Pilot-Aided Joint-Channel Carrier-Phase Estimation in Space-Division Multiplexed Multicore Fiber Transmission

(article starts on next page)

Pilot-Aided Joint-Channel Carrier-Phase Estimation in Space-Division Multiplexed Multicore Fiber Transmission

Arni F. Alfredsson, *Student Member, IEEE*, Erik Agrell, *Fellow, IEEE*,
Henk Wymeersch, *Member, IEEE*, Benjamin J. Puttnam, *Member, IEEE*, Georg Rademacher, *Member, IEEE*,
Ruben S. Luís, and Magnus Karlsson, *Senior Member, IEEE, Fellow, OSA*

Abstract—The performance of pilot-aided joint-channel carrier-phase estimation (CPE) in space-division multiplexed multicore fiber (MCF) transmission with correlated phase noise is studied. To that end, a system model describing uncoded MCF transmission where the phase noise comprises a common laser phase noise, in addition to core- and polarization-specific phase drifts, is introduced. It is then shown that the system model can be regarded as a special case of a multidimensional random-walk phase-noise model. A pilot-aided CPE algorithm developed for this model is used to evaluate two strategies, namely joint-channel and per-channel CPE. To quantify the performance differences between the two strategies, their respective phase-noise tolerances are assessed through Monte Carlo simulations of uncoded transmission for different modulation formats, pilot overheads, laser linewidths, numbers of spatial channels, and degrees of phase-noise correlation across the channels. For 20 GBd transmission with 200 kHz combined laser linewidth and 1% pilot overhead, joint-channel CPE yields up to 3.4 dB improvement in power efficiency or 25.5% increased information rate. Moreover, through MCF transmission experiments, the system model is validated and the strategies are compared in terms of bit-error-rate performance versus transmission distance for uncoded transmission of different modulation formats. Up to 21% increase in transmission reach is observed for 1% pilot overhead through the use of joint-channel CPE.

Index Terms—Carrier phase estimation, coherent communications, multicore fiber, space-division multiplexing

I. INTRODUCTION

In response to the ever-increasing throughput demands on fiber-optical networks, space-division multiplexed (SDM) systems have become a topic of interest worldwide [1]. They are believed to have the potential to meet the demands in a cost-effective manner through, e.g., the integration of optical hardware components, sharing of digital signal processing (DSP) resources [2], specialized detection techniques [3],

A. F. Alfredsson, E. Agrell, and H. Wymeersch are with the Department of Electrical Engineering, Chalmers University of Technology, SE-41296 Göteborg, Sweden (e-mail: arnia@chalmers.se; agrell@chalmers.se; henkw@chalmers.se).

M. Karlsson is with the Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology SE-41296 Göteborg, Sweden (e-mail: magnus.karlsson@chalmers.se).

B. J. Puttnam, G. Rademacher, and R. S. Luís are with the Photonic Network System Laboratory, National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan (e-mail: ben@nict.go.jp; georg.rademacher@nict.go.jp; rluis@nict.go.jp).

and the use of spatial superchannels [4]. Moreover, in order to maximize the information rate of the system, multilevel modulation formats, such as dual-polarization M-ary quadrature amplitude modulation (DP-MQAM) or more advanced multidimensional formats [5], are being increasingly utilized. However, such higher-order formats typically come at the cost of higher signal-to-noise ratio (SNR) requirements. They can also increase sensitivity to various transmission impairments, in particular laser phase noise (LPN), which calls for effective carrier-phase estimation (CPE). In addition, symbol-rate optimization implemented through subcarrier multiplexing has been a topic of interest in recent years as gains in transmission reach have been observed by using symbol rates on the order of 2–6 GBd, which is much lower than standard symbol rates [6]. The impact of LPN increases as the symbol rate is decreased, and thus, powerful CPE is even more crucial for symbol-rate-optimized systems.

Traditionally, CPE in optical transmission systems is performed on a per-channel basis, using blind methods such as the Viterbi–Viterbi algorithm [7] or blind phase search (BPS) [8]. However, due to the $\pi/2$ rotation invariance that is inherent to the most commonly used modulation formats, blind methods suffer from ambiguity in the estimated carrier phase and are thus susceptible to cycle slips, which can lead to bursts of errors. Differential encoding can be used to convert the burst errors into a finite number of errors, but it increases the bit error rate (BER) by a constant factor compared to Gray coding in the absence of cycle slips [9]. Alternatively, the CPE can be carried out with the help of pilot symbols whose phases are known unambiguously. This greatly reduces the probability of cycle slips [10] and eliminates the need for differential encoding. Although this comes at the cost of reduced information rate, pilot-aided CPE methods have garnered attention in recent years due to their high performance [11]–[14].

CPE has been experimentally demonstrated for various superchannel transmission scenarios, e.g., SDM transmission using multicore fibers (MCFs) [2] or multimode fibers [15], and wavelength-division multiplexed transmission using frequency combs [16] or electrically generated subcarriers [17]. These demonstrations were possible due to the spatial correlation in the phase noise that is inherent to these systems [2], [15]–[17]. In particular, for SDM transmission using MCFs, the LPN will be common among the spatial channels if all cores share a light

source on the transmitter side and a local oscillator (LO) on the receiver side [2]. However, temperature variations, other external perturbations, and imperfections or certain properties in system components will cause phase drifts that are specific to cores and polarizations [2], [18]. These phase drifts are normally orders of magnitude slower than the LPN, and thus, the phase noise will have a high degree of spatial correlation across the channels.

The correlation can be exploited to lower the required computational complexity in DSP through optical techniques [3], [19], or with the help of DSP-based methods, such as master–slave CPE (MS-CPE) [2]. These strategies rely on CPE using a single spatial channel, whose outcome is shared amongst all spatial channels. Although this may substantially reduce the required CPE resources, any phase differences between the channels will reduce their effectiveness. In contrast, by performing joint-channel CPE (JC-CPE) where all channels are used collectively, phase-noise tolerance can be improved. This can be used to benefit system performance in terms of power efficiency, information rate, hardware requirements, or transmission reach, at the cost of added computational complexity. A comparison between pilot-aided JC-CPE, per-channel CPE (PC-CPE), and MS-CPE was made in [20], [21] in terms of BER versus transmission distance for MCF transmission of DP-16QAM, and it was shown that JC-CPE can cope more effectively than MS-CPE with phase differences between the cores.

Algorithms that perform JC-CPE have been extensively investigated for wireless multiple-input multiple-output transmission [22], [23]. Furthermore, we recently proposed pilot-aided algorithms that perform JC-CPE for optical transmission using forward error correction (FEC) in the presence of arbitrarily correlated phase noise for any number of channels [24]. We further showed that they can significantly outperform the typical CPE approach of using BPS on a per-channel basis in terms of post-FEC BER performance.

In this paper, we investigate the performance of JC-CPE for SDM transmission through uncoupled-core MCFs. The contributions are summarized as follows: (i) We introduce a general phase-noise model for uncoupled-core MCF transmission that comprises a common LPN, in addition to core- and polarization-specific phase drifts that are independent of each other and decorrelate the common phase noise among the spatial channels. (ii) We show that this phase-noise model can be regarded as a multidimensional random walk and utilize a pilot-aided algorithm to perform JC-CPE for arbitrarily correlated phase noise and any number of channels. (iii) Using Monte Carlo simulations, we compare the performance differences between two strategies, namely JC-CPE and PC-CPE, in terms of the resulting power efficiency, information rate, and laser-linewidth requirements of the system. This part of the paper refines preliminary results that were presented in [25]. (iv) Finally, we experimentally validate the system model and evaluate the two strategies in terms of transmission reach for DP quadrature phase-shift keying (DP-QPSK), DP-16QAM, and DP-64QAM transmission using different pilot overheads (OHs). This part extends the PC-CPE and JC-CPE comparison presented in [26].

II. SYSTEM MODEL

Consider uncoded single-wavelength DP transmission in $D/2$ cores resulting in a total of D spatial channels where, without loss of generality, D is assumed to be an even integer. The transmitted symbol block in each channel is modelled as a vector of N independent random variables, where every random variable corresponding to a data symbol is drawn uniformly from a set $\mathcal{X}$ of constellation points that corresponds to the used modulation format. Moreover, the constellation is normalized such that the mean of the constellation points is zero and the average symbol energy is E_s.

