Overcoming degradation in spatial multiplexing systems with stochastic nonlinear impairments

Filipe M. Ferreira, Christian S. Costa, Stylianos Sygletos & Andrew D. Ellis

Single-mode optical fibres now underpin telecommunication systems and have allowed continuous increases in traffic volume and bandwidth demand whilst simultaneously reducing cost- and energy-per-bit over the last 40 years. However, it is now recognised that such systems are rapidly approaching the limits imposed by the nonlinear Kerr effect. To address this, recent research has been carried out into mitigating Kerr nonlinearities to increase the nonlinear threshold and into spatial multiplexing to offer additional spatial pathways. However, given the complexity associated with nonlinear transmission in spatial multiplexed systems subject to random inter-spatial-path nonlinearities it is widely believed that these technologies are mutually exclusive. By investigating the linear and nonlinear crosstalk in few-mode fibres based optical communications, we numerically demonstrate, for the first time, that even in the presence of significant random mixing of signals, substantial performance benefits are possible. To achieve this, the impact of linear mixing on the Kerr nonlinearities should be taken into account using different compensation strategies for different linear mixing regimes. For the optical communication systems studied, we demonstrate that the performance may be more than doubled with the appropriate selection of compensation method for fibre characteristics which match those presented in the literature.

The Kerr nonlinear limit has imposed an ever-growing capacity gap between the technologies generating/processing data and the technologies transporting it – namely, optical fibre communication systems. The first has consistently grown at 40% compound annual growth rate (CAGR), but the latter has slowed to 20% CAGR since late 1990s. Such large scaling disparity is expected to lead to a full exhaustion of system capacity within the next 5 to 15 years. By 2024, optical networks are projected to require 1 Pb/s transmission capacity which with current technological limits can be expected to be met using 10 parallel line systems each carrying 100 Tb/s per fibre. A trend that would potentially increase the cost- and energy-per-bit by 10 times except for efficiency gains as ancillary functions overhead is reduced via sub-system integration. However, with the number of required spatial paths projected to double every 2-years and the current communications infrastructure accounting for 1–2% of global energy the current paradigm is exhausted. Thus, research effort must be directed towards the development of transformative means for achieving spatial parallelism that can ensure sublinear scaling of the total system cost and energy consumption. Otherwise, the dooming capacity exhaustion will lead to a dramatic increase of the bandwidth price and ultimately bring the information revolution to a halt.

Mode-division multiplexing (MDM) over few-mode fibres (FMFs) holds one the greatest potential to deliver future cost- and energy-effective high-capacity systems with spatial parallelism. Figure 1 shows the basic system concept of a multi-span MDM-FMF system, composed by integrated arrays of transmitter and receiver units, mode multiplexers/de-multiplexers (e.g. photonic lanterns), and multimode amplifiers. The information is carried over a set of orthogonal spatial modes overlapping on a single fibre core. Compared to alternative technologies, such as uncoupled multi-core fibres or single-mode fibre (SMF) bundles, MDM-FMF systems offer a number of advantages, such as lower nonlinear coefficients; higher pump efficiency for their optical amplifiers (similar to core pumped SMF); and higher spatial-density level of optical integration for transponders, amplifiers, and add-drop multiplexers (multiple spatial modes can be routed together). Nevertheless, coupled-core multi-core fibres (CC-MCFs) offer similar potential to that of FMFs when designed to have similar spatial mode densities. Finally, the techniques presented in this paper apply to all SDM fibre types, including CC-MCFs.

Aston Institute of Photonic Technologies, Aston University, Birmingham, B47ET, United Kingdom. Correspondence and requests for materials should be addressed to F.M.F. (email: f.ferreira@aston.ac.uk)

