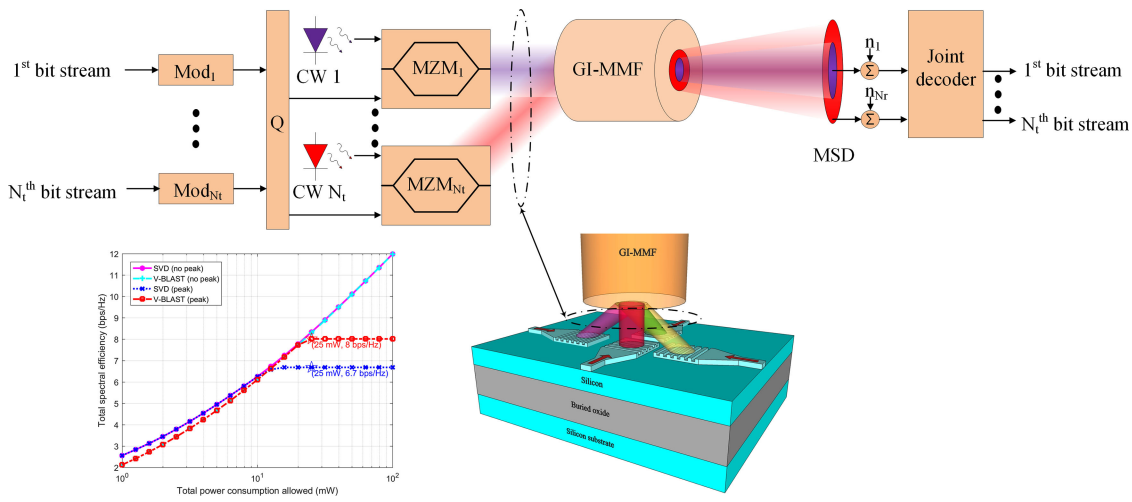
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Power Loading in Peak Limited Intensity Modulation Direct Detection Mode Group Division Multiplexing

Volume 10, Number 1, February 2018

Nir Sheffi
Dan Sadot, ***Senior Member, IEEE***



DOI: 10.1109/JPHOT.2017.2775144
1943-0655

Power Loading in Peak Limited Intensity Modulation Direct Detection Mode Group Division Multiplexing

Nir Sheffi and Dan Sadot, *Senior Member, IEEE*

Electrical and Computer Engineering Department, Ben Gurion University of the Negev, Beer Sheva 84105, Israel

DOI:10.1109/JPHOT.2017.2775144

Manuscript received October 14, 2017; revised November 13, 2017; accepted November 14, 2017. Date of publication November 17, 2017; date of current version January 5, 2018. (Corresponding author: Nir Sheffi (e-mail: nirshef@post.bgu.ac.il).

Abstract: The exponential growth of network traffic within data centers and high-performance computing generates a growing volume of fiber cabling. This raises the need for spectrally efficient communication scheme, which allows easy integration of optics and electronics. In this paper, we suggest a mode group division multiplexing scheme based upon intensity-modulated silicon photonics (SiP) Mach–Zehnder modulators (MZMs) coupled to standard graded-index multi-mode fiber directly detected using multi-segment concentric photo-detector. In addition, we have theoretically derived the bit and power loading of such a system in a closed-loop multiple-input multiple-output (MIMO) formation using convex optimization with two possible types of architectures: the vertical Bell Labs layered space-time architecture (V-BLAST) with minimum mean square error and successive interference cancellation (MMSE-SIC); and singular value decomposition (SVD). Our optimization problem maximizes the system's capacity under peak amplitude, power consumption, and BER constraints. Simulation results have shown the superiority of a V-BLAST MMSE-SIC over the SVD architecture in terms of total spectral efficiency for a 4×4 MIMO short reach low driving voltage conventional SiP-based MZM system.

Index Terms: Graded-index multi-mode fiber (GI-MMF), mode group division multiplexing (MGDM), multiple-input-multiple-output (MIMO), high performance computing (HPC), data-centers.

1. Introduction

Bandwidth demands in a data-center doubles every 12–15 months [1]. In turn, the power consumed by the data-center continues to grow exponentially, with most power being consumed by the interconnects within and between racks of servers [2]. Consequently, the fiber volume becomes a major design constraint [3].

The main rack-to-rack interconnect technology that is used in todays' data-center is either copper-based or optical [2]. Due to the limited bandwidth-distance product, the copper-based interconnect technology is less efficient and slower [2] as compared to its optical counterpart.

Consequently, in order to address the fiber volume constraint, the optical rack-to-rack interconnect technology offers the following main alternatives. One alternative is an optical interconnect based upon silicon photonics (SiP) modulator with a single mode fiber (SMF) using parallel optics [4], or wavelength division multiplexing [5]. Another alternative is an optical interconnect based upon

Vertical Cavity Surface Emitting Laser (VCSEL) and GaAs PIN photodiode coupled to a standard 50 μm core graded-index multi-mode fiber (GI-MMF) using parallel optics [6].

Generally, SiP allows monolithic integration of optics and electronics using mature silicon process technology [7]. Moreover, SiP allows generation of high bandwidth modulators, as opposed to directly modulated VCSEL array, which is limited to lower bandwidth. VCSEL array is bias-current bandwidth dependent, which results in higher power consumption at high bandwidth applications.

Recently, it has been suggested to alleviate the fiber volume constraint using a two-dimensional (2D) array of 850 nm VCSELs with multi-core fibers (MCFs), where each core is a GI-MMF [3]. Moreover, in [8] it has been suggested to use these VCSEL arrays along with GI-MMF for optical SDM scheme in short-reach application.

In the broader perspective, it should be noted that the concept of spatial-division multiplexing (SDM) is almost as old as the fiber optics communication itself reported as far back as the early 80's [9], [10]. Lately, as part of the "capacity crunch" problem, the traffic growth rate has increased at around 60% per year for transport links over 1500 km [11]. While current optical communication technologies do not meet the demand rate, SDM along with wavelength division multiplexing, polarization-division multiplexing, and coherent detection, using advanced modulation formats, e.g., phase modulation or quadrature-amplitude-modulation, attracts major academic attention as a possible solution [12]. Although SDM possesses many advantages, it also introduces some major challenges. One of the main challenges in the SDM technology is a generation of a compact and efficient spatial multiplexer and demultiplexer [13]–[18].

In short-reach applications, there is a plethora of SDM research primarily targeting GI-MMF using intensity modulation (IM) and direct detection (DD) [10], [19]–[26]. The main interest in SDM over GI-MMF is significantly higher spectral efficiency turning modal dispersion to multiplexing gain, as the spatial multiplexing allows transmission of multiple data streams over multiple spatial channels. Furthermore, it effectively selects subsets of modes essentially decreasing the dispersion per subset. In addition, compared to SMF, the GI-MMF's large cross-section reduces the Kerr effect caused by light intensity, thus decreasing non-linear effects [27]. In such short-reach links, each transmitter excites different group of modes in the fiber, and at the receiver side, each detector selectively respond to a different part of the near-field pattern (NFP) at the fiber output [22]. Such SDM short-reach applications, which are based upon group of modes, are usually termed mode group division (or diversity) multiplexing (MGDM).

In this paper, we suggest an MGDM system model based upon IM SiP Mach-Zehnder modulators (MZMs) coupled to a standard 50 μm GI-MMF with a multi-segment concentric photo-detector (MSD) [28]. Such a system allows us to benefit from monolithic integration of optics and electronics [7]. The spatial multiplexer model is generated using different tilt output for each MZM along with its corresponding spot size [29] using an SiP integrated circuit. One should notice that unlike conventional mode pattern launching, here, we excite a large number of modes in a non-precise matter for each MZM. Moreover, it also handles the fiber volume problem using already deployed GI-MMF, maximizing its inherent spatial degree of freedom with multiple-input-multiple-output (MIMO) digital signal processing (DSP). In addition, the compactness of such a system with the integration of complementary metal-oxide-semiconductor (CMOS) process can allow advanced sub-1V CMOS circuits as drivers [30].

Most MGDM with IM/DD systems were analyzed using computer simulation and/or experimental results [19], [20], [22], [23], [25]. Some recent works on optical wireless channels used an analytical approach to optimally allocate system parameters by formulating it as a rate adaptive problem with a nonnegative constraint [31], [32]. Lately, a GI-MMF MGDM IM/DD system using a VCSEL array has been analyzed with the aid of convex optimization, minimizing electrical power consumption [24].

Here, inspired by [24], we propose a power loading method based upon an analytical approach using convex optimization. We maximize the system capacity under two types of closed-loop architectures: the vertical Bell Labs layered space-time architecture (V-BLAST) with minimum mean square error and successive interference cancellation (MMSE-SIC) [33]–[35]; and the singular value decomposition (SVD) architecture [34]. The V-BLAST architecture was originally proposed as a

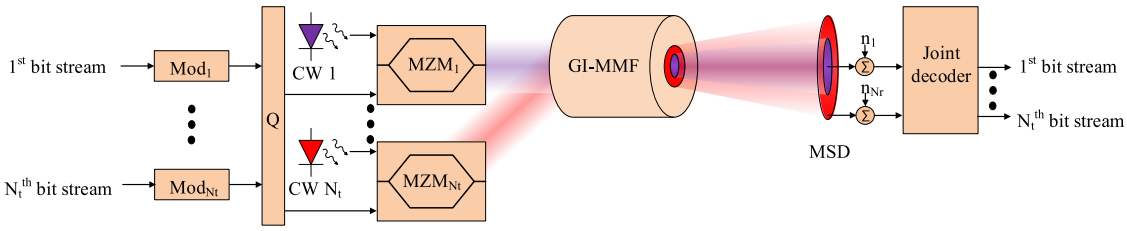


Fig. 1. Mode group division multiplexing (MGDM) system model with Mach-Zehnder modulators (MZMs), tilted Gaussian beams multiplexer, graded-index multi-mode fiber (GI-MMF), and a multi-segment concentric photo-detector (MSD).

simple spatial multiplexing scheme when there is no channel knowledge available at the transmitters [33], also in optical SDM applications [19], [25], [36]. Here, we consider the closed-loop extension of the V-BLAST, which allows power and bit loading in each transmitter via a rate feedback channel [35]. This closed-loop extension is known in wireless communication to attain an achievable rate very close to SVD architecture including optimal water-filling-based power and rate assignments [35]. One should notice that in comparison to [24], the introduction of SiP MZMs instead of a VCSEL array, along with different closed-loop architectures (V-BLAST), and capacity maximization as opposed to margin adaptive, leads to a different mathematical analysis and optimization problem formulation.

Generally, throughout this paper, the proposed power loading method maximizes the system's capacity under the following set of constraints:

1) *Peak amplitude:* The modulated voltage of each *i*-th SiP MZM should be in the linear regime of the MZM, allowing the use of silicon driver [30].
2) *Power consumption:* The entire power consumption due to all MZMs is limited. It is assumed that the MZMs' power consumption is mostly RF dominant [37] and is governed by a capacitor load [30], [38].
3) *BER requirement:* The BER should be below a target BER level to guarantee the required quality of service.

The paper is organized as follows. In Section 2, we describe the system and channel model, which includes a model for the spatial multiplexer, the MZM, and the MIMO channel. In Section 3, we formally derive the power loading optimization problem for V-BLAST MMSE-SIC and SVD architectures. The derivation starts with a closed-loop MIMO formulation without the peak amplitude constraint for both architectures. It is followed by a general formulation of the peak amplitude constraint. Afterwards, using the general formulation of the peak constraint, a complete mathematical description of the power loading optimization problem under all constraints is shown for both architectures. In Section 4, a quantitative analysis of the power loading optimization problem for both architectures is presented for a typical scenario. Finally, we draw our conclusions in Section 5.