All signal distortions are assumed to have been ideally compensated with the exception of phase noise and amplified spontaneous emission noise, which are approximated as a random walk and additive white Gaussian noise (AWGN), resp. Moreover, the adaptive equalization is assumed to have been carried out using a phase-immune equalizer, such as a radially-directed equalizer (RDE) [27]. As already mentioned, the phase noise can be highly correlated, albeit not identical, across the cores and polarizations in an MCF system where all cores share the light-source and LO lasers [2], [18]. In the absence of a model that accurately describes the phase-noise statistics across all channels, we resort to a simplified model in which the phase noise comprises multiple components. The dominant component is common to all the spatial channels and is assumed to contain the combined LPN of the light source and LO. The other components, which account for effects that decorrelate the phase noise across the channels, are core- and polarization-specific phase drifts that are statistically independent of each other. Also assuming one sample per symbol, the discrete-time baseband model is written as

$$r_{i,k} = s_{i,k}e^{j\theta_{i,k}} + n_{i,k}, \tag{1}$$

where $k = 1, \ldots, N$ is a time index and $i = 1, \ldots, D$ is a channel index. The received signals, transmitted symbols, and AWGN samples are denoted with $r_{i,k}$, $s_{i,k}$, and $n_{i,k}$, resp. Each $n_{i,k}$ is the realization of a zero-mean Gaussian random variable with variance σ_i^2 per real dimension, which can be different for each channel. A specific mapping between the channel indices and polarization–core combinations is assumed. Denoting polarization $w \in \{\mathsf{x}, \mathsf{y}\}$ on the jth core with (w, j), the channel indices $1, 2, 3, \ldots, D-1, D$ correspond to $(\mathsf{x}, 1), (\mathsf{y}, 1), (\mathsf{x}, 2), \ldots, (\mathsf{x}, D/2), (\mathsf{y}, D/2)$. The sets of pilot and data symbol indices are denoted with $\mathcal{P}$ and $\mathcal{D}$, resp. Certain symbols within the transmitted blocks are designated as pilots such that if $(i, k) \in \mathcal{P}$ then $s_{i,k} = \rho_{i,k}$, where $\rho_{i,k} \in \mathbb{C}$ is known to the transmitter and receiver.

The phase noise $\theta_{i,k}$ is a sum of three statistically independent components, i.e., $\theta_{i,k} = \theta_k^{\mathrm{L}} + \theta_{i,k}^{\mathrm{C}} + \theta_{i,k}^{\mathrm{P}}$, where θ_k^{L}, $\theta_{i,k}^{\mathrm{C}}$, and $\theta_{i,k}^{\mathrm{P}}$ are the LPN, core-specific phase drift, and polarization-specific phase drift, resp. The statistical nature of $\theta_{i,k}^{\mathrm{C}}$ and $\theta_{i,k}^{\mathrm{P}}$ will be highly dependent on the system involved. For simplicity, all the components are approximated as random walks, i.e.,

$$\theta_k^{\mathrm{L}} = \theta_{k-1}^{\mathrm{L}} + \Delta\theta_k^{\mathrm{L}}, \tag{2}$$

$$\theta_{i,k}^{\mathrm{C}} = \begin{cases} \theta_{i,k-1}^{\mathrm{C}} + \Delta\theta_{i,k}^{\mathrm{C}}, & i \text{ odd}, \\ \theta_{i-1,k}^{\mathrm{C}}, & i \text{ even}, \end{cases} \tag{3}$$

$$\theta_{i,k}^{\mathrm{P}} = \theta_{i,k-1}^{\mathrm{P}} + \Delta\theta_{i,k}^{\mathrm{P}}, \tag{4}$$

for $k = 2, \ldots, N$, where θ_1^{L}, $\theta_{i,1}^{\mathrm{C}}$, and $\theta_{i,1}^{\mathrm{P}}$ are uniformly distributed on $[0, 2\pi)$, and $\Delta\theta_k^{\mathrm{L}}$, $\Delta\theta_{i,k}^{\mathrm{C}}$, and $\Delta\theta_{i,k}^{\mathrm{P}}$ are zero-mean Gaussian random variables with variances σ_{L}^2, σ_{C}^2, and σ_{P}^2, resp. In particular, $\sigma_{\mathrm{L}}^2 \triangleq 2\pi\Delta\nu T_{\mathrm{s}}$, where $\Delta\nu$ is the combined linewidth of the light-source and LO lasers, and T_{s} is the symbol duration. Moreover, σ_{C}^2 and σ_{P}^2 are defined in relation to σ_{L}^2. These variances determine the speed of their corresponding phase drifts. The phase noise is statistically independent of the transmitted symbols and AWGN, and unknown to both the transmitter and receiver. Finally, the variances σ_{L}^2, σ_{C}^2, σ_{P}^2, and $(\sigma_1^2, \ldots, \sigma_D^2)$ are assumed to be known to the receiver.

The phase noise can alternatively be described as a multi-dimensional random walk, i.e.,

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \Delta\boldsymbol{\theta}_k, \tag{5}$$

where $\boldsymbol{\theta}_1$ is uniformly distributed on $[0, 2\pi)^D$ and $\Delta\boldsymbol{\theta}_k$ is a multivariate zero-mean Gaussian random variable with covariance matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 & \mathbf{Q}_2 & \cdots & \mathbf{Q}_2 \\ \mathbf{Q}_2 & \mathbf{Q}_1 & & & \vdots \\ \mathbf{Q}_2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \mathbf{Q}_2 \\ \mathbf{Q}_2 & \cdots & \cdots & \mathbf{Q}_2 & \mathbf{Q}_1 \end{bmatrix} \in \mathbb{R}^{D \times D}, \tag{6}$$

where

$$\mathbf{Q}_1 = \begin{bmatrix} \sigma_{\mathrm{L}}^2 + \sigma_{\mathrm{C}}^2 + \sigma_{\mathrm{P}}^2 & \sigma_{\mathrm{L}}^2 + \sigma_{\mathrm{C}}^2 \\ \sigma_{\mathrm{L}}^2 + \sigma_{\mathrm{C}}^2 & \sigma_{\mathrm{L}}^2 + \sigma_{\mathrm{C}}^2 + \sigma_{\mathrm{P}}^2 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} \sigma_{\mathrm{L}}^2 & \sigma_{\mathrm{L}}^2 \\ \sigma_{\mathrm{L}}^2 & \sigma_{\mathrm{L}}^2 \end{bmatrix}. \tag{7}$$

The covariance matrix in (6) is specific to the structure of the phase noise in (1). The multidimensional random-walk description in (5) and (6) is utilized by the CPE algorithm, which is detailed in the next section.

III. CPE ALGORITHM

JC-CPE involves using all channels collectively in order to estimate the phase noise simultaneously across the channel domain. This strategy can be implemented using either blind or pilot-aided algorithms. If the phase noise is assumed identical in all channels, existing PC-CPE algorithms be extended in a simple manner to perform estimate averaging across the channel domain, which will reduce the impact of the additive noise that corrupts the CPE. However, to track interchannel phase drifts, more involved algorithms are required. Here, we make use of an iterative pilot-aided algorithm that is developed using the sum–product algorithm, which operates in a factor graph. It approximates the maximum *a posteriori* (MAP) symbol detector for multichannel transmission in the presence of arbitrarily correlated phase noise. Moreover, the algorithm is designed to be used after adaptive equalization has taken place, i.e., the CPE performed by the algorithm is not embedded in the loop of, e.g., a least mean square (LMS) equalizer. Finally, it can perform CPE on any number of channels. Hence, although it is designed for JC-CPE, it can

also be used on a channel-by-channel basis to perform PC-CPE. In this section, the main derivation results are presented. For more details on the derivations, refer to [24], where an analogous bit-detection algorithm for coded transmission is derived using the same techniques.

Let $\boldsymbol{r}$, $\boldsymbol{s}$, and $\boldsymbol{\theta}$ contain all the received samples, transmitted symbols, and phase-noise samples, respectively. The MAP symbol-detection strategy is optimal in the sense that it minimizes the symbol error rate [28, Ch. 5.1]. It is performed on a symbol-by-symbol basis according to

$$\hat{s}_{i,k} = \underset{x \in \mathcal{X}}{\arg\max}\, P(s_{i,k}|\boldsymbol{r}). \tag{8}$$

The *a posteriori* symbol probability in (8) is hard to compute exactly for the considered system model. However, it can be expressed as the marginalization of the joint *a posteriori* distribution of $\boldsymbol{s}$ and $\boldsymbol{\theta}$, i.e.,

$$P(s_{i,k}|\boldsymbol{r}) = \int_{\mathbb{R}^{D \times N}} \sum_{\boldsymbol{s} \in \mathcal{S}_{i,k}} p(\boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r}) \mathrm{d}\boldsymbol{\theta}, \tag{9}$$

where $\mathcal{S}_{i,k} = \{\boldsymbol{s}' \in \mathcal{X}^{D \times N} : s'_{i,k} = s_{i,k}\}$. Moreover, $p(\boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r})$ can be factorized as

$$p(\boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{r}) \propto \prod_{k=2}^{N} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) \prod_{i,k} p(r_{i,k}|s_{i,k}, \theta_{i,k}), \tag{10}$$

where $\propto$ denotes proportionality with respect to $\boldsymbol{r}$.

