Digital Signal Processing for MDG Estimation in Long-Haul SDM Transmission

Digital Signal Processing for MDG Estimation in Long-Haul SDM Transmission

Ruby S. B. Ospina, *Student Member, OSA*, Darli A. A. Mello, *Member, IEEE*, Lucas Zischler, *Student Member, OSA*, Ruben S. Luís, *Senior Member, IEEE*, Benjamin J. Puttnam, *Senior Member, IEEE*, Hideaki Furukawa, *Senior Member, IEEE*, Menno van den Hout, *Student Member, IEEE*, Sjoerd van der Heide, *Student Member, IEEE*, Chigo Okonkwo, *Senior Member, IEEE*, Roland Ryf, *Senior Member, IEEE*, and Georg Rademacher, *Senior Member, IEEE*

(Invited Paper)

***Abstract*—In space-division multiplexing (SDM) transmission, mode-dependent gain (MDG) can be estimated by digital signal processing (DSP) from the transfer matrix of the multiple-input multiple-output (MIMO) equalizer. However, previous works have shown that the estimation performance using MIMO equalizers based on the minimum mean square error (MMSE) is significantly degraded, particularly at low SNRs. In this paper, we experimentally assess DSP-based MDG estimation techniques in long-haul SDM transmission. Besides the intrinsic MDG of the optimized setup configuration, an additional power imbalance between modes is artificially introduced to generate a controlled amount of MDG. Two approaches for MDG estimation employing MMSE MIMO equalization are discussed. The direct method readily computes the channel eigenvalues from the MIMO equalizer coefficients. Conversely, the reverse method requires equalizer training dedicated for MDG estimation, which is suitable for laboratory experiments but, eventually, overly computationally expensive for real-time DSP. Moreover, an ANN model is evaluated as an ML-based technique for MDG estimation. The results corroborate previous works showing that the direct method saturates the estimated MDG for high values of accumulated MDG. The reverse method offers accurate results for an increased range of MDG values, confirming its suitability for experimental works. Finally, the ANN-based method tends to overestimate the MDG for high accumulated values but has promising prospects for real-time operation considering its low complexity.**

***Index Terms*—Digital signal processing, mode-dependent gain, mode-dependent loss, space division multiplexing.**

Manuscript received 13 July 2023; revised 27 October 2023; accepted 9 November 2023. Date of publication 16 November 2023; date of current version 2 February 2024. This work was supported in part by FAPESP, under Grants 2017/25537-8, 2015/24341-7, and 2022/07488-8, and in part by the National Institute of Information and Communications Technology (NICT). *(Corresponding author: Darli A. A. Mello.)*

Ruby S. B. Ospina, Darli A. A. Mello, and Lucas Zischler are with the School of Electrical and Computer Engineering, State University of Campinas, Campinas 13083-970, Brazil (e-mail: r163653@dac.unicamp.br; darli@unicamp.br; l257176@dac.unicamp.br).

Ruben S. Luís, Benjamin J. Puttnam, and Hideaki Furukawa are with the National Institute of Information and Communications Technology (NICT), Tokyo 184-8795, Japan (e-mail: rluis@nict.go.jp; ben@nict.go.jp; furukawa@nict.go.jp).

Menno van den Hout, Sjoerd van der Heide, and Chigo Okonkwo are with the High Capacity Optical Transmission Laboratory, Electro-Optical Communications Group, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: m.v.d.hout@tue.nl; s.p.v.d.heide@tue.nl; c.m.okonkwo@tue.nl).

Roland Ryf is with the Nokia Bell Labs, Holmdel, NJ 07733 USA (e-mail: roland.ryf@nokia-bell-labs.com).

Georg Rademacher is with the University of Stuttgart, 70174 Stuttgart, Germany (e-mail: georg.rademacher@int.uni-stuttgart.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2023.3333668.

Digital Object Identifier 10.1109/JLT.2023.3333668

I. Introduction

TO SUSTAIN the exponential growth of data traffic, the infrastructure of transport optical networks must implement new disruptive technologies. Space-division multiplexing (SDM) is being studied as a promising solution to further scale the capacity of optical networks [1]. For SDM to be adopted for the next generation of optical networks, the technology should provide capacity increase while preserving the low cost of transmission per bit through device sharing and integration at the component, fiber, and system levels.

In coupled SDM transmission over multi-mode fibres (MMFs) [2], few-mode fibres (FMFs) [3], [4], or coupled-core multi-core fibers (CC-MCFs) [5], linear crosstalk and modal dispersion do not represent fundamental limitations for the system capacity as they can be partially compensated for by digital signal processing (DSP) employing multiple-input multiple-output (MIMO) equalizers. In contrast, in amplified SDM transmission, propagation modes may be subject to imbalanced amplification, an effect known as mode-dependent gain (MDG), reducing the average channel capacity and increasing the system outage probability [6], [7], [8]. The unequal amplification over multiple amplifiers can generate a large accumulated MDG, introducing a prohibitive loss in the channel capacity [9], [10], [11]. As the compensation of MDG by receiver DSP is not trivial [12], its effect must be handled at the component design level or, alternatively, during transmission [13]. Therefore, the MDG accumulated over a transmission link should be accurately estimated for system monitoring, performance evaluation, and troubleshooting.

MDG can be computed directly from the overall channel transfer matrix [14]. In the classical theory of MIMO communications, channel estimation is usually accomplished by methods

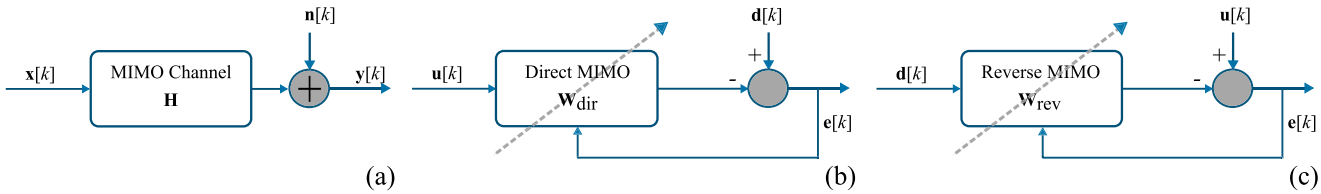

Fig. 1. (a) Linear MIMO transmission. (b) and (c) MIMO equalizer operated in direct mode and reverse mode, respectively. In the direct mode, the input vector, $\mathbf{u}[k]$, is injected into the MIMO equalizer. After filtering, the reference signal, $\mathbf{d}[k]$, is employed for error signal computation and filter updating. In the reverse mode, the MIMO equalizer is fed at its input with the reference signal vector, $\mathbf{d}[k]$, and the signal, $\mathbf{u}[k]$, is employed as a training sequence for error calculation and filter updating.

based on the least-squares (LS) and minimum mean square error (MMSE) [15], [16]. Employing these classic channel estimation algorithms, designed for linear channels, directly in SDM optical transmission systems is not straightforward due to the severe conditions of phase noise and frequency drift generated in semiconductor lasers typically used in fiber optic communications [17]. Alternatively, in optical SDM, channel estimation is usually performed heuristically through a MIMO equalizer dynamically adapted based on the MMSE criterion, which, at typical phase noise levels, is operated in a loop with phase recovery algorithms to prevent the impact of phase noise on the equalizer performance [18].

It was shown in [19], [20] that, due to noise, the MDG computed directly from the transfer matrix of the MMSE MIMO equalizer is significantly underestimated at low signal-to-noise ratio (SNR) and high MDG. To alleviate this problem, an analytical correction factor that significantly improves the MDG estimates was proposed in [20], [21]. As a drawback, the method requires knowledge of the optical SNR at the input of the receiver front-end. In [22], a method based on artificial neural networks (ANNs) for MDG estimation was also proposed. The ANN-based estimation technique was validated in short-reach SDM transmission setups with suitable performance. In [23] and [24], an alternative approach for MDG estimation employs the MMSE MIMO equalizer operated in reverse mode, in which the roles of the training sequence and received signals are exchanged. Notwithstanding, the operating fundamentals of such technique applied to SDM transmission have not been discussed.

