Shaping Lightwaves in Time and Frequency for Optical Fiber Communication

Junho Cho (✉ junho.cho@nokia-bell-labs.com)

Nokia Bell Labs

Xi Chen

Nokia Bell Labs

Greg Raybon

Nokia Bell Labs

Di Che

Nokia Bell Labs

Ellsworth Burrows

Nokia Bell Labs

Samuel Olsson

Nokia

Robert Tkach

Nokia Bell Labs

Article

Keywords: Kerr nonlinearity, optical communications, lightwaves

Posted Date: August 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-452651/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on February 10th, 2022. See the published version at <https://doi.org/10.1038/s41467-022-28349-x>.

Shaping Lightwaves in Time and Frequency for Optical Fiber Communication

Junho Cho^{1*}, Xi Chen¹, Greg Raybon¹, Di Che¹, Ellsworth Burrows¹, Samuel Olsson², and
Robert Tkach¹

¹ Nokia Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974, USA

² Nokia, 600 Mountain Ave, Murray Hill, NJ 07974, USA

*e-mail: junho.cho@nokia-bell-labs.com

Abstract

In optical communications, sphere shaping is used to limit the energy of lightwaves to within a certain value over a period. This minimizes the energy required to contain information, allowing the rate of information transmission to approach the theoretical limit if the transmission medium is linear. In optical fiber, however, the sphere shaping induces Kerr nonlinearity in a peculiar way that makes analysis of transmission performance difficult, potentially lowering the communications capacity. In this article, we show how the impact of sphere shaping on Kerr nonlinearity varies with the chromatic dispersion and the structure of shaped lightwaves in time and frequency, and give insights into why the structure matters. As a practical consequence, by optimally controlling the structure of lightwaves in time and frequency, it is experimentally demonstrated that the information rate can be increased by up to 25% in low-dispersion channels on a 2824-km dispersion-managed wavelength-division multiplexed optical fiber link.

Introduction

The bandwidth of lightwaves was once considered an almost limitless resource for optical communications. However, due to the ever-increasing demand for communication capacity, the efficient use of bandwidth has become an important goal for today's optical communications. This has led to the use of coherent transmission technology and high-order quadrature amplitude modulation (QAM), which allow more information to be contained in a single communication symbol. Lately, an information theoretic approach has been adopted for optical QAM systems to improve bandwidth efficiency by applying forward error correction (FEC) coding and probabilistic constellation shaping (PCS)¹⁻³ technology. This has created an intriguing phenomenon in nonlinear optical fiber communications as presented in this article, which was never seen in other communications systems.

An implementation of PCS called *sphere shaping*⁴⁻¹⁰, notably important in both theory and real-world applications, only creates the blocks of amplitudes with the total energy not exceeding a certain limit. This fundamentally minimizes the average energy of a spectro-temporal block of lightwaves to contain a given amount of information. When the lightwaves are transmitted through a linear medium, increasing the shaping block length towards infinity allows the information rate to approach the theoretical limit¹¹. However, when the lightwaves are transmitted over optical fiber, the sphere shaping can increase Kerr nonlinearity¹²⁻¹⁵ and lower the nonlinear capacity¹⁴ while retaining its fundamental energy efficiency⁴⁻¹⁰. The unique way that the sphere shaping constrains energy changes the *statistical properties* and *temporal structure* of the shaped lightwaves differently depending on the shaping block length, and accordingly the manifestation of Kerr nonlinearity varies^{10,15}.

In addition to the shaping block length, the symbol rate at which communication symbols can change values also affects Kerr nonlinearity. This is because the symbol rate determines the way a continuum of the lightwaves, in time and frequency, is divided into small blocks of symbols in today's densely packed wavelength-division multiplexing (WDM) systems. Namely, the frequency bandwidth of symbols increases in proportion to the symbol rate while their time duration decreases inversely. Therefore, changing the symbol rate changes both the *spectral* and *temporal* properties of the lightwaves. For the traditional unshaped lightwaves that carry statistically independent and identically distributed (i.i.d.) symbols, the influence of the symbol rate has been well established through simulation, experiment, and analysis^{16,17}. However, in the case of sphere-shaped lightwaves, the presence of a temporal structure of energy invalidates the i.i.d. assumption and makes it difficult to study the Kerr nonlinearity analytically.

In this article, we provide a comprehensive picture showing the relationship between the spectro-temporal block size of sphere-shaped lightwaves, chromatic dispersion of fiber, and Kerr nonlinearity. Not only does this finding contribute to an intrinsic understanding of the peculiar characteristics of sphere-shaped lightwaves, it also allows the most efficient use of optical bandwidth by optimizing sphere shaping in WDM systems. By adjusting the spectro-temporal block size of sphere shaping, we experimentally demonstrate that the effective signal-to-noise ratio (SNR) increases by up to 1.1 dB and the net data rate (NDR) increases by up to 25% in low-performing channels in a 3.7-THz-wide full C-band transmission system.

