 (a)–(c) is that QPSK exhibits the best effective SNR for the distances under consideration, as it has the smallest EDI. Secondly, the effective SNRs of i.i.d. 64QAM symbol sequences and the shaped symbol sequence with $n = 10000$ almost coincide for all distances. Meanwhile, their EDIs have marginal differences. Thirdly, the effective SNR of uniform 64QAM falls between the effective SNRs of shaped 64QAM using $n = 10$ and $n = 10000$, so does their EDIs. Lastly, Fig. 12 shows that the SNR gains offered by CCDDM symbol sequences using $n = 10$ instead of $n = 10000$ decrease from 0.91 dB to 0.64 dB as transmission distance increases.

VI. CONCLUSIONS

This paper proposed a new heuristic metric called energy dispersion index (EDI) to predict the impact of blocklength on the effective SNR for CCDDM-coded systems. EDI is a measure of the windowed energy dispersion which captures the interaction between the statistical properties of a CCDDM input sequence and the channel memory with respect to the received NLI magnitude. Numerical results show that the effective SNR is highly correlated to the EDI of the transmitted symbol sequence, with correlation coefficients greater than 99%.

Being a heuristic metric, the EDI requires future analytical substantiation, possibly leading to a refined version thereof. One possible improvement of the EDI accuracy could be appropriately weighting the symbol energies within the EDI window to reflect their uneven contributions to the NLI. In this paper, we only studied EDI for CCDDM, and thus, a performance analysis of EDI to other shaping algorithms or

other constellations is still pending. All these open problems are left for future work. Nevertheless, we believe that EDI can facilitate the development of NLI-tolerant signaling schemes that aim to optimize time-varying statistical properties of the input symbol sequence.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yunus Can Gültekin and Sebastiaan Goossens (Eindhoven University of Technology) for fruitful discussions on shaping techniques.

APPENDIX A PROOF OF THEOREM 2

In this Appendix, we start by computing $R_E(i, \tau)$ in (21). For $i = 0, 1, \dots, n-1$ (one period), $R_E(i, \tau)$ is given by

$$R_E(i, \tau) = \begin{cases} \mathbb{E}[|X|^4], & \text{if } \tau = 0 \\ \rho, & \text{if } \tau \neq 0, 0 < i + \tau < n \\ \mathbb{E}[|X|^2]^2, & \text{if } \tau \neq 0, i + \tau \geq n \text{ or } i + \tau \leq 0 \end{cases} \quad (55)$$

The second case in (55) is when two different symbol energies belong to the same block. In what follows we will prove that in this case symbol energies are equally correlated and yields the same autocorrelation ρ in (30), and also autocovariance. By using Lemma 1, (16) and (18), for $\forall i, j \in \{0, 1, \dots, n-1\}$ and $i \neq j, \tau = j - i$, we can write

$$\text{Cov}[E_i, E_j] = R_E(i, j - i) - \mathbb{E}[E]^2 \quad (56)$$

$$= \mathbb{E}[(A_{I,i}^2 + A_{Q,i}^2)(A_{I,j}^2 + A_{Q,j}^2)] - 4\mathbb{E}[A^2]^2 \quad (57)$$

$$= 2\mathbb{E}[A_i^2 A_j^2] - 2\mathbb{E}[A^2]^2, \quad (58)$$

$$= 2 \sum_{a,b \in \mathcal{A}} \mathbb{P}_{A_i, A_j}(a, b) a^2 b^2 - 2\mathbb{E}[A^2]^2, \quad (59)$$

where (57) follows from (20), (22) and (24), and (58) from the fact that amplitudes in the I/Q branches in Fig. 1 are independent from each other. It can be seen in (59) that $R_E(i, j - i)$ and $\text{Cov}[E_i, E_j]$ is determined by the joint probability \mathbb{P}_{A_i, A_j} , i.e.,

$$\mathbb{P}_{A_i, A_j}(a, b) = \mathbb{P}_{A_i}(a) \cdot \mathbb{P}_{A_j|A_i}(b|a), \quad (60)$$

where $\mathbb{P}_{A_i}(a) = n_a/n$ (see (15)). For a given $A_i = a$, the amplitude composition at time j is updated, and thus,

$$\mathbb{P}_{A_j|A_i}(b|a) = \begin{cases} \frac{n_a - 1}{n - 1}, & \text{if } b = a \\ \frac{n_b}{n - 1}, & \text{if } b \neq a \end{cases}. \quad (61)$$

The joint probability \mathbb{P}_{A_i, A_j} in (60) is thus, independent of i and j , and so is $R_E(i, j - i) = \rho$ and $\text{Cov}[E_i, E_j]$.