Throughout the paper, we use the following notations. Upper case and lower case boldface are used to employ matrices and vectors, respectively. For an arbitrary matrix, $\mathbf{A}$, $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$, $|\mathbf{A}|$ is the determinant of $\mathbf{A}$, and $\mathrm{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$. For an arbitrary vector $\mathbf{a}$, $[\mathbf{a}]_i$ depicts the *i*-th element of vector $\mathbf{a}$. $\|\mathbf{a}\|_p$ denotes the p-norm of vector $\mathbf{a}$. $\mathbf{a}^*$ designates the complex conjugate. $\mathrm{diag}(\mathbf{a}^T)$ is a diagonal matrix with entries given by the vector $\mathbf{a}$. When the inequality symbol $\geqslant$ is used between vectors, it denotes element-wise inequality. $\mathbb{E}[\cdot]$ designates the expectation operator.

2. System and Channel Model

The MGDM system model contains a spatial multiplexer, a GI-MMF, and a spatial demultiplexer, as illustrated in Fig. 1. The spatial multiplexer is based upon several incoherent MZM outputs, where

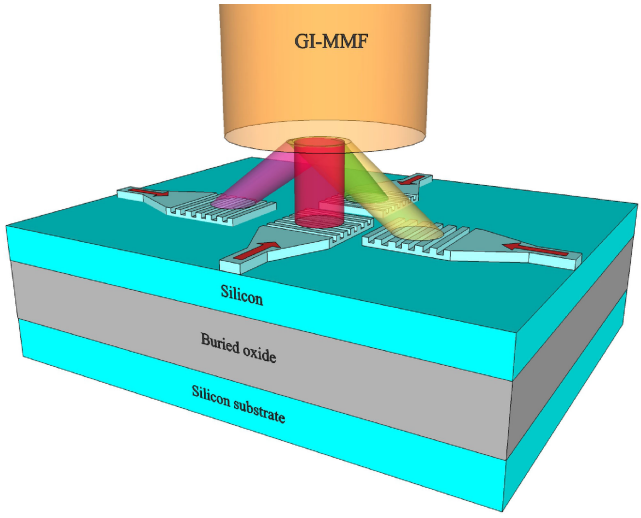


Fig. 2. Illustration of a silicon photonics tilted Gaussian beams multiplexer using spot conversion and apodized waveguide grating couplers of 4 Mach-Zehnder modulators, each with a different tilt and spot size.

each MZM has a different beam tilt with respect to the GI-MMF, thus forming a tilted Gaussian beams multiplexer, as in [24], [29]. The spatial demultiplexer is performed using an MSD [28].

2.1 Spatial Multiplexer Model

The spatial multiplexer model suggested in this paper is based upon use of a grating coupler for each SiP MZM. One should notice that unlike conventional mode pattern launching, here, we excite a large number of modes in a non-precise matter for each grating coupler. An illustration of such a spatial multiplexer is given in Fig. 2. Generally, in order to achieve the required tilt coupling angle for each MZM, a phase matching condition is required, i.e., [39]

$$n_{\text{eff}} k_0 + \frac{2\pi}{\Lambda} = k_0 \sin\psi \tag{1}$$

where, k_0 is the wavenumber in free space, Λ is the grating period, n_{eff} is the mean effective index along one grating period, and ψ is the tilt coupling angle.

Furthermore, while the analysis in [29] assumes a Gaussian-shaped field profile in the entrance to the GI-MMF. Such a profile can be theoretically achieved using apodized waveguide grating coupler, as seen in [40]–[42].

2.2 Mach-Zehnder Modulator Model

The model of the *i*-th MZM is composed of two parts: the intensity output and its power consumption.

First, the intensity at the output of an *i*-th MZM using a dual drive modulation is given by

$$I_{i,\text{MZM}}\left(\tilde{v}_{i,a}(t)\right) = \frac{E_0^2}{2}\left(1 + \sin\left(\frac{2\pi}{v_\pi}\tilde{v}_{i,a}(t)\right)\right), \tag{2}$$

where E_0^2 corresponds to the continuous-wave (CW) laser power including insertion loss (IL), $v_{i,a}(t)$ is the modulated voltage component across an MZM arm. v_π is a MZM halfwave voltage parameter, which indicates the amount of voltage required for a π phase shift. Throughout this paper, we assume the maximum allowed modulated voltage amplitude is due to the use of silicon driver [30]

and it is in the linear regime of the MZM, i.e.,

$$-\tilde{v}_{\text{peak}} \leqslant \tilde{v}_{i,a}(t) \leqslant \tilde{v}_{\text{peak}}. \tag{3}$$

Thus, using the linear approximation, $\sin(x) \approx x$, we deduce that the intensity at the i-th MZM's output is

$$I_{i,\text{MZM}}(\tilde{v}_{i,a}(t)) = \frac{E_0^2}{2}\left(1 + \frac{2\pi}{v_\pi}\tilde{v}_{i,a}(t)\right). \tag{4}$$

The modulated voltage at the i-th MZM arm, $\tilde{v}_{i,a}(t)$, is given by

$$\begin{gathered}\tilde{v}_{i,a}(t) = \sum_{k=-\infty}^{\infty} x_i[k]\phi(t - kT), \\ i = 1, 2, \ldots, N_t,\end{gathered} \tag{5}$$

where $x_i[k]$ is the symbol loaded into the k-th sample, T is the symbol duration, and $\phi(t)$ is the pulse shaping function. In this paper, we use a sinc pulse shaping function with energy $\mathcal{E}_\phi$.

Second, the i-th MZM power consumption is mostly RF dominant [37] and governed by a capacitor load [30], [38]. Thus,

$$p_i = \frac{C_{\text{MZM}}}{2T}\mathbb{E}[\tilde{v}_{i,a}(T) - \tilde{v}_{i,a}(0)]^2, \tag{6}$$

where C_{MZM} is the MZM capacitance. Moreover, due to the independent nature of the symbols loaded into the k-th sample, their zero mean and the sinc pulse shaping function, we can re-write (6) as

$$p_i = \frac{C_{\text{MZM}}}{T}\mathbb{E}[x_i[k]]^2. \tag{7}$$

Thus, the entire power consumption due to all MZMs is given by

$$p_{total} = \sum_{i=1}^{N_t} p_i = \sum_{i=1}^{N_t} \frac{C_{\text{MZM}}}{T}\mathbb{E}[x_i[k]]^2 = \frac{C_{\text{MZM}}}{T}\mathbb{E}\left\|\mathbf{x}_k\right\|_2^2, \tag{8}$$

where N_t is the number of MZM transmitters, and the transmit vector elements $[\mathbf{x}_k]_i = x_i[k]$

2.3 MIMO Model

In this section, we describe the MIMO relation between the transmitted signal into the MZMs and the received signal after the MSD. First, we assume the channel has a flat frequency response. Second, using (4) and (5), the received match-filtered k-th sample on each j-th photo-detector segment after AC coupling can be written as

$$y_j[k] = \sum_{i=1}^{N_t} \alpha_{\text{loss}} h_{ji} x_i[k] + n_j[k], \tag{9}$$

where h_{ji} is the channel's gain associated with the i-th transmitter and j-th receiver; α_{loss} is a common attenuation factor; $n_j[k]$ is spatially and temporally white Gaussian noise, i.e., $n_j[k] \sim \mathcal{N}(0, \sigma_n^2)$. It can be easily shown that

$$\sigma_n^2 = \frac{i_c^2}{2}\mathcal{E}_\phi, \tag{10}$$

where i_c is the noise current spectral density. α_{loss} includes: the CW laser power including insertion loss, E_0^2; pulse shaping energy, $\mathcal{E}_\phi$; link budget power margin, α_{margin}; MSD responsivity, R_{PD}; MZM's

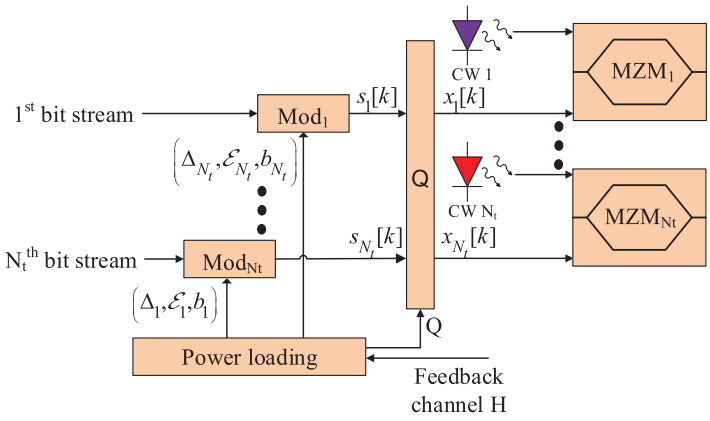


Fig. 3. Block diagram of the proposed transmitter. In V-BLAST MMSE-SIC, $\mathbf{Q} = \mathbf{I}_{N_t}$, whereas in SVD architecture, $\mathbf{Q} = \mathbf{V}$. For a specific i-th modulator: Δ_i is the maximum PAM symbol (PAM gain), $\mathcal{E}_i$ corresponds to the modulator energy, b_i denotes the number of transmitted bits per dimension, $s_i[k]$ is the modulator (PAM) output at the k-th sample. $x_i[k]$ is the signal (voltage) at the k-th sample after precoding matrix, $\mathbf{Q}$.

halfwave voltage, v_π. α_{loss} is calculated via

$$\alpha_{\text{loss}} = \frac{\pi E_0^2 \mathcal{E}_\phi \alpha_{\text{margin}} R_{\text{PD}}}{v_\pi}. \tag{11}$$

(9) can be written as an additive white Gaussian noise (AWGN) MIMO channel in matrix form, as

$$\mathbf{y}_k = \alpha_{\text{loss}} \mathbf{H} \mathbf{x}_k + \mathbf{n}_k, \tag{12}$$

where the received vector $\mathbf{y}_k \in \mathbb{R}^{N_r}$, the channel matrix $\mathbf{H} \in \mathbb{R}^{N_r \times N_t}$, the transmit vector $\mathbf{x}_k \in \mathbb{R}^{N_t}$, and the noise vector is $\mathbf{n}_k \in \mathbb{R}^{N_r}$. N_t and N_r correspond to the number of transmitters and receivers, respectively.

The channel matrix itself is a random matrix due to the speckles phenomenon introduced by the GI-MMF statistical propagation model. In order to properly calculate the channel matrix coefficients, h_{ji}, we use the narrow-band channel model seen in [24], as presented in Appendix A.

3. Power Loading Optimization Problem

In this section, we formulate a power loading optimization problem, which maximizes the system's capacity under the following set of constraints:

1) *Peak amplitude:* The modulated voltage, $\tilde{v}_{i,a}(t)$, of each i-th MZM should be in the linear regime of the MZM, i.e., $-\tilde{v}_{\text{peak}} \leqslant \tilde{v}_{i,a}(t) \leqslant \tilde{v}_{\text{peak}}$, $i = 1, 2, \ldots, N_t$.
2) *Power consumption:* The entire power consumption due to all MZMs is limited, i.e., $p_{\text{total}} \leqslant p_{\text{budget}}$.
3) *BER requirement:* $BER_i \leqslant BER_T$, $i = 1, 2, \ldots, N_t$, where BER_i is the bit error rate for the i-th bit stream, and BER_T denotes the target BER.