Results

Sphere Shaping of Lightwaves. We consider a polarization-division multiplexed (PDM) M^2 -ary QAM system. A digital shaping encoder at the transmitter produces blocks of amplitudes chosen from an equally spaced numerical alphabet $\mathcal{A} = \{1, 3, \dots, 2M - 1\}$ in arbitrary units (a.u.). Each amplitude is multiplied by an equiprobable sign in $\{+1, -1\}$, resulting in a probability distribution over $\{\pm 1, \pm 3, \dots, \pm(2M - 1)\}$ symmetric around the origin. The symmetry of the probability allows for legitimate analysis with only positive amplitudes, and hence we omit the sign throughout this article. In our system, four *consecutive* amplitudes constitute one dual-polarization symbol that maintains the value over a period of T_{sym} (s) and can change the value at a rate of $R_{sym} = 1/T_{sym}$ (Bd). We therefore set up the encoder to produce a $4n$ -long amplitude block over n symbol periods. Let us denote this amplitude block as $\mathbf{a} = [a_1, \dots, a_{4n}]$ with $a_i \in \mathcal{A}$ for $i = 1, \dots, 4n$ and the corresponding symbol block as $\mathbf{x} = [x_1, \dots, x_n]$ with dual-polarization symbols $x_i = [a_{4i-3}, a_{4i-2}, a_{4i-1}, a_{4i}] \in \mathcal{A}^4$ for $i = 1, \dots, n$. Then, in traditional QAM systems, the amplitude block $\mathbf{a}$ (and hence the symbol block $\mathbf{x}$) can have a maximum total energy of $E_{Unshaped}^* = \max \|\mathbf{a}\|^2 = 4n(2M - 1)^2$ when all amplitudes are at their maximum. On the other hand, sphere shaping imposes a limit E_{Shaped}^* on the total energy such that only amplitudes that jointly fulfill $\|\mathbf{a}\|^2 \leq E_{Shaped}^*$ are created by the encoder, where $E_{Shaped}^* < E_{Unshaped}^*$ in general. This method is called *sphere shaping*⁴⁻¹⁰ because if every shaped block is plotted as a point in $4n$ -dimensional signal space, with the i -th amplitude being the position of the point on the i -th coordinate axis, the points uniformly fill (due to the symmetry by equiprobable signs) a $4n$ -dimensional (hyper-) sphere of radius $\sqrt{E_{Shaped}^*}$. As we decrease the limit E_{Shaped}^* , more combinations of amplitudes are not allowed to be created, hence less information can be contained per block. This is because the maximum

number of information bits per block is given by $\lfloor \log_2 N \rfloor$ if there are N possible combinations of amplitudes, where $\lfloor \cdot \rfloor$ denotes the floor function. The maximum information rate in bits per dual-polarization symbol is then given by $H = \lfloor \log_2 N \rfloor / n + 4$, where the addition by 4 accounts for four sign bits per symbol. Fig. 1(a) shows the probability distributions of the total energy in a block of length $n = 5$. Compared to the unshaped PDM 16-QAM (left figure) with $H = 8$ and $E_{\text{Unshaped}}^* = 180$, sphere shaping (right figure) substantially reduces the maximum energy to $E_{\text{Shaped}}^* = 52$ (71% reduction) at the expense of only 1.6 bit reduction in H (20% reduction). In a densely packed WDM system with identical channel configurations, such probabilistic energy distributions as in Fig. 1(a) represent each rectangular block that divides lightwaves in the time-frequency plane as shown in Fig. 1(b), where the width and height of the block are determined by both R_{Sym} and n . In the following sections, we will minimize the nonlinear interference (NLI) by optimally dividing the time-frequency plane.

Power Deviations in Sphere-Shaped Lightwaves. In long-haul optical fiber links, the NLI generated from i.i.d. symbols behaves similarly to additive white Gaussian noise (AWGN)¹⁸⁻²⁰. The average NLI power increases in proportion to the cube of the average symbol power as¹⁸⁻²⁰

$$\langle P_{\text{NLI}} \rangle = \eta \langle \|x\|^2 \rangle^3, \quad (1)$$

where $\langle \cdot \rangle$ denotes statistical averaging, η is the *NLI coefficient* determined mostly by link parameters and to some extent by modulation. Denoting the symbol power *normalized* to have unit mean by $\mathcal{p} \triangleq \|x\|^2 / \langle \|x\|^2 \rangle$, η increases as the *central moment* μ_n of $\mathcal{p}$ increases, where μ_n is defined as

$$\mu_n = \langle (\mathcal{p} - 1)^n \rangle. \quad (2)$$

Namely, μ_n quantifies the n -th order deviation of instantaneous $\mathcal{p}$ from average $\langle \mathcal{p} \rangle = 1$, and the more $\mathcal{p}$ deviates from $\langle \mathcal{p} \rangle$, the greater the NLI. Note that the instantaneous and average powers are implicitly measured over a symbol period T_{Sym} and infinite time, respectively.

A question arises here as to whether measuring the instantaneous power over a *single* symbol period is appropriate to study the light propagation effect, since the shaped symbols have a unique temporal energy structure that, while propagating through optical fiber, results in a different evolution compared to i.i.d. symbols. The presence of the temporal energy structure in dispersive medium suggests that a longer time period than T_{Sym} may better characterize the light propagation effect^{21,22}. To find an appropriate time to measure the instantaneous power, we define a new statistical figure of merit called the *windowed central moment* of $\mathcal{p}$ as

$$\bar{\mu}_n = \langle (\langle p \rangle_w - 1)^n \rangle \cdot \underbrace{(2w)^{n-1}}_{(a)}, \quad (3)$$

where $\langle \cdot \rangle_w$ denotes a moving average filter with a sliding window of length w symbols. Namely, $\bar{\mu}_n$ quantifies the n -th order power deviation by the instantaneous power measured over w symbol periods, with w being a free parameter that will be optimized in the following section. Since $\langle p \rangle_w$ approaches 1 as w increases, the term (a) on the right-hand side ensures that $\bar{\mu}_n$ remains the same regardless of w , if p is i.i.d. Here, the factor 2 compensates for $\langle p \rangle_w$ being averaged over 2 polarizations. Without the term (a), the second windowed central moment $\bar{\mu}_2$ equals the energy dispersion index that was lately suggested²².