Finally, based on the fact that ρ is a constant, we compute ρ and the autocovariance. Similar to (12), the total energy of a symbol energy block is a constant, i.e.,

$$\sum_{i=0}^{n-1} E_i = 2 \sum_{i=0}^{n-1} A_i^2 = 2 \sum_{a \in \mathcal{A}} a^2 n_a. \quad (62)$$

In analogy to (13), (62) shows that these n symbol energies also satisfy a linear relationship. By taking expectation on both sides of (62), we have

$$\sum_{i=0}^{n-1} \mathbb{E}[E_i] = 2 \sum_{a \in \mathcal{A}} a^2 n_a. \quad (63)$$

Subtracting (63) from (62) yields

$$\sum_{i=0}^{n-1} (E_i - \mathbb{E}[E_i]) = 0, \quad (64)$$

which can be rewritten as

$$E_j - \mathbb{E}[E_j] = - \sum_{\substack{j'=0 \\ j' \neq j}}^{n-1} (E_{j'} - \mathbb{E}[E_{j'}]). \quad (65)$$

By using (17) and (65), $\text{Cov}[E_i, E_j]$ is expanded, i.e.,

$$\text{Cov}[E_i, E_j] = \mathbb{E} \left[(E_i - \mathbb{E}[E_i]) \left(- \sum_{\substack{j'=0 \\ j' \neq j}}^{n-1} (E_{j'} - \mathbb{E}[E_{j'}]) \right) \right] \quad (66)$$

$$= - \mathbb{E} \left[(E_i - \mathbb{E}[E_i])^2 \right] - \sum_{\substack{j'=0 \\ j' \neq j, i}}^{n-1} \mathbb{E}[(E_i - \mathbb{E}[E_i])(E_{j'} - \mathbb{E}[E_{j'}])] \quad (67)$$

$$= - \text{Var}[E_i] - \sum_{\substack{j'=0 \\ j' \neq j, i}}^{n-1} \text{Cov}[E_i, E_{j'}]. \quad (68)$$

Therefore, (68) can be rewritten as

$$\text{Cov}[E_i, E_j] + (n-2)\text{Cov}[E_i, E_j] = -\text{Var}[E_i]. \quad (69)$$

By using (23) in (69), the equality in (31) is obtained. The inequality in (31) clearly follows from the fact that the variance $\text{Var}[|X|^2]$ is positive. The last step in the proof is to show that the autocorrelation $R_E(i, \tau) = \rho$ is given by (30). This follows from substituting (31) into (18). Because of the negative autocovariance, it can be observed that ρ is smaller than $\mathbb{E}[|X|^2]^2$. This completes the proof.

APPENDIX B PROOF OF THEOREM 5

To prove (42), we use (36) in (40) to obtain

$$\mathbb{E}[G^W] = \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=i-W/2}^{i+W/2} \mathbb{E}[|X_j|^2] \quad (70)$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=i-W/2}^{i+W/2} \mathbb{E}[|X|^2] \quad (71)$$

$$= (W+1)\mathbb{E}[|X|^2], \quad (72)$$

where (70) uses the linearity of expectation, and (71) follows from (22).

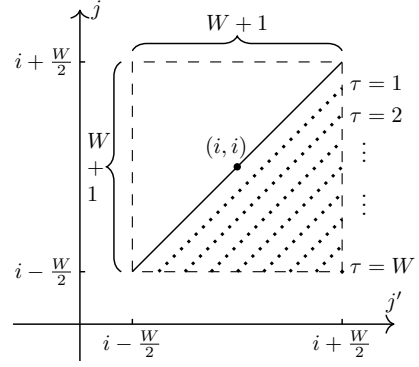


Fig. 13. An illustration of possible combinations of j and j' for the double summations of the last term in (76). The dots at each diagonal direction represent $W - \tau + 1$ possible pairs of j and j' when $j - j' > 0$.