This paper studies DSP-based MDG estimation in long-haul 3-mode transmission with controlled MDG. Firstly, we model the MIMO transmission in the linear regime to discuss MDG estimation based on the MMSE MIMO equalizer operated in direct and reverse modes. Moreover, the fundamentals of the ANN-based MDG estimator are presented. Later, we assess the estimation performance of these three DSP-based MDG estimation methods in a simulated long-haul 3-mode transmission setup and an experimental long-haul 3-mode transmission setup with controlled MDG. We do not investigate the correction-factor-based method as the optical SNR is not available in the experimental setup.

The remainder of this paper is divided as follows. Section II describes the mathematical modeling of a linear MIMO transmission channel. Section III reviews the DSP-based techniques employed for MDG estimation. This section covers the fundamentals of dynamic MIMO equalization based on the MMSE criterion, deriving the equalizer stochastic steepest-descent cost function for both MIMO equalizers operated in direct and reverse mode. In addition, this section presents the fundamentals of the ANN-based MDG estimator. Section IV presents the MDG estimated in a simulated 3-mode transmission setup by the three investigated methods: direct MIMO equalization, reverse MIMO equalization and ANN. Section V studies the MDG estimated in an experimental long-haul 3-mode transmission setup with controlled MDG based on MMSE equalization and ANN. Lastly, Section VI concludes the paper.

II. Mathematical Modeling of a Linear MIMO Transmission

Assuming linear operation and additive white Gaussian noise (AWGN), a MIMO channel model is depicted in Fig. 1(a). Considering a narrowband channel (i.e. with flat channel response within the channel bandwidth), the model can be mathematically expressed as [25]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{1}$$

At the k^{th} instant, the zero mean unit power transmitted signal vector, $\mathbf{x}[k]$, and received signal vector, $\mathbf{y}[k]$, consist of $2\mathrm{M_p}$ samples as

$$\mathbf{x}[k] = \left[x_1[k]\ \ x_2[k]\ \ x_3[k]\ \ \ldots\ \ x_{2\mathrm{M_p}}[k]\right]^T, \tag{2}$$

$$\mathbf{y}[k] = \left[y_1[k]\ \ y_2[k]\ \ y_3[k]\ \ \ldots\ \ y_{2\mathrm{M_p}}[k]\right]^T, \tag{3}$$

where $\mathrm{M_p}$ accounts for the supported spatial modes, each supporting two polarization modes. The narrowband channel transfer matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = \begin{bmatrix} h_{1,1}[k] & h_{1,2}[k] & h_{1,3}[k] & \cdots & h_{1,2\mathrm{M_p}}[k] \\ h_{2,1}[k] & h_{2,2}[k] & h_{2,3}[k] & \cdots & h_{2,2\mathrm{M_p}}[k] \\ h_{3,1}[k] & h_{3,2}[k] & h_{3,3}[k] & \cdots & h_{3,2\mathrm{M_p}}[k] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{2\mathrm{M_p},1}[k] & h_{2\mathrm{M_p},2}[k] & h_{2\mathrm{M_p},3}[k] & \cdots & h_{2\mathrm{M_p},2\mathrm{M_p}}[k] \end{bmatrix} \tag{4}$$

The elements $h_{i,j}[k]$ of $\mathbf{H}$ denote the impulse response coefficients from transmitter j to receiver i at the k^{th} instant. The AWGN signal vector, $\mathbf{n}[k]$, consists of $2\mathrm{M_p}$ noisy samples that

are added to the received signal as

$$\mathbf{n}[k] = \left[n_1[k]\ \ n_2[k]\ \ n_3[k]\ \ \ldots\ \ n_{2\mathrm{M_p}}[k]\right]^T. \tag{5}$$

In the scope of coherent optical SDM transmission systems, the received signal vector, $\mathbf{y}[k]$, is detected by the coherent receiver front-end for optical to electrical conversion. The DSP module digitizes and processes the electrical signal matrix. After pre-processing stages (deskew, timing recovery, static equalization, and frequency offset compensation), whose description is omitted here for clarity, the signal is passed to the MIMO equalizer for dynamic equalization and MDG estimation.

III. DSP-Based MDG Estimation

The MDG introduced by an SDM channel with transfer matrix $\mathbf{H}$ can be quantified from the eigenvalues λ_i^2 of the operator $\mathbf{H}\mathbf{H}^H$, using the eigenvalues standard deviation $\sigma_{\mathrm{mdg}} = \mathrm{std}(\log(\lambda_i^2))$ or the subtraction of the largest and the lowest eigenvalue in dB $\mathrm{MDG_{peak-peak}} = 10\log_{10}(\lambda_i^{\max})^2 - 10\log_{10}(\lambda_i^{\min})^2$ [14]. Although this paper is focused on the standard deviation metric, the investigated methods also apply to the peak-to-peak metric. In practical transmission systems, $\mathbf{H}$ is unknown, but can be estimated from the received data. In general, the MDG is estimated from the coefficients of the MIMO adaptive equalizer after convergence. In this paper, we evaluate experimentally and by simulations the three techniques presented below.

A. MDG Estimation Based on the MMSE MIMO Equalizer Operated in Direct Mode

The block diagram of the MIMO equalizer operated in direct mode is shown in Fig. 1(b). The equalizer is fed with the pre-processed signal vector, $\mathbf{u}[k]$, given by

$$\mathbf{u}[k] = \left[u_1[k]\ \ u_2[k]\ \ u_3[k]\ \ \ldots\ \ u_{2\mathrm{M_p}}[k]\right]^T. \tag{6}$$

Filtering is carried out by the direct equalizer matrix $\mathbf{W}_{\mathrm{dir}}$ $(2\mathrm{M_p} \times 2\mathrm{M_p})$. Assuming supervised training, an error signal vector, $\mathbf{e}[k]$, is computed as the difference between the reference signal vector, $\mathbf{d}[k]$, and the filtered signal $\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k]$. The cost function to be minimized by the MIMO equalizer operated in direct mode, $J(\mathbf{W}_{\mathrm{dir}})$, assuming the MMSE criterion, is given by

$$\begin{aligned}
J(\mathbf{W}_{\mathrm{dir}}) &= \mathbb{E}\left\{\mathbf{e}[k]^H\mathbf{e}[k]\right\}\\
&= \mathbb{E}\left\{(\mathbf{d}[k]-\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k])^H(\mathbf{d}[k]-\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k])\right\}\\
&= \mathbb{E}\left\{(\mathbf{d}^H[k]-\mathbf{u}^H[k]\mathbf{W}_{\mathrm{dir}}^H)(\mathbf{d}[k]-\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k])\right\}\\
&= \mathbb{E}\left\{\mathbf{d}^H[k]\mathbf{d}[k]-\mathbf{u}^H[k]\mathbf{W}_{\mathrm{dir}}^H\mathbf{d}[k]\right.\\
&\quad\left.-\mathbf{d}^H[k]\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k]+\mathbf{u}^H[k]\mathbf{W}_{\mathrm{dir}}^H\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k]\right\}\\
&= 2\mathrm{M_p}\sigma_d^2-\mathbb{E}\left\{\mathbf{u}^H[k]\mathbf{W}_{\mathrm{dir}}^H\mathbf{d}[k]\right\}\\
&\quad-\mathbb{E}\left\{\mathbf{d}^H[k]\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k]\right\}\\
&\quad+\mathbb{E}\left\{\mathbf{u}^H[k]\mathbf{W}_{\mathrm{dir}}^H\mathbf{W}_{\mathrm{dir}}\mathbf{u}[k]\right\},
\end{aligned} \tag{7}$$

where $\mathbb{E}$ is the expectation operation, $(.)^H$ denotes Hermitian transpose operator, and σ_d^2 is the variance of the reference signal.