Applying the sum–product algorithm on a factor graph associated with the right-hand side of (10) yields messages $P_{i,k}(s_{i,k})$ and $p_k(\boldsymbol{\theta}_k)$ that approximate the *a posteriori* probability of $s_{i,k}$ and probability density function of $\boldsymbol{\theta}_k$, resp., for all i, k. The factor graph does not contain any cycles, and hence, it does not yield an iterative algorithm. However, to improve performance, the messages $P_{i,k}(s_{i,k})$ and $p_k(\boldsymbol{\theta}_k)$ are computed in an iterative fashion. In each iteration, $p_k(\boldsymbol{\theta}_k)$ is first updated for all k through the use of extended Kalman smoothing (EKS) [29, Ch. 9] and soft symbols, i.e., the first and second moments of $s_{i,k}$ with respect to $P_{i,k}(s_{i,k})$ from a previous iteration, for all $i, k,$. Then, $P_{i,k}(s_{i,k})$ is updated using the current estimates of $p_k(\boldsymbol{\theta}_k)$ for all $(i, k) \in \mathcal{D}$. This iterative process reduces the resulting BER after symbol-to-bit mapping until the iterations converge, after which the algorithm performance cannot be improved further. The convergence speed depends on various system parameters such as the modulation format, pilot rate, and laser linewidth. In this paper, a fixed number of iterations is used as a criterion for stopping the iterations.

This CPE algorithm will be referred to as factor-graph based EKS (FGK) hereafter. It is described in a high-level manner in Fig. 1 and detailed in the form of a pseudocode in Algorithm 1. The first and second moments of $s_{i,k}$, denoted as $\bar{s}_{i,k}$ and $\bar{\sigma}_{i,k}^2$, are initialized in lines 1–4 such that $\bar{s}_{i,k} = \rho_{i,k}$ and $\bar{\sigma}_{i,k}^2 = \sigma_i^2$ for all $(i, k) \in \mathcal{P}$, whereas $\bar{s}_{i,k} = 0$ and $\bar{\sigma}_{i,k}^2 = \sigma_i^2 + E_{\mathrm{s}}/2$ for all $(i, k) \in \mathcal{D}$. The EKS equations, used to estimate $p_k(\boldsymbol{\theta}_k)$ for all k, are then listed in lines 6–20, where $\boldsymbol{\theta}_k^{\mathrm{s}}$ and $\mathbf{M}_k^{\mathrm{s}}$ represent the estimated mean and covariance of $p_k(\boldsymbol{\theta}_k)$ at time k, resp. Finally, the logarithm of $P_{i,k}(s_{i,k})$ is computed[1] for

---

[1]The reason for computing the logarithm of $P_{i,k}(s_{i,k})$ is to ensure numerical stability.
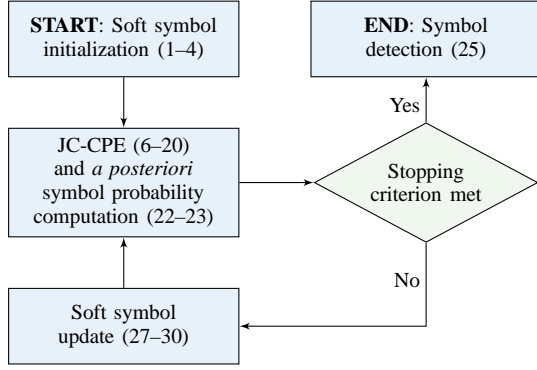
Fig. 1. Flowchart illustrating the high-level structure of FGK. The numbers in the blocks correspond to line numbers in Algorithm 1.
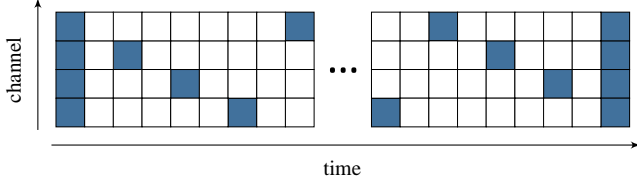

Fig. 2. The pilot symbol arrangement where the pilot symbols, spaced at every other time index, are placed on a "wrapped diagonal" line [30]. Dark and white squares indicate pilot and data symbols, resp.

all $(i,k) \in \mathcal{D}$ in lines 22–23, where $\theta_{i,k}^{\mathrm{s}}$ is the ith component of $\boldsymbol{\theta}_k^{\mathrm{s}}$ and $M_{i,k}^{\mathrm{s}}$ is the ith element on the diagonal line of $\mathbf{M}_k^{\mathrm{s}}$. If the stopping criterion has been met, symbol detection is performed in line 25. Otherwise, the soft symbols are updated in lines 27–30 and another iteration is run.

IV. SIMULATION RESULTS

In this section, FGK and BPS are first compared for PC-CPE in terms of phase-noise tolerance in order to put the performance of FGK into perspective. Thereafter, PC-CPE and JC-CPE, which are both implemented using FGK, are compared in terms of the resulting power efficiency, information rate, and laser-linewidth requirements of the system. To that end, uncoded transmission of Gray-mapped DP-MQAM, for $M = 16, 64, 256, 1024$, is carried out using Monte Carlo simulations for different numbers of cores, pilot OHs, laser linewidths, and degrees of spatial correlation in the phase noise. FGK is run for 2 iterations in all cases unless otherwise stated.

The length of the transmitted block in each channel is $N = 10\,000$ data and pilot symbols. The pilot symbols all take on the same point, i.e., $\rho_{i,k} = \sqrt{E_{\mathrm{s}}} \ \forall \ (i,k) \in \mathcal{P}$. For PC-CPE, they are distributed identically in each channel, i.e., uniformly throughout the transmitted block. For JC-CPE, we showed in [30] that a particular pilot symbol arrangement, where the pilot symbols are placed on a diagonal line that wraps around the top and bottom rows of the matrix, performs well in general for a wide range of phase-noise correlation degrees. The average pilot OH across the channels, $\mathrm{OH_P}$, can be adjusted by changing the temporal distance between the pilot symbols. Finally, all channels have a pilot symbol at the beginning and end of the transmitted block. This arrangement

---

Algorithm 1 FGK

Input: $\boldsymbol{r}$, D, N, $\mathcal{P}$, $\mathcal{D}$, $\mathbf{Q}$, $\boldsymbol{\sigma}^2$, $\mathcal{X}$
Output: $\hat{s}_{i,k}, \forall \ (i,k) \in \mathcal{D}$
1: $\bar{s}_{i,k} = \rho_{i,k}, \forall \ (i,k) \in \mathcal{P}$
2: $\bar{s}_{i,k} = 0, \forall \ (i,k) \in \mathcal{D}$
3: $\bar{\sigma}_{i,k}^2 = \sigma_i^2, \forall \ (i,k) \in \mathcal{P}$
4: $\bar{\sigma}_{i,k}^2 = \sigma_i^2 + E_{\mathrm{s}}/2, \forall \ (i,k) \in \mathcal{D}$
5: **for all** iterations **do**
6: $\quad \hat{\boldsymbol{\theta}}_1^{\mathrm{f}} = \left[\angle(r_{1,1}\bar{s}_{1,1}^*), \dots, \angle(r_{D,1}\bar{s}_{D,1}^*)\right]^T$
7: $\quad \mathbf{M}_1^{\mathrm{f}} = \mathrm{diag}(\bar{\sigma}_{1,k}^2/E_{\mathrm{s}}, \dots, \bar{\sigma}_{D,k}^2/E_{\mathrm{s}})$
8: $\quad$ **for** $k = 2, \dots, N$ **do**
9: $\qquad h_{i,k} = \Im\{r_{i,k}\bar{s}_{i,k}^* e^{-j\hat{\theta}_{i,k-1}^{\mathrm{f}}}\}/\bar{\sigma}_{i,k}^2, \forall \ i = 1, \dots, D$
10: $\qquad \mathbf{V}_k = \mathrm{diag}(|\bar{s}_{1,k}|^2/\bar{\sigma}_{1,k}^2, \dots, |\bar{s}_{D,k}|^2/\bar{\sigma}_{D,k}^2)$
11: $\qquad \mathbf{M}_{k|k-1}^{\mathrm{f}} = \mathbf{M}_{k-1}^{\mathrm{f}} + \mathbf{Q}$
12: $\qquad \mathbf{M}_k^{\mathrm{f}} = (\mathbf{I}_D + \mathbf{M}_{k|k-1}^{\mathrm{f}}\mathbf{V}_k)^{-1}\mathbf{M}_{k|k-1}^{\mathrm{f}}$
13: $\qquad \hat{\boldsymbol{\theta}}_k^{\mathrm{f}} = \hat{\boldsymbol{\theta}}_{k-1}^{\mathrm{f}} + \mathbf{M}_k^{\mathrm{f}}\boldsymbol{h}_k$
14: $\quad$ **end for**
15: $\quad \hat{\boldsymbol{\theta}}_N^{\mathrm{s}} = \hat{\boldsymbol{\theta}}_N^{\mathrm{f}}$ and $\mathbf{M}_N^{\mathrm{s}} = \mathbf{M}_N^{\mathrm{f}}$
16: $\quad$ **for** $k = N-1, N-2, \dots, 1$ **do**
17: $\qquad \mathbf{A}_k = \mathbf{M}_k^{\mathrm{f}}(\mathbf{M}_{k+1|k}^{\mathrm{f}})^{-1}$
18: $\qquad \hat{\boldsymbol{\theta}}_k^{\mathrm{s}} = \hat{\boldsymbol{\theta}}_k^{\mathrm{f}} + \mathbf{A}_k(\hat{\boldsymbol{\theta}}_{k+1}^{\mathrm{s}} - \hat{\boldsymbol{\theta}}_k^{\mathrm{f}})$
19: $\qquad \mathbf{M}_k^{\mathrm{s}} = \mathbf{M}_k^{\mathrm{f}} + \mathbf{A}_k(\mathbf{M}_{k+1}^{\mathrm{s}} - \mathbf{M}_{k+1|k}^{\mathrm{f}})\mathbf{A}_k^T$
20: $\quad$ **end for**
21: $\quad$ **for all** $(i,k) \in \mathcal{D}$ **do**
22: $\qquad \xi_{i,k}(x) = \frac{e^{j\hat{\theta}_{i,k}^{\mathrm{s}}}}{M_{i,k}^{\mathrm{s}}} + \frac{r_{i,k}x^*}{\sigma_i^2} - \frac{r_{i,k}\bar{s}_{i,k}^*}{\bar{\sigma}_{i,k}^2}, \forall \ x \in \mathcal{X}$
23: $\qquad f_{i,k}(x) = |\xi_{i,k}(x)| - \frac{|x|^2}{2\sigma_i^2} - \frac{1}{2}\ln|\xi_{i,k}(x)|, \forall \ x \in \mathcal{X}$
24: $\qquad$ **if** stopping criterion met **then**
25: $\qquad\quad \hat{s}_{i,k} = \mathrm{argmax}_{x \in \mathcal{X}} f_{i,k}(x)$
26: $\qquad$ **else**
27: $\qquad\quad f_{i,k}^{\max} = \max_{x \in \mathcal{X}} f_{i,k}(x)$
28: $\qquad\quad P_{i,k}(x) = \frac{\exp(f_{i,k}(x) - f_{i,k}^{\max})}{\sum_{x' \in \mathcal{X}} \exp(f_{i,k}(x') - f_{i,k}^{\max})}, \forall \ x \in \mathcal{X}$
29: $\qquad\quad \bar{s}_{i,k} = \sum_{x \in \mathcal{X}} x P_{i,k}(x)$
30: $\qquad\quad \bar{\sigma}_{i,k}^2 = \sigma_i^2 + \frac{1}{2}\sum_{x \in \mathcal{X}} |x - \bar{s}_{i,k}|^2 P_{i,k}(x)$
31: $\qquad$ **end if**
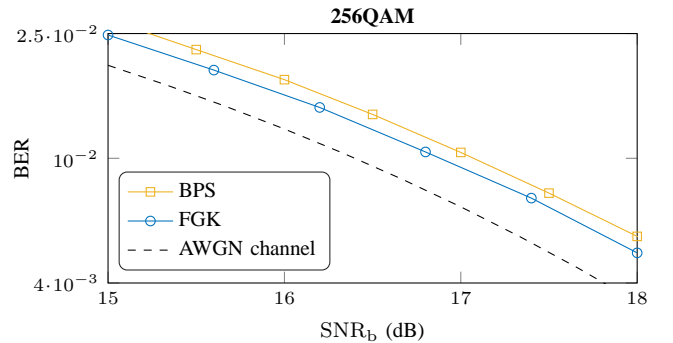32: $\quad$ **end for**
33: **end for**