To prove (43), we begin with $\text{Var}[G_i^W]$. Using (36), Lemma 1, and [48, Thm. 9.2], we have

$$\text{Var}[G_i^W] = \sum_{j=i-W/2}^{i+W/2} \text{Var}[|X_j|^2] + 2 \sum_{j=i-W/2}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} \text{Cov}[E_j, E_{j'}] \quad (73)$$

$$= (W+1) \text{Var}[|X|^2] + 2 \sum_{j=i-W/2}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} \text{Cov}[E_j, E_{j'}] \quad (74)$$

$$= (W+1) \text{Var}[|X|^2] + 2 \sum_{j=i-W/2}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} [R_E(j, j' - j) - \mathbb{E}[|X|^2]^2] \quad (75)$$

$$= (W+1) \text{Var}[|X|^2] - W(W+1)\mathbb{E}[|X|^2]^2 + 2 \sum_{j=i-W/2}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} R_E(j, j' - j). \quad (76)$$

where (74) follows from (23), and (75) from (18).

The first two terms in (76) are known, since $\text{Var}[|X|^2]$ and $\mathbb{E}[|X|^2]$ are given in (25) and (24). Meanwhile, $R_E(j, j' - j)$ depends on the time instant j and delay $\tau = j' - j > 0$. Therefore, based on (41), only the last term in (76) needs to be considered for averaging over n time instants, i.e.,

$$\frac{1}{n} \sum_{i=0}^{n-1} \left(2 \sum_{j=i-W/2}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} R_E(j, j' - j) \right) \quad (77)$$

$$= 2 \sum_{\tau=1}^W \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=i-W/2}^{i+W/2} R_E(j, \tau) \quad (78)$$

$$= 2 \sum_{\tau=1}^W (W - \tau + 1) \bar{R}_E(\tau), \quad (79)$$

where (78) can be explained by observing Fig. 13, which shows possible pairs of j and j' and the corresponding values

of τ . To obtain (79), we observe in Fig. 13 that for each value of τ , there are $W - \tau + 1$ pairs in the corresponding diagonal, whose $R_E(j, \tau)$ are averaged. Hence, (79) is obtained. Finally, $\overline{\text{Var}}[G^W]$ in (43) is simply the sum of the first two terms in (76) and (79), which completes the proof.

APPENDIX C PROOF OF THEOREM 6

In this Appendix, we show that in the special case of $n \leq W + 2$, the computation of $\text{Var}[G_i^W]$ can be simplified by considering the symbol energy “pattern” inside the window. This leads to a simpler expression for the EDI in (49).

Given a window length W , the window has $W + 1$ symbol energies. This window in general covers multiple symbol energy blocks, and thus we can write

$$W + 1 = un + r, \quad (80)$$

where $u \in \mathbb{N}$ is the maximum number of complete symbol energy blocks covered by the window, and r is the remainder when $W + 1$ is divided by n ($0 \leq r \leq n - 1$).

As the window slides, the symbol energy “pattern” covered by the window varies cyclically with a period n , which is illustrated in Fig. 14. The pattern shows that the sum of symbol energies with a shaded background is constant, whereas the energy of the incomplete blocks located at two edges of the windows is random. Therefore, the constant part of energy has no contribution to window energy variance, and thus can be removed in the computation of the variance. For example, for the pattern at the top of Fig. 14, it consists of u complete blocks at the right side, while at its left edge there are r symbols from the adjacent block. Hence, the variance of the top pattern is simplified by only considering these r symbols. Such simplification can be done by using (76). After substituting $W + 1$ with r , and $R_E(j, j' - j)$ with ρ , we have

$$\begin{aligned} \text{Var}[G_i^W] &= r \text{Var}[|X|^2] - r(r-1) \mathbb{E}[|X|^2]^2 \\ &\quad + r(r-1)\rho. \end{aligned} \quad (81)$$