Employing the properties of the derivative of complex matrices[1], and commonly defined derivatives[2], the global minimum of the paraboloid $J(\mathbf{W}_{\mathrm{dir}})$ can be obtained through

$$\frac{\partial J(\mathbf{W}_{\mathrm{dir}})}{\partial \mathbf{W}_{\mathrm{dir}}} = -2\mathbf{P}_{du} + 2\mathbf{W}_{\mathrm{dir}}\mathbf{P}_{uu} = \mathbf{0}, \tag{8}$$

where $\mathbf{P}_{du}$ $(2\mathrm{M_p} \times 2\mathrm{M_p})$ is the cross-correlation matrix between the reference signal and the equalizer input, and $\mathbf{P}_{uu}$ $(2\mathrm{M_p} \times 2\mathrm{M_p})$ is the correlation matrix of the equalizer inputs. Its main diagonal terms are the auto-correlation function of the input modes, and the off-diagonal terms are the cross-correlation between different input modes. From (8), the optimal solution is given by the classic Wiener filter

$$\mathbf{W}_{\mathrm{dir}} = \mathbf{P}_{du}\mathbf{P}_{uu}^{-1}. \tag{9}$$

The correlation matrix, $\mathbf{P}_{uu}$, is computed as

$$\begin{aligned}
\mathbf{P}_{uu} &= \mathbb{E}\left\{\mathbf{u}[k]\mathbf{u}^H[k]\right\}\\
&= \mathbb{E}\left\{(\mathbf{H}\mathbf{x}[k]+\mathbf{n}[k])(\mathbf{H}\mathbf{x}[k]+\mathbf{n}[k])^H\right\}\\
&= \mathbb{E}\left\{\mathbf{H}\mathbf{x}[k]\mathbf{x}^H[k]\mathbf{H}^H+\mathbf{H}\mathbf{x}[k]\mathbf{n}^H[k]+\mathbf{n}[k]\mathbf{x}^H[k]\mathbf{H}^H\right.\\
&\quad\left.+\mathbf{n}[k]\mathbf{n}^H[k]\right\}.
\end{aligned} \tag{10}$$

For zero mean unit power transmitted signals, $\mathbf{P}_{uu}$ is redefined as

$$\mathbf{P}_{uu} = \mathbf{H}\mathbf{H}^H + \sigma_N^2\mathbf{I}, \tag{11}$$

where $\mathbf{I}$ is the identity matrix and σ_N^2 is the noise variance.

The cross-correlation matrix, $\mathbf{P}_{du}$, is given by

$$\mathbf{P}_{du} = \mathbb{E}\left\{\mathbf{d}[k]\mathbf{u}[k]^H\right\}. \tag{12}$$

Assuming that the equalizer input, $\mathbf{u}[k]$, keeps the statistics of the received vector, $\mathbf{y}[k]$, the matrix $\mathbf{P}_{du}$ can be expressed as

$$\begin{aligned}
\mathbf{P}_{du} &= \mathbb{E}\left\{\mathbf{d}[k]\mathbf{u}^H[k]\right\}\\
&= \mathbb{E}\left\{\mathbf{d}[k](\mathbf{H}\mathbf{x}[k]+\mathbf{n}[k])^H\right\}\\
&= \mathbb{E}\left\{\mathbf{d}[k]\mathbf{x}^H[k]\mathbf{H}^H+\mathbf{d}[k]\mathbf{n}^H[k]\right\}.
\end{aligned} \tag{13}$$

In a MIMO equalizer operated in direct mode and assuming a fully supervised equalization, the reference training vector, $\mathbf{d}[k]$, corresponds to the transmitted signal at the k^{th} instant, $\mathbf{x}[k]$. For zero mean unit power reference signal, $\mathbf{P}_{du}$ is finally expressed as

$$\mathbf{P}_{du} = \mathbf{H}^H. \tag{14}$$

Replacing (11) and (14) in (9), the optimal $\mathbf{W}_{\mathrm{dir}}$ is given by

$$\mathbf{W}_{\mathrm{dir}} = \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \sigma_N^2\mathbf{I}\right)^{-1}. \tag{15}$$

For a zero mean unit power AWGN, $\mathbf{W}_{\mathrm{dir}}$ can be redefined as

$$\mathbf{W}_{\mathrm{dir}} = \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \frac{\mathbf{I}}{\mathrm{SNR}}\right)^{-1}. \tag{16}$$

[1] Complex derivative: $\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \frac{1}{2}\left(\frac{\partial f(\mathbf{X})}{\partial \mathbb{R}\{\mathbf{X}\}} - j\frac{\partial f(\mathbf{X})}{\partial \mathbb{I}\{\mathbf{X}\}}\right)$

[2] $\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$, $\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$, $\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = 2\mathbf{X}\mathbf{a}\mathbf{a}^T$

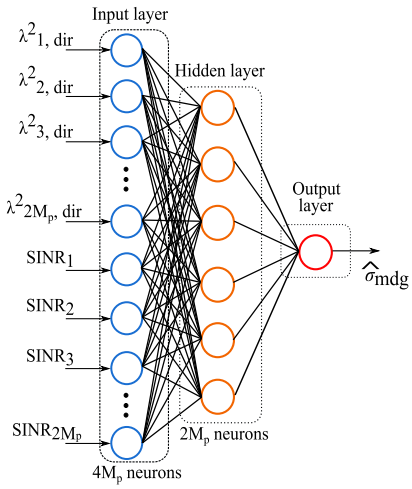

Fig. 2. ANN-based MDG estimator. The ANN is fed at its input with the vector of eigenvalues $\lambda^2_{i,\text{dir}}$ and the vector of SINRs. At the output, the ANN provides an estimate of σ_{mdg}.

Using the push-through identity[3], (16) can be rewritten as [26]

$$\mathbf{W}_{\text{dir}} = \left(\mathbf{H}^H\mathbf{H} + \frac{\mathbf{I}}{\text{SNR}}\right)^{-1}\mathbf{H}^H. \tag{17}$$

For a MMSE MIMO equalizer operated in direct mode, the standard deviation of the overall MDG can be estimated from the eigenvalues, $\lambda^2_{i,\text{dir}}$, of the operator $\mathbf{W}^{-1}_{\text{dir}}(\mathbf{W}^{-1}_{\text{dir}})^H$ as

$$\widehat{\sigma}_{\text{mdg,dir}} = \text{std}(\log(\lambda^2_{i,\text{dir}})). \tag{18}$$

The relationship between the actual eigenvalues, λ^2_i, and the eigenvalues estimated from the MIMO equalizer operated in direct mode, $\lambda^2_{i,\text{dir}}$, can be obtained from the eigendecomposition of $\mathbf{W}^{-1}_{\text{dir}}(\mathbf{W}^{-1}_{\text{dir}})^H$, as

$$\begin{aligned}\mathbf{W}^{-1}_{\text{dir}}(\mathbf{W}^{-1}_{\text{dir}})^H &= \frac{(\mathbf{H}\mathbf{H}^H)^{-1}}{\text{SNR}^2} + \frac{2\mathbf{I}}{\text{SNR}} + \mathbf{H}\mathbf{H}^H \\ &= \mathbf{Q}\left[\frac{\mathbf{\Lambda}_{\text{H}}{}^{-1}}{\text{SNR}^2} + \frac{2\mathbf{I}}{\text{SNR}} + \mathbf{\Lambda}_{\text{H}}\right]\mathbf{Q}^{-1},\end{aligned} \tag{19}$$

where $\mathbf{\Lambda}_{\text{H}}$ is a diagonal matrix whose main diagonal has elements λ^2_i, and $\mathbf{Q}$ is a matrix whose columns are the eigenvectors of $\mathbf{H}\mathbf{H}^H$. From (19), the actual eigenvalues, λ^2_i, and the eigenvalues obtained by DSP, $\lambda^2_{i,\text{dir}}$, are related by [20]

$$\lambda^2_{i,\text{dir}} = \left[\frac{(\lambda^2_i)^{-1}}{\text{SNR}^2} + \frac{2}{\text{SNR}} + \lambda^2_i\right]. \tag{20}$$

The presence of the SNR term in (17) and (20) shows the effect of noise on the MDG estimation technique based on the direct MIMO equalization scheme [20].