Fig. 3. BER versus $\mathrm{SNR_b}$ for transmission of 256QAM in the presence of phase noise, comparing FGK and BPS for PC-CPE. As a reference, the theoretical BER of transmission over the AWGN channel is included.

is illustrated in Fig. 2 for 4 spatial channels and the pilot symbols placed at every other time index.
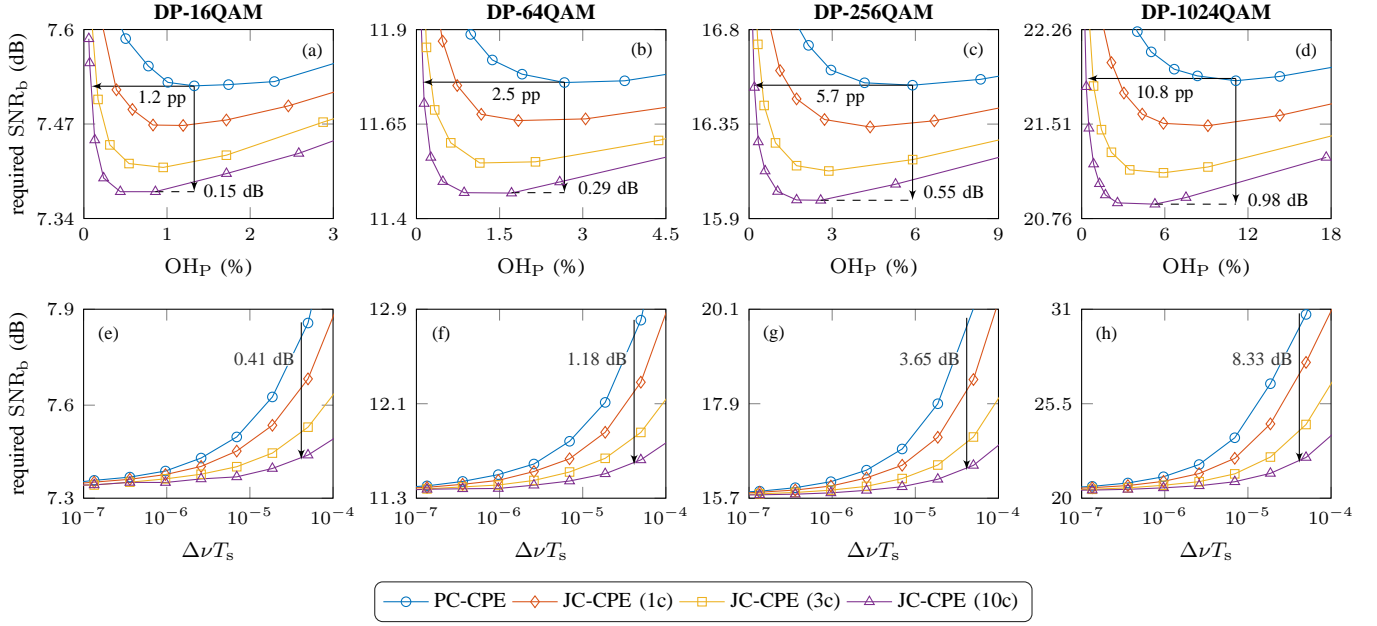
Performance is assessed by estimating BER or achievable

Fig. 4. The required $\mathrm{SNR_b}$ to attain the target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$ versus $\mathrm{OH_P}$ (top) and $\Delta\nu T_s$ (bottom), comparing PC-CPE and JC-CPE for different modulation formats and numbers of cores.

information rate (AIR) for different $\mathrm{OH_P}$, laser linewidth and symbol duration products, $\Delta\nu T_s$, and SNRs per bit [31, Ch. 1], $\mathrm{SNR_b}$, defined as

$$\mathrm{SNR_b} \triangleq \frac{E_s(1 + \mathrm{OH_P})}{2\sigma^2 \log_2 |\mathcal{X}|}, \tag{11}$$

where σ^2 is the complex AWGN variance per real dimension, which is identical for all channels in this section. From (11), it can be seen that the rate reduction due to the use of pilots is penalized through an increased AWGN variance, since for a fixed E_s, $|\mathcal{X}|$, and $\mathrm{SNR_b}$, σ^2 increases with $\mathrm{OH_P}$. For each BER estimate, bit errors are accumulated through repeated transmission until the total number of bit errors reaches at least 10^4. The data bits, phase noise, and AWGN are realized according to the system model in (1), independently for each transmission.

A. Comparison Between FGK and BPS for PC-CPE

To ensure high performance for both algorithms, BPS uses 128 test phases and the filter length is optimized for each tested SNR to minimize the resulting BER, whereas FGK is run for 20 iterations to allow for proper convergence of the CPE. Moreover, BPS is provided with perfect knowledge of the initial value of the phase noise. Fig. 3 shows BER as a function of $\mathrm{SNR_b}$ for transmission of 256QAM at 20 GBd, with $\Delta\nu = 200$ kHz and $\mathrm{OH_P} = 1\%$ for FGK. The theoretical BER performance for uncoded, pilot-free transmission of Gray-mapped 256QAM over the AWGN channel [32] is included as a reference. FGK outperforms BPS across all tested SNRs, which is consistent with the coded-transmission results presented in [24].

B. Power Efficiency

To assess the gains in power efficiency that JC-CPE enables, the required $\mathrm{SNR_b}$ to achieve a target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$, corresponding to a staircase code with a FEC OH of $\mathrm{OH_{FEC}} = 20\%$ [33], is estimated for different $\mathrm{OH_P}$, $\Delta\nu T_s$, and degrees of spatial correlation in the phase noise. Since the system model in (1) does not account for nonlinearities, the results in this subsection are mostly relevant to short-haul transmission where the system operates in the linear regime.