As long as $W + 1 \geq n - 1$, it can be concluded from Fig. 14 that: (i) two edges with random symbol energies contribute to the variance, and (ii) these n patterns keep $0, 1, \dots, n - 1$ mutually correlated symbol energies at both edges, respectively. Under these circumstances, $\overline{\text{Var}}[G^W]$ is obtained by averaging the variance over n patterns, i.e.,

$$\begin{aligned} \overline{\text{Var}}[G^W] &= \frac{2}{n} \sum_{i=0}^{n-1} \left[i \text{Var}[|X|^2] - i(i-1) \mathbb{E}[|X|^2]^2 \right. \\ &\quad \left. + i(i-1)\rho \right] \end{aligned} \quad (82)$$

$$= \frac{2 \text{Var}[|X|^2]}{n(n-1)} \sum_{i=0}^{n-1} (ni - i^2) \quad (83)$$

$$= \frac{(n+1) \text{Var}[|X|^2]}{3}, \quad (84)$$

where (83) follows from (18) and (31). By substituting (84) and (42) into (39), EDI in (49) is obtained.

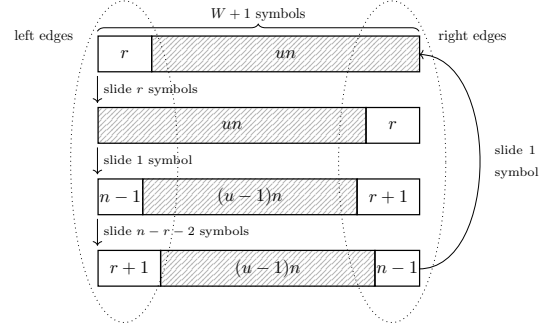


Fig. 14. An illustration of four symbol energy patterns as window slides to the right through the symbol energy sequence. The shaded area represents u complete blocks.

APPENDIX D PROOF OF COROLLARY 7

We first discuss the bounds on $\overline{\text{Var}}[G^W]$, which directly determine the bounds on EDI. For any finite blocklength n , the $\overline{\text{Var}}[G^W]$ of the CCDM QAM symbol sequence satisfies

$$0 \leq \overline{\text{Var}}[G^W] \leq (W + 1) \text{Var}[|X|^2]. \quad (85)$$

The variance is always positive, hence the left inequality in (85) is true. Based on (74), due to negative autocovariance from (31), we have,

$$\begin{aligned} \text{Var}[G_i^W] &= (W + 1) \text{Var}[|X|^2] \\ &\quad + 2 \sum_{j=i-W/2-1}^{i+W/2-1} \sum_{j'=j+1}^{i+W/2} \text{Cov}[E_j, E_{j'}], \end{aligned} \quad (86)$$

$$\leq (W + 1) \text{Var}[|X|^2], \quad (87)$$

where (87) holds with equality when $W = 0$. After using (41), we have the right inequality in (85). By using (85) and (42) in (39), the inequalities for the EDI in (50) is obtained.

We now prove how the bounds in (50) are achieved asymptotically by W . In terms of the upper bound in (51), by setting $W = 0$, the window only encompasses one symbol energy, thereby $\overline{\text{Var}}[G^W] = \text{Var}[|X|^2]$ and $\mathbb{E}[G^W] = \mathbb{E}[|X|^2]$. With (39) and using (10), (51) is obtained. From (49), it can be seen that for $n < W + 2$, the EDI tends to 0 for $W \rightarrow \infty$. This completes the proof.

REFERENCES

- [1] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.
- [2] J. Cho and P. J. Winzer, “Probabilistic constellation shaping for optical fiber communications,” *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
- [3] P. Schulte and G. Böcherer, “Constant composition distribution matching,” *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [4] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, “Multiset-partition distribution matching,” *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1885–1893, Mar. 2019.
- [5] F. Steiner, P. Schulte, and G. Bocherer, “Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping,” in *52nd Annual Conference on Information Sciences and Systems (CISS)*. Princeton, NJ, USA, 21–23 Mar. 2018.
- [6] F. M. J. Willems and J. Wuijts, “A pragmatic approach to shaped coded modulation,” in *Proc. Symp. Commun. Veh. Technol. Benelux*. Delft, The Netherlands, Oct. 1993.