Moreover, the optimization problem is based upon a closed-loop MIMO architecture, i.e., the channel matrix is known to both transmitter and receiver, as seen in Fig. 3. In our analysis, we examine two types of architectures:

1) *The V-BLAST MMSE-SIC receiver:* In this case, the precoding matrix $\mathbf{Q}$ illustrated in Fig. 3 is identity, i.e., $\mathbf{Q} = \mathbf{I}_{N_t}$. The joint decoder is a bank of linear MMSE receivers, used for the detection of the corresponding parallel bit streams. During the detection process the receivers are activated sequentially, i.e., all decisions made by all prior receivers are used to remove from the input signal, as illustrated in Fig. 4 [34].

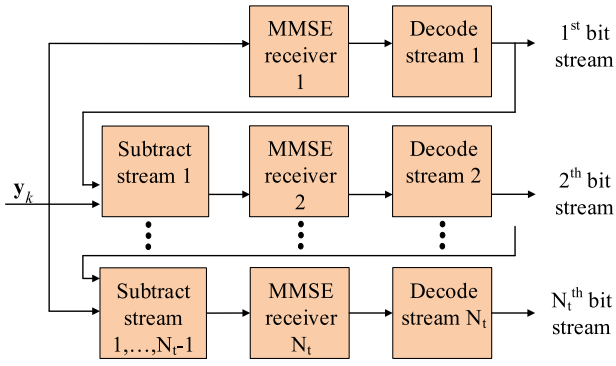


Fig. 4. Block diagram of the minimum mean square error and successive interference cancellation (MMSE-SIC) receiver, which is the joint decoder in the V-BLAST MMSE-SIC. During the detection process the receivers are activated sequentially, i.e., all decisions made by all prior receivers are used to remove from the input signal [34].

2) *The SVD architecture* [34]: In this case, the precoding matrix $\mathbf{Q}$ illustrated in Fig. 3 is $\mathbf{V} \in \mathbb{R}^{N_t \times N_t}$, i.e., $\mathbf{Q} = \mathbf{V}$ which is the right-singular vector of the channel matrix $\mathbf{H}$.

We start our capacity analysis without the peak amplitude constraint in a closed-loop MIMO architecture. Then, the general formulation of the peak amplitude constraint is presented. Afterwards, we continue to the complete problem formulation, which includes the peak amplitude constraint, for both suggested architectures.

3.1 Capacity in Closed Loop MIMO Without Peak Amplitude Constraint

In order to calculate the capacity of a block with N_t transmitters, we first write the mutual information between $\mathbf{x}_k$ and $\mathbf{y}_k$ given the AWGN MIMO model seen in (12), as [43]

$$I\left(\mathbf{x}_k; \mathbf{y}_k\right) = h\left(\mathbf{y}_k\right) - h\left(\mathbf{n}_k\right), \tag{13}$$

where $h(\cdot)$ is the differential entropy. Since $h\left(\mathbf{n}_k\right)$ is determined only by the noise distribution, maximizing $I\left(\mathbf{x}_k; \mathbf{y}_k\right)$ corresponds to maximizing $h\left(\mathbf{y}_k\right)$. Moreover, the Gaussian distribution maximizes the entropy over all distributions with the same covariance matrix, thus $\mathbf{y}_k$ is a Gaussian random vector. Consequently, $\mathbf{x}_k$ has Gaussian distribution, when we already know that $\mathbf{n}_k$ has Gaussian distribution. In addition, since the transmit vector, $\mathbf{x}_k$, and the noise, $\mathbf{n}_k$, are independent, the covariance of the output $\mathbf{y}_k$ is the sum of $\mathbf{H}\mathbf{x}_k$ and $\mathbf{n}_k$ covariance matrices, i.e.,

$$\mathbf{\Sigma_y} = \alpha_{\text{loss}}^2 \mathbf{H} \mathbf{\Sigma_x} \mathbf{H}^T + \mathbf{\Sigma_n}, \tag{14}$$

where $\mathbf{\Sigma_y}$, $\mathbf{\Sigma_x}$, and $\mathbf{\Sigma_n}$ are the covariance matrices of $\mathbf{y}_k$, $\mathbf{x}_k$, and $\mathbf{n}_k$, respectively. Thus, we can explicitly write the capacity as,

$$\begin{aligned} C &= \max I\left(\mathbf{x}_k; \mathbf{y}_k\right) \\ &= \frac{1}{2}\log_2\left((2\pi e)^{N_r}\left|\mathbf{\Sigma_n} + \alpha_{\text{loss}}^2 \mathbf{H}\mathbf{\Sigma_x}\mathbf{H}^T\right|\right) - \frac{1}{2}\log_2\left((2\pi e)^{N_r}\left|\mathbf{\Sigma_n}\right|\right), \end{aligned} \tag{15}$$

or, using the spatial-temporal whiteness of the noise, i.e., $\mathbf{\Sigma_n} = \mathbf{I}_{N_r} \cdot \sigma_n^2$ as

$$C = \frac{1}{2}\log_2\left|\mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\sigma_n^2}\mathbf{H}\mathbf{\Sigma_x}\mathbf{H}^T\right|. \tag{16}$$

A practical system must transmit at a bit rate lower than capacity due to modulator type, coding scheme and BER requirement. We adopt multi-level pulse-amplitude-modulator (PAM) as a

convenient standard modulator. Thus, the achievable spectral efficiency can be expressed as

$$b_{\max} = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma \sigma_n^2} \mathbf{H} \mathbf{\Sigma}_{\mathbf{x}} \mathbf{H}^T \right|, \tag{17}$$

where Γ is defined as the SNR gap. The SNR gap, introduced by Forney and Ungerboeck [44], Cioffi *et al.* [45], represents a loss with respect to Shannon capacity. The SNR gap analysis allows us to separate the coding gain from the power allocation. We assume the same SNR gap for all transmitted bit streams in both architectures. This relies on the concept that if one bit stream has significantly higher error probability, then it would dominate the BER [46]. Thus, our BER requirement is incorporated within the SNR gap assuming equal error probability in all the bit streams. For instance, for any uncoded AWGN single-input-single-output (SISO) channel, the SNR gap can be approximated by

$$\Gamma = \frac{\zeta^2}{3}, \tag{18}$$

where $\zeta = Q^{-1}(P_e)$, P_e is the symbol-error probability, and $Q^{-1}(\cdot)$ is the inverse Q function. We assume that the allowed symbol-error probability $P_e \approx BER_T$ for all parallel SISO channels. As an example, in the case of $BER_T = 10^{-6}$, the required SNR gap, Γ, is 8.8 dB. It is worth to mention that the pre-FEC BER requirement of the Ethernet 400GBASE-SR16 standard is a BER below 2.4×10^{-4} [47].

Besides the BER requirement, we are also limited by the allowed budget for power consumption due to all MZMs, i.e., $p_{\text{total}} \leqslant p_{\text{budget}}$. Using (8), we can re-write the power consumption constraint as

$$\mathbb{E}\,\|\mathbf{x}_k\|_2^2 \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}. \tag{19}$$

Or equivalently in matrix notation as,

$$\text{tr}\,(\mathbf{\Sigma}_{\mathbf{x}}) \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}, \tag{20}$$

Thus, we can define our capacity problem without the peak amplitude constraint, but with the power consumption constraint and BER requirement as

$$\text{maximize} \quad b_{\max} = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma \sigma_n^2} \mathbf{H} \mathbf{\Sigma}_{\mathbf{x}} \mathbf{H}^T \right| \tag{21a}$$

$$\text{subject to} \quad \text{tr}\,(\mathbf{\Sigma}_{\mathbf{x}}) \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}. \tag{21b}$$

Our analysis of achievable spectral efficiency splits into two possible architectures: the SVD, and the V-BLAST MMSE-SIC.

In the SVD architecture case, we perform an SVD of the matrix $\mathbf{H}$, i.e.,

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \tag{22}$$

where $\mathbf{U} \in \mathbb{R}^{N_r \times N_r}$ and $\mathbf{V} \in \mathbb{R}^{N_t \times N_t}$ are unitary matrices and $\mathbf{\Lambda} \in \mathbb{R}^{N_r \times N_t}$ is a rectangular matrix, whose diagonal elements are non-negative real numbers and whose off-diagonal elements are zero. The diagonal elements $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{N_t}$ are the ordered singular values of the matrix $\mathbf{H}$,

where we assume that $N_t \leqslant N_r$. Thus (17) turns into

$$\begin{aligned} b_{\max} &= \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2} \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T \right| \\ &= \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2} \boldsymbol{\Lambda}\mathbf{V}^T \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{V}\boldsymbol{\Lambda} \right| \\ &= \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2} \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathbf{s}}\boldsymbol{\Lambda} \right|, \end{aligned} \tag{23}$$

where in the last equation we defined

$$\boldsymbol{\Sigma}_{\mathbf{s}} \triangleq \mathbf{V}^T \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{V}. \tag{24}$$

Moreover, one should notice that

$$\begin{aligned} \text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{s}}\right) &= \text{tr}\left(\mathbf{V}^T \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{V}\right) \\ &= \text{tr}\left(\mathbf{V}\mathbf{V}^T \boldsymbol{\Sigma}_{\mathbf{x}}\right) \\ &= \text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}}\right). \end{aligned} \tag{25}$$

Thus, we have transformed the capacity problem without the peak amplitude constraint into

$$\text{maximize } b_{\max} = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2} \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathbf{s}}\boldsymbol{\Lambda} \right| \tag{26a}$$

$$\text{subject to } \text{tr}\left(\boldsymbol{\Sigma}_{\mathbf{x}}\right) \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}. \tag{26b}$$

It is worth mentioning that in general, if we apply Hadamard's inequality on (23), which states that the determinant of any positive definite matrix is less than the product of its diagonal elements [43], we receive

$$\begin{aligned} b_{\max} &= \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2} \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\mathbf{s}}\boldsymbol{\Lambda} \right| \\ &\leqslant \frac{1}{2}\sum_{n_t=1}^{N_t} \log_2 \left(1 + \frac{\alpha_{\text{loss}}^2 \lambda_{n_t}^2}{\Gamma\sigma_n^2} \mathcal{E}_{n_t,\mathbf{s}} \right), \end{aligned} \tag{27}$$

where $\mathcal{E}_{n_t,\mathbf{s}}$ is the n_t-th element on the diagonal of $\boldsymbol{\Sigma}_{\mathbf{s}}$. However, in the SVD architecture case, if we denote the modulator output vector at the k-th sample as $\mathbf{s}_k$, which is related to $\mathbf{x}_k$ in the SVD architecture through

$$\mathbf{x}_k = \mathbf{V}\mathbf{s}_k, \tag{28}$$

we receive that $\boldsymbol{\Sigma}_{\mathbf{s}}$ is the covariance matrix of the modulator output vector, $\mathbf{s}_k$. Moreover, in the SVD architecture case, $\boldsymbol{\Sigma}_{\mathbf{s}}$ is diagonal and denoted in the form

$$\boldsymbol{\Sigma}_{\mathbf{s},\text{SVD}} = \text{diag}\left(\left[\mathcal{E}_{1,\text{SVD}} \cdots \mathcal{E}_{N_t,\text{SVD}} \right]^T \right), \tag{29}$$

where $\mathcal{E}_{n_t,\text{SVD}}$ corresponds to the n_t-th modulator energy. In this special case, i.e., SVD architecture, for which $\boldsymbol{\Sigma}_{\mathbf{s}}$ is diagonal, the inequality in (27) becomes an equality. Thus, using (27) as equality in

the SVD architecture, we can re-write (26a)–(26b) as

$$\text{maximize} \quad b_{\max} = \frac{1}{2}\sum_{n_t=1}^{N_t} \log_2\left(1 + \frac{\alpha_{\text{loss}}^2 \lambda_{n_t}^2}{\Gamma\sigma_n^2}\mathcal{E}_{n_t,\text{SVD}}\right) \tag{30a}$$

$$\text{subject to} \quad \sum_{n_t=1}^{N_t} \mathcal{E}_{n_t,\text{SVD}} \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}, \tag{30b}$$

$$\left[\mathcal{E}_{1,\text{SVD}} \cdots \mathcal{E}_{N_t,\text{SVD}}\right]^T \geqslant 0. \tag{30c}$$

where (30c) explicitly denotes the former implicit assumption that modulator energy is positive. We have reached a classical water-filling problem with the SVD architecture without the peak amplitude constraint.