B. MDG Estimation Based on the MMSE MIMO Equalizer Operated in Reverse Mode

For a MIMO equalizer operating in reverse mode, as in Fig. 1(c), the received signal vector at the k^{th} instant, $\mathbf{u}[k]$, is used as the training sequence. The reference signal, $\mathbf{d}[k]$, is applied to the input of the MIMO equalizer. In this case, the error signal, $\mathbf{e}[k]$, is computed as the difference between $\mathbf{u}[k]$ and $\mathbf{W}_{\text{rev}}\ \mathbf{d}[k]$, where $\mathbf{W}_{\text{rev}}$ is the reverse MIMO transfer matrix ($2\text{M}_\text{p} \times 2\text{M}_\text{p}$).

For the MIMO equalizer operated in reverse mode, the cost function to be minimized, $J(\mathbf{W}_{\text{rev}})$, is given by

$$\begin{aligned}J(\mathbf{W}_{\text{rev}}) &= \mathbb{E}\left\{\mathbf{e}[k]^H\mathbf{e}[k]\right\} \\ &= \mathbb{E}\left\{(\mathbf{u}[k] - \mathbf{W}_{\text{rev}}\mathbf{d}[k])^H(\mathbf{u}[k] - \mathbf{W}_{\text{rev}}\mathbf{d}[k])\right\} \\ &= \mathbb{E}\left\{(\mathbf{u}^H[k] - \mathbf{d}^H[k]\mathbf{W}^H_{\text{rev}})(\mathbf{u}[k] - \mathbf{W}_{\text{rev}}\mathbf{d}[k])\right\} \\ &= \mathbb{E}\left\{\mathbf{u}^H[k]\mathbf{u}[k] - \mathbf{d}^H[k]\mathbf{W}^H_{\text{dir}}\mathbf{u}[k]\right. \\ &\quad \left. - \mathbf{u}^H[k]\mathbf{W}_{\text{rev}}\mathbf{d}[k] + \mathbf{d}^H[k]\mathbf{W}^H_{\text{rev}}\mathbf{W}_{\text{rev}}\mathbf{d}[k]\right\} \\ &= 2\text{M}_\text{p}\sigma^2_u - \mathbb{E}\left\{\mathbf{d}^H[k]\mathbf{W}^H_{\text{rev}}\mathbf{u}[k]\right\} \\ &\quad - \mathbb{E}\left\{\mathbf{u}^H[k]\mathbf{W}_{\text{rev}}\mathbf{d}[k]\right\} \\ &\quad + \mathbb{E}\left\{\mathbf{d}^H[k]\mathbf{W}^H_{\text{rev}}\mathbf{W}_{\text{rev}}\mathbf{d}[k]\right\},\end{aligned} \tag{21}$$

where σ^2_u is the variance of one input signal assuming the same statistical characteristics for all input modes.

The global minimum of the paraboloid cost function, $J(\mathbf{W}_{\text{rev}})$, is obtained through

$$\frac{\partial J(\mathbf{W}_{\text{rev}})}{\partial \mathbf{W}_{\text{rev}}} = -2\mathbf{P}_{ud} + 2\mathbf{W}_{\text{rev}}\mathbf{P}_{dd} = \mathbf{0}, \tag{22}$$

where the main diagonal of the correlation matrix $\mathbf{P}_{dd}$ ($2\text{M}_\text{p} \times 2\text{M}_\text{p}$) accounts for the auto-correlation between reference signals for the same mode, and the off-diagonal terms are the cross-correlation between reference signals for different modes. From (22), the optimal filter can be expressed as

$$\mathbf{W}_{\text{rev}} = \mathbf{P}_{ud}\mathbf{P}^{-1}_{dd}, \tag{23}$$

where $\mathbf{P}_{dd} = \mathbf{I}$ for a zero mean unit power signal $\mathbf{d}[k] = \mathbf{x}[k]$. The cross-correlation matrix between $\mathbf{u}[k]$ and $\mathbf{d}[k]$ is given by

$$\begin{aligned}\mathbf{P}_{ud} &= \mathbb{E}\left\{(\mathbf{H}\mathbf{x}[k] + \mathbf{n}[k])\mathbf{d}^H[k]\right\} \\ &= \mathbb{E}\left\{\mathbf{H}\mathbf{x}[k]\mathbf{d}^H[k] + \mathbf{n}[k]\mathbf{d}^H[k]\right\} \\ &= \mathbf{H}.\end{aligned} \tag{24}$$

Applying $\mathbf{P}_{dd} = \mathbf{I}$ and (24) in (23), $\mathbf{W}_{\text{rev}}$ is obtained as

$$\mathbf{W}_{\text{rev}} = \mathbf{H}. \tag{25}$$

Using a MMSE MIMO equalizer operated in reverse mode, the standard deviation of the overall MDG can be estimated from the eigenvalues, $\lambda^2_{i,\text{rev}}$, of the operator $\mathbf{W}_{\text{rev}}(\mathbf{W}_{\text{rev}})^H$ as

$$\widehat{\sigma}_{\text{mdg,rev}} = \text{std}(\log(\lambda^2_{i,\text{rev}})). \tag{26}$$

Unlike the MIMO equalizer operated in direct mode, the MIMO equalizer trained in reverse mode recovers the channel transfer matrix without involving the SNR in the cost function. Generally, very low SNR levels can cause convergence problems to the equalizer, affecting the MDG estimates. Still, we expect the reverse equalizer to yield a much lower error than that generated by the direct equalizer, at the cost of substantially

[3]Push-through identity: $\mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1} = (\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A}$

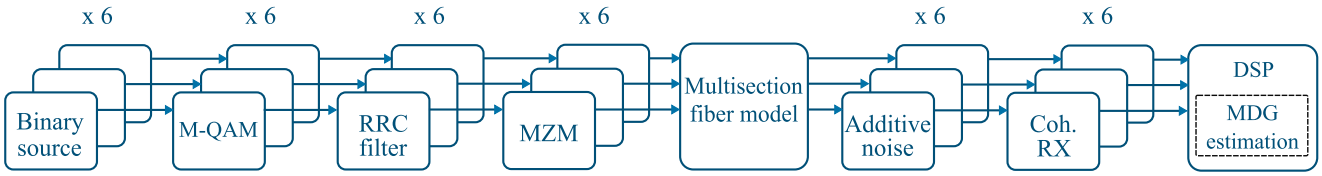

Fig. 3. Simulation setup of a coupled long-haul 3-mode transmission link. The transmitter generates sequences of 400 000 16-QAM symbols at 30 GBd. The multisection model simulates a 5000-km fiber link. The group delay standard deviation is set to 3.1 ps/$\sqrt{\text{km}}$. The MDG of the link is controlled by a per-amplifier MDG standard deviation, σ_g. Nonlinear transmission effects are not simulated. Direct and reverse 6×6 MIMO equalization is carried out by 36 finite impulse response filters with 800 taps each, updated by a fully supervised least mean squares (LMS) algorithm. MDG estimation is performed by DSP based on the MIMO MMSE transfer function.