- [7] Y. C. Gültekin, W. J. van Houtum, S. Şerbetli, and F. M. J. Willems, "Constellation shaping for IEEE 802.11," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Commun.* Montreal, QC, Canada, Oct. 2017.
- [8] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces—Part I: Optimum shaping, shell mapping," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1799–1808, Nov. 1993.
- [9] J. Cho, S. Chandrasekhar, R. Dar, and P. J. Winzer, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. Eur. Conf. Opt. Commun.* Düsseldorf, Germany, Sep. 2016, Paper W1C.2.
- [10] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, Nov. 2016.
- [11] A. Amari, S. Goossens, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, C. M. Okonkwo, F. M. J. Willems, and A. Alvarado, "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5926–5936, Dec. 2019.
- [12] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.
- [13] A. Ghazisaeidi, I. F. de Jauregui Ruiz, R. Rios-Müller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet *et al.*, "Advanced C+L-band transoceanic transmission systems based on probabilistically shaped PDM-64QAM," *J. Lightw. Technol.*, vol. 35, no. 7, pp. 1291–1299, Apr. 2017.
- [14] J. Cho, X. Chen, S. Chandrasekhar, G. Raybon, R. Dar, L. Schmalen, E. Burrows, A. Adamiecki, S. Corteselli, Y. Pan *et al.*, "Trans-atlantic field trial using high spectral efficiency probabilistically shaped 64-QAM and single-carrier real-time 250-Gb/s 16-QAM," *J. Lightw. Technol.*, vol. 36, no. 1, pp. 103–113, Jan. 2018.
- [15] S. L. Olsson, J. Cho, S. Chandrasekhar, X. Chen, E. C. Burrows, and P. J. Winzer, "Record-high 17.3-bit/s/Hz spectral efficiency transmission over 50 km using probabilistically shaped PDM 4096-QAM," in *Proc. Opt. Fiber Commun. Conf.* San Diego, CA, USA, Mar. 2018, Paper Th4C.5.
- [16] G. Agrawal, *Nonlinear Fiber Optics, Third Edition*. Academic Press, Jan. 2001.
- [17] Z. Qu and I. B. Djordjevic, "Geometrically shaped 16QAM outperforming probabilistically shaped 16QAM," in *Proc. Eur. Conf. Opt. Commun.* IEEE, Sep. 2017, Paper Th.2.F.4.
- [18] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information," in *Proc. Opt. Fiber Commun. Conf.* San Diego, CA, USA, Mar. 2020, Paper W3D.4.
- [19] J. Renner, T. Fehenberger, M. P. Yankov, F. Da Ros, S. Forchhammer, G. Böcherer, and N. Hanik, "Experimental comparison of probabilistic shaping methods for unrepeatable fiber transmission," *J. Lightw. Technol.*, vol. 35, no. 22, pp. 4871–4879, Nov. 2017.
- [20] E. Sillekens, D. Semrau, G. Liga, N. A. Shevchenko, Z. Li, A. Alvarado, P. Bayvel, R. I. Killely, and D. Lavery, "A simple nonlinearity-tailored probabilistic shaping distribution for square QAM," in *Proc. Opt. Fiber Commun. Conf.* San Diego, CA, USA, Mar. 2018, Paper M3C.4.
- [21] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.
- [22] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *IEEE Int. Symp. Inform. Theory*. Honolulu, HI, USA, June 2014, pp. 2794–2798.
- [23] —, "Time varying ISI model for nonlinear interference noise," in *Optical Fiber Communication Conference*. San Diego, CA, USA, Mar. 2014, Paper W2A.62.
- [24] M. P. Yankov, K. J. Larsen, and S. Forchhammer, "Temporal probabilistic shaping for mitigation of nonlinearities in optical fiber systems," *J. Lightw. Technol.*, vol. 35, no. 10, pp. 1803–1810, May 2017.
- [25] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, K. Parsons, and H. Griesser, "Analysis of nonlinear fiber interactions for finite-length constant-composition sequences," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 457–465, Jan. 2020.
- [26] S. Goossens, S. Van der Heide, M. van den Hout, A. Amari, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, F. M. J. Willems, A. Alvarado *et al.*, "First experimental demonstration of probabilistic enumerative sphere shaping in optical fiber communications," in *24th OptoElectronics and Communications Conference (OECC)*. Fukuoka, Japan, July 2019.