On the other hand, in the V-BLAST MMSE-SIC case, because the transmitted bit streams are parallel and the precoding matrix $\mathbf{Q} = \mathbf{I}_{N_t}$, the covariance matrix of $\mathbf{x}_k$ is diagonal and denoted in the form

$$\mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}} = \text{diag}\left(\left[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}\right]^T\right), \tag{31}$$

where $\mathcal{E}_{n_t,\text{V-BLAST}}$ corresponds to the n_t-th modulator energy. As shown in Appendix B, (17) has a relation to the signal-to-interference-plus-noise ratio (SINR) in each V-BLAST MMSE-SIC's i-th iteration [34]

$$\begin{aligned} b_{\text{V-BLAST}} &= \frac{1}{2}\log_2\left|\mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2}\mathbf{H}\mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}}\mathbf{H}^T\right| \\ &= \frac{1}{2}\sum_{i=1}^{N_t}\log_2\left(1 + \text{SINR}_i\right). \end{aligned} \tag{32}$$

This means that V-BLAST MMSE-SIC, achieves the optimal sum-rate of a block N_t as it demodulates each i-th bit stream with the SINR figure of merit. Moreover, (32) is a concave function with respect to a diagonal covariance $\mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}}$ as shown in Appendix C. Thus, we can re-write our capacity problem without the peak amplitude constraint in the V-BLAST MMSE-SIC case based upon (21a)–(21b), as

$$\text{maximize} \quad b_{\text{V-BLAST}} = \frac{1}{2}\log_2\left|\mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma\sigma_n^2}\mathbf{H}\mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}}\mathbf{H}^T\right| \tag{33a}$$

$$\text{subject to} \quad \sum_{n_t=1}^{N_t} \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}, \tag{33b}$$

$$\left[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}\right]^T \geqslant 0. \tag{33c}$$

This is a convex optimization problem with respect to the optimization vector $[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}]^T$.

To summarize this section, it is evident that capacity in closed loop MIMO without the peak amplitude constraint in the SVD architecture is equal or superior to the V-BLAST MMSE-SIC architecture with the same system parameters. This originates from the "architectural" constraint that $\mathbf{\Sigma}_\mathbf{x}$ is diagonal in the V-BLAST MMSE-SIC architecture, whereas in the SVD architecture the diagonal architectural constraint is on $\mathbf{\Sigma}_\mathbf{s}$ and leads to no capacity loss due to Hadamard's inequality.

3.2 Peak Amplitude Constraint - General Formulation

In this section, we take a closer look at the peak amplitude constraint as stated in (3). As it can be easily noticed that using (5) and sinc pulse shaping function, it can be written as

$$\forall k, \quad i = 1, 2, \ldots, N_t : \qquad -\tilde{v}_{\text{peak}} \leqslant x_i[k] \leqslant \tilde{v}_{\text{peak}}. \tag{34}$$

Thus, the peak constraint becomes a discrete PAM symbol gain constraint for all the transmitters after precoding. In order to understand this statement better, we first, for notational convenience, turn into matrix form and re-write (34) as

$$\forall k : \qquad -\tilde{\mathbf{v}}_{\text{peak}} \leqslant \mathbf{x}_k \leqslant \tilde{\mathbf{v}}_{\text{peak}}, \tag{35}$$

where $\tilde{\mathbf{v}}_{\text{peak}} \in \mathbb{R}^{N_t}$ is the peak constraint vector with $\tilde{v}_{\text{peak}}$ in each of its entries. In addition, in our n_t-th PAM, we denote the number of signal constellation points as M_{n_t}, the number of transmitted bits per dimension as b_{n_t}, and the maximum PAM symbol value, or PAM gain, as Δ_{n_t}. In matrix form, we denote the PAM gain vector as

$$\boldsymbol{\Delta} \triangleq \left[\Delta_1 \cdots \Delta_{N_t}\right]^T. \tag{36}$$

Generally, our transmitted vector, $\mathbf{x}_k$, is related to the modulator output vector, $\mathbf{s}_k$, through a general precoding matrix, $\mathbf{Q}$, thus receiving,

$$\mathbf{x}_k = \mathbf{Q}\mathbf{s}_k. \tag{37}$$

Here, it should be noted that the maximum possible output voltage into each n_t-th MZM originates from the cumulative effect of all PAMs' gain multiplied by their associated precoder matrix coefficients. Thus, one can re-write the peak amplitude constraint in (35) as

$$\text{abs}(\mathbf{Q}) \cdot \boldsymbol{\Delta} \leqslant \tilde{\mathbf{v}}_{\text{peak}}. \tag{38}$$

where abs $(\cdot)$ denotes an element-wise absolute operator. In order to express (38), in terms of the modulators' energy, one should first notice that there is a simple relation between the n_t-th modulator energy, $\mathcal{E}_{n_t,\text{PAM}}$, to its gain, Δ_{n_t}, and number of signal constellation points, M_{n_t}, which is [24], [31]

$$\mathcal{E}_{n_t,\text{PAM}} = \frac{\Delta_{n_t}^2}{3}\left(\frac{M_{n_t}+1}{M_{n_t}-1}\right). \tag{39}$$

Thus, we can re-phrase (38) using (39), in terms of the modulators' energy, which yields

$$\forall n_t : \qquad \sum_{i=1}^{N_t} \left|[\mathbf{Q}]_{n_t,i}\right| \sqrt{3\mathcal{E}_{i,\text{PAM}}\left(\frac{M_i-1}{M_i+1}\right)} \leqslant \tilde{v}_{\text{peak}}. \tag{40}$$

3.3 Optimal Power Loading Using SVD With Peak Amplitude Constraint

In this section, we formulate the power loading optimization problem under all constraints including the peak amplitude constraint for the SVD architecture. We have already received a convex optimization problem using the power consumption constraint and BER requirement, as stated in (30a)–(30c) in terms of the modulators' energy. Moreover, we have described for a general precoding matrix, $\mathbf{Q}$, the peak amplitude constraint in terms of the modulators' energy in (40). Thus, we receive that for the SVD architecture case, for which, $\mathbf{Q} = \mathbf{V}$, the power loading optimization

problem under all constraints is

$$\text{maximize} \quad \frac{1}{2}\sum_{n_t=1}^{N_t}\log_2\left(1+\frac{\alpha_{\text{loss}}^2\lambda_{n_t}^2}{\Gamma\sigma_n^2}\mathcal{E}_{n_t,\text{SVD}}\right) \tag{41a}$$

$$\text{subject to} \quad \sum_{n_t=1}^{N_t}\mathcal{E}_{n_t,\text{SVD}} \leqslant \frac{T}{C_{\text{MZM}}}p_{\text{budget}}, \tag{41b}$$

$$\left[\mathcal{E}_{1,\text{SVD}}\cdots\mathcal{E}_{N_t,\text{SVD}}\right]^T \geqslant 0, \tag{41c}$$

$$\forall n_t : \sum_{i=1}^{N_t}\left|[\mathbf{V}]_{n_t,i}\right|\sqrt{3\mathcal{E}_{i,\text{SVD}}\left(\frac{M_i-1}{M_i+1}\right)} \leqslant \tilde{v}_{\text{peak}}. \tag{41d}$$

This problem is not convex under the modulators' energy optimization variable due to the peak amplitude constraint in (41d). In order to make this problem mathematically tractable, we assume that the number of transmitted bits per dimension, b_{n_t} of the n_t-th modulator is related to the number of signal constellation points, M_{n_t} using

$$M_{n_t} = 2^{b_{n_t}}, \tag{42}$$

as assumed in [24], [31]. Considering the fact that the capacity objective function is the sum of all transmitted bits per dimension over all N_t bit streams and using (41a)

$$b_{n_t} = \frac{1}{2}\log_2\left(1+\frac{\alpha_{\text{loss}}^2\lambda_{n_t}^2}{\Gamma\sigma_n^2}\mathcal{E}_{n_t,\text{SVD}}\right), \tag{43}$$

for each n_t-th modulator. Furthermore, using (39), (42) and (43) and some simple mathematical manipulations, it can be shown that

$$b_{n_t} = \log_2\left(1+\frac{\alpha_{\text{loss}}\lambda_{n_t}}{\sqrt{3\Gamma}\sigma_n}\Delta_{n_t,\text{SVD}}\right), \tag{44}$$

$$\mathcal{E}_{n_t,\text{SVD}} = \frac{\Delta_{n_t}^2}{3} + 2\Delta_{n_t,\text{SVD}}\sqrt{\frac{\Gamma}{3}}\frac{\sigma_n}{\alpha_{\text{loss}}\lambda_{n_t}}. \tag{45}$$

Thus, we can reformulate the power loading optimization problem, (41a)–(41d), using the PAM gain vector, $\boldsymbol{\Delta}$, optimization variable and receive

$$\text{maximize} \quad \sum_{n_t=1}^{N_t}\log_2\left(1+\frac{\alpha_{\text{loss}}\lambda_{n_t}}{\sqrt{3\Gamma}\sigma_n}\Delta_{n_t,\text{SVD}}\right) \tag{46a}$$

$$\text{subject to} \quad \sum_{n_t=1}^{N_t}\left(\frac{\Delta_{n_t,\text{SVD}}^2}{3} + 2\Delta_{n_t,\text{SVD}}\sqrt{\frac{\Gamma}{3}}\frac{\sigma_n}{\alpha_{\text{loss}}\lambda_{n_t}}\right) \leqslant \frac{T}{C_{\text{MZM}}}p_{\text{budget}}, \tag{46b}$$

$$\left[\Delta_{1,\text{SVD}}\cdots\Delta_{N_t,\text{SVD}}\right]^T \geqslant 0, \tag{46c}$$

$$\text{abs}(\mathbf{V})\cdot\boldsymbol{\Delta}_{\text{SVD}} \leqslant \tilde{\mathbf{v}}_{\text{peak}}. \tag{46d}$$

This problem is a convex optimization problem with respect to the optimization vector $\boldsymbol{\Delta}_{\text{SVD}}$.