Received: 30 April 2018
Accepted: 23 October 2018
Published online: 03 December 2018
Despite the potential of MDM-FMF technology, commercialisation and deployment is not imminent yet because whilst the fibres are now filling up, there are spare fibres in the network, giving a few years grace before an expensive installation of new fibre becomes necessary. By the end of that period, considering the 45% CAGR of CMOS processing capabilities, it is likely that the DSP ASICs will be able to handle 10 Tb/s worth of traffic by directly coupling to arrays of 10 (or more) parallel integrated high-speed (~Tb/s) optical and opto-electronic components (such as modulators or coherent receivers). Processing multiple tributaries simultaneously would allow to compensate the combined effects of mode-mixing and walk-off between the different parallel fibre paths shown in Fig. 2. Presently, differential mode delay (DMD) and linear mode-crosstalk (XT) have been successfully mitigated with offline multi-input multi-output (MIMO) based DSP techniques and DMD compensation maps after transmission over thousands of kilometres. However, whilst SDM systems have seen a significant growth of their bit-rate distance product (e.g. 115 Petabit/s∙km using a 6-mode 19-core fibre and 166 Petabit/s∙km using a 3-mode single-core fibre), they have yet to match the records established using SMFs (e.g. 535 Petabit/s∙km in). This is due to the increased mode dependent loss of the prototype components used in a research lab environment, see Fig. 1, and due to the intermodal nonlinear interactions that occur in the transmission medium. However, given the engineering development of prototype components (for example, monolithic mode-selective few-mode multicore fibre multiplexers with insertion loss < 2 dB and mode dependent loss < 0.5 dB have been recently demonstrated over the C + L band with identified margin for improvement), the impact of the former will become smaller and the latter will therefore become dominant. The intermodal nonlinear distortion is most relevant when the system is loaded with a sufficiently high number of channels such that the total system bandwidth allows exact cancelling of chromatic dispersion and DMD walk-offs for all pairs of modes. Thus, in order to realise the full potential of FMF system by simultaneously increasing the information spatial density and maintaining the system reach it is of paramount importance to address the nonlinear penalties.

Digital-back propagation (DBP) is a nonlinear mitigation method originally proposed for SMFs that compensates for the deterministic linear and nonlinear fibre impairments by numerically back-propagating the received optical field with inverted channel parameters, using the split-step Fourier method (SSFM) to solve the nonlinear propagation equation, as shown in Fig. 3. However, effectiveness of this technique is reduced in the presence of random processes such as the group delay spread induced by DMD and XT, like polarisation mode dispersion (PMD) in SMFs. A brute force approach, following recent attempts to mitigate PMD, would be to estimate in the digital domain the slowly varying differential delays and the rapidly varying random mixing on relatively short length scales. However, this would increase the complexity, due to the short step size required, and would accumulate an increasing numerical error due to the large number of estimates. In FMFs with large DMD
between non-degenerate modes this problem can become much more pronounced depending on the strength of the random linear mode coupling since in its absence the optical signal evolution is fully predictable and DBP could compensate for the nonlinear distortion perfectly (ignoring transceiver noise). In this way, one can expect the amount of DMD tolerable by DBP systems to be much higher for FMFs operating in the weak linear coupling regime than for FMFs operating in the strong linear coupling regime (depending on the DBP implementation).

In this paper, we introduce new simplified DBP methods for the different operational regimes of MDM-FMF systems as determined by the strength of DMD and XT. To illustrate their benefit, we consider the transmission of wavelength-division multiplexed (WDM) multi-level quadrature amplitude modulated (QAM) signals. Some of the proposed methods have been considered in a preliminary study under simplified conditions that include the absence of amplified spontaneous noise, and constant amplitude signals.

In this paper, we use an extension of the single-mode split-step Fourier method (SSFM) to numerically solve the generalized nonlinear Schrödinger equation (GNLSE) that governs multimode propagation over FMFs (including all linear and nonlinear fibre effects), provided that the split-step size is compatible with the additional requirements, see Methods. To achieve higher accuracy in the modelling of forward propagation over a real fibre, we consider a symmetric implementation of the SSFM in which the effect of nonlinearity is included in the middle of the segment rather than at the segment boundary, see Fig. 4. Finally, for backward propagation over the virtual fibre (with inverted parameters), GNLSE can be simplified depending on the strength of the linear mode coupling, see Methods, reducing the computational requirements.