Figs. 4 (a)–(d) show the required $\mathrm{SNR_b}$ as a function of $\mathrm{OH_P}$ at 20 GBd for $\Delta\nu = 200$ kHz, $\sigma_C^2 = \sigma_L^2/10^3$, and $\sigma_P^2 = \sigma_L^2/10^6$. As can be seen, there exists an optimal $\mathrm{OH_P}$ for all modulation formats, denoted with $\mathrm{OH_P^*}$, that requires the minimum $\mathrm{SNR_b}$ to attain the pre-FEC BER threshold. The phase-noise tolerance of JC-CPE increases with the number of cores, which leads to a lower minimum required $\mathrm{SNR_b}$ compared to PC-CPE. This reduction, marked with a vertical arrow in each plot, is up to 0.98 dB for transmission of DP-1024QAM through 10 cores. However, note that $\mathrm{OH_P^*}$ is higher for PC-CPE than JC-CPE. Moreover, the difference in required $\mathrm{SNR_b}$ highly depends on $\mathrm{OH_P}$. As an example, a difference of 0.15 dB, 0.41 dB, 1.12 dB, and 3.38 dB is observed at $\mathrm{OH_P} = 1\%$ for 10-core transmission of DP-16QAM, DP-64QAM, DP-256QAM, DP-1024QAM, resp.

Figs. 4 (e)–(h) show the required $\mathrm{SNR_b}$ as a function of $\Delta\nu T_s$ for $\mathrm{OH_P} = 1\%$, $\sigma_C^2 = \sigma_L^2/10^3$, and $\sigma_P^2 = \sigma_L^2/10^6$. As before, the phase-noise tolerance increases for JC-CPE with the number of cores, and thus, the required $\mathrm{SNR_b}$ to attain the pre-FEC BER threshold is less for JC-CPE than PC-CPE. Furthermore, the difference in the required $\mathrm{SNR_b}$ between PC-CPE and JC-CPE grows with increasing $\Delta\nu T_s$, i.e., with increasing laser linewidth and/or decreasing symbol rate. The difference for $\Delta\nu T_s = 5 \cdot 10^{-5}$, again marked with
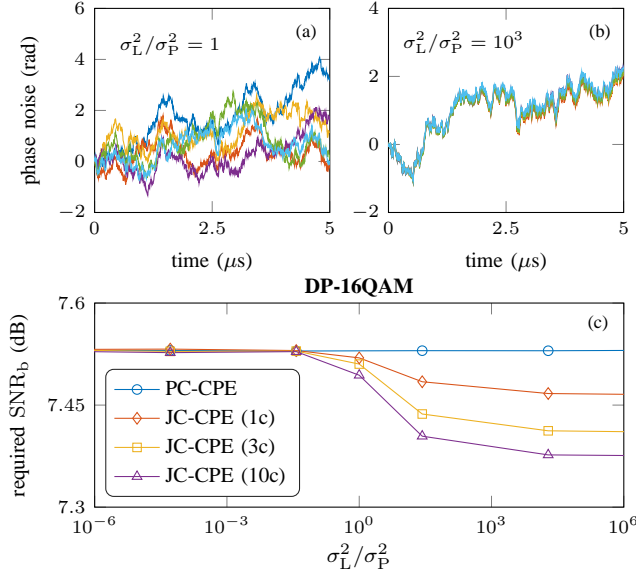
Fig. 5. (a)–(b) Phase-noise realizations for transmission through 6 spatial channels with two different correlation degrees. (c) The required $\mathrm{SNR_b}$ to attain the target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$ versus $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2$, comparing PC-CPE and JC-CPE for different numbers of cores.

TABLE I
MAXIMUM FEC OH REDUCTION AND RESULTING SE INCREASE FOR
DIFFERENT MODULATION FORMATS AND NUMBERS OF CORES (IR:
INFORMATION RATE, PP: PERCENTAGE POINT)

	Max. OH reduction (pp)			IR increase (%)		
	1c	3c	10c	1c	3c	10c
DP-16QAM	0	1.90	1.90	0	1.54	1.54
DP-64QAM	1.90	3.33	8.33	1.54	2.85	6.66
DP-256QAM	4.29	10.23	13.33	3.90	8.31	11.11
DP-1024QAM	13.33	23.33	27.08	11.11	21.21	25.49

a vertical arrow in the plots, is up to 8.33 dB in the case of DP-1024QAM transmission through 10 cores. The value $\Delta\nu T_\mathrm{s} = 5 \cdot 10^{-5}$ corresponds to, e.g., $\Delta\nu = 1$ MHz at 20 GBd or $\Delta\nu = 100$ kHz at 2 GBd.

For illustration purposes, Figs. 5 (a) and (b) show phase-noise realizations in 6 channels for two different degrees of phase-noise correlation, namely $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2 = 1$ and $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2 = 10^3$. Moreover, Fig. 5 (c) shows the required $\mathrm{SNR_b}$ as a function of phase-noise correlation for transmission of DP-16QAM with $\mathrm{OH_P} = 1\%$. The correlation is quantified by fixing $\sigma_\mathrm{L}^2 + \sigma_\mathrm{P}^2 = 6.3 \cdot 10^{-5}$ (corresponding to $\Delta\nu = 200$ kHz at 20 GBd) and $\sigma_\mathrm{C}^2 = 0$, while varying the ratio $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2$. As the ratio tends to 0, the phase noise becomes independent between the spatial channels, and JC-CPE has no performance gain over PC-CPE. However, as the ratio tends to infinity, the phase noise becomes identical in all channels. For $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2 > 10^{-1}$, JC-CPE outperforms PC-CPE, and the performance gains grow with increasing correlation. At approximately $\sigma_\mathrm{L}^2/\sigma_\mathrm{P}^2 = 10^4$, JC-CPE reaches a point where a greater correlation yields marginal gains. Identical results are found for DP-64QAM, DP-256QAM, and DP-1024QAM.

C. Information Rate

The information rate of the system can be increased by, e.g., lowering $\mathrm{OH_{FEC}}$ and/or $\mathrm{OH_P}$. In Figs. 4 (a)–(d), it can be seen that the value of $\mathrm{OH_P}$ corresponding to a particular required $\mathrm{SNR_b}$ decreases for JC-CPE with an increasing number of cores. As an example, the difference in $\mathrm{OH_P}$ between PC-CPE and JC-CPE to achieve the minimum required $\mathrm{SNR_b}$ of PC-CPE, marked with a horizontal arrow in each plot, is up to 10.8 percentage points for transmission of DP-1024QAM through 10 cores. Given the assumption of $\mathrm{OH_{FEC}} = 20\%$, this corresponds to a 12.1% rate increase (from 7.41 b/symbol to 8.31 b/symbol in each polarization).

Fig. 6 shows BER versus $\mathrm{SNR_b}$, comparing PC-CPE and JC-CPE for transmission at 20 GBd through 10 cores with $\Delta\nu = 200$ kHz, $\mathrm{OH_P} = 1\%$, $\sigma_\mathrm{C}^2 = \sigma_\mathrm{L}^2/10^3$, and $\sigma_\mathrm{P}^2 = \sigma_\mathrm{L}^2/10^6$. To determine possible rate improvements that can be achieved by lowering $\mathrm{OH_{FEC}}$ and performing JC-CPE, several staircase codes from [33] are considered with their corresponding pre-FEC BER thresholds and $\mathrm{OH_{FEC}}$ values marked in the plot. As in Section IV-A, the theoretical BER performance for uncoded, pilot-free transmission of Gray-mapped QAM over the AWGN channel is included as a reference. As JC-CPE yields lower BER than PC-CPE, it allows for the use of a FEC code with lower pre-FEC threshold and $\mathrm{OH_{FEC}}$ than what could otherwise be used in the case of PC-CPE. By considering $\mathrm{SNR_b}$ values where the PC-CPE performance crosses a pre-FEC BER threshold, the greatest reduction in $\mathrm{OH_{FEC}}$, marked with a vertical arrow for each modulation format at the corresponding $\mathrm{SNR_b}$, is up to 27.08 percentage points in the case of DP-1024QAM transmission. Since $\mathrm{OH_P} = 1\%$, this corresponds to a 25.5% rate increase (from 7.50 b/symbol to 9.41 b/symbol in each polarization). Table I further details the maximum percentage-point reduction in $\mathrm{OH_{FEC}}$ and the corresponding rate gain for different modulations formats and numbers of cores, again by looking at $\mathrm{SNR_b}$ values where PC-CPE crosses a pre-FEC BER threshold. Overall, increasing the modulation format order and the number of cores yields a bigger rate improvement.