3.4 Optimal Power Loading Using V-BLAST MMSE-SIC With Peak Amplitude Constraint

Following our power loading optimization problem formulation for SVD architecture, here, we formulate the power loading optimization problem under all constraints including the peak amplitude constraint for the V-BLAST MMSE-SIC. We have already received a convex optimization problem using the power consumption constraint and BER requirement, as stated in (33a)–(33c) in terms

of the modulators' energy. Moreover, we have described for a general precoding matrix, $\mathbf{Q}$, the peak amplitude constraint in terms of the modulators' energy in (40). Thus, we receive that for the V-BLAST MMSE-SIC case, for which, $\mathbf{Q} = \mathbf{I}_{N_t}$, the power loading optimization problem under all constraints is

$$\text{maximize } b_{\text{V-BLAST}} = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma \sigma_n^2} \mathbf{H} \mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}} \mathbf{H}^T \right| \tag{47a}$$

$$\text{subject to } \sum_{n_t=1}^{N_t} \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}, \tag{47b}$$

$$\left[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}\right]^T \geqslant 0, \tag{47c}$$

$$\forall n_t : \sqrt{3\mathcal{E}_{n_t,\text{V-BLAST}} \left(\frac{M_{n_t} - 1}{M_{n_t} + 1} \right)} \leqslant \tilde{v}_{\text{peak}}. \tag{47d}$$

One can rephrase the peak amplitude constraint in (47d) as

$$\forall n_t : \quad \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{\tilde{v}_{\text{peak}}^2}{3} \left(\frac{M_{n_t} + 1}{M_{n_t} - 1} \right). \tag{48}$$

For the sake of mathematical tractability, we will assume that $M_{n_t} \gg 1$. Thus, we make the peak amplitude constraint a bit tighter than necessary, i.e.,

$$\forall n_t : \quad \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{\tilde{v}_{\text{peak}}^2}{3}. \tag{49}$$

To conclude, the power loading optimization problem using V-BLAST MMSE-SIC under all constraints is given by

$$\text{maximize } b_{\text{V-BLAST}} = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma \sigma_n^2} \mathbf{H} \mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}} \mathbf{H}^T \right| \tag{50a}$$

$$\text{subject to } \sum_{n_t=1}^{N_t} \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{T}{C_{\text{MZM}}} p_{\text{budget}}, \tag{50b}$$

$$\left[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}\right]^T \geqslant 0, \tag{50c}$$

$$\forall n_t : \mathcal{E}_{n_t,\text{V-BLAST}} \leqslant \frac{\tilde{v}_{\text{peak}}^2}{3}. \tag{50d}$$

This problem is convex with respect to the optimization vector $\left[\mathcal{E}_{1,\text{V-BLAST}} \cdots \mathcal{E}_{N_t,\text{V-BLAST}}\right]^T$.

4. System Simulation

In this section, we perform quantitative analysis of the power loading optimization problem, which maximizes the system's capacity under the following set of constraints: peak amplitude, power consumption, and BER requirement. The analysis of both V-BLAST MMSE-SIC and SVD architectures is performed using 1310 nm and 1550 nm wavelengths with the MGDM system presented in Fig. 1. The system simulation parameters are given in Table 1.

The channel matrices for both 1310 nm and 1550 nm are calculated as described in Appendix A. Due to the statistical nature of the channel matrix, its mean and standard deviation are based upon

TABLE 1
System Simulation's Baseline Parameters

Parameter	Value	Description
Γ	8.8 dB ($BER_T = 10^{-6}$)	SNR gap
i_c	10 pA/$\sqrt{\text{Hz}}$	Noise current spectral density
N_t	4	Number of transmitters
N_r	4	Number of receivers
T	100 psec	Symbol duration
R_{PD}	0.45 A/W [28]	MSD responsivity
α_{margin}	0.25 [48]	Link budget power margin
E_0^2	1 mW	CW laser power inc. IL
v_π	3 V	MZM's halfwave voltage
C_{MZM}	27 pF [38]	MZM capacitance
$\tilde{v}_{\text{peak}}$	0.25 V	MZM peak voltage amplitude allowed
p_{budget}	0–100 mW	Total power consumption allowed

10000 random realizations for 1310 nm and 1550 nm are given by

$$\mathbf{H}_{1310} = \begin{bmatrix} 0.7921 & 0.3192 & 0.1326 & 0.0823 \\ 0.2004 & 0.4881 & 0.2884 & 0.1791 \\ 0.0075 & 0.1824 & 0.4559 & 0.3501 \\ 0.0000 & 0.0102 & 0.1231 & 0.3886 \end{bmatrix}, \tag{51}$$

$$\mathbf{H}_{1550} = \begin{bmatrix} 0.7601 & 0.3217 & 0.1414 & 0.0978 \\ 0.2236 & 0.4818 & 0.2974 & 0.2055 \\ 0.0161 & 0.1843 & 0.4220 & 0.3535 \\ 0.0002 & 0.0122 & 0.1392 & 0.3432 \end{bmatrix}, \tag{52}$$

$$\mathbf{H}_{1310\text{-std}} = \begin{bmatrix} 0.0338 & 0.0897 & 0.0375 & 0.0220 \\ 0.0285 & 0.0708 & 0.0463 & 0.0286 \\ 0.0059 & 0.0496 & 0.0543 & 0.0361 \\ 0.0000 & 0.0039 & 0.0267 & 0.0492 \end{bmatrix}, \tag{53}$$

$$\mathbf{H}_{1550\text{-std}} = \begin{bmatrix} 0.0829 & 0.0917 & 0.0405 & 0.0270 \\ 0.0666 & 0.0695 & 0.0489 & 0.0336 \\ 0.0179 & 0.0494 & 0.0551 & 0.0388 \\ 0.0003 & 0.0046 & 0.0306 & 0.0509 \end{bmatrix}, \tag{54}$$

where $\mathbf{H}_{1310}$ and $\mathbf{H}_{1550}$ denote the 1310 nm and 1550 nm mean channel matrices, respectively. Furthermore, $\mathbf{H}_{1310\text{-std}}$ and $\mathbf{H}_{1550\text{-std}}$ denote the 1310 nm and 1550 nm standard-deviation channel matrices, respectively.

As mentioned in [24], the channel matrices are dominated by their mean values, while their standard-deviation is relatively small. There are several possible approaches to deal with the standard-deviation in practice, such as increasing the SNR gap, or employing a fast tracking mechanism. The following results will be based upon the ensemble of possible channel matrices, similar to [22], [24].

Moreover, all the results of the convex power optimization problems shown in this section are solved using CVX, a package for specifying and solving convex problems [49].

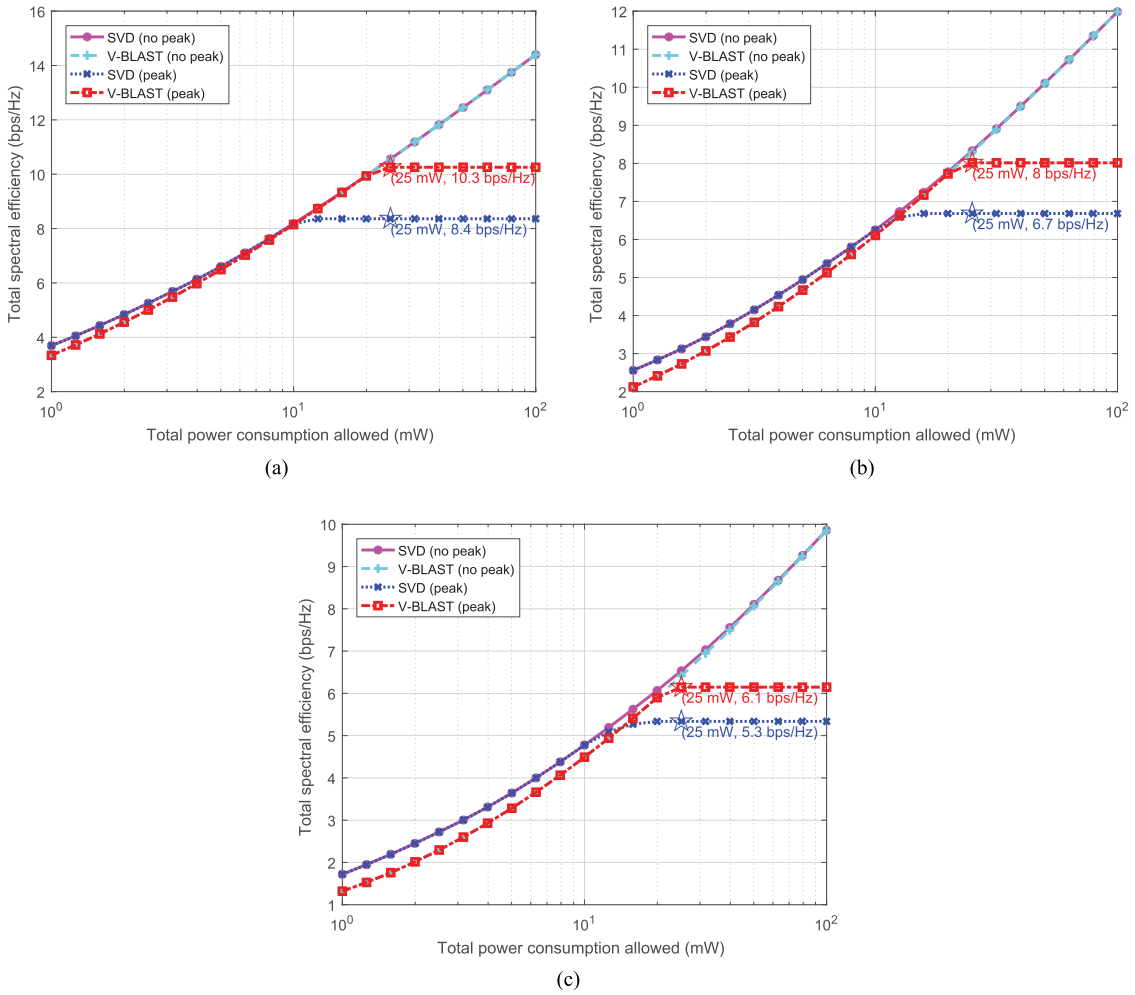


Fig. 5. Total spectral efficiency according to the parameters in Table 1 with 1310 nm wavelength for 4 different power optimization problems: SVD without peak amplitude constraint, (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, (33a)–(33c); SVD with all constraints, (46a)–(46d); V-BLAST MMSE-SIC with all constraints, (50a)–(50d); where (a) $BER_T = 10^{-3}$, (b) $BER_T = 10^{-6}$, (c) $BER_T = 10^{-12}$.

Generally, the results shown in Figs. 5–9 are based upon a solution to 4 different power optimization problems: SVD without peak amplitude constraint, given by (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, given by (33a)–(33c); SVD with all constraint, given by (46a)–(46d); V-BLAST MMSE-SIC with all constraints, given by (50a)–(50d). Moreover, in Figs. 5–9, as mentioned in Section 3-1, the maximum achievable spectral efficiency is upper bounded by the SVD without peak amplitude constraint.

In Fig. 5, we show the total spectral efficiency vs. the total power consumption allowed, p_{budget}, for 3 different pre-FEC BER requirements, i.e., 10^{-3}, 10^{-6}, 10^{-12}. Generally, as expected, stringent BER requirements leads to lower total spectral efficiency in all architectures. This is quite intuitive, as a stricter BER requirement means a higher SNR requirement. In turn, this leads to the usage of lower modulation format alphabet size or stronger FEC, which lowers the total spectral efficiency.

A different observation seen in Fig. 5 is that the total spectral efficiency using SVD with and without the peak amplitude constraint are almost equal in the low power consumption region, i.e., up to about 10–15 mW; similar behavior is shown for the V-BLAST MMSE-SIC. This means that in the low power consumption region the peak amplitude constraint is not a dominant factor.

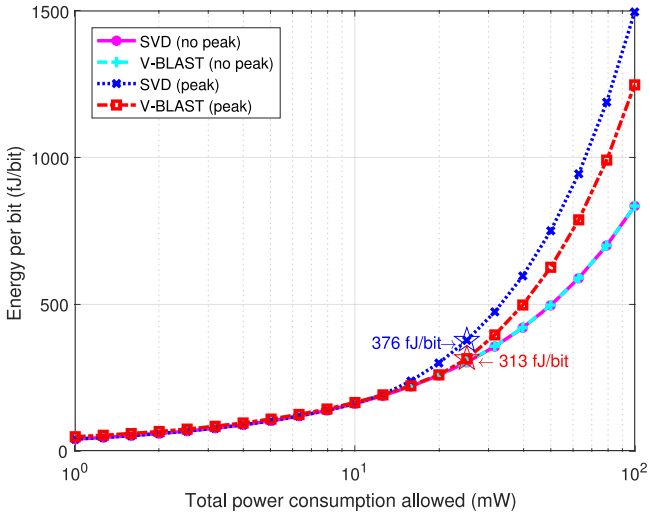


Fig. 6. Energy per bit according to the parameters in Table 1 with 1310 nm wavelength for 4 different power optimization problems: SVD without peak amplitude constraint, (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, (33a)–(33c); SVD with all constraints, (46a)–(46d); V-BLAST MMSE-SIC with all constraints, (50a)–(50d).