Fig. 1(c) shows the probability distributions of $\langle p \rangle_w$ for PDM 16-QAM symbols without (left figures) and with (right figures) sphere shaping ($n = 5$). When the instantaneous power is measured with $w = 1$ as conventionally done (upper figures), it *appears to be* more spread around the average for the shaped symbols than for the unshaped symbols. According to the statistical analysis developed for i.i.d. symbols^{20,23}, this implies that *sphere shaping increases NLI* compared to the unshaped lightwaves. On the contrary, recent simulations and experiments have shown that, if n is small as in Fig. 1(c), *sphere shaping reduces NLI*^{7,15} compared to the unshaped lightwaves. This contradiction may be attributed to the fact that, when there exist small structures of power in time, the power deviation can be measured differently depending on w . For example, Fig. 1(c) shows that, as we increase w from 1 (upper figures) to 40 (lower figures), $\langle p \rangle_w$ becomes concentrated near $\langle p \rangle$ for the shaped symbols (right figures) due to the block-wise energy constraint, whereas it maintains the variance for the unshaped symbols (left figures). Depicted in Fig. 1(d) is $\bar{\mu}_2$ as a function of w . As w increases, $\bar{\mu}_2$ remains the same for *all* types of i.i.d. symbols (dashed lines), but it decreases for the finite-length shaped symbols (solid lines). Therefore, if the manifestation of Kerr nonlinearity depends on the power deviation measured over a longer period than T_{sym} , short sphere shaping can reduce NLI, as we will investigate further in the following section.

Optimization of Sphere Shaping in the Time-Frequency Plane. We first study the influence of sphere shaping on NLI using split-step simulations^{24,25}. We transmit symbols in a total bandwidth of 100 GHz centered at 193.4 THz (1550.1 nm) over N_{ch} channels that evenly divide the total bandwidth, with $N_{ch} = 1, 2, 4, 8, 16, 32, 64$. The symbols are transmitted in each channel at a rate of $R_{sym} = 88/N_{ch}$ GBd, ranging from 1.375 to 88 GBd, leaving 12% spectral margins for root-raised cosine (RRC) pulse shaping with a roll-off factor of 0.1. Sphere shaping is

performed in each channel with $n = 5, 10, 20, 40, 80, 320, 1280, 5120$, with a fixed $H = 6.4$ bits per dual-polarization symbol. Therefore, all the WDM configurations under test send data at the same rate in the same total bandwidth. We assume, throughout the article, the use of a rate-0.8 field-programmable gate array (FPGA)-verified spatially-coupled low-density parity-check (LDPC) code²⁶ for forward error correction (FEC), which has a normalized generalized mutual information (NGMI) threshold^{27,28} of $NGMI^* = 0.86$ for error-free decoding. With this, the same total information rate of 422.4 Gb/s can be achieved by all the WDM configurations. Four links with regular 60-km spans are constructed by standard single-mode fiber (SSMF) of length L_{SSMF} km followed by dispersion compensating fiber (DCF) of length $L_{DCF} = 60 - L_{SSMF}$ km in each span (with dispersion coefficients of 17.24 and -80 ps/nm/km, respectively), as shown in Table 1. An Erbium doped fiber amplifier (EDFA) recovers the launch power after every span with 4.5 dB noise figure. Complete details of the simulation setup are given in the Methods section.

Fig. 2(a) shows the effective SNR as a function of the symbol rate (on x -axis) and shaping block duration (on y -axis), defined as

$$SNR_{Eff} = \frac{\langle \|x\|^2 \rangle}{\langle P_{ASE} \rangle + \langle P_{NLI} \rangle} \quad (4)$$

where $\langle P_{ASE} \rangle$ is the average amplified spontaneous emission (ASE) noise power. The signal, ASE, and NLI powers are all measured over the total bandwidth of 100 GHz. The contour lines are obtained by interpolating 56 simulation points (red dots). The launch power is optimized for each point to maximize SNR_{Eff} . The top and bottom points at each R_{Sym} represent $n = 5120$ and 5, respectively, but for the same n , the shaping block duration in *nanoseconds* on y -axis varies with R_{Sym} . The effective SNR is evaluated at the transmission distances where the NGMI is near the threshold. With our system settings, $n = 5120$ approximately represents i.i.d. shaping from the dispersion perspective, since each symbol is dispersed over less than 5120 symbol periods in all conditions except for $R_{Sym} = 88$ GBd in Link D. Looking at Fig. 2(a), we notice the following: (i) like the i.i.d. uniform QAM¹⁷, the optimal R_{Sym} for i.i.d. shaping to maximize SNR_{Eff} (yellow stars) decreases as the total net dispersion D_{Total} increases, (ii) reducing n improves SNR_{Eff} if R_{Sym} is near optimal (red stars), and (iii) the optimal R_{Sym} is higher for finite-length shaping than for i.i.d. shaping (compare the yellow and red stars), since increasing R_{Sym} reduces the shaping block duration in absolute time for the same n . The highest SNR_{Eff} at the smallest n , however, does not necessarily

maximize the end-to-end communication performance measured in NGMI, as shown in Fig. 2(b), due to the complex interplay with the fundamental shaping efficiency that decreases as n decreases. In general, (i) the optimal n for NGMI tends to increase with D_{Total} , and (ii) the optimal R_{Sym} for NGMI is higher than the optimal R_{Sym} for SNR_{Eff} .

We scrutinize the nonlinear effect by extracting the NLI coefficient η from the simulation results as

$$\eta = \left[\frac{\langle \|x\|^2 \rangle}{SNR_{Eff}} - \langle P_{ASE} \rangle \right] / \langle \|x\|^2 \rangle^3, \quad (5)$$

derived from equations (1) and (4). The result is shown in Fig. 2(c-1). We factorize η as

$$\eta = \eta^\infty \times \Delta\eta, \quad (6)$$

where η^∞ represents the NLI coefficient of i.i.d. shaping, approximated by $n = 5120$, cf. Fig. 2(c-2), and $\Delta\eta \leq 1$ is a correction factor for finite-length shaping, cf. Fig. 2(c-3). Furthermore, from another set of simulations using the same WDM settings as above, yet with only one channel being transmitted in each setting, we evaluate η_{SPM} produced solely by self-phase modulation (SPM)^{29,30}, cf. Fig. 2(c-4). Then, by subtracting the SPM contribution from the full-fledged simulation results, we obtain η_{XPM} produced solely by cross-phase modulation (XPM)^{31,32}, cf. Fig. 2(c-7). Looking at Figs. 2(c-5), (c-8), as R_{Sym} increases, η_{SPM}^∞ increases whereas η_{XPM}^∞ decreases, as a consequence of the relative increase in channel bandwidth within a fixed total bandwidth (see also ¹⁷), and thus the combined effect η^∞ changes just marginally, cf. Fig. 2(c-2). Looking at the finite-length effect in Figs. 2(c-6), (c-9), the influence of $\Delta\eta_{SPM}$ and $\Delta\eta_{XPM}$ is prominent only near the red circled areas, where their base coefficients η_{SPM}^∞ and η_{XPM}^∞ are large. Areas far from the circled areas have a negligible impact on SNR_{Eff} . Combining $\Delta\eta_{SPM}$ and $\Delta\eta_{XPM}$, $\Delta\eta$ changes much more than η^∞ , indicating that n has a much greater impact on SNR_{Eff} than R_{Sym} near the optimal point. This leads to SNR_{Eff} of Fig. 2(a). More results are provided in Supplementary Figs. 1-4.