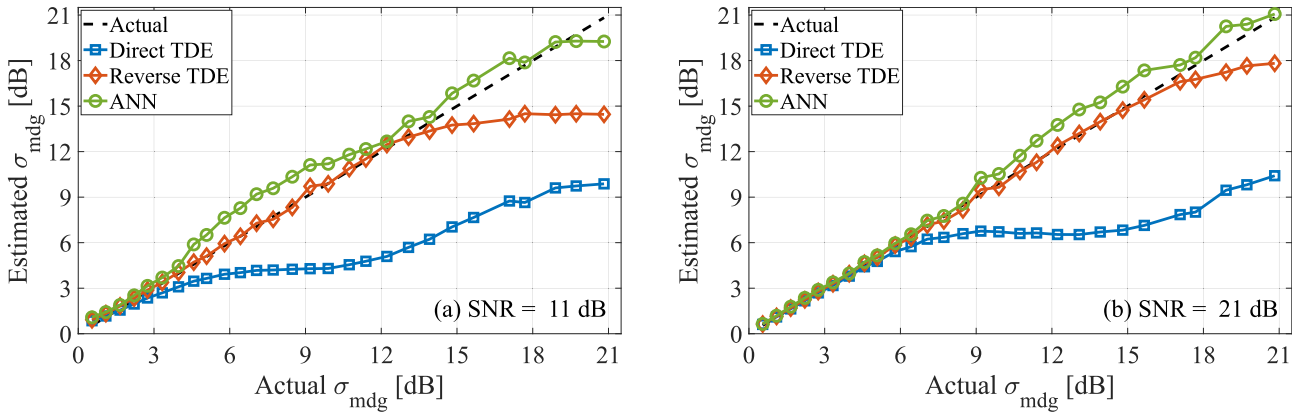

Fig. 4. σ_{mdg} estimated using different methods as a function of the actual σ_{mdg}, for (a) SNR = 11 dB, and (b) SNR = 21 dB. Each point of the curve results from averaging over five iterations for the same configuration of actual σ_{mdg}.

increase in complexity. Although this higher complexity may be tolerable for offline processing of experimental laboratory data, it may be undesirable for real-time transmission.

C. Artificial Neural Network-Based MDG Estimation

Currently, machine learning (ML) techniques are being considered for optical performance monitoring in mode-multiplexed systems [27]. In [28], we employ a multi-layer perceptron ANN to estimate the MDG and the SNR in an experimental 3-mode SDM transmission setup over a 73-km MMF. The ANN-based MDG estimator is depicted in Fig. 2. For a transmission setup with 2M_p propagation modes, the ANN receives 2M_p $\lambda^2_{i,\text{dir}}$ values computed from the transfer matrix of the MIMO equalizer operated in direct mode and 2M_p values of signal-to-interference-plus-noise ratio (SINR), SINR_i. The SINR corresponds to the SNR of the digital constellations estimated after MIMO equalization. From the input features, the ANN provides an estimate of σ_{mdg}.

The ANN is trained with semi-analytical data generated by numerical simulations. For generating the training data, the SDM channel is simulated using the multisection model for strongly-coupled SDM transmission presented in [14]. Later, analytic formulas are employed to generate the training samples comprising a set of features and the corresponding target as $[\lambda^2_{i,\text{dir}}\ \ \text{SINR}_i\,,\ \ \sigma_{\text{mdg}}]$ with dimension $4\text{M}_\text{p}+1$. The ANN is trained using the Adam optimizer [29] with batches of 5 samples. The number of hidden layers, number of neurons, batch size, and other hyper-parameters are optimized to maximize an accuracy metric. Once the ANN is trained with semi-analytic data, the resulting model can be applied to experimental setups for MDG estimation. More details on the ANN-based MDG estimator can be found in [28].

IV. MDG Estimation in a Simulated Long-Haul 3-Mode Transmission Setup

We simulate a coupled long-haul transmission link with $\text{M}_\text{p} = 3$ polarization-multiplexed spatial modes depicted in Fig. 3 .

At the transmitter, 2M_p sequences of 400,000 16-QAM symbols are generated at 30 GBd. The complex constellations are fed into 0.01 roll-off factor root-raised-cosine (RRC) shaping filters. The shaped signals are then sent to I/Q Mach-Zehnder modulator (MZM) models for electro-optical conversion. The 2M_p optical signals are then launched into the transmission fiber model with strong mode coupling. The fiber is modeled using the multisection scheme presented in [14] for 100 spans of 50 km for a total length of 5,000 km. The channel is represented by 1,000 frequency bins spread over the simulation bandwidth. The resolution of the channel in the frequency domain is adjusted by replicating channel matrices between simulated frequency bins. The group delay standard deviation is set to 3.1 ps/$\sqrt{\text{km}}$ [30]. The MDG of the link is controlled by the per-amplifier MDG standard deviation, σ_g. Nonlinear transmission effects are

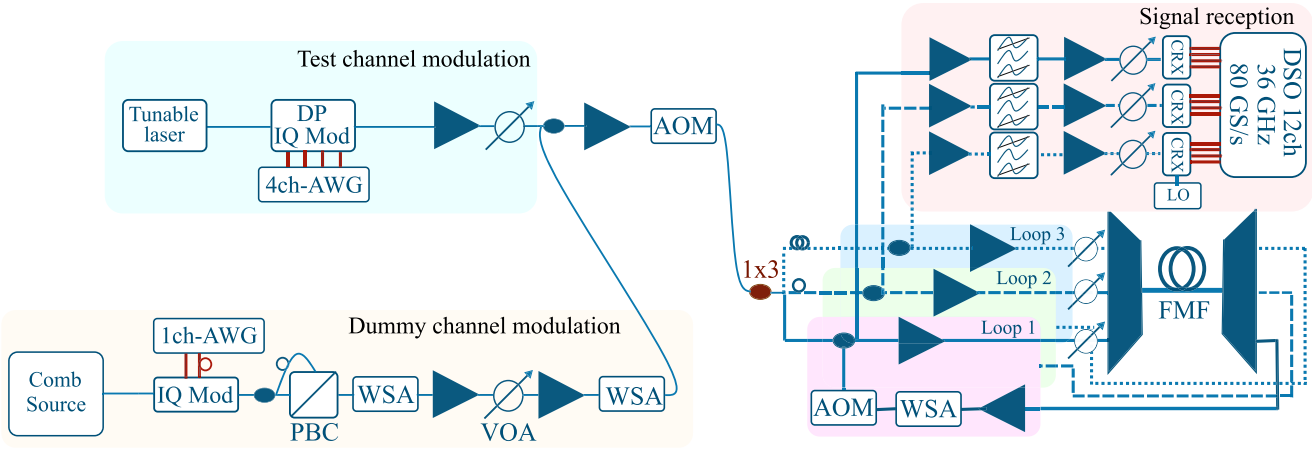

Fig. 5. Experimental 3-mode transmission setup with recirculating loop. At the transmitter, test and dummy bands are modulated, amplified, and combined. Three independent data streams are emulated in a split and delay configuration. The spatial channels are injected into three recirculating loops comprising a 55 km FMF. After propagation, the signals are detected, amplified, and filtered in an SDM receiver. After coherent reception, the signals are digitized and re-sampled. Off-line DSP is applied to the digital electrical signals. MDG is estimated from the transfer matrix of the 6×6 MIMO equalizer.

not simulated. After propagation, the receiver front-end model converts the received signals from the optical to the electrical domain. No phase noise has been considered for the simulations. The electric signals are down-sampled to two samples per symbol and fed into the time domain equalizer (TDE). For data reception, 6×6 MIMO time-domain equalization in direct mode is carried out by 36 finite impulse response filters with 800 taps each, updated by a fully supervised least mean square (LMS) algorithm. 6×6 MIMO equalization in reverse mode is also performed exclusively for MDG estimation. The MDG standard deviation σ_{mdg} is computed at each frequency of the MIMO transfer function and averaged across the signal band.