In the presence of extreme linear mode coupling regimes, it has been shown that some or all the linear mode coupling terms in GNLSE can be assumed to vary rapidly and seemingly randomly on a length scale that is expected to be short compared to the effective lengths associated with chromatic dispersion and the various manifestations of nonlinearity. Thus, like in single-mode fibres and the well-known Manakov-PMD equations, one can average the propagation equation itself over all spatial modes. New Manakov-like equations were derived for FMFs with nonlinear coefficients averaged for the two extreme coupling regimes. In the weak coupling (WC) regime, only the averaging over birefringence fluctuations must be considered, reducing the intramodal degeneracy factor to 8/9 and the intermodal degeneracy factor to 4/3, see Methods. In the strong coupling (SC) regime the averaging includes all propagation modes, see Methods.

In the intermediate linear coupling regime, the linear mode coupling introduced by GNLSE needs to be explicitly considered in the SSFM. It has been shown that the semi-analytical solution of this term is possible by discretising the fibre imperfections responsible for linear coupling as fibre sections with a random displacement of the core centre position. In this way, the authors proposed a multi-section model where the coupling strength is set using a given radial displacement and a uniformly distributed azimuthal displacement for each section. More importantly, the proposed model proved to be accurate against analytical predictions for the statistics of group-delays in FMF links. Finally, the linear coupling step in the SSFM used in this paper is implemented following the virtual back-propagation concept.

To compensate for the GD spread imposed by the interplay of XT and DMD, see Fig. 2, multi-input multi-output (MIMO) digital signal processing (DSP) is used with training-symbol-based frequency-domain equalisers offering the lowest complexity. However, as the effective transmission rate is reduced by the additional training symbols overhead, low DMD links must be used. It has been shown that a DMD lower than 12 ps/km is required for 2000 km MDM transmission at 100 Gb/s in order to restrict the training sequences overhead to 10%. In this way, significant effort has been directed to the optimisation of FMFs, with refractive-index profiles typically composed of a graded-index core (for DMD reduction) and a cladding trench (for macro-bend loss reduction via field confinement). Ferreira et al. have optimised such profile to guide six linearly polarized
transmission 44) and macro-bend losses as low as in SMFs 46. Here, we use the same optimum profile. Table 1
impact, the GD compensation length must be smaller than the correlation length set by the coupling 49, which
zero in the absence of linear mode coupling, otherwise there is a residual GD spread. To minimise mode coupling
DMD might not be practical when correlation length is on the order of a km. In this paper, when considering
>
10 ps/km, we will consider one compensation segment per span.

Orthogonal polarisations. The simulation setup is shown in Fig. 5, where 12.8 Tbit/s are transmitted over 19
factors of 0.001 were used for pulse shaping. For the simulations we considered 216 symbols per polarisation mode,
(CAZAC) sequences for time synchronization and channel estimation. Root raised cosine filters with a roll-off
with the information data, a preamble was transmitted consisting of constant amplitude zero autocorrelation
array of 6 erbium doped fibre amplifiers19, with noise figures of 3 dB and negligible mode dependent loss since the
Fibre Linear Characteristics at 1550 nm.

Table 1. Fibre Linear Characteristics at 1550 nm.

| Mode#1 | xN-spans |
|--------|----------|
| Mode#6 |

Figure 5. Simulation setup for 6 LP modes each with 2 orthogonal polarisations, 19 WDM channels, including:
optical transmitter, mode multiplexer (MUX) and demultiplexer (DEMUX), FMF, coherent receiver, and DSP blocks.

(LP) modes (LP01, LP02, LP11a, LP11b, LP21a and LP21b) with low DMD (12 ps/km for GD unmanaged long-haul
transmission 44) and macro-bend losses as low as in SMFs 46. Here, we use the same optimum profile. Table 1 shows the optimum profile linear characteristics at 1550 nm, the DMD defined as max(GD)-min(GD) is 5.19 ps/km. Table 2 shows the uncoupled nonlinear coefficients whilst the uncoupled degeneracy factors are found in the GNLSE, see Methods. In this paper, when considering different DMD values, we simply scale the GD vector in Table 1 instead of reoptimizing the fibre profile (as in 14) to avoid fluctuations of the other fibre characteristics, such that a direct performance assessment of the proposed DBP methods can be accomplished.