The window size for $\bar{\mu}_2$ is then determined in a similar way to ²² by calculating the Pearson correlation coefficient $\rho(\eta, \bar{\mu}_2)$ that quantifies how much η is correlated with $\bar{\mu}_2$. This is done separately for SPM and XPM, as shown in Fig. 3(a). The red lines show the linear fits of the optimal window sizes w_{SPM}^* and w_{XPM}^* that produce the greatest $\rho(\eta_{SPM}, \bar{\mu}_2)$ and $\rho(\eta_{XPM}, \bar{\mu}_2)$, respectively, which are given by

$$w_{SPM}^* \approx 2R_{Sym}B_{Ch}|\beta_2|L_{Span}N_{Span} \quad (7)$$

and

$$w_{XPM}^* \approx 2R_{Sym}B_{Ch}\sqrt{N_{Ch}/0.88}|\beta_2|L_{Span}N_{Span} \quad (8)$$

in number of symbols. Here, R_{Sym} is in s^{-1} , $B_{Ch} \approx R_{Sym}$ is the channel bandwidth in s^{-1} , β_2 is the dispersion coefficient in s^2/m , L_{Span} is the span length in m, and N_{Span} is the number of spans. In equation (8), the division by 0.88 is due to channel spacing, and the factor $\sqrt{N_{Ch}}$ indicates that w_{XPM}^* increases with the number of channels but only in proportion to its square root. This is consistent with the fact that more symbols from the copropagating channels are involved in nonlinearity as the number of channels increases but more distant channels contribute less to nonlinearity. Note that equations (7) and (8) are valid for all the systems under test (cf. Supplementary Figs. 5-8). Plugging the windowed central moments $\bar{\mu}_2$ and $\bar{\mu}_3$ obtained with w_{SPM}^* and w_{XPM}^* into a state-of-the-art analytic model known as the enhanced Gaussian noise (EGN) model^{20,23}, we obtain SNR_{Eff} as shown in Fig. 3(b). Although the EGN model assumes i.i.d. symbols, given the intractability of designing an analytic model for lightwaves with local energy structures, we use the EGN model with optimized $\bar{\mu}_2$ and $\bar{\mu}_3$ (green solid lines), which provides good agreement with the split-step simulation results (black dashed lines) in a wide range of conditions (cf. Supplementary Figs. 9-10). This shows that the temporal energy structure should and can be taken into account when analyzing the light propagation effect.

Demonstration of Optimal Sphere Shaping through Full C-Band Transmission Experiment. We

experimentally demonstrated the optimal sphere shaping in a full C-band transmission system shown in Fig. 4(a). The dependence of the optimal sphere shaping on dispersion is conveniently verified in a recirculation loop that accumulates varying dispersions over frequency. The loop consists of 7 spans of 40.3 km (on average) fiber, 1 span of which (span #4) is SSMF (with 0.092 ps/nm²/km dispersion slope) and the rest are anomalous-dispersion fiber (with -2.47 ps/nm/km dispersion and -0.1026 ps/nm²/km dispersion slope, at 1550.1 nm). We transmit 37 100-GHz-wide channels in the C-band over 10 loops (2824 km). Fig. 4(b) shows the accumulated dispersion as a function of distance for several selected channels. We denote the channels at 192.1 to 195.7 THz sequentially by Ch#1 to Ch#37. Then, Ch#8 at 192.8 THz (solid red line) undergoes zero net dispersion after every loop, and the farther away from Ch#8 the channel undergoes greater absolute net dispersion, reaching up to $|D_{Total}| = 4.85$ ns/nm at Ch#37.

Three independent streams of symbols are transmitted over the loop, one of which is loaded on the channel under test (CUT), the other two on the even- and odd-indexed interfering channels that are fully decorrelated by delay

fibers of distinct lengths (10, 20, ..., 190 m for odd channels, and 10, 20, ..., 180 m for even channels, cf. Fig. 4(a)).

At the transmitter (TX), all interfering channels are loaded with the same E_{Shaped}^* , n , and R_{Sym} as the CUT. At the receiver (RX), commonly used coherent digital signal processing (DSP) is performed offline to recover the transmitted symbols. Complete details of the experimental setup are given in the Methods section.

We implement up to 20 combinations of n and R_{Sym} by using $n = 5, 10, 20, 80, 320$, and $R_{\text{Sym}} \approx 9.8, 19.7, 39.4, 78.8$ GBd. On each of the 100-GHz-wide optical channels in our experimental system, we emulate the four R_{Sym} above by performing digital subcarrier multiplexing (DSM)^{33,34} with $N_{\text{SC}} = 1, 2, 4, 8$ digital subcarriers on an optical carrier modulated at 78.8 GBd, as shown in Fig. 4(c). RRC pulse shaping is performed on each subcarrier with a roll-off factor flexibly adjusted between 0.05 and 0.1 depending on N_{SC} . With this, we ensure in a back-to-back experiment that SNR_{Eff} averaged over subcarriers differs by no more than 0.2 dB between DSM of all N_{SC} at the same optical SNR (OSNR), thereby circumventing the bandwidth limitations of the transmitter favoring $N_{\text{SC}} = 1$ (see the Methods section for more details).