Fig. 4 shows the estimated σ_{mdg} as a function of the actual σ_{mdg} for the three investigated techniques. Each point of the curves results from averaging over 5 iterations for the same configuration of actual σ_{mdg}. The σ_{mdg} is swept from 0.2 dB to 21 dB, and the SNR is set at 11 dB or 21 dB in Fig. 4(a) and (b), respectively. In addition to the MDG estimation methods based on direct and reverse MIMO equalization, the ANN-based technique estimates the MDG from the set of eigenvalues, $\lambda^2_{i,\text{dir}}$, computed from the direct MIMO equalizer transfer matrix and the electrical SNRs obtained from the equalized constellations [22], [28]. In Fig. 4, the blue curve corresponds to the MDG estimated from the transfer matrix of the MIMO operated in direct mode based on (18). The estimates track the actual MDG only in the low MDG region, largely underestimating the actual value for middle and high MDG regimes with higher estimation errors at lower SNR. In contrast, the low complexity ANN-based estimator provides more accurate estimates for high levels of MDG, outperforming the direct equalizer technique, although overestimating the MDG over certain regions. The accuracy of the ANN estimator is observed to be higher at higher SNRs. Finally, the orange curve corresponds to the case where σ_{mdg} is estimated from the transfer matrix of the MIMO equalizer operated in reverse mode, as indicated in (26). Ideally, according to (25), the performance of the MDG estimation technique based on reverse MIMO equalization should be independent of the noise. Nonetheless, the estimates deviate from the actual values from $\sigma_{\text{mdg}} = 12$ dB and 16 dB for SNR = 11 dB and 21 dB, respectively. These results can be explained by the influence of noise in the equalizer convergence, affecting the MDG estimation process.

The results suggest the reverse equalizer method as a preferential method for MDG estimation in laboratory experiments. However, training an equalizer exclusively for MDG estimation may be too computationally expensive for real-time hardware implementation. In these cases, the low-complexity ANN-based method may be the preferred choice [11].

V. MDG Estimation in an Experimental Long-Haul 3-Mode Transmission Setup

A. Experimental 3-Mode Transmission Setup

The investigated MDG estimation techniques are evaluated in the 3-mode experimental setup with recirculating loop shown in Fig. 5. The generated signals were divided into a test channel and a dummy band. The test-channel is generated by modulating the light from a tunable laser source in a dual polarization IQ-modulator that is driven by a four-channel arbitrary wave form generator (AWG) operating a 49 Gs/s, which generates 24.5 GBaud 16-QAM signals shaped with a 0.01 roll-off factor RRC shaping filter. The dummy band was generated based on a single optical comb source that generated 25 GHz spaced optical carrier lines over the C-band that are modulated by a single-polarization IQ-modulator. Polarization multiplexing is generated in a split and delay configuration. A wavelength selective attenuator (WSA) selected 20 wavelength-division multiplexing (WDM) channels from the dummy channel spectrum and was used to flatten the optical spectrum and to carve a notch into the dummy band to accommodate the test channel. Only one test channel wavelength was used for this experiment and the 20 dummy channels were used to provide a sufficiently high total power level required for proper operation of

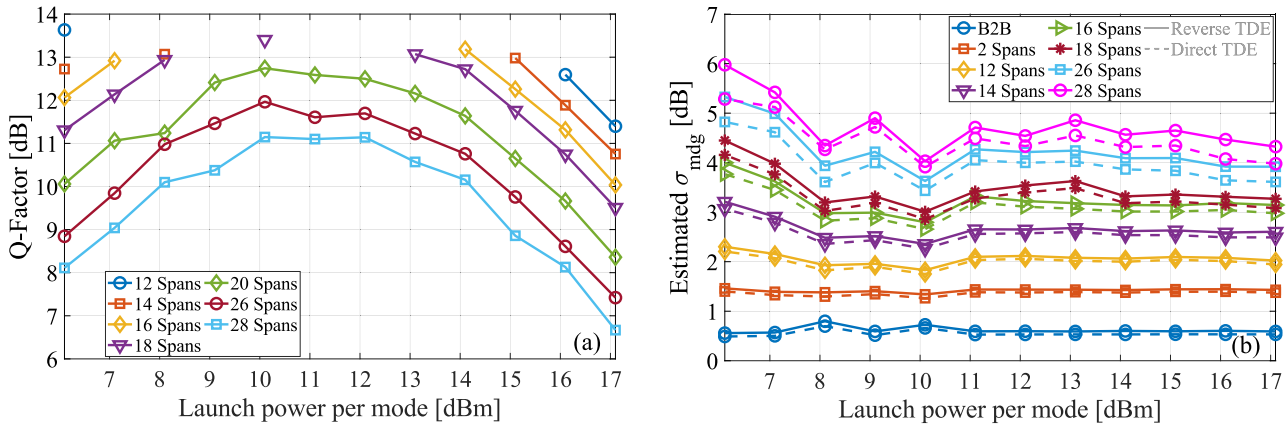

Fig. 6. (a) Average Q-factor as a function of the launch power per mode in dBm equally distributed over the 21 wavelength-division multiplexing (WDM) channels. (b) σ_{mdg} estimated at different transmission distances as a function of the launch power per mode in dBm. The dashed lines correspond to the direct approach and the solid lines to the reverse approach.

the erbium-doped fibre amplifiers (EDFAs) in the transmission loop.

Three independent and identical transmitters are emulated in a split and delay setup. Two paths are optically delayed by 100 ns and 200 ns, approximately, to obtain three de-correlated data streams, one for each spatial channel. To transmit over long-haul distances, three recirculating loop setups are used. The recirculating loops comprise EDFAs, variable optical attenuators (VOAs), a spatial multiplexer, a span of 55 km FMF, a spatial de-multiplexer, and WSAs to equalize spectral gain variations coming from the amplification stages. An arrangement of acoustic optical modulators (AOMs) at the transmitter and the end of the loop is employed to load the loop and define the total number of recirculations. The three VOAs located before the spatial multiplexer allow to independently set the launch power per mode to get a certain MDG per span, resulting in a controlled amount of accumulated MDG. The 55-km FMF link was manufactured by fusion splicing three fiber segments for almost ideal compensation of the differential-mode delay (DMD). The LP01 and LP11 modes have attenuation below 0.2 dB/km and 0.22 dB/km, respectively, at 1550 nm wavelength. The effective area of the fundamental mode was measured at 125 μm^2 [31].

After transmission, signal reception is performed by a 3-channel SDM receiver composed of EDFAs, optical filters (to select the channel under test), VOAs (to control the receiver power), and coherent receivers (for optical-to-electrical conversion). A 60 kHz-linewidth local oscillator (LO) laser is employed for coherent mixing. The electrical signals are digitized by a 12-channel real-time oscilloscope operating at 80 GS/s and with a 36-GHz electrical bandwidth. The digital signals are re-sampled at 2 samples per symbol and processed in an off-line DSP module. Front-end compensation, frequency offset compensation, chromatic dispersion removal, 6×6 dynamic MIMO equalization, and carrier phase recovery are performed. MDG estimation is carried out from the transfer matrix of the MIMO equalizer operated in direct and reverse modes. In addition, the ANN model is trained and later applied to the experimental data for MDG estimation. The samples to feed the ANN comprise the vector of eigenvalues computed from the transfer matrix of the MIMO equalizer operated in direct mode and the SNRs calculated from the digital constellations after DSP.

B. *Experimental Results Without Additional MDG*

Data transmission is carried out in back-to-back configuration and up to 28 recirculations to achieve transmission distances between 0 km and 1,540 km. The launch power per mode, equally distributed over the 21 WDM channels, is swept by employing three VOAs located before the spatial multiplexer. Fig. 6(a) shows the Q-factor in dB averaged over the 6 propagation modes as a function of the launch power per mode for different transmission distances. The discontinuity of the curves corresponds to an infinite Q-factor calculated from the limited length of the sequence employed to count bit errors. For distances shorter than 12 spans (660 km), the transmission is error-free for all the launch powers evaluated. From 12 spans, the Q-factor increases with the increment of the launch power per mode until reaching the optimum value of around 10 dBm. For higher launch power values, the fiber starts working in the non-linear regime, and the performance decreases.