In order to estimate possible rate improvements through JC-CPE when a soft-decision bit-wise FEC decoder is used, an appropriate performance metric to consider is a particular AIR: the generalized mutual information (GMI) including rate loss due to pilot symbols. The GMI is computed according to [34, Eq. (26)] with the bit-wise log-likelihood ratios calculated in exact form. Fig. 7 shows a comparison between PC-CPE and JC-CPE in terms of AIR for different values of $\mathrm{SNR_b}$. Transmission of DP-16QAM and DP-1024QAM at 20 GBd through 10 cores is considered, with $\Delta\nu = 200$ kHz, $\mathrm{OH_P} = 1\%$, $\sigma_\mathrm{C}^2 = \sigma_\mathrm{L}^2/10^3$, and $\sigma_\mathrm{P}^2 = \sigma_\mathrm{L}^2/10^6$. As a reference, the GMI and Shannon capacity [35] of the AWGN channel, are also shown. As Fig. 7 shows, the performance gains due to joint processing are marginal for DP-16QAM but significant for DP-1024QAM with up to 0.59 b/symbol increase per polarization in AIR.
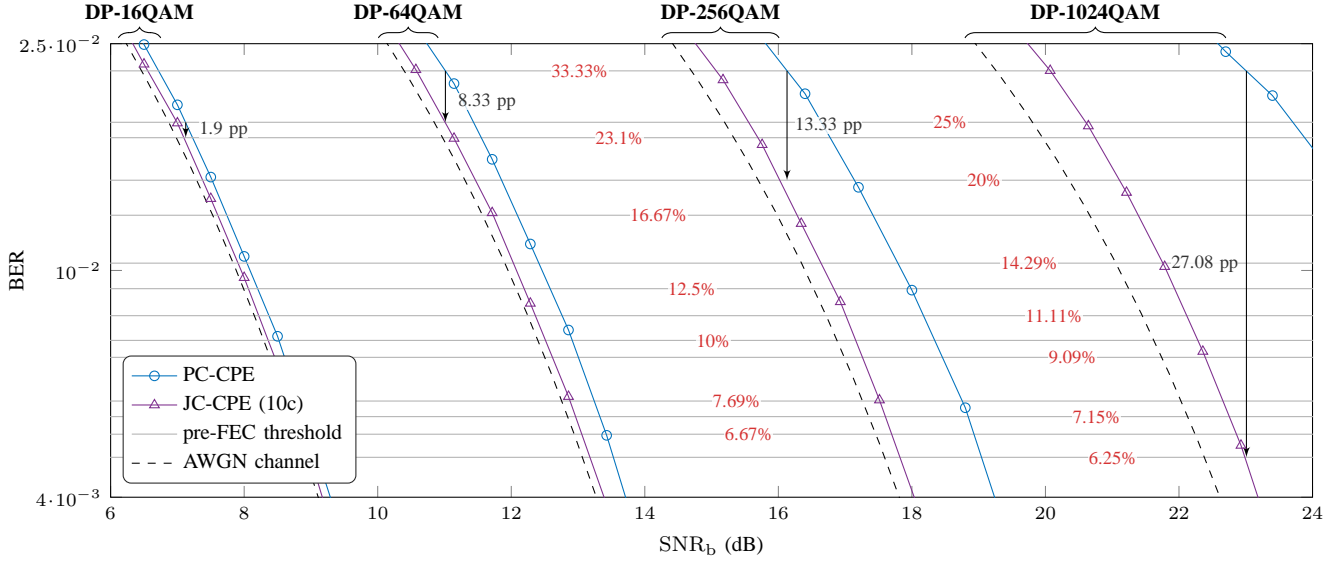
Fig. 6. BER versus SNR_b for transmission of different modulation formats through 10 cores, comparing PC-CPE and JC-CPE. The pre-FEC BER thresholds and the corresponding OH_{FEC} of the considered staircase codes are marked in the plot with gray horizontal lines and red text, resp. As a reference, the theoretical BER of transmission in the absence of phase noise is included. (pp: percentage point)
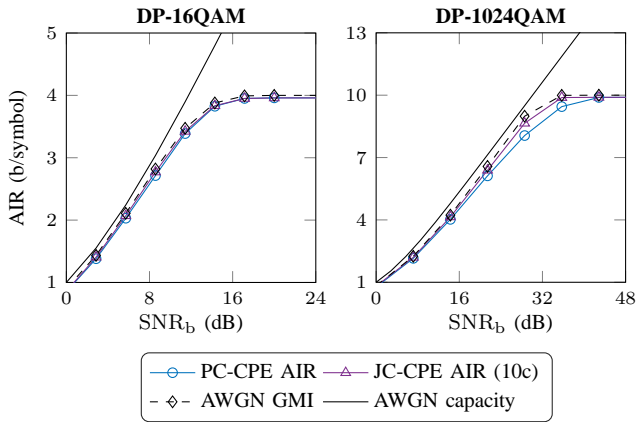

Fig. 7. AIR in bits per symbol per polarization versus SNR_b for transmission of DP-16QAM and DP-1024QAM through 10 cores, comparing PC-CPE and JC-CPE. The channel capacity and GMI for the AWGN channel are included for reference.
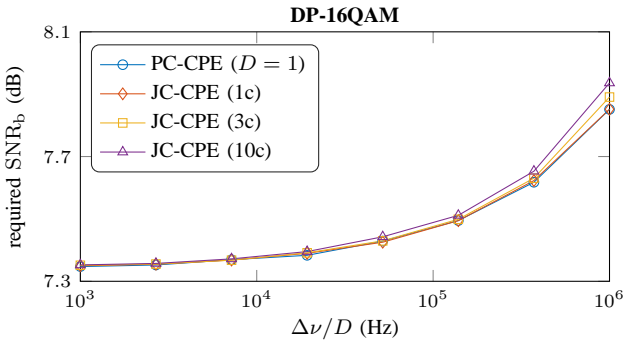

Fig. 8. The required SNR_b to attain the target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$ versus $\Delta\nu/D$, comparing PC-CPE and JC-CPE for different numbers of cores.

D. Laser-Linewidth Requirements

Fig. 8 shows the required SNR_b to attain a target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$ as a function of $\Delta\nu/D$, i.e.,

the laser linewidth normalized by the number of channels, for transmission of DP-16QAM at 20 GBd with $OH_P = 1\%$, $\sigma_C^2 = \sigma_L^2/10^3$, and $\sigma_P^2 = \sigma_L^2/10^6$. All the curves essentially overlap, which shows that JC-CPE for D channels tolerates approximately D times more laser linewidth compared to PC-CPE. Thus, in the case of SDM DP transmission through $D/2$ cores that share lasers with linewidths $\Delta\nu$, JC-CPE performs close to standard transmission through $D/2$ cores, where each core has an independent laser with linewidth $\Delta\nu/D$, provided that the phase noise has sufficient spatial correlation.

V. EXPERIMENTAL RESULTS

In this section, PC-CPE and JC-CPE are experimentally compared in terms of BER versus transmission distance, and the results are used to validate the system model in (1). The experimental setup consisted of 3 synchronized recirculating loops running through 3 adjacent cores of a 7-core, 53.7 km, weakly-coupled, single-mode, homogeneous MCF. Transmission of DP-QPSK, DP-16QAM, and DP-64QAM was carried out at 20 GBd, with the corresponding transmitter output power set to -4 dBm, -3.5 dBm, and 0 dBm, resp. An external cavity laser (ECL) with 100 kHz linewidth operating at 1550 nm was used as a light source and shared for all cores at the transmitter, while a separate 100 kHz ECL acting as an LO was shared for all cores at the receiver. All detected signals were simultaneously digitized in a 12-channel oscilloscope operating at 80 GS/s for offline DSP. For a more detailed description of the setup, refer to [36].

All stages in the DSP chain except the CPE were performed on a per-core basis. The first stages consisted of resampling to 2 samples per symbol, dispersion estimation and compensation, orthonormalization, timing recovery, and blind 2×2 equalization using the constant modulus algorithm. The orthonormalization was performed using the Gram–Schmidt algorithm, but alternatively, it could be achieved in the equal-
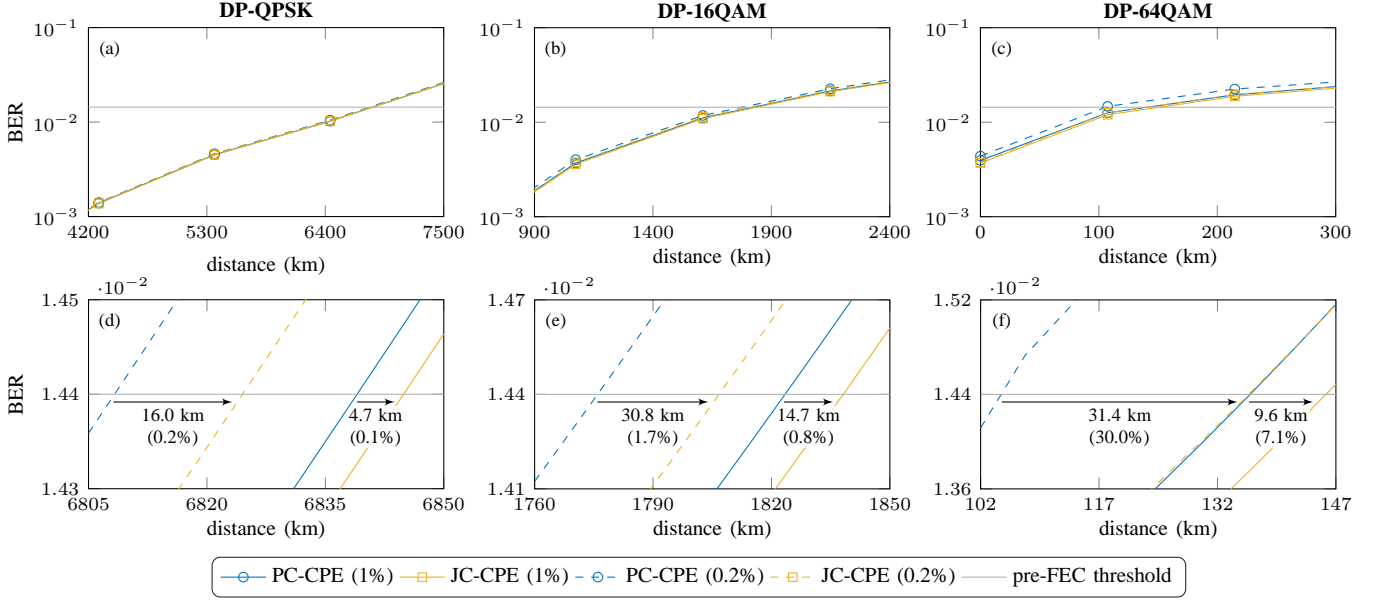
Fig. 9. Experimental results showing BER versus transmission distance, comparing JC-CPE and PC-CPE for transmission through 3 cores of a weakly-coupled single-mode MCF, using different modulation formats and $\mathrm{OH_P}$. The bottom plots show the same results zoomed in around the target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$.