For each pair of n and R_{Sym} , the achieved NDR is determined by finding the maximum $H^* \in \{5.2, 5.6, \dots, 7.6\}$ varied in increments of 0.4, which yields an NGMI greater than $NGMI^* = 0.86$. The NDR per optical carrier is then given by

$$NDR = 4 \times [H^*/4 - m(1 - R_c)]R_{\text{Sym}}N_{\text{SC}}R_{\text{Pi}} \quad (9)$$

in Gb/s, where the multiplication by 4 in the right-hand side accounts for 4 amplitudes per symbol time, $m = \log_2 M$, and $R_{\text{Pi}} = 47/48$ is the pilot ratio used to assist coherent DSP. Instead of $H = 4.8$ for shaped 16-QAM, we use uniform quadrature phase-shift keying (QPSK), since they produce the same NDR. Combined with up to 20 configurations of n and R_{Sym} , changing H for each channel to maximize the NDR implies that a huge number of measurements are made across the 37 channels in C-band for experimental verification of optimal sphere shaping.

We present the result obtained with the launch power (defined as the total power for the full C-band) of 13 dBm in this article, although the results are obtained with various launch powers from 8 to 13 dBm in 1 dBm steps. The power excursion across the C-band is maintained within 4 dB after 10 loops by adjusting the inline dynamic gain equalizers (DGEs, cf. Fig. 4(a)). Typical recovered constellations are shown in Fig. 4(a) next to the RX. Fig. 5(a) shows the optimal n^* that maximizes NDR in each channel. In general, n^* increases as $|D_{\text{Total}}|$ (green solid line)

increases, which is consistent with the simulation result in Fig. 2(b). Fig. 5(b) shows that R_{Sym}^* decreases from 78.8 to 39.4 GBd when $|D_{Total}|$ increases from approximately 2 to 4 ns/nm, and this trend of change also matches the simulation result in Fig. 2(a). Using the optimal n^* and R_{Sym}^* of Figs. 5(a) and (b), SNR_{Eff} of Fig. 4(c) (orange circles) is obtained. Note that greater SNR_{Eff} than those in Fig. 5(c) are obtained with n smaller than n^* , but they provide smaller NDRs due to the reduced shaping efficiency. Also shown in Fig. 5(c) is the SNR_{Eff} of the shaping with the longest $n = 320$ (blue triangles), chosen as a benchmark for its best performance in linear channels, at a symbol rate of $R_{Sym} = 39.4$ GBd that represents a widely used value in deployed coherent systems. Near the zero-dispersion regime at 192.8 THz, the benchmark shaping does not produce an NGMI greater than $NGMI^*$ with any H , hence the channels are modulated with QPSK (blue pluses in Fig. 5(c)). Fig. 5(d) shows the NDRs achieved by the optimal n^* and R_{Sym}^* (orange circles), and by the benchmark shaping (blue triangles), where the NLI-optimized shaping offers higher NDRs than the linear-channel optimal benchmark widely in low-dispersion channels. The NDR increase by the optimized shaping reaches 25% near the zero-dispersion regime (cf. red arrow in Fig. 5(d)) and 12.1% on average over the underperforming 20 channels with low dispersion. The total NDR increase in the C-band reaches 6.4 % (12.86 Tb/s as compared to 12.09 Tb/s of the benchmark, obtained as the sum of the NDRs of all 37 channels in Fig. 5(d)).

Discussion

As seen from the comparison of the finite-length and i.i.d. shaped lightwaves under various conditions, finite-length sphere shaping reduces Kerr nonlinearity. To explain this phenomenon, we hypothesized that there is a specific time window for quantifying the power deviation of lightwaves to best describe the propagation effect and that, when measured with this time window, the power deviation is reduced by the small spherical energy structures in the shaped lightwaves. We provided three evidences that support this hypothesis. First, the many horizontal line segments of the NLI contour $\Delta\eta$ observed in wide areas of Fig. 2 and Supplementary Figs. 1-4 indicate that Kerr nonlinearity depends on the spherical energy structure present in absolute time rather than in the number of symbols. Second, equations (7) and (8) show that the measurement window size for the instantaneous power agrees remarkably well with the current theoretical understanding of SPM and XPM in dispersive medium. Third, the analytical results obtained with such an optimized time window fairly match the numerical simulation results, as

shown in Fig. 3 and Supplementary Figs. 9 and 10. Nevertheless, we cannot completely rule out the possibility that there may be other, yet undiscovered, reasons that explain the phenomenon. Also, while equations (7) and (8) are valid for all system configurations tested in this article, it remains for future work to see if the optimal measurement window size changes in different system configurations, such as when the total system bandwidth changes.

The finding of this article that there is a specific time window for measuring power deviations that allows the manifestation of Kerr nonlinearity to be most accurately envisioned, has significant practical implications for optical fiber communications. For the sphere shaping demonstrated in this article, identifying such an optimal time window allows us to analytically determine the optimal spectro-temporal sphere shaping configuration for maximum communication performance. It also opens up the possibility of finding new modulation techniques for lightwaves that minimize power deviations using the exact measurement window, which will be different from what are incorrectly found using a single symbol window.

It should be noted that the simulation and experiment presented in this paper focus on the study of the light propagation effects by excluding the impact of imperfect transceivers. In real world applications, however, determining the optimal shaping is a much more complex problem since the implementation cost and technological limitations must be considered. For example, in the presence of the spectral roll-off of high-bandwidth transceiver components, or due to the complexities associated with multi-carrier DSP, single-carrier transmission at the maximum bandwidth allowed by the technology can find an advantage over DSM.