Due to current estimations of the fibre manufacturing limitations 14 it is challenging to achieve DMD < 12 ps/km for more than 3 LP modes, therefore GD managed spans are often used to minimise the total GD spread by cascading fibres with opposite sign DMD 19,47,48. Here we define a GD managed span of length L/2 with the same characteristics but opposite sign GD (in practice for fibres with more than 2 non-degenerate LP modes, more than two different fibres are needed 19,47,48 since exact opposite sign GD is hard to achieve). The GD spread at the end of a managed span is only zero in the absence of linear mode coupling, otherwise there is a residual GD spread. To minimise mode coupling impact, the GD compensation length must be smaller than the correlation length set by the coupling 49, which might not be practical when correlation length is on the order of a km. In this paper, when considering DMD > 10 ps/km, we will consider one compensation segment per span.

Along this study we consider a WDM-MDM system with 6 linearly polarized (LP) modes each with two orthogonal polarisations. The simulation setup is shown in Fig. 5, where 12.8 Tbit/s are transmitted over 19 WDM channels (in each spatial mode) modulated with 14 Gbaud polarmisation-multiplexed 16QAM. Together with the information data, a preamble was transmitted consisting of constant amplitude zero autocorrelation (CAZAC) sequences for time synchronisation and channel estimation. Root raised cosine filters with a roll-off factor of 0.001 were used for pulse shaping. For the simulations we considered 216 symbols per polarisation mode, with a 216 symbols CAZAC preamble. The in-phase and quadrature components of each signal drove the optical field of an ideal laser through an optical IQ modulator, and the optical signals were fed into the FMF link (non-GD-managed and GD-managed links are considered). Fibre attenuation was fully compensated using an array of 6 erbium doped fibre amplifiers 44, with noise figures of 3 dB and negligible mode dependent loss since the aim of this paper is to assess the isolated impact of the FMF mode coupling and mode delay on DBP performance. The optical signals were coupled in and out of the FMF using a mode multiplexer (MUX) and a mode demultiplexer (DEMUX), respectively, assumed ideal since their transfer function can be fully compensated through DSP given prior characterization (as in conventional transceivers 49). After homodyne detection, the baseband
electrical signals were sampled at 56 GS/s, yielding 12 digital signals at 2 samples/symbol. DBP was then implemented by launching the coherently received signals into a virtual fibre with characteristics of opposite-sign values of those in the transmission channel, except that no mode coupling was considered. Back-propagation was implemented using the modified SSFM proposed with a fixed step size and considering the nonlinear degeneracy factors derived for WC and SC, Eqs (3) and (4), respectively. As a reference, for linear compensation, the coherently received signals were compensated for chromatic dispersion in the frequency domain using the parameter values of Table 1. In all cases, mode coupling and (residual) DMD were subsequently compensated for using training-symbol-based channel estimation and equalization, as shown in Fig. 5. Coarse time synchronization was performed using the Schmidl & Cox autocorrelation metric. Subsequently, fine-time synchronization and channel impulse response (CIR) estimation were performed by cross-correlating with the training CAZAC sequences. The $12 \times 12$ CIR estimations were converted into the frequency domain. The MIMO frequency domain equalizer was calculated by inverting the channel matrix, and, finally, the Q-factor for each received signal was estimated using the mean and standard deviation of the received symbols$^{43}$.

Results

DBP performance was studied on an optical super-channel consisting of 19 channels (in each of the 6 LP modes), 14 Gbaud polarisation-multiplexed 16QAM (with a frequency spacing of 14.1 GHz), corresponding to a total line-rate of 12.8 Tb/s (a net-data-rate of 9.8 Tb/s given 20% FEC and 3.1% training sequences overheads), over 12 spans of 20 km (the required energy per bit is minimised for span lengths going from 35 km at 6 bits/s/Hz to 20 km at 12 bits/s/Hz$^5$). With a total WDM bandwidth of 268 GHz, the centre WDM channels will experience intermodal error$^{38}$, a method found to be more computationally efficient at high accuracy than other common methods such as nonlinear phase rotation. The local error was bound to be lower than $10^{-5}$ as smaller values led to negligible performance change. Conversely, the maximum step size is kept much smaller than the dispersion length, the walk-off length, and the correlation length, as explained in Methods. For backward propagation, a constant step size is used, its value is determined in the following. In all cases, backpropagation considered the total number of channels being transmitted since multi-channel DBP was found to be required to achieve effective nonlinearity mitigation for SMF systems$^{26}$ and this paper aims at exploring the full potential of DBP in SDM systems.