ization stage through the use of a widely linear equalizer [37]. At this point, frame synchronization was performed. A training stage followed where approximately 10^5 symbols were used for LMS 2×2 equalization and estimation of the complex additive noise variance in each spatial channel. The rationale for performing per-core 2×2 equalization as opposed to 6×6 equalization was to save processing time, as the latter variant has a much higher complexity. Moreover, it was shown in [36] using the same setup as in this paper that per-core 2×2 equalization does not yield significantly worse performance than 6×6 equalization. After the training stage, an RDE with a slow convergence rate was used, initialized using the filter taps from the trained equalization. The RDE is immune to the signal phase and does not compensate for frequency offsets and phase noise. Therefore, an explicit carrier recovery stage was needed following the RDE. As FGK is designed to be used after adaptive equalization, the choice the RDE is reasonable. The equalization was followed by blind carrier-frequency offset (CFO) compensation, down-conversion, matched filtering, and down-sampling to 1 sample per symbol.

Effective joint-core CFO compensation was not possible due to the acousto-optic modulators used in the experimental setup, which introduced different frequency shifts to the signals running through the different cores. However, relative CFOs between the cores were minimized as follows. Coarse CPE was performed on a per-core basis using the Viterbi–Viterbi algorithm [7]. The reason for using this algorithm was to save processing time and the fact that it sufficed for coarse CPE. By subtracting the phase-noise estimates associated with a specific core from the estimates associated with the other two cores, the relative phases between the cores were obtained. Then, through linear least-squares fitting of the relative phases, the relative CFOs were estimated and mitigated using the resulting

TABLE II
GAINS IN TOTAL TRANSMISSION REACH FOR DIFFERENT MODULATION FORMATS, PRE-FEC BER THRESHOLDS, AND 1% PILOT OH

BER	DP-QPSK	DP-16QAM	DP-64QAM
$5.16 \cdot 10^{-3}$	5.5 km (0.1%)	13.4 km (1.1%)	3.9 km (21.0%)
$7.04 \cdot 10^{-3}$	5.4 km (0.1%)	11.8 km (0.8%)	3.4 km (8.5%)
$9.29 \cdot 10^{-3}$	5.4 km (0.1%)	10.3 km (0.7%)	6.5 km (10.0%)
$1.44 \cdot 10^{-2}$	4.7 km (0.1%)	14.7 km (0.8%)	9.6 km (7.1%)
$1.71 \cdot 10^{-2}$	4.7 km (0.1%)	14.0 km (0.7%)	11.9 km (6.8%)
$2.24 \cdot 10^{-2}$	4.7 km (0.1%)	17.9 km (0.8%)	15.5 km (5.5%)

fit. Moreover, a second orthonormalization step to correct for transmitter I/Q imbalances, which cannot be compensated before the carrier recovery stage due to the presence of CFOs and phase noise, was carried out before symbol detection. Finally, each BER estimate was computed by counting bit errors out of at least 10^7 bits.

FGK uses the covariance matrix $\mathbf{Q}$ in (6), and hence, depends on σ_L^2, σ_C^2, and σ_P^2. These parameters were tweaked in order to minimize the resulting BER estimates. For DP-QPSK and DP-16QAM, different values of σ_L^2 and σ_C^2 in the ranges $[50, 300]$ kHz and $[\sigma_L^2/10^4, \sigma_L^2]$, resp., were found to minimize the BER. As the phase noise was highly correlated in the two polarizations in each core, $\sigma_P^2 = 10^6$ in all cases. Possible reasons for the wide ranges of optimal values for σ_L^2 and σ_C^2 are different amounts of residual CFOs, inaccurate estimates of the complex additive noise variance, and the presence of nonlinear phase noise. However, for DP-64QAM, $\sigma_L^2 = 40$ kHz, $\sigma_C^2 = \sigma_L^2/10^4$ and $\sigma_P^2 = \sigma_L^2/10^6$ were overall the best out the tested values.

Figs. 9 (a)–(c) show the estimated BER as a function of transmission distance for PC-CPE and JC-CPE, evaluated at

$OH_P = 1\%$ and $OH_P = 0.2\%$, with a target pre-FEC BER threshold of $1.44 \cdot 10^{-2}$. Moreover, Figs. 9 (d)–(f) show the same results, zoomed in around the pre-FEC threshold. As can be seen, the performance difference between JC-CPE and PC-CPE grows with increasing modulation order and decreasing OH_P. The gains in transmission reach, marked with horizontal arrows in the plots, are determined in terms of absolute distance and percentage-wise gain. The largest percentage-wise gain of 30.0% corresponds to the case of DP-64QAM and $OH_P = 0.2\%$. Furthermore, Table II details the gains in transmission reach for $OH_P = 1\%$ and different pre-FEC BER thresholds, corresponding to selected staircase codes from [33]. For all the transmitted modulation formats, the percentage-wise gain in transmission reach is larger for lower pre-FEC BER thresholds. Overall, the gains are negligible for DP-QPSK for DP-16QAM, but in the case of DP-64QAM, between 5.5% and 21% increased transmission reach is observed.

It is worth noting that these results depend on the signal quality following the DSP stages prior to the CPE. However, as mentioned above, all relative CFOs were minimized before the CPE was carried out, but skipping this step and letting FGK track the residual CFOs yields an insignificant performance penalty. For example, in the case of DP-64QAM transmission using $OH_P = 1\%$, the estimated BER increases at most by 0.4% (from $8.2 \cdot 10^{-3}$ to $8.23 \cdot 10^{-3}$).

The system model in (1), (5), and (6), which was used to develop FGK, does not capture transmitter I/Q imbalances, and the random-walk assumptions for the different phase-noise components may not be accurate if, e.g., phase differences between the spatial channels are caused by residual CFOs or nonlinearities. However, the transmission reach improvements suggest that the model is able to sufficiently describe the processed signal after all stages in the DSP chain prior to the CPE. Moreover, Figs. 10 (a) and (b) show the estimated BER as a function of transmission distance for DP-64QAM based on the experiments and simulations, resp. For the simulations, the AWGN variance was based on the estimated complex additive noise variance of the experimental data. Furthermore, $\sigma_L^2 = 2 \cdot 10^{-6}$, which corresponds to $\Delta\nu = 40$ kHz at 20 GBd, $\sigma_C^2 = \sigma_L^2/10^4$ and $\sigma_P^2 = \sigma_L^2/10^6$. A strong agreement is observed between the simulations and the experimental results. Considering the same pre-FEC BER thresholds as in Table II, the simulated transmission-reach gains in Fig. 10 (b) range from 9.5% to 24.3% for $OH_P = 1\%$, which is on the same order as the experimental gains.