Methods

Sphere shaping. The digital sphere shaping encoder is implemented by enumerative sphere shaping (ESS)^{6,7} for $n = 5, 10, 20, 40, 80$, and by constant composition distribution matching (CCDM)^{8,9} for $n = 320, 1280, 5120$. The sphere shaping encoder produces $4n$ -long amplitude blocks to conform with the amplitude-to-PDM symbol mapping rule used in the article. For the case of $H = 6.4$ bits per dual-polarization symbol, the finite-length shaping blocks used in the split-step simulation and C-band transmission experiment consume 0.583, 0.346, 0.209, 0.137, 0.099, 0.025, 0.009, 0.002 dB more average symbol energy than ideal shaping, respectively, for $n = 5, 10, 20, 40, 80, 320, 1280, 5120$. Here, the ideal shaping refers to an i.i.d. process that creates dual-polarization symbols according to a Maxwell-Boltzmann distribution over the given support at the same entropy rate as H .

Split-step simulation. The fiber loss parameters are $\alpha_{SSMF} = 0.2$ dB/km and $\alpha_{DCF} = 0.45$ dB/km, where the subscripts denote the fiber type. The fiber dispersion parameters are $\beta_{2,SSMF} \approx -2.199 \times 10^{-26}$ s²/m and $\beta_{2,DCF} \approx 1.020 \times 10^{-25}$ s²/m. The fiber nonlinearity parameters are $\gamma_{SSMF} = 1.45 \times 10^{-3}$ W⁻¹/m and $\gamma_{DCF} = 5.02 \times 10^{-3}$ W⁻¹/m. All fiber parameters are quantified at a wavelength of 1550.1 nm. We transmit 2^{18} , 2^{17} , 2^{16} , 2^{15} , 2^{14} , 2^{14} , 2^{14} dual-polarization symbols in *each* channel, respectively, for the WDM systems with $N_{Ch} = 1, 2, 4, 8, 16, 32, 64$. The total number of dual-polarization symbols transmitted over *all* channels is therefore 2^{18} for $N_{Ch} \leq 16$, 2^{19} for $N_{Ch} = 32$, and 2^{20} for $N_{Ch} = 64$. The modulated symbols are oversampled with factors of 2, 4, 8, 16, 32, 64, 128, respectively, for $N_{Ch} = 1, 2, 4, 8, 16, 32, 64$; i.e., the oversampling factor increases linearly with the symbol period T_{sym} . The step size is adaptively determined by the distance at which the nonlinear phase shift in a single step is 0.1 degrees, or 10 km, whichever is smaller. Every EDFA along the optical fiber links has a noise figure of 4.5 dB and produces a constant output power equal to the launch power such that the ASE-induced signal droop is accurately evaluated³⁵. For all system configurations, we calculate SNR_{Eff} by dividing the *total* signal power within the entire 100-GHz-wide frequency band by the *total* ASE plus NLI power within the same band. While the experiment is performed with full C-band transmission for practical significance, the split-step simulation is performed over a total bandwidth of 100 GHz due to the simulation time required for the many configurations under test; seven symbol rates, eight shaping block lengths, six launch powers over four different links correspond to 1344 simulation runs.

Full C-band transmission experiment. The laser for the TX at CUT is a semiconductor external cavity laser (ECL). The ECL's wavelength is fully tunable in C-band, typical output power is 13 dBm, and linewidth is 40 kHz. The same type of laser is used as the local oscillator (LO) of the coherent RX. For the other TXs for 36 interfering channels, distributed feedback (DFB) lasers are used. The DFB lasers' typical output power is 8 dBm and linewidth is 10 MHz. The modulators are LiNbO₃ modulators with 3-dB bandwidth of 35 GHz. The half-wave voltage (V_π) of the modulator is 3 V. The lightwaves from the 36 DFB lasers are multiplexed into even and odd frequency bands using polarization maintaining silica based arrayed waveguide grating multiplexers (AWGs). We modulate the lightwaves with dual-polarization I/Q modulators for all channels. At the TX, $336896/N_{SC}$ sphere-shaped dual-polarization symbols are generated by using a computer for each of N_{SC} subcarriers. To assist coherent DSP of the RX, one QPSK symbol is inserted as a pilot for every 47 sphere-shaped symbols, using the same average power as

the sphere-shaped symbols. The number of modulation symbols is limited by the memory size of the digital-to-analog converters (DACs). The generated N_{SC} symbol streams are filtered by RRC filters with a pass bandwidth of R_{Sym}/N_{SC} GHz and roll-off factors between 0.05 and 0.1, digitally frequency-shifted and multiplexed to build a comb of N_{SC} digital subcarriers in the baseband. Then, the combined N_{SC} symbol streams are filtered by another RRC filter with a pass bandwidth of R_{Sym} GHz and a roll-off factor of 0.1. Digital pre-distortion³⁶ in the frequency domain follows and compensates for the bandwidth limitations of the TX components such that the optical signal launched into fiber has a flat power spectral density (PSD) in the pass band. Four 17-nm CMOS DACs with 8-bit resolution transform the generated digital symbol streams to electrical fields at a sampling rate of 120 GSa/s. Each subcarrier is modulated at a symbol rate of $R_{Sym} = 78.75/N_{SC} = 9.84375, 19.6875, 39.375, 78.75$ GBd, respectively, for $N_{SC} = 8, 4, 2, 1$. The maximum symbol rate is determined to $R_{Sym} = 78.75$ GBd such that DSM (with $N_{SC} \geq 2$) can perform similar to $N_{SC} = 1$ in a back-to-back configuration, since at higher symbol rates the DSM performs worse than $N_{SC} = 1$ due to the bandwidth limitations of the transmitter components. The oversampling factors correspond to $120/R_{Sym} \approx 12.19, 6.10, 3.05, 1.52$, respectively, for $N_{SC} = 8, 4, 2, 1$. At the TX, interfering channels are split by a liquid crystal on silicon (LCoS)-based wavelength selective switch (WSS), propagate through different delay fibers for decorrelation with each other, and recombined by another WSS of the same type.

In the recirculating loop, EDFAs with a noise figure around 4.7 dB are used. A variable optical attenuator (VOA) is attached after each EDFA, such that the optical power at the output of the VOA equals the launch power. The PSD of the optical signal at the EDFA output is flattened by an LCoS-based or micro-electro-mechanical systems (MEMS)-based programmable optical filter that acts as a DGE.