Figure 6 shows the Q-factor gain over linear equalisation for WC-, SC-, and Intra-DBP as a function of the DBP step size after 240 km transmission with a launch power of 0 dBm/ch, for two DMD free fibre links, (a) one with very low XT and (b) one with very high XT, and for (c) one high DMD fibre link GD-managed (one segment). These cases are representative of the broader range of DMD and XT values considered in the following.

In all cases, results suggest that the simulation had converged by 100 m (similar results to 10 m). Thus, from this point, the step was kept at 100 m as this paper aims at exploring the full potential of digital nonlinear mitigation.

Figure 7 shows the Q-factor as a function of the power per channel ($P_{\text{in}}$) after 240 km for two DMD free fibre links, (a) one with very low XT and (b) one with very high XT, and for (c) one high DMD fibre link GD-managed (one segment). From Fig. 7(a,b), it can be seen that WC- and SC-DBP provide substantial Q-factor improvement in particular for low XT and high XT respectively, as Manakov approximations are applicable. And, that Intra-DBP provides performance improvement only for low XT but of smaller order as intermodal nonlinear interactions (strong for low DMD) are not accounted while DBP. Moreover, one can observe that WC-DBP only provides gain for transmission over the weakly coupled fibre, while SC-DBP provides gain for both fibres. WC-DBP is particularly penalizing for high XT values as the nonlinear coefficients in Eq. (2) are larger than the actual channel coefficients leading to large overcompensation. SC-DBP provides gain even for low XT as the nonlinear coefficients in Eq. (3) are smaller than the actual channel coefficients leading to undercompensation. From Fig. 7(c), it can be concluded that when using GD management with high DMD and low XT (GD management with high XT is not effective$^{43}$), WC-DBP allows to achieve maximum Q-factors of the same order of those for low DMD non-GD-managed spans (Fig. 7(a)). Also in Fig. 7(c), as intermodal FWM efficiency is reduced for high DMD (and a given system bandwidth, as explained previously), Intra-DBP produces similar gains to those of WC-DBP. Finally, in Fig. 7(a,c), the large errors bars for WC-DBP are a consequence of having extremely low strength mode mixing ($\sim$60 dB/100 m) providing a few additional pathways to intermodal four-wave mixing phase matching without introducing sufficiently fast random rotations of the field polarisation state along the fibre length which would have led to an averaging of the efficiency of the overall nonlinear process. Note that in practice the lower bound of the Q-factor variation would be considered in the dimensioning of the system. Note that in practice the lower bound of the Q-factor variation would be considered in the dimensioning of the system.
a range of low $XT$ in which sufficient coupling events randomise a sufficient share of the intermodal nonlinear distortion. Thus, for sufficiently low $XT$, Intra-DBP gain rolls-off as can be seen in Fig. 8(a). In this way, for the WC-regime, and for fibres with $-55 < XT [\text{dB/100 m}] < -35$ and $DMD > 30 \text{ ps/km}$ Intra-DBP provides the highest improvement between 1 and 3 dB, and for fibres with $XT < -40 \text{ dB/100 m}$ and $DMD < 30 \text{ ps/km}$ WC-DBP provides an improvement between 1 and 4 dB. These $XT$ and $DMD$ ranges cover many of the fibres presented in literature\(^{53-58}\). For the SC-regime, SC-DBP provides significant compensation for small $DMD$ values ($\leq 4 \text{ ps/km}$) and $XT \geq -10 \text{ dB/100 m}$, a regime that can be achieved using for example a fibre similar to the one in\(^{58}\). For intermediate $XT$ values ($-30 < XT [\text{dB/100 m}] < -15$) none of the DBP approaches work even for negligible $DMD$ values. This is because for significant transmission distances (240 km, in this case) linear mode coupling leads to evolutions of the nonlinear operator that differ significantly from that of the uncoupled operators in the Manakov approximation. Outside the operational regime identified for WC- and SC-DBP, the evolution of the GD operator.
is no longer well approximated using the uncoupled GD coefficients as in Table 1, thus the nonlinear distortion is either overcompensated or undercompensated when using the coefficients in Table 2 or their direct average41,42.