Furthermore, in the low power consumption region, stringent BER requirement leads to more pronounced total spectral efficiency SVD superiority over V-BLAST MMSE-SIC. This is the power consumption region where the precoding matrix, $\mathbf{V}$, available in the SVD architecture is helping to achieve the maximum theoretical capacity when the peak amplitude constraint is not a factor, as described in Section 3-1. Nevertheless, as the total power consumption allowed is larger, the precoding matrix, $\mathbf{V}$, becomes less important, closing the spectral efficiency gap between SVD and V-BLAST MMSE-SIC architectures.

Examining the power consumption region between 10–15 mW and up to 25 mW in Fig. 5 raises a different issue. In this power consumption region, the V-BLAST MMSE-SIC is superior to the SVD with peak amplitude constraint and is quite close to the SVD without peak amplitude constraint upper bound. This is the peak amplitude constraint dominant region. In this region, the SVD with peak amplitude constraint is limited by the abs($\mathbf{V}$) in (46d), which makes the feasibility region of this optimization problem more restrictive than in the V-BLAST MMSE-SIC case. It is quite intriguing to notice that while there is no peak amplitude constraint, we can only benefit from the use of a precoding matrix, $\mathbf{V}$; on the other hand, when there is a peak amplitude constraint, it becomes a nuisance and limits our total spectral efficiency.

Moreover, in Fig. 5, increasing the power consumption allowed above 25 mW cannot improve the total spectral efficiency as we are limited by the peak amplitude constraint. Thus, as seen in Fig. 6, the energy per bit in the region above 25 mW deteriorates dramatically, and is mostly pronounced in the SVD peak amplitude constraint case. This is the peak amplitude constraint limited power consumption region.

The same analysis and classification using 3 power consumption regions, i.e., low power consumption region, peak amplitude constraint dominant power consumption region, and peak amplitude constraint limited power consumption region, is valid for other scenarios as well. For example, in Fig. 7, using a channel matrix of 1550 nm shows similiar behavior to a channel matrix of 1310 nm. Similarly, in Fig. 8, using a MZM's halfwave voltage, $v_\pi = 6$ V, with MZM peak voltage amplitude allowed, $\tilde{v}_{\text{peak}} = 0.5$ V, also reveals the 3 power consumption regions.

An interesting case was analyzed in Fig. 9, as we lowered the CW laser power inc. IL, E_0^2, to 0.25 mW. In this case, the SNR in the receiver became so low as compared to the other analyzed scenarios causing the precoding matrix, $\mathbf{V}$, to be beneficial in spite of the peak amplitude

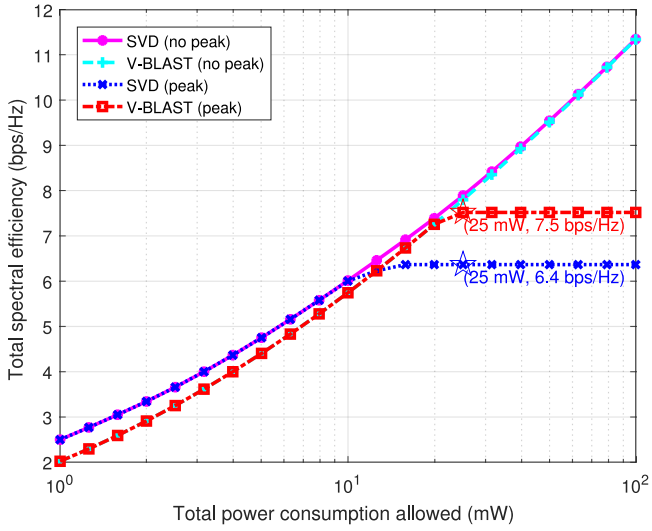


Fig. 7. Total spectral efficiency according to the parameters in Table 1 with 1550 nm wavelength for 4 different power optimization problems: SVD without peak amplitude constraint, (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, (33a)–(33c); SVD with all constraints, (46a)–(46d); V-BLAST MMSE-SIC with all constraints, (50a)–(50d).

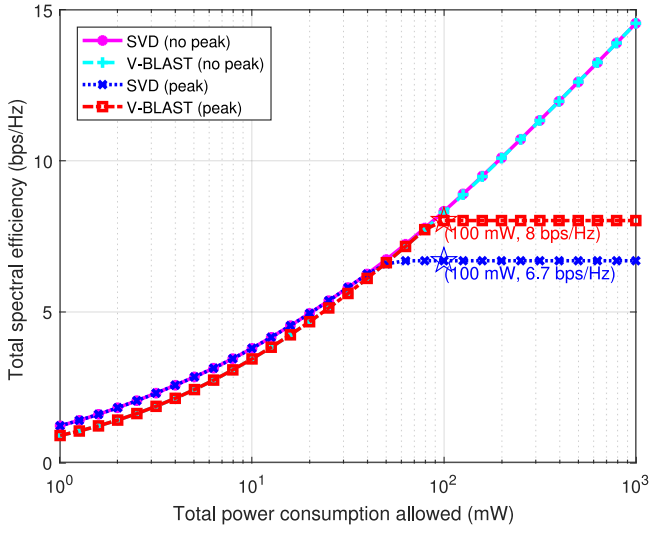


Fig. 8. Total spectral efficiency according to the parameters in Table 1 with 1310 nm wavelength, $v_\pi = 6$ V, and $\tilde{v}_{\text{peak}} = 0.5$ V, for 4 different power optimization problems: SVD without peak amplitude constraint, (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, (33a)–(33c); SVD with all constraints, (46a)–(46d); V-BLAST MMSE-SIC with all constraints, (50a)–(50d).

constraint's dominance. Consequently, the SVD architecture is superior over the entire total allowed power consumption range for this extreme case.

To conclude, for a conventional scenario using the parameters in Table 1 along with 1310 nm channel matrix, the maximum spectral efficiency with total power consumption of 25 mW is an 8 bps/Hz (313 fJ/bit) for a V-BLAST MMSE-SIC compared to a 6.7 bps/Hz (376 fJ/bit) using an SVD architecture.

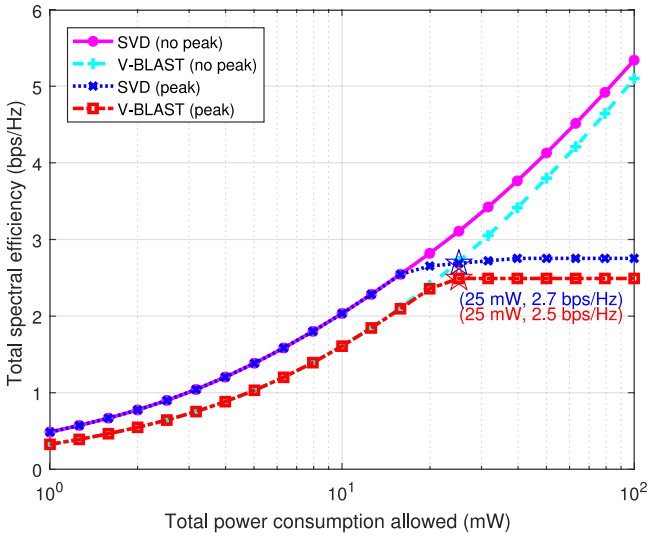


Fig. 9. Total spectral efficiency according to the parameters in Table 1 with 1310 nm wavelength, $E_0^2 = 0.25$ mW, for 4 different power optimization problems: SVD without peak amplitude constraint, (30a)–(30c); V-BLAST MMSE-SIC without peak amplitude constraint, (33a)–(33c); SVD with all constraints, (46a)–(46d); V-BLAST MMSE-SIC with all constraints, (50a)–(50d).

5 Conclusion

In this paper, we have formulated an analytical framework to optimize the power loading in a peak limited IM/DD MGDM optical link. It relaxes performance requirements from the spatial multiplexer, allowing the channel matrix to be non-diagonal. The compensation for the associated imperfection is handled via MIMO digital signal processing.

Furthermore, we have formulated convex power loading optimization problems for V-BLAST MMSE-SIC and SVD architectures. In addition, we have shown the superiority of a V-BLAST MMSE-SIC over the SVD architecture for a 4×4 MIMO short reach low driving voltage conventional SiP based MZM system.

Appendix A
Channel Matrix Coefficients

The channel matrix used in this paper is a random matrix due to the speckles phenomenon introduced by the GI-MMF statistical propagation model. In order to properly calculate the channel matrix coefficients, h_{ji}, we use the aforementioned model [24], as described in this appendix.

A.1 Transmitter to GI-MMF Power Coupling

First, an overlap integral is used between the tilted beams transmission and the GI-MMF's Laguerre-Gaussian (LG) mode functions. This calculation results in a power-coupling coefficient (PCC), i.e., the expected power in the *n*-th output mode, upon launching unity power into the *i*-th input. More specifically, if we denote the PCC's overlap integral of the *i*-th MZM for GI-MMF *n*-th mode (ν, μ, o) as $\eta_{i,\nu\mu o}$ it equals

$$\eta_{i,\nu\mu o} = \frac{\left|\iint_{A\,\text{core}} E_{i,\text{MZM}}(r,\phi)\cdot E^{*}_{\nu\mu o}(r,\phi,z=0) r dr d\phi\right|^2}{\iint_{A\,\text{core}} \left|E_{i,\text{MZM}}\right|^2 r dr d\phi \iint_{A\,\text{core}} \left|E_{\nu\mu o}(r,\phi,z=0)\right|^2 r dr d\phi}, \tag{55}$$

where $E_{i,\text{MZM}}$ describes the incident electrical field from the *i*-th MZM, $E_{\nu\mu o}$ represents the GI-MMF's mode with an infinite parabolic refractive index profile, A_{core} is the fiber facet indicating

the truncation of the Gaussian beam, and $(\cdot)^*$ designates complex conjugate. The LG mode functions in Cylindrical coordinates (r, ϕ, z) obeys

$$E_{\nu\mu o} = \psi_{\nu\mu}(r) \begin{Bmatrix} \cos(\nu\phi) \\ \sin(\nu\phi) \end{Bmatrix} e^{-j\beta_{\nu\mu} z}, \tag{56}$$

introducing the modal propagation constant $\beta_{\nu\mu}$; the azimuthal order $\nu \geqslant 0$; radial order $\mu \geqslant 1$; and the orientation $o \in \{a, b\}$, where a stands for a cosine solution, and b for a sine solution. The spatial dependence of the transverse modal electric field is given by

$$\psi_{\nu\mu}(r) = C_{\nu\mu}(r/r_{\text{co}})^{\nu} \exp(-0.5V(r/r_{\text{co}})^2) L_{\mu-1}^{\nu}(V(r/r_{\text{co}})^2), \tag{57}$$

where $L_{\mu-1}^{\nu}$ is generalized Laguerre polynomial, and normalization constant is set to

$$C_{\nu\mu} = \left(\frac{2V^{\nu+1}(\mu-1)!}{\pi(1+\delta_{0\nu}) r_{\text{co}}^2 (\mu+\nu-1)!} \right)^{1/2}. \tag{58}$$

where $\delta_{0\nu}$ is the Kronecker delta function, r_{co} is the core radius of the GI-MMF, and $V = k_0 n_{\text{co}} r_{\text{co}} \sqrt{2\Delta_{\text{GI-MMF}}}$ is the normalized frequency. k_0 is wavenumber in free space; the profile height parameter is given by $\Delta_{\text{GI-MMF}} = (n_{\text{co}}^2 - n_{\text{cl}}^2)/2n_{\text{co}}^2 \approx (n_{\text{co}} - n_{\text{cl}})/n_{\text{co}}$, where n_{co} and n_{cl} are the refractive indices of the core and cladding, respectively. We will refer to LG modes with the same or nearly the same propagation constant as degenerate mode group, which obeys the relationship $m = 2\mu + \nu - 1$, where m is the degenerate mode group number.