At the RX, an LCoS-based WSS filters out interfering channels, leaving only the CUT. Four balanced photo diodes (BPDs) with 75 GHz bandwidth detect the optical signals on the CUT. The electrical signals produced by the BPDs, containing N_{SC} symbol streams inside, are sampled at 256 GSa/s each by four analog-to-digital converters (ADCs) in a real-time oscilloscope. The digitized waveforms are resampled at 157.5 GSa/s, which will eventually lead to an oversampling factor of 2 for any N_{SC} . Chromatic dispersion is digitally compensated for in the combined N_{SC} symbol streams. Each of the N_{SC} symbol streams is then frequency-shifted to baseband, match-filtered by an RRC filter with a pass bandwidth of R_{Sym}/N_{SC} GHz, and down-sampled to 2 samples per symbol. For each separated

symbol stream, after recovery of the symbol timing³⁷, real-valued PDM equalization³⁸ is performed in the frequency domain with carrier phase recovery based solely on pilot symbols. We calculate SNR_{Eff} in Fig. 5(c) by dividing the *total* signal power for N_{SC} symbol streams by the *total* ASE plus NLI power for N_{SC} symbol streams. Likewise, the NDR in Fig. 5(c) is calculated on an optical carrier basis.

Author Contributions

J. C. proposed the idea. J. C. and S. O. performed the numerical simulations. J. C., X. C., G. R., D. C., and E. B. performed the experiments. J. C., X. C., G. R., D. C., E. B., S. O., and R. T. discussed the idea and analyzed the results. R. T. supervised the project. J. C. wrote the draft. X. C., G. R., D. C., S. O., and R. T. substantially improved the manuscript.

Competing Interests

The authors declare no competing financial interests.

References

1. Buchali, F. et al. Rate adaptation and reach increase by probabilistically shaped 64-QAM: an experimental demonstration. *J. Lightwave Technol.* **34**, 1599–1609 (2016).
2. Maher, R., Alvarado, A., Lavery, D. & Bayvel, P. Increasing the information rates of optical communications via coded modulation: a study of transceiver performance. *Sci. Rep.* **6**, 21278 (2016).
3. Cho, J. & Winzer, P. J. Probabilistic constellation shaping for optical fiber communications. *J. Lightwave Technol.* **37**, 1590–1607 (2019).
4. Forney, G. D. & Wei, L.-F. Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations. *IEEE J. Sel. Areas Commun.* **7**, 877–892 (1989).
5. Laroia, R., Farvardin, N. & Tretter, S. A. On optimal shaping of multidimensional constellations. *IEEE Trans. Inform. Theory* **40**, 1044–1056 (1994).
6. Willems, F. M. J. & Wuijts, J. J. A pragmatic approach to shaped coded modulation. In *Symp. Commun. Veh. Technol.* 1–6 (1993).

7. Amari, A. et al. Introducing enumerative sphere shaping for optical communication systems with short blocklengths. *J. Lightwave Technol.* **37**, 5926–5936 (2019).
8. Ramabadran, T. V. A coding scheme for m-out-of-n codes. *IEEE Trans. Commun.* **38**, 1156–1163 (1990).
9. Schulte P. & Böcherer B. Constant composition distribution matching. *IEEE Trans. Inform. Theory* **62**, 430–434 (2016).
10. Dar, R., Feder, M., Mecozzi, A. & Shtaif, M. On shaping gain in the nonlinear fiber-optic channel. In *IEEE Int. Symp. Inform. Theory*, 2794–2798 (2014).
11. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
12. Kerr, J., XL. A new relation between electricity and light: dielectrified media birefringent. *The London, Edinburgh, and Dublin Philosoph. Mag. and J. Science* **50**, 337–348 (1875).
13. Stolen, R. H. & Ashkin., A. Optical Kerr effect in glass waveguide. *Appl. Phys. Lett.* **22**, 294–296 (1973).
14. Mitra, P. P. & Stark, J. B. Nonlinear limits to the information capacity of optical fibre communications. *Nature* **411**, 1027–1030 (2001).
15. Fehenberger, T. et al. Analysis of nonlinear fiber interactions for finite-length constant-composition sequences. *J. Lightwave Technol.* **38**, 457–465 (2020).
16. Du., L. B. & Lowery, A. J. Optimizing the subcarrier granularity of coherent optical communications systems. *Opt. Express* **19**, 8079–8084 (2011).
17. Poggiolini, P. et al. Analytical and experimental results on system maximum reach increase through symbol rate optimization. *J. Lightwave Technol.* **34**, 1872–1885 (2016).
18. Carena, A., Curri, V., Bosco, G., Poggiolini, P. & Forghieri, F. Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links. *J. Lightwave Technol.* **30**, 1524–1539 (2012).
19. Poggiolini, P. The GN model of non-linear propagation in uncompensated coherent optical systems. *J. Lightwave Technol.* **30**, 3857–3879 (2012).
20. Dar, R., Feder, M., Mecozzi, A. & Shtaif M. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Opt. Express* **21**, 25685–25699 (2013).
21. Agrell, E., Alvarado, A., Durisi, G. & Karlsson, M. Capacity of a nonlinear optical channel with finite memory. *J. Lightwave Technol.* **32**, 2862–2876 (2014).