Discussion

Even for the most complex MDM systems significant performance improvement is possible using DBP provided that appropriate approximations for the effect of the stochastic nature of the linear crosstalk are taken into account. For example, fibres optimised primarily for low XT (and with intermediate-to-high DMD), including trench-assisted graded-index fibres19,53 or multiple-step index fibres 54,60, allow a significant DBP gain if the crosstalk is neglected. However, for such high XT fibres, if the instantaneous crosstalk is averaged, the so-called generalised Manakov approach, high performance gains are again possible. Whilst a small range of possible fibre parameters exist where the approximate models considered here failed to provide significant gain, compensation would require continuous estimation of the random linear coupling, significant performance gains were possible for all possible XT and DMD regimes in which real fibres operate.

To extend the operational range of Manakov-DBP active tracking of the GD operator is needed, as observed for SMF systems impacted by polarisation mode dispersion (PMD) 31,34. The DBP algorithm proposed in 31, takes into account PMD by simply considering GD accumulating in staircase fashion, span-by-span, in such a way that the total PMD accumulated in the forward propagation is reversed (total PMD is available at conventional linear channel equalizers). In34, instead of considering GD accumulating over the same principal states of polarisation in the backward propagation, a blind optimization of each section Jones matrix improved performance by reducing the DBP gain variability at the expense of additional complexity. Future research to develop methods similar to34 applicable to MDM systems are expected to deliver significant additional gains (>1 dB) over a broad range of XT and DMD regimes.

Methods

Generalized Nonlinear Schrödinger Equation. The generalized nonlinear Schrödinger equation (GNLSE) for FMFs can be written as37:

$$\begin{align*}
\frac{\partial}{\partial z} A_{ni} &+ \frac{j\omega_{ni}}{2} A_{ni}^2 + \frac{j\beta_{ni}^{(2)}}{2} A_{ni}^2 + \cdots + \frac{\alpha_{ni}}{2} A_{ni}^2 \\
&= -\frac{j}{\gamma_{nii}} |A_{ni}|^2 + 2\gamma_{niv} \sum_{v=1}^{n} |A_{vi}|^2 + 2\gamma_{niv} \sum_{v=1}^{n} |A_{vi}|^2 A_{vi} \\
&- j \sum_{v=1}^{n} C_{niv} A_{vi} e^{j(\beta^{(1)}_{niv} - \beta^{(0)}_{niv}) z},
\end{align*}$$

(1)
where $i$ and $j$ are the orthogonal states of polarisation of each mode $u$ and $v$, with $u, v = (1, \ldots, N)$ for $N$ linearly polarised modes each two orthogonal polarisations. In this way, (1) is a set of $2N$ coupled equations, one for each polarisation mode $u, i, A_u(z,t), \beta_u^{(1)}, \beta_u^{(2)}$ and $\alpha_u$ are the slowly varying field envelope, group delay, group delay dispersion and attenuation, respectively. $\gamma_{uvij}$ is the nonlinear coefficient between $u$ and $v$, which depends on the nonlinear refractive index $n_2$ of the silica, approximately $2.6 \times 10^{-19} \text{m}^2/\text{W}$, and on the intermodal effective area, and is given by:

$$
\gamma_{uvij} = n_2 \frac{\omega_0}{c} \frac{\left( \int_{-\infty}^{+\infty} |E_u|^2 \frac{\partial}{\partial x} E_v \, dx \right)}{\left( \int_{-\infty}^{+\infty} |E_u|^2 \, dx \right) \left( \int_{-\infty}^{+\infty} |E_v|^2 \, dx \right)}
$$

where $E_u(x, y)$ is the mode field transverse distribution for the $i$th polarisation of mode $u$.

In Eq. (1), $\hat{D}$ is the differential operator that accounts for dispersion and attenuation, and $\hat{N}$ is the nonlinear operator that accounts for all the intramodal and intermodal nonlinear effects. The last term on the right-hand side accounts for the linear mode coupling arising from fibre structure imperfections, where $C_{uvij}$ are the coupling coefficients as derived in Ref. [17].