A.2 GI-MMF Statistical Propagation Model

The PCCs between each transmitter to the LG mode functions are the basis for the statistical properties of the GI-MMF propagation model.

Our statistical model assumes that each modal power has a negative-exponential distribution independent of the other modes, and the modal phase is uniformly distributed [26], [29], [50]. Thus, the spatial distribution of each mode is expressed using a zero-mean circular Gaussian coefficient associated with each LG mode function [26], [29]. Moreover, the LG coefficient's variance is governed by the PCC and GI-MMF propagation properties.

The propagation through the GI-MMF can be described using the notion of degenerate mode groups, where a degenerate mode group is all the modes sharing a common propagation constant, β. All modes sharing the same propagation constant are fully mixed [51], i.e., they have the same LG coefficient's variance. On the other hand, different degenerate mode groups are well separated in terms of propagation constants, thus have negligible crosstalk between them [22], [52]. Practically, as mentioned in [52] and [53], these assumptions are valid within a fiber length of some hundreds or even thousands of meters.

The aforementioned model allows us to calculate the LG coefficient's variance of each mode using the PCCs. We can calculate the PCC to each degenerate mode group as the average of PCCs associated with the specific degenerate mode group, which will be denoted $\tilde{\eta}_{i,m}$.

Once all the modes' statistical parameters are known, we can generate random spatial distributions associated with each i-th MZM post GI-MMF propagation, i.e.,

$$E_{i,\text{GI-MMF}}(r, \phi) = \sum_{\forall m} C_m \cdot \left(\sum_{\forall \nu\mu o \in \{m = 2\mu+\nu-1\}} a_{\nu\mu o}(\tilde{\eta}_{i,m}) \cdot E_{\nu\mu o}(r, \phi) \right), \tag{59}$$

where $a_{\nu\mu o}(\tilde{\eta}_{i,m})$ is the LG zero mean circular Gaussian coefficient with a variance $\tilde{\eta}_{i,m}$. C_m is a normalization factor to maintain energy conservation within each m-th degenerate mode group, i.e.,

$$C_m = \sqrt{\tilde{\eta}_{i,m}} \left(\iint_{A\,\text{core}} \left| \sum_{\forall \nu\mu o \in \{m = 2\mu+\nu-1\}} a_{\nu\mu o}(\tilde{\eta}_{i,m}) \cdot E_{\nu\mu o}(r, \phi) \right|^2 r dr d\phi \right)^{-1}. \tag{60}$$

Thus, the power distribution of the NFP caused by the i-th MZM on the fiber output is given by

$$I_{i,\text{GI-MMF}}(r,\phi) = E_{i,\text{GI-MMF}}(r,\phi) \cdot E^{*}_{i,\text{GI-MMF}}(r,\phi). \tag{61}$$

Due to the incoherency of the MZMs, we can analyze each i-th MZM spatial distribution separately.

A.3 Spatial Demultiplexing

Once we have the power distribution of each i-th MZM on the fiber output, spatial demultiplexing is carried out using an MSD [22], [28]. Thus, the channel matrix coefficients, h_{ji}, can be estimated by

$$h_{ji} = \frac{\int\int_{R_j} I_{i,\text{GI-MMF}}(r,\phi)\, r dr d\phi}{\int_0^{2\pi}\int_0^{r_{\text{co}}} I_{i,\text{GI-MMF}}(r,\phi)\, r dr d\phi}, \tag{62}$$

where R_j is the area of the j-th receiver photo-detector segment. The MSD's radiuses are chosen in the intersection of the beams' normalized power distribution, as used in [22], [24].

A.4 Modal Group Dispersion

In this paper, we assume that the modal dispersion caused by different degenerate mode groups is small compared to the symbol time. This is a narrow-band model used in [36]. The delay propagation for each degenerate mode group, τ_m, is calculated via: [54]

$$\begin{aligned}\tau_m &= \frac{1}{c}\frac{\partial \beta_m}{\partial k_0} \\ &\approx \frac{n_{\text{co}}}{c}\left[1 + \frac{\alpha_{\text{GI-MMF}}-2}{\alpha_{\text{GI-MMF}}+2}\Delta_{\text{GI-MMF}}\left(\frac{m}{M_{\text{GI-MMF}}}\right)^{2\alpha_{\text{GI-MMF}}/(\alpha_{\text{GI-MMF}}+2)}\right. \\ &\quad \left. + \frac{3\alpha_{\text{GI-MMF}}-2}{\alpha_{\text{GI-MMF}}+2}\Delta^2_{\text{GI-MMF}}\left(\frac{m}{M_{\text{GI-MMF}}}\right)^{4\alpha_{\text{GI-MMF}}/(\alpha_{\text{GI-MMF}}+2)}\right],\end{aligned} \tag{63}$$

where $\alpha_{\text{GI-MMF}}$ characterizes the shape of the GI-MMF core index profile, c is the speed of light, and $M_{\text{GI-MMF}}$ is the total number of degenerate mode groups. For example, a pure alpha profile of $\alpha_{\text{GI-MMF}} = 1.97$, with $M_{\text{GI-MMF}} = 12$ corresponding to a standard 50 μm GI-MMF with 1310 nm wavelength, and $n_{\text{co}} = 1.45$, resulting in a delay spread of 0.12 ps/m, which theoretically allows a 10 GBaud with 300 m link. Nevertheless, a practical link will show a much larger delay spread as seen in [51] using the 108 fiber model [55]. On the other hand, each receiver segment in the MSD may be designed with a different delay, thus reducing the overall delay spread.

Appendix B
V-BLAST MMSE-SIC Capacity to SINR Relation

In this appendix, we prove the relation between the achievable spectral efficiency in a V-BLAST MMSE-SIC architecture to the summation of the signal-to-interference-plus-noise ratio (SINR) in each V-BLAST MMSE-SIC's i-th iteration.

If $\mathbf{A}$ is an invertible square matrix and $\mathbf{x}$ is a column vector. Then the matrix determinant lemma states [56] that

$$\left|\mathbf{A} + \mathbf{x}\mathbf{x}^{T}\right| = \left(1 + \mathbf{x}^{T}\mathbf{A}^{-1}\mathbf{x}\right)|\mathbf{A}|. \tag{64}$$

Thus,

$$\log_2\left|\mathbf{A} + \mathbf{x}\mathbf{x}^{T}\right| - \log_2|\mathbf{A}| = \log_2\left(1 + \mathbf{x}^{T}\mathbf{A}^{-1}\mathbf{x}\right). \tag{65}$$

If we take

$$\mathbf{A} = \mathbf{I}_{N_r} + \sum_{i=k+1}^{N_t} \frac{\alpha_{\text{loss}}^2 \mathcal{E}_{i,\text{V-BLAST}}}{\Gamma \sigma_n^2} \mathbf{h}_i \mathbf{h}_i^T. \tag{66}$$

And,

$$\mathbf{x} = \sqrt{\frac{\alpha_{\text{loss}}^2 \mathcal{E}_{k,\text{V-BLAST}}}{\Gamma \sigma_n^2}} \mathbf{h}_k, \tag{67}$$

then we get

$$\begin{aligned} \log_2 \left| \mathbf{I}_{N_r} + \sum_{i=k}^{N_t} \frac{\alpha_{\text{loss}}^2 \mathcal{E}_{i,\text{V-BLAST}}}{\Gamma \sigma_n^2} \mathbf{h}_i \mathbf{h}_i^T \right| - \log_2 (1 + \text{SINR}_k) = \\ \log_2 \left| \mathbf{I}_{N_r} + \sum_{i=k+1}^{N_t} \frac{\alpha_{\text{loss}}^2 \mathcal{E}_{i,\text{V-BLAST}}}{\Gamma \sigma_n^2} \mathbf{h}_i \mathbf{h}_i^T \right|, \end{aligned} \tag{68}$$

where

$$\text{SINR}_k = \alpha_{\text{loss}}^2 \mathcal{E}_{k,\text{V-BLAST}} \cdot \mathbf{h}_k^T \left(\Gamma \sigma_n^2 \cdot \mathbf{I}_{N_r} + \sum_{i=k+1}^{N_t} \alpha_{\text{loss}}^2 \mathcal{E}_{i,\text{V-BLAST}} \cdot \mathbf{h}_i \mathbf{h}_i^T \right)^{-1} \mathbf{h}_k. \tag{69}$$

Adding (68) for all *k*, we receive

$$\log_2 \left| \mathbf{I}_{N_r} + \sum_{i=1}^{N_t} \frac{\alpha_{\text{loss}}^2 \mathcal{E}_{i,\text{V-BLAST}}}{\Gamma \sigma_n^2} \mathbf{h}_i \mathbf{h}_i^T \right| = \sum_{i=1}^{N_t} \log_2 (1 + \text{SINR}_i). \tag{70}$$

Defining $\mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}}$ as a diagonal matrix with *i*-th diagonal entry of $\mathcal{E}_{i,\text{V-BLAST}}$, and $\mathbf{H}$ as the channel matrix with columns $\mathbf{h}_i$, we receive that

$$\frac{1}{2} \log_2 \left| \mathbf{I}_{N_r} + \frac{\alpha_{\text{loss}}^2}{\Gamma \sigma_n^2} \mathbf{H} \mathbf{\Sigma}_{\mathbf{x},\text{V-BLAST}} \mathbf{H}^T \right| = \frac{1}{2} \sum_{i=1}^{N_t} \log_2 (1 + \text{SINR}_i). \tag{71}$$

Appendix C
Concavity of the V-BLAST MMSE-SIC Capacity Objective Function

In this appendix, we prove the concavity of the V-BLAST MMSE-SIC capacity objective function with respect to the a diagonal covariance matrix of a transmitted vector $\mathbf{x}$, denoted by a matrix $\mathbf{\Sigma}$. For notational clarity and without loss of generality, the function is given by

$$f(\mathbf{\Sigma}) = \log_2 \left| \mathrm{I} + \mathbf{H} \mathbf{\Sigma} \mathbf{H}^T \right|. \tag{72}$$

By definition of concavity, we would like to prove that

$$f(\alpha \mathbf{\Sigma}_1 + (1 - \alpha) \mathbf{\Sigma}_0) \geqslant \alpha f(\mathbf{\Sigma}_1) + (1 - \alpha) f(\mathbf{\Sigma}_0), \tag{73}$$

for all $\alpha \in (0, 1)$. If we denote,

$$\mathbf{A} \triangleq \mathrm{I} + \mathbf{H} \mathbf{\Sigma}_1 \mathbf{H}^T, \tag{74}$$

$$\mathbf{B} \triangleq \mathrm{I} + \mathbf{H} \mathbf{\Sigma}_0 \mathbf{H}^T, \tag{75}$$