- [27] T. Fehenberger, H. Griesser, and J.-P. Elbers, "Mitigating fiber nonlinearities by short-length probabilistic shaping," in *Proc. Opt. Fiber Commun. Conf.* San Diego, CA, USA, Mar. 2020, Paper Th11.2.
- [28] S. Civelli, E. Forestieri, and M. Secondini, "Interplay of probabilistic shaping and carrier phase recovery for nonlinearity mitigation," Sep. 2020. [Online] Available: <https://arxiv.org/abs/2009.01135>.
- [29] P. Skvortcov, I. D. Phillips, W. Forysiak, T. Koike-Akino, K. Kojima, K. Parsons, and D. Millar, "Huffman-coded sphere shaping for extended-reach single-span links," *IEEE J. Sel. Top. Quantum Electron.*, Jan. 2021, (early access).
- [30] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," *IEEE Photon. Technol. Lett.*, vol. 23, no. 11, pp. 742–744, June 2011.
- [31] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.
- [32] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.
- [33] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of nonlinear noise in long, dispersion-uncompensated fiber links," *Opt. Express*, vol. 21, no. 22, pp. 25 685–25 699, Nov. 2013.
- [34] —, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.
- [35] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "On the accuracy of the GN-model and on analytical correction terms to improve it," Jan. 2014. [Online] Available: <https://arxiv.org/abs/1401.6946>.
- [36] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution," *J. Lightw. Technol.*, vol. 33, no. 2, pp. 459–473, Jan. 2015.
- [37] T. Fehenberger and A. Alvarado, "Analysis and optimisation of distribution matching for the nonlinear fibre channel," in *Proc. Eur. Conf. Opt. Commun.* Dublin, Ireland, Sep. 2019, Poster Session 1.
- [38] T. Fehenberger, "On the impact of finite-length probabilistic shaping on fiber nonlinear interference," June 2020. [Online] Available: <https://arxiv.org/abs/2006.07004>.
- [39] A. Mecozzi and R. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, June 2012.
- [40] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," *J. Lightw. Technol.*, vol. 29, no. 17, pp. 2570–2576, Sep. 2011.
- [41] J. Armstrong, "OFDM for optical communications," *J. Lightw. Technol.*, vol. 27, no. 3, pp. 189–204, Feb. 2009.
- [42] B. Chen, A. Alvarado, S. van der Heide, M. van den Hout, H. Hafermann, and C. Okonkwo, "Analysis and experimental demonstration of orthant-symmetric four-dimensional 7 bit/4D-sym modulation for optical fiber communication," Mar. 2020. [Online] Available: <https://arxiv.org/abs/2003.12712>.
- [43] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *J. Lightw. Technol.*, vol. 34, no. 16, pp. 3884–3889, May 2016.
- [44] F. Gray, "Pulse code communication," US Patent 2 632 058, Mar. 1953.
- [45] Y. C. Gültekin, T. Fehenberger, A. Alvarado, and F. M. J. Willems, "Probabilistic shaping for finite blocklengths: Distribution matching and sphere shaping," *Entropy*, vol. 22, no. 5, p. 581, May 2020.
- [46] Y. C. Gültekin, W. J. Van Houtum, A. Koppelaar, and F. M. J. Willems, "Comparison and optimization of enumerative coding techniques for amplitude shaping," *IEEE Commun. Lett.*, Dec. 2020, (early access).
- [47] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes, Fourth Edition*. Tata McGraw-Hill Education, 2002.
- [48] R. D. Yates and D. J. Goodman, *Probability and stochastic processes: a friendly introduction for electrical and computer engineers, Third Edition*. John Wiley & Sons, 2014.
- [49] G. B. Giannakis, "Cyclostationary signal analysis," in *Digital Signal Processing Handbook, First Edition*. V. K. Madiseti and D. Williams, Eds, Boca Raton, FL: CRC, 1998.
- [50] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference: Fourth Edition, Revised and Expanded*. CRC press, May 2003.