22. Wu, K., Liga, G., Sheikh, A., Willems, F. M. J. & Alvarado, A. Temporal energy analysis of symbol sequences for fiber nonlinear interference modelling via energy dispersion index. Preprint at <https://arxiv.org/abs/2102.12411> (2021).
23. Carena, A. et al. EGN model of non-linear fiber propagation. *Opt. Express* **22**, 16335–16362 (2014).
24. Agrawal, G. P. *Nonlinear fiber optics* (Academic, New York, NY, USA, 2007).
25. Sinkin, O. V., Holzlohner, R., Zweck, J. & Menyuk, C. R. Optimization of the split-step Fourier method in modeling optical-fiber communications systems. *J. Lightwave Technol.* **21**, 61–68 (2003).
26. Schmalen, L. et al. Spatially coupled soft-decision error correction for future lightwave systems. *J. Lightwave Technol.* **33**, 1109–1116 (2015).
27. Alvarado, A., Agrell, E., Lavery, D., Maher, R. & Bayvel P. Replacing the soft-decision FEC limit paradigm in the design of optical communication systems. *J. Lightwave Technol.* **33**, 4338–4352 (2015).
28. Cho, J., Schmalen, L. & Winzer, P. Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM. In *European Conf. Opt. Commun.*, Paper M.2.D.2 (2017).
29. Shimizu, F. Frequency broadening in liquids by a short light pulse. *Phys. Rev. Lett.* **19**, 1097 (1967).
30. Stolen, R. H. & Lin, C. Self-phase-modulation in silica optical fibers. *Phys. Rev. A* **17**, 1448 (1978).
31. Islam, M. N. et al. Cross-phase modulation in optical fibers. *Opt. Lett.* **12**, 625–627 (1987).
32. Matsuda, N. et al. Observation of optical-fibre Kerr nonlinearity at the single-photon level. *Nature Photon.* **3**, 95–98 (2009).
33. Hui, R. et al. Subcarrier multiplexing for high-speed optical transmission. *J. Lightwave Technol.* **20**, 417–427 (2002).
34. Qiu, M. et al. Digital subcarrier multiplexing for fiber nonlinearity mitigation in coherent optical communication systems. *Opt. Express* **22**, 18770–18777 (2014).
35. Antona, J.-C., Meseguer, A. C. & Letellier, V. Transmission systems with constant output power amplifiers at low SNR values: a generalized droop model. In *Opt. Fiber Commun. Conf.*, Paper M1J-6 (2019).
36. Rafique, D., Napoli, A., Calabro, S. & Spinnler, B. Digital preemphasis in optical communication systems: on the DAC requirements for terabit transmission applications. *J. Lightwave Technol.* **32**, 3247–3256 (2014).

37. Lee, S. J. A new non-data-aided feedforward symbol timing estimator using two samples per symbol. *IEEE Commun. Lett.* **6**, 205–207 (2002).
38. Randel, S. et al. 6× 56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization. *Opt. Express* **19**, 16697–16707 (2011).

Tables

Table 1 Link configurations in which the impact of sphere shaping is evaluated.

Link Name	L_{SSMF}	L_{DCF}	Net dispersion D_{span} per span
Link A	49.36 km	10.64 km	0 ps/nm
Link B	49.98 km	10.02 km	60 ps/nm
Link C	51.83 km	8.17 km	240 ps/nm
Link D	60 km	0 km	1033.8 ps/nm

Figures

Figure 1 Comparison of the power deviation between unshaped and shaped lightwaves.

a, Probability distribution of the total energy in a block of the unshaped (left figure) and shaped (right figure) symbols. The maximum energy is significantly lower in the shaped symbols than in the unshaped symbols. **b**, Division of the continuum of lightwaves in time and frequency. The probability distributions shown in **(a)** appear in each of the rectangular blocks that divide the lightwaves. **c**, Probability distribution of the normalized power $\langle p \rangle_w$ of the unshaped (left figures) and shaped (right figures) symbols, measured with a sliding window of length $w = 1$ (upper figures) and 40 (lower figures). **d**, The windowed second central moment $\bar{\mu}_2$ as a function of the measurement window size w . As w increases, the moment decreases only for the shaped symbols.

Figure 2 Split-step simulation with various spectro-temporal division of lightwaves. a, Effective SNR drawn as a function of the symbol rate (spectral division) and shaping block duration (temporal division). **b,** NGMI showing the end-to-end performance, produced by the complex interaction of the

propagation effect and the fundamental energy efficiency of the underlying sphere shaping. $\mathbf{c}$, NLI coefficient η and its decompositions into SPM contribution η_{SPM} and XPM contribution η_{XPM} , and into factors due to i.i.d. shaping η^∞ and finite-length shaping $\Delta\eta$.

Figure 3 The optimal window size for instantaneous power measurement and SNR_{Eff} predicted by the EGN model using it. **a**, Pearson correlation coefficients between η_{SPM} and $\bar{\mu}_2$ (left figure), and between η_{XPM} and $\bar{\mu}_2$ (right figure) with various window sizes w . The red lines are obtained using equations (7) and (8), estimated to produce the greatest correlation. There is no XPM at $R_{Sym} = 88$ GBd, since only one channel is transmitted. **b**, SNR_{Eff} predicted by the EGN model, obtained with the optimal w_{SPM}^* and w_{XPM}^* (green solid lines) and traditional $w = 1$ (red dotted lines), in comparison with the split-step simulation results (black dashed lines) at 60 and 240 spans in Link D.

Figure 4 Experimental system to verify the performance of optimal sphere shaping in full C-band transmission. a, Schematic of the experimental setup. Inset: typical recovered constellations at the receiver. **b**, Dispersion map of various WDM channels over the optical fiber link. **c**, Typical received power spectral densities when a single (upper figure) and eight (lower figure) digital subcarriers are transmitted.

Figure 5 Experimental results obtained with the nonlinear fiber-channel-optimized sphere shaping.

a, Optimal shaping length n^* that tends to increase with the total net dispersion $|D_{Total}|$. **b**, Optimal symbol rate R_{Sym}^* that decreases from 78.8 to 39.4 GBd when $|D_{Total}|$ increases from approximately 2 to 4 ns/nm. **c**, The effective SNR achieved by the fiber-channel-optimized n^* and R_{Sym}^* (orange circles) and by the linear-channel-optimized $n = 320$ at $R_{Sym} = 39.4$ GBd (blue triangles and pluses). **d**, The NDRs achieved by the fiber-channel- optimized and linear-channel- optimized sphere shaping configurations. Near the zero-dispersion regime, an NDR increase of up to 25% is achieved by the fiber-channel-optimized shaping.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NC2021Supplementaryv19.pdf](#)