MDG estimation based on direct and reverse MIMO equalization is performed over the linear and non-linear transmission regimes. Fig. 6(b) shows the estimated σ_{mdg} as a function of the launch power per mode. The dashed lines correspond to the direct equalization approach, and the solid lines correspond to the MDG estimated when the MIMO equalizer is operated in reverse mode. From Fig. 6(b) and consistently with the simulation results of Section IV, the direct approach tends to provide lower estimates than the reverse approach. The gap between the evaluated methods increases with the increment of the transmission distance. It is interesting to note that non-linearities do not significantly affect the MDG estimation performance for either of the two methods. However, at low launch power levels, both estimation methods provide higher MDG estimates

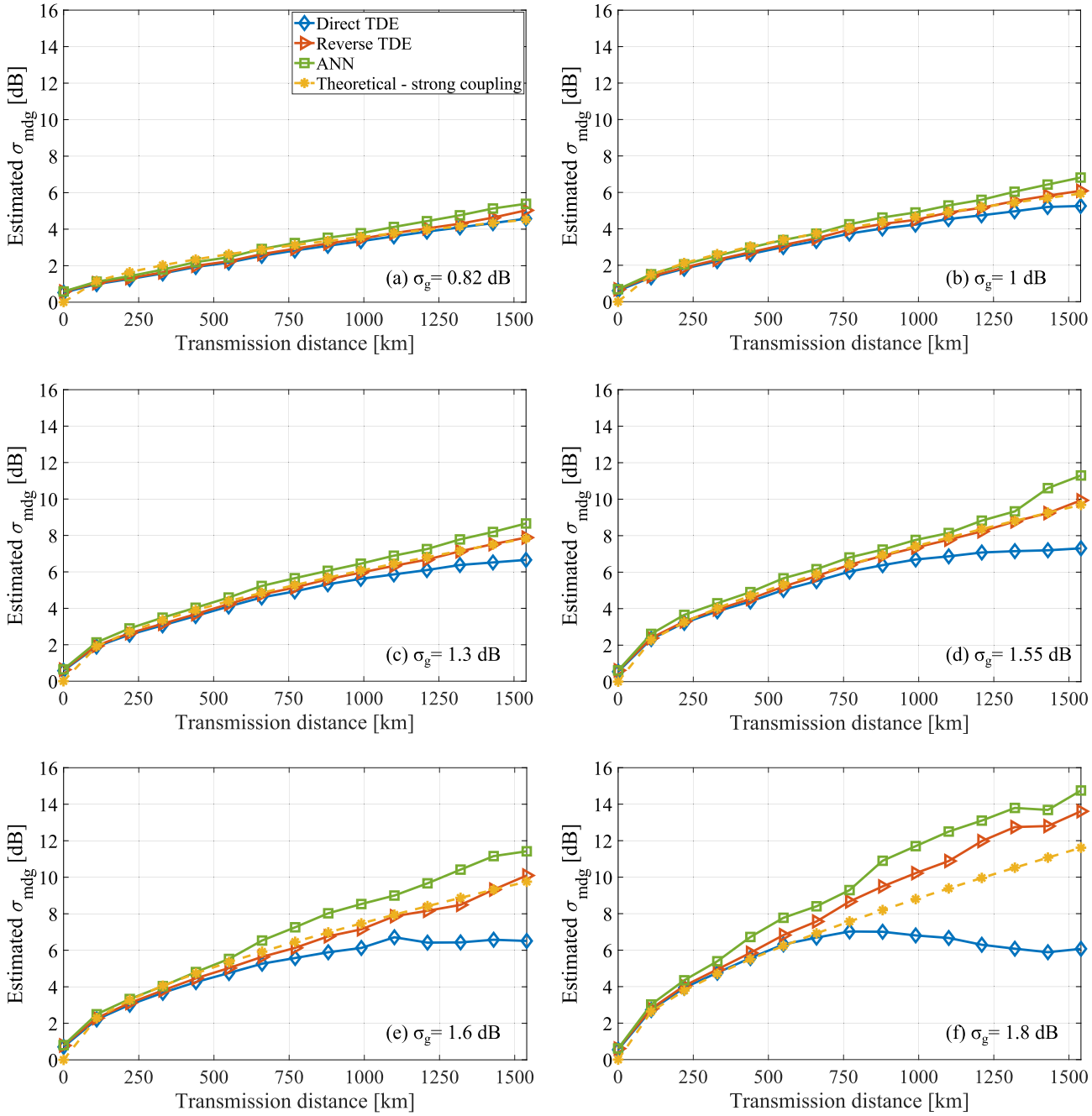

Fig. 7. σ_{mdg} estimated through different methods as a function of the transmission distance for (a) $\sigma_{\text{mdg}} = 0.82$ dB. (b) $\sigma_{\text{mdg}} = 1$ dB. (c) $\sigma_{\text{mdg}} = 1.3$ dB. (d) $\sigma_{\text{mdg}} = 1.55$ dB. (e) $\sigma_{\text{mdg}} = 1.6$ dB. (f) $\sigma_{\text{mdg}} = 1.8$ dB. The per-span MDG (σ_g) is controlled by attenuating and de-attenuating VOAs according to Table I. The theoretical curve of the accumulated MDG (σ_{mdg}) for strongly-coupled SDM transmission (dashed line) is plotted for reference.

in comparison to other launch power values, a behavior that is more evident for increasing transmission distances.

C. Experimental Results With Additional MDG

In the previous section, the same power was launched into the LP01, LP11a, and LP11b modes. Therefore, the MDG presented in Fig. 6(b) corresponds to the inherent MDG of the transmission setup at different fiber lengths. To study the estimation methods over a wide range of overall MDG, the per-span MDG is increased by intentionally unbalancing the per-mode launch power. Three VOAs located before the spatial multiplexer are employed to attenuate or de-attenuate the mode powers around a reference point of 12 dBm, keeping the total launch power constant. The attenuation introduced by the VOAs into the LP01, LP11a, and LP11b modes, for six different attenuation cases, are depicted in Table I.

For case (a), an attenuation of 5 dB is applied equally to the 3 modes. This configuration still corresponds to the inherent MDG of the transmission link since no additional gain imbalances

TABLE I
VARIABLE OPTICAL ATTENUATORS (VOAS) CONFIGURATION

Case	Attenuation sweep		
	LP01	LP11a	LP11b
(a)	5 dB	5 dB	5 dB
(b)	4.14 dB	5.5 dB	5.5 dB
(c)	3.18 dB	6.3 dB	6.3 dB
(d)	2.60 dB	7 dB	7 dB
(e)	10 dB	3.72 dB	3.72 dB
(f)	11 dB	3.62 dB	3.62 dB

are introduced. In cases (b), (c), and (d), the LP01 mode is de-attenuated, and the LP11a and LP11b modes are equally attenuated. In cases (e) and (f), to achieve very high values of accumulated MDG, the LP01 mode is severely attenuated while the LP11a and LP11b modes are de-attenuated. Fig. 7 shows the estimated σ_{mdg} as a function of the transmission distance for the six attenuation cases of Table I. The MDG is estimated using the three investigated techniques: the direct approach, the reverse approach, and the ANN-based estimator. Moreover, the theoretical curve of the accumulated σ_{mdg} for strongly-coupled SDM transmission is plotted for reference [14] (dashed line). Fig. 7(a), in which no extra power imbalances are applied, corresponds to the link inherent MDG. For back-to-back transmission, σ_{mdg} is 0.58 dB. After 1 span, $\sigma_{\mathrm{mdg}} = \sigma_{\mathrm{g}} = 0.82$ dB. For short distances, at high SNR and low MDG, the three MDG estimation methods provide roughly the same σ_{mdg} estimates. From 12 spans, the gap between estimates starts being evident, but without exceeding 0.8 dB at 28 spans (1,540 km). In Fig. 7(b)–(d), with σ_{g} equal to 1 dB, 1.3 dB, and 1.55 dB, respectively, the difference between the estimates provided by the methods is more evident, as the MDG is higher and the SNR is lower. As observed in previous studies, the performance of the direct approach saturates at intermediate MDG levels. The reverse and ANN-based approaches, however, continue estimating increasing MDG values. The reverse approach tracks the theoretical curve for a strongly-coupled transmission up to $\sigma_{\mathrm{g}} = 1.6$ dB. For $\sigma_{\mathrm{g}} = 1.8$ dB, the estimated MDG grows faster, deviating from the theoretical curve after approximately 8 spans (440 km). We conjecture that, at this high MDG level, the assumptions implied in the theoretical model for strongly-coupled transmission may start to break down. The ANN, in turn, tends to follow the reverse approach, but always overestimates the MDG. These results indicate that a more optimized ANN has potential for accurate and low-complexity MDG estimation.