For the extreme coupling regimes, the nonlinear coefficients and degeneracy factors in (1) can be assumed being averaged by the linear coupling theory, obtaining the so-called few-mode Manakov equations. In the weak coupling (WC) regime, only the averaging over birefringence fluctuations must be considered, reducing the intramodal degeneracy factor to $8/9$ and the intermodal degeneracy factor to $4/3$. Thus, the nonlinear operator in (1) becomes:

$$
\hat{N} = -\frac{8}{9} \sum_{k=\{u,v\}} \gamma_{uak} |A_k|^2 + \frac{4}{3} \sum_{k=\{u,v\}} \sum_{k'=\{u,v\}} \gamma_{uakl} |A_k A_{k'}|^2.
$$

In the strong coupling (SC) regime, the averaging includes all propagation modes. For $N$-modes with two polarisations, the nonlinear operator in (1) becomes:

$$
\hat{N} = -j \sum_{k=\{u,v\}} \kappa |A_k|^2, \quad \kappa = \frac{4}{3} \frac{2N}{N^2 + 1} \left( \frac{1}{N} \sum_{k=\{u,v\}} \gamma_{uakl} \right).
$$

**Few-Mode Symmetric Split-Step Fourier Method.** The single-mode split-step Fourier method (SSFM) obtains an approximate solution of the Schrödinger equation by assuming that over a small distance $\Delta x$ the dispersive and nonlinear effects act independently. For FMFs, we extend such an approach by assuming that the mode coupling also acts independently. Such approximations require $\Delta x$ to be much shorter than the dispersion length $T_D/2 |\beta_0^{(1)}|$, and the walk-off length $T_0/|\beta_0^{(1)} - \beta_0^{(2)}|$, where $T_D$ is the pulse width, and shorter than the correlation length $L_c$ defined as $X_T(L_c) = [e^x - 1]/[e^{x/2} + 1]$.

Figure 1 presents a schematic illustration of the few-mode symmetric SSFM proposed. In a symmetric SSFM, the effect of nonlinearity is included in the middle of the segment rather than at the segment boundary, providing higher accuracy. Finally, the step-size was selected by bounding the local error, more computationally efficient at high accuracy than the other methods, e.g. nonlinear phase rotation.