One can notice, that both $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices and Hermitian. Thus, there exist a non-singular matrix such that [57]

$$\mathbf{I} = \mathbf{C}^* \mathbf{A} \mathbf{C}, \tag{76}$$

$$\mathbf{\Lambda} = \mathbf{C}^* \mathbf{B} \mathbf{C}. \tag{77}$$

Where $\mathbf{\Lambda} = \text{diag}(\lambda_1 \cdots \lambda_n)$ with all $\lambda_i > 0$. Then,

$$\begin{aligned} f(\alpha\mathbf{\Sigma}_1 + (1-\alpha)\mathbf{\Sigma}_0) &= \log_2 \left|\mathbf{I} + \mathbf{H}\left[\alpha\mathbf{\Sigma}_1 + (1-\alpha)\mathbf{\Sigma}_0\right]\mathbf{H}^T\right| \\ &= \log_2 \left|\mathbf{I} + \alpha\mathbf{H}\mathbf{\Sigma}_1\mathbf{H}^T + (1-\alpha)\mathbf{H}\mathbf{\Sigma}_0\mathbf{H}^T\right| \\ &= \log_2 \left|\alpha\mathbf{A} + (1-\alpha)\mathbf{B}\right| \\ &= \log_2 \left|\mathbf{C}\left[\alpha\mathbf{I} + (1-\alpha)\mathbf{\Lambda}\right]\mathbf{C}^*\right| \\ &= \log_2 \left|\mathbf{C}^*\mathbf{C}\right| + \log_2 \left|\alpha\mathbf{I} + (1-\alpha)\mathbf{\Lambda}\right| \\ &= \log_2 |\mathbf{A}| + \sum_{i=1}^{n} \log_2 (\alpha + (1-\alpha)\lambda_i). \end{aligned} \tag{78}$$

On the other hand,

$$\begin{aligned} \alpha f(\mathbf{\Sigma}_1) + (1-\alpha) f(\mathbf{\Sigma}_0) &= \alpha \log_2 |\mathbf{A}| + (1-\alpha)\log_2 |\mathbf{B}| \\ &= \alpha \log_2 |\mathbf{A}| + (1-\alpha)\log_2 \left|\mathbf{C}\mathbf{\Lambda}\mathbf{C}^*\right| \\ &= \alpha \log_2 |\mathbf{A}| + (1-\alpha)\log_2 |\mathbf{A}| + (1-\alpha)\log_2 |\mathbf{\Lambda}| \\ &= \log_2 |\mathbf{A}| + (1-\alpha)\log_2 |\mathbf{\Lambda}|. \end{aligned} \tag{79}$$

Thus, using the concavity property of the log function,

$$\begin{aligned} f(\alpha\mathbf{\Sigma}_1 + (1-\alpha)\mathbf{\Sigma}_0) &= \log_2 |\mathbf{A}| + \sum_{i=1}^{n} \log_2 (\alpha + (1-\alpha)\lambda_i) \\ &\geqslant \log_2 |\mathbf{A}| + \sum_{i=1}^{n} \alpha \log_2 (1) + (1-\alpha)\log_2 (\lambda_i) \\ &= \log_2 |\mathbf{A}| + (1-\alpha)\log_2 \left(\prod_{i=1}^{n} \lambda_i\right) \\ &= \log_2 |\mathbf{A}| + (1-\alpha)\log_2 |\mathbf{\Lambda}| \\ &= \alpha\, f(\mathbf{\Sigma}_1) + (1-\alpha) f(\mathbf{\Sigma}_0). \end{aligned} \tag{80}$$

Acknowledgment

The authors would like to thank A. Gorshtein for helpful discussions, as well as insightful suggestions.

References

[1] A. Singh *et al.*, "Jupiter rising: A decade of clos topologies and centralized control in Google's datacenter network," *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 183–197, Aug. 2015. [Online]. Available: http://doi.acm.org/10.1145/2829988.2787508

[2] W. H. Hofmann, P. Moser, and D. Bimberg, "Energy-efficient VCSELs for interconnects," *IEEE Photon. J.*, vol. 4, no. 2, pp. 652–656, Apr. 2012.

[3] B. G. Lee *et al.*, "End-to-end multicore multimode fiber optic link operating up to 120 Gb/s," *J. Lightw. Technol.*, vol. 30, no. 6, pp. 886–892, Mar. 2012.

[4] A. Ghiasi, "Large data centers interconnect bottlenecks," *Opt. Exp.*, vol. 23, no. 3, pp. 2085–2090, Jan. 2015.

[5] Y. A. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communications beyond 100G," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. S67–S72, Feb. 2012.

[6] F. E. Doany *et al.*, "Terabit/sec VCSEL-based 48-channel optical module based on holey CMOS transceiver IC," *J. Lightw. Technol.*, vol. 31, no. 4, pp. 672–680, Feb. 2013.

[7] E. Temporiti *et al.*, "Insights into silicon photonics Mach–Zehnder-based optical transmitter architectures," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 3178–3191, Dec. 2016.

[8] H. Roscher, F. Rinaldi, and R. Michalzik, "Small-pitch flip-chip-bonded VCSEL arrays enabling transmitter redundancy and monitoring in 2-D 10-Gbit/s space-parallel fiber transmission," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 5, pp. 1279–1289, Sep./Oct. 2007.

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

[10] U. Levy, H. Kobrinsky, and A. Friesem, "Angular multiplexing for multichannel communication in a single fiber," *IEEE J. Quantum Electron.*, vol. QE-17, no. 11, pp. 2215–2224, Nov. 1981.

[11] P. J. Winzer, "Energy-efficient optical transport capacity scaling through spatial multiplexing," *IEEE Photon. Technol. Lett.*, vol. 23, no. 13, pp. 851–853, Jul. 2011.

[12] V. A. J. M. Sleiffer *et al.*, "73.7 Tb/s (96 $\times$ 3 $\times$ 256-Gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA," *Opt. Exp.*, vol. 20, no. 26, pp. B428–B438, Dec. 2012.

[13] R. Ryf *et al.*, "Mode-division multiplexing over 96 km of few-mode fiber using coherent 6 $\times$ 6 MIMO processing," *J. Lightw. Technol.*, vol. 30, no. 4, pp. 521–531, Feb. 2012.

[14] J. Carpenter, B. C. Thomsen, and T. D. Wilkinson, "Degenerate mode-group division multiplexing," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3946–3952, Dec. 2012.

[15] M. Blau and D. M. Marom, "Optimization of spatial aperture-sampled mode multiplexer for a three-mode fiber," *IEEE Photon. Technol. Lett.*, vol. 24, no. 23, pp. 2101–2104, Dec. 2012.

[16] S. Yerolatsitis, I. Gris-Sánchez, and T. A. Birks, "Adiabatically-tapered fiber mode multiplexers," *Opt. Exp.*, vol. 22, no. 1, pp. 608–617, Jan. 2014.

[17] R. R. Thomson, T. A. Birks, S. G. Leon-Saval, A. K. Kar, and J. Bland-Hawthorn, "Ultrafast laser inscription of an integrated photonic lantern," *Opt. Exp.*, vol. 19, no. 6, pp. 5698–5705, Mar. 2011.

[18] Y. Ding, H. Ou, J. Xu, and C. Peucheret, "Silicon photonic integrated circuit mode multiplexer," *IEEE Photon. Technol. Lett.*, vol. 25, no. 7, pp. 648–651, Apr. 2013.

[19] K. Appaiah, R. Salas, S. Vishwanath, and S. R. Bank, "Offset coupling, feedback, and spatial multiplexing in 4 $\times$ 4 incoherent-MIMO multimode fiber links," *J. Lightw. Technol.*, vol. 31, no. 17, pp. 2926–2939, Sep. 2013.

[20] H. S. Chen, H. P. A. van den Boom, and A. M. J. Koonen, "30-Gb/s 3 $\times$ 3 optical mode group-division-multiplexing system with optimized joint detection," *IEEE Photon. Technol. Lett.*, vol. 23, no. 18, pp. 1283–1285, Sep. 2011.

[21] G. Stepniak, L. Maksymiuk, and J. Siuzdak, "Influence of mode coupling on mode group diversity multiplexing in multimode fibers," *Opt. Quantum Electron.*, vol. 41, no. 3, pp. 203–213, Nov. 2009.

[22] C. P. Tsekrekos, A. Martinez, F. M. Huijskens, and A. M. J. Koonen, "Design considerations for a transparent mode group diversity multiplexing link," *IEEE Photon. Technol. Lett.*, vol. 18, no. 22, pp. 2359–2361, Nov. 2006.

[23] M. Kowalczyk and J. Siuzdak, "Four-channel incoherent MIMO transmission over 4.4-km MM fiber," *Microw. Opt. Technol. Lett.*, vol. 53, no. 3, pp. 502–506, Mar. 2011.

[24] N. Sheffi and D. Sadot, "Energy-efficient VCSEL array using power and offset allocation of spatial multiplexing in graded-index multimode fiber," *J. Lightw. Technol.*, vol. 35, no. 11, pp. 2098–2108, Jan. 2017.

[25] H. R. Stuart, "Dispersive multiplexing in multimode optical fiber," *Science*, vol. 289, no. 5477, pp. 281–283, Jul. 2000.

[26] M. Greenberg, M. Nazarathy, and M. Orenstein, "Data parallelization by optical MIMO transmission over multimode fiber with intermodal coupling," *J. Lightw. Technol.*, vol. 25, no. 6, pp. 1503–1514, Jun. 2007.

[27] F. Ferreira, S. Jansen, P. Monteiro, and H. Silva, "Nonlinear semi-analytical model for simulation of few-mode fiber transmission," *IEEE Photon. Technol. Lett.*, vol. 24, no. 4, pp. 240–242, Feb. 2012.

[28] S. H. Murshid and J. Iqbal, "Array of concentric CMOS photodiodes for detection and de-multiplexing of spatially modulated optical channels," *J. Opt. Laser Technol.*, vol. 41, no. 6, pp. 764–769, Feb. 2009.

[29] N. Sheffi and D. Sadot, "Tilted Gaussian beams multiplexer for graded-index multimode fiber in data-centers interconnections," *IEEE Photon. J.*, vol. 7, no. 3, Jun. 2015, Art. no. 7901816.

[30] X. Wu *et al.*, "A 20 Gb/s NRZ/PAM-4 1V transmitter in 40nm CMOS driving a Si-photonic modulator in 0.13 μm CMOS," in *Proc. IEEE IEEE Int. Solid-State Circuits Conf.*, Feb. 2013, pp. 128–129.

[31] K. H. Park, Y. C. Ko, and M. S. Alouini, "On the power and offset allocation for rate adaptation of spatial multiplexing in optical wireless MIMO channels," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1535–1543, Apr. 2013.

[32] X. Gao, Y. Zhang, Z. Wang, X. Wang, and L. Duan, "Optimal power efficiency of spatial multiplexing in optical wireless MIMO channels," in *Proc. 9th Int. Conf. Commun. Netw. China*, Aug. 2014, pp. 422–426.

[33] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. Int. Symp. Signals, Systems, Electron.*, Sep. 1998, pp. 295–300.

[34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[35] R. Zhang, Y. C. Liang, R. Narasimhan, and J. M. Cioffi, "Approaching MIMO-OFDM capacity with per-antenna power and rate feedback," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1284–1297, Sep. 2007.

[36] A. R. Shah, R. C. Hsu, A. Tarighat, A. H. Sayed, and B. Jalali, "Coherent optical MIMO (COMIMO)," *J. Lightw. Technol.*, vol. 23, no. 8, pp. 2410–2418, Aug. 2005.

[37] W. M. J. Green, M. J. Rooks, L. Sekaric, and Y. A. Vlasov, "Ultra-compact, low RF power, 10 Gb/s silicon Mach-Zehnder modulator," *Opt. Exp.*, vol. 15, no. 25, pp. 17106–17113, Dec. 2007.

[38] L. Liao *et al.*, "High speed silicon Mach-Zehnder modulator," *Opt. Exp.*, vol. 13, no. 8, pp. 3129–3135, Apr. 2005.

[39] C. Kopp *