VI. CONCLUSION

We have investigated the DSP-based MDG estimation in long-haul SDM transmission by simulation and experiments with controlled MDG. Confirming previous studies, we showed that the widely known method that employs the transfer matrix of the MMSE MIMO equalizer operated in direct mode tends to underestimate the MDG at low SNR and high MDG levels. In contrast, the results show that the MDG estimation method based on the MMSE MIMO equalizer operated in reverse mode presents accurate estimation performance, at the cost of requiring a MIMO equalizer exclusively for MDG estimation. Alternatively, the low-complexity ANN-based estimator is able to track the true MDG for a wide range of values, but overestimates the MDG for high accumulated values. The results indicated the reverse approach as the preferred choice for laboratory experiments where computational complexity is not a limiting factor. For real-time applications, the ANN-based approach offers a promising solution combining reasonable accuracy and low-complexity.

REFERENCES

[1] B. J. Puttnam, G. Rademacher, and R. S. Luís, "Space-division multiplexing for optical fiber communications," *Optica*, vol. 8, no. 9, pp. 1186–1203, 2021.

[2] J. v. Weerdenburg et al., "138 Tbit/s transmission over 650km graded-index 6-mode fiber," in *Proc. Eur. Conf. Opt. Commun.*, 2017, pp. 1–3.

[3] G. Rademacher et al., "Long-haul transmission over few-mode fibers with space-division multiplexing," *J. Lightw. Technol.*, vol. 36, no. 6, pp. 1382–1388, Mar. 2018.

[4] G. Rademacher et al., "High capacity transmission with few-mode fibers," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 425–432, Jan. 2019.

[5] R. Ryf et al., "Coupled-core transmission over 7-core fiber," in *Proc. Opt. Fiber Commun. Conf.*, 2019, pp. 1–3.

[6] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Exp.*, vol. 19, no. 17, pp. 16680–16696, 2011.

[7] D. A. A. Mello, H. Srinivas, K. Choutagunta, and J. M. Kahn, "Impact of polarization- and mode-dependent gain on the capacity of ultra-long-haul systems," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 303–318, Jan. 2020.

[8] P. J. Winzer, H. Chen, R. Ryf, K. Guan, and S. Randel, "Mode-dependent loss, gain, and noise in MIMO-SDM systems," in *Proc. Eur. Conf. Opt. Commun.*, 2014, pp. 1–3.

[9] C. Antonelli, A. Mecozzi, M. Shtaif, and P. J. Winzer, "Modeling and performance metrics of MIMO-SDM systems with different amplification schemes in the presence of mode-dependent loss," *Opt. Exp.*, vol. 23, no. 3, pp. 2203–2219, 2015.

[10] K.-P. Ho and J. M. Kahn, "Frequency diversity in mode-division multiplexing systems," *J. Lightw. Technol.*, vol. 29, no. 24, pp. 3719–3726, Dec. 2011.

[11] D. A. A. Mello, R. S. B. Ospina, H. Srinivas, K. Choutagunta, E. Chou, and J. M. Kahn, "Impact and mitigation of mode-dependent gain in ultra-long-haul SDM systems," in *Proc. Opt. Fiber Commun. Conf.*, 2023, pp. 1–3.

[12] E. S. Chou and J. M. Kahn, "Successive interference cancellation on frequency-selective channels with mode-dependent gain," *J. Lightw. Technol.*, vol. 40, no. 12, pp. 3729–3738, Jun. 2022.

[13] A. Lobato et al., "Impact of mode coupling on the mode-dependent loss tolerance in few-mode fiber transmission," *Opt. Exp.*, vol. 20, no. 28, pp. 29776–29783, 2012.

[14] K.-P. Ho and J. M. Kahn, "Mode-dependent loss and gain: Statistics and effect on mode-division multiplexing," *Opt. Exp.*, vol. 19, no. 17, pp. 16612–16635, 2011.

[15] J.-J. v. d. Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Veh. Technol. Conf. Countdown Wireless Twenty-First Century*, 1995, pp. 815–819.

[16] O. Edfors, M. Sandell, J.-J. v. d. Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[17] S. Hoffmann et al., "Frequency and phase estimation for coherent QPSK transmission with unlocked DFB lasers," *IEEE Photon. Technol. Lett.*, vol. 20, no. 18, pp. 1569–1571, Sep. 2008.

[18] Y. Mori, C. Zhang, and K. Kikuchi, "Novel configuration of finite-impulse-response filters tolerant to carrier-phase fluctuations in digital coherent optical receivers for higher-order quadrature amplitude modulation signals," *Opt. Exp.*, vol. 20, no. 24, pp. 26236–26251, 2012.

[19] R. S. Ospina, C. Okonkwo, and D. A. Mello, "DSP-based mode-dependent loss and gain estimation in coupled SDM transmission," in *Proc. Opt. Fiber Commun. Conf.*, 2020, pp. 1–3.
[20] R. S. Ospina et al., "Mode-dependent loss and gain estimation in SDM transmission based on MMSE equalizers," *J. Lightw. Technol.*, vol. 39, no. 7, pp. 1968–1975, Apr. 2021.
[21] M. v. d. Hout et al., "Experimental validation of MDL emulation and estimation techniques for SDM transmission systems," in *Proc. Eur. Conf. Opt. Commun.*, 2020, pp. 1–4.
[22] R. S. B. Ospina, M. v. d. Hout, S. v. d. Heide, C. Okonkwo, and D. A. A. Mello, "Neural-network-based MDG and optical SNR estimation in SDM transmission," in *Proc. Opt. Fiber Commun. Conf.*, 2021, pp. 1–3.
[23] R. Ryf et al., "White Gaussian noise based capacity estimate and characterization of fiber-optic links," in *Proc. Opt. Fiber Commun. Conf.*, 2018, pp. 1–3.
[24] G. Rademacher et al., "Highly spectral efficient C+ L-band transmission over a 38-core-3-mode fiber," *J. Lightw. Technol.*, vol. 39, no. 4, pp. 1048–1055, Aug. 2020.
[25] S. S. Haykin, *Adaptive Filter Theory*, 5th ed. London, U.K.: Pearson, 2002.
[26] M. R. McKay, I. B. Collings, and A. M. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 396–410, Jan. 2010.
[27] W. S. Saif, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Optical performance monitoring in mode division multiplexed optical networks," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 491–504, Jan. 2021.
[28] R. S. Ospina et al., "MDG and SNR estimation in SDM transmission based on artificial neural networks," *J. Lightw. Technol.*, vol. 40, no. 15, pp. 5021–5030, Aug. 2022.
[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[30] T. Hayashi, Y. Tamura, T. Hasegawa, and T. Taru, "Record-low spatial mode dispersion and ultra-low loss coupled multi-core fiber for ultra-long-haul transmission," *J. Lightw. Technol.*, vol. 35, no. 3, pp. 450–457, Feb. 2017.
[31] G. Rademacher et al., "Investigation of wideband distributed Raman amplification in a few-mode fiber link," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2022, pp. 1–3.