**References**

1. Ellis, A. D., Suibhne, N. M., Saad, D. & Payne, D. N. Communication networks beyond the capacity crunch. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, https://doi.org/10.1098/rsta.2015.0191 (2016).
2. Tkach, R. W. Scaling optical communications for the next decade and beyond. *Bell Labs Technical Journal* 14, 3–9, https://doi.org/10.1002/bltj.20400 (2010).
3. Tucker, R. S. Green Optical Communications®x2014; Part I: Energy Limitations in Transport. *IEEE Journal of Selected Topics in Quantum Electronics* 17, 245–260, https://doi.org/10.1109/JSTQE.2010.2051216 (2011).
4. Kilper, D. C. & Rastegarfar, H. Energy challenges in optical access and aggregation networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, https://doi.org/10.1098/rsta.2014.0435 (2016).
5. Richardson, D. J. New optical fibres for high-capacity optical communications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, https://doi.org/10.1098/rsta.2014.0441 (2016).
6. Li, G., Bai, N., Zhao, N. & Xia, C. Space-division multiplexing: the next frontier in optical communication. *Adv. Opt. Photon.* 6, 413–487, https://doi.org/10.1038/AOP.2014.00013 (2014).
7. Birks, T. A., Gris-Sánchez, I., Yerolatsitis, S., Leon-Saval, S. G. & Thomson, R. R. The photonic lantern. *Adv. Opt. Photon.* 7, 107–167, https://doi.org/10.1038/AOP.2015.00107 (2015).
8. Chen, H. et al. Integrated cladding-pumped multicore few-mode erbium-doped fibre amplifier for space-division-multiplexed communications. *Nature Photonics* 10, 529, https://doi.org/10.1038/nphoton.2016.125, https://www.nature.com/articles/nphoton.2016.125#supplementary-information (2016).
9. Bigot, L., Coçq, G. L. & Quaquempe, Y. Few-Mode Eribium-Doped Fiber Amplifiers: A Review. *Journal of Lightwave Technology* 33, 588–596, https://doi.org/10.1109/JLT.2014.2376975 (2015).
10. Winzer, P. J. & Neilson, D. T. From Scaling Disparities to Integrated Parallelism: A Decathlon for a Decade. *Journal of Lightwave Technology* 35, 1099–1115, https://doi.org/10.1109/JLT.2017.2862082 (2017).
48. Sleiffer, V. A. J. M. et al. 73.7 Tb/s (96 × 3 × 256-Gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA. Optics Express 20, B428–B438, https://doi.org/10.1364/OE.20.00B428 (2012).
49. Arik, S. O., Ho, K.-P. & Kahn, J. M. Delay Spread Reduction in Mode-Division Multiplexing: Mode Coupling Versus Delay Compensation. Journal of Lightwave Technology 33, 4504–4512, https://doi.org/10.1109/JLT.2015.2475422 (2015).
50. Napoli, A. et al. Digital Compensation of Bandwidth Limitations for High-Speed DACs and ADCs. Journal of Lightwave Technology 34, 3053–3064, https://doi.org/10.1109/JLT.2016.2535487 (2016).
51. Schmogrow, R. et al. Error Vector Magnitude as a Performance Measure for Advanced Modulation Formats. IEEE Photonics Technology Letters 24, 61–63, https://doi.org/10.1109/LPT.2011.2172405 (2012).
52. Doran, N. J. & Ellis, A. D. Minimising total energy requirements in amplified links by optimising amplifier spacing. Optics Express 22, 19810–19817, https://doi.org/10.1364/OE.22.019810 (2014).
53. Gruner-Nielsen, L. et al. Few Mode Transmission Fiber With Low DGD, Low Mode Coupling, and Low Loss. Journal of Lightwave Technology 30, 3693–3698, https://doi.org/10.1109/JLT.2012.2227243 (2012).
54. Li, A., Amin, A. A., Chen, X. & Shiob, W. Transmission of 107-Gb/s mode and polarization multiplexed CO-OFDM signal over a two-mode fiber. Opt. Express 19, 8808–8814, https://doi.org/10.1364/OE.19.008808 (2011).
55. Ryf, R. et al. Mode-Division Multiplexing Over 96 km of Few-Mode Fiber Using Coherent 6 × 6 MIMO Processing. Journal of Lightwave Technology 30, 521–531 (2012).
56. Mori, T. et al. Low DMD Four LP Mode Transmission Fiber for Wide-band WDM-MIMO System, Proc. OFC 2013, p. OTh3K.1, https://doi.org/10.1364/OFC.2013.OTh3K.1.
57. Ryf, R. et al. Space-division multiplexed transmission over 4200-km 3-core microstructured fiber, In Proc. OFC 2012, p. PDP5C.2, IEEE, Los Angeles, USA.
58. Hayashi, T., Tamura, Y., Hasegawa, T. & Taru, T. Record-Low Spatial Mode Dispersion and Ultra-Low Loss Coupled Multi-Core Fiber for Ultra-Long-Haul Transmission. Journal of Lightwave Technology 35, 450–457, https://doi.org/10.1109/JLT.2016.2614000 (2017).
59. Maher, R. et al. Spectrally Shaped DP-16QAM Super-Channel Transmission with Multi-Channel Digital Back-Propagation. Scientific Reports 5, 8214, https://doi.org/10.1038/srep08214 (2015).
60. Sakamoto, T., Morii, T., Yamamoto, T. & Tomita, S. Differential Mode Delay Managed Transmission Line for WDM-MIMO System Using Multi-Step Index Fiber. Journal of Lightwave Technology 30, 2783–2787, https://doi.org/10.1109/JLT.2012.2208095 (2012).

Acknowledgements
This work has been partially supported by the EU (654809-HSPACE and 659950-INVENTION), and by EPSRC (EP/L000091/1-PEACE and EP/R024057/1-FPA-ROCS).

Author Contributions
F.F. and C.C. conceived, designed, and performed the simulations; All analysed the data and wrote the paper.

Additional Information
Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018