VI. CONCLUSION

The performance of JC-CPE for SDM transmission via MCFs in the presence of correlated phase noise was studied. To that end, a phase-noise model was introduced that describes DP transmission through an MCF, where all cores share lasers on the transmitter and receiver sides, giving rise to a common LPN in addition to core- and polarization-specific phase drifts, which decorrelate the phase noise across the spatial channels. It was further shown that this model can be regarded as a multidimensional random walk, and an algorithm developed
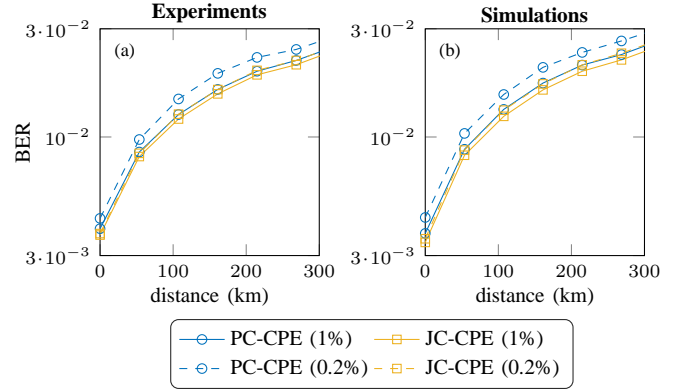

Fig. 10. Comparison of (a) experimental and (b) Monte Carlo simulation results pertaining to BER versus transmission distance for DP-64QAM using $OH_P = 1\%$ and $OH_P = 0.2\%$.

for this model was used to compare the performance of PC-CPE and JC-CPE for MCF transmission of various modulation formats through different numbers of cores, pilot OHs, laser linewidths, and degrees of phase-noise correlation. Through Monte Carlo simulations of 20 GBd transmission with 200 kHz laser linewidth and 1% pilot OH, JC-CPE led to a higher phase-noise tolerance. This could be exploited to improve power efficiency by up to 3.4 dB or increase information rate by up to 25.5%. Furthermore, JC-CPE can relax laser-linewidth requirements by up to a factor equal to the number of spatial channels. Finally, using data from MCF transmission experiments, the system model was validated, and PC-CPE and JC-CPE were compared in terms of transmission reach. Up to 21% increase in transmission distance was observed for 1% pilot OH through the use of JC-CPE.

REFERENCES

[1] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nat. Photon.*, vol. 7, no. 5, pp. 354–362, Apr. 2013.

[2] M. D. Feuer, L. E. Nelson, X. Zhou, S. L. Woodward, R. Isaac, B. Zhu, T. F. Taunay, M. Fishteyn, J. M. Fini, and M. F. Yan, "Joint digital signal processing receivers for spatial superchannels," *IEEE Photon. Technol. Lett.*, vol. 24, no. 21, pp. 1957–1960, Nov. 2012.

[3] B. J. Puttnam, R. S. Luís, J.-M. Delgado Mendinueta, J. Sakaguchi, W. Klaus, Y. Awaji, N. Wada, A. Kanno, and T. Kawanishi, "Long distance transmission in a multi-core fiber with self-homodyne detection," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2015, p. Th1D.5.

[4] M. D. Feuer, L. E. Nelson, K. Abedin, X. Zhou, T. F. Taunay, J. F. Fini, B. Zhu, R. Isaac, R. Harel, G. Cohen, and D. M. Marom, "ROADM system for space division multiplexing with spatial superchannels," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2013, p. PDP5B.8.

[5] M. Karlsson and E. Agrell, "Multidimensional modulation and coding in optical transport," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 876–884, Feb. 2017.

[6] P. Poggiolini, A. Nespola, Y. Jiang, G. Bosco, A. Carena, L. Bertignono, S. M. Bilal, S. Abrate, and F. Forghieri, "Analytical and experimental results on system maximum reach increase through symbol rate optimization," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1872–1885, Apr. 2016.

[7] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, Jul. 1983.

[8] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations," *J. Lightw. Technol.*, vol. 27, no. 8, pp. 989–999, Apr. 2009.

[9] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2675–2692, Sep. 2007.

[10] H. Cheng, Y. Li, F. Zhang, J. Wu, J. Lu, G. Zhang, J. Xu, and J. Lin, "Pilot-symbols-aided cycle slip mitigation for DP-16QAM optical communication systems," *Opt. Express*, vol. 21, no. 19, pp. 22 166–22 172, Sep. 2013.

[11] M. Morsy-Osman, Q. Zhuge, M. Chagnon, X. Xu, and D. V. Plant, "Experimental demonstration of pilot-aided polarization recovery, frequency offset and phase noise mitigation," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2013, p. OTu3I.6.

[12] M. Pajovic, D. S. Millar, T. Koike-Akino, R. Maher, D. Lavery, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, "Experimental demonstration of multi-pilot aided carrier phase estimation for DP-64QAM and DP-256QAM," in *Proc. European Conf. Opt. Commun.*, Sep. 2015, p. Mo.4.3.3.

[13] C. Zhu and N. Kaneda, "Discrete cosine transform based pilot-aided phase noise estimation for high-order QAM coherent optical systems," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2017, p. Th4C.1.

[14] S. T. Le, T. Kanesan, M. E. McCarthy, E. Giacoumidis, I. D. Phillips, M. F. C. Stephens, M. Tan, N. J. Doran, A. D. Ellis, and S. K. Turitsyn, "Experimental demonstration of data-dependent pilot-aided phase noise estimation for CO-OFDM," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2014, p. Tu3G.4.

[15] R. G. H. van Uden, C. M. Okonkwo, V. A. J. M. Sleiffer, M. Kuschnerov, H. de Waardt, and A. M. J. Koonen, "Single DPLL joint carrier phase compensation for few-mode fiber transmission," *IEEE Photon. Technol. Lett.*, vol. 25, no. 14, pp. 1381–1384, Jul. 2013.

[16] L. Lundberg, M. Mazur, A. Lorences-Riesgo, M. Karlsson, and P. A. Andrekson, "Joint carrier recovery for DSP complexity reduction in frequency comb-based superchannel transceivers," in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. Th.1.D.3.

[17] D. V. Souto, B.-E. Olsson, C. Larsson, and D. A. A. Mello, "Joint-polarization and joint-subchannel carrier phase estimation for 16-QAM optical systems," *J. Lightw. Technol.*, vol. 30, no. 20, pp. 3185–3191, Oct. 2012.

[18] R. S. Luís, B. J. Puttnam, J.-M. Delgado Mendinueta, W. Klaus, Y. Awaji, and N. Wada, "Comparing inter-core skew fluctuations in multi-core and single-core fibers," in *Proc. Conf. Lasers and Electro-Opt.*, May 2015, p. SM2L.5.

[19] M. Mazur, A. Lorences-Riesgo, J. Schröder, P. A. Andrekson, and M. Karlsson, "High spectral efficiency PM-128QAM comb-based superchannel transmission enabled by a single shared optical pilot tone," *J. Lightw. Technol.*, vol. 36, no. 6, pp. 1318–1325, Mar. 2018.

[20] E. Agrell, A. F. Alfredsson, B. J. Puttnam, R. S. Luís, G. Rademacher, and M. Karlsson, "Modulation and detection for multicore superchannels with correlated phase noise," in *Proc. Conf. Lasers and Electro-Opt.*, May 2018, p. SM4C.3.

[21] B. J. Puttnam, R. S. Luís, G. Rademacher, A. Alfredsson, W. Klaus, J. Sakaguchi, Y. Awaji, E. Agrell, and N. Wada, "Characteristics of homogeneous multi-core fibers for SDM transmission," *APL Photon. J.*, vol. 4, no. 2, p. 022804, Feb. 2019.

[22] T. C. W. Schenk, X.-J. Tao, P. F. M. Smulders, and E. R. Fledderus, "Influence and suppression of phase noise in multi-antenna OFDM," in *Proc. Vehicular Technol. Conf.*, vol. 2, Sep. 2004, pp. 1443–1447.

[23] R. Krishnan, G. Colavolpe, A. Graell i Amat, and T. Eriksson, "Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3360–3375, Jul. 2015.

[24] A. F. Alfredsson, E. Agrell, and H. Wymeersch, "Iterative decoding and phase-noise compensation for multichannel optical transmission," *arXiv e-prints*, Dec. 2018. [Online]. Available: arxiv.org/abs/1804.02263

[25] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, "Phase-noise compensation for spatial-division multiplexed transmission," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2017, p. Th4C.7.

[26] A. F. Alfredsson, E. Agrell, H. Wymeersch, B. J. Puttnam, and R. S. Luís, "Joint phase tracking for multicore transmission with correlated phase noise," in *Proc. IEEE Summer Topicals Meeting Series*, Jul. 2018, p. MF1.2.

[27] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3042–3049, Aug. 2009.

[28] J. G. Proakis, *Digital communications*, 4th ed. Boston, MA, USA: McGraw-Hill, 2000.

[29] S. Särkkä, *Bayesian Filtering and Smoothing*, 1st ed. Cambridge, UK: Cambridge University Press, 2013.

[30] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, "Pilot distributions for phase tracking in space-division multiplexed systems," in *Proc. European Conf. Opt. Commun.*, Sep. 2017, p. P1.SC3.48.

[31] C. Heegard and S. B. Wicker, *Turbo Coding*, 1st ed. Kluwer Academic Publishers, 1999.

[32] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

[33] L. M. Zhang and F. R. Kschischang, "Staircase codes with 6% to 33% overhead," *J. Lightw. Technol.*, vol. 32, no. 10, pp. 1999–2002, May 2014.

[34] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, Oct. 2015.

[35] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Technical J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[36] R. S. Luís, B. J. Puttnam, G. Rademacher, Y. Awaji, and N. Wada, "On the use of high-order MIMO for long-distance homogeneous single-mode multicore fiber transmission," in *Proc. Opt. Fiber Commun. Conf.*, Sep. 2017, p. Th.2.F.2.

[37] E. P. da Silva and D. Zibar, "Widely linear equalization for IQ imbalance and skew compensation in optical coherent receivers," *J. Lightw. Technol.*, vol. 34, no. 15, pp. 3577–3586, Aug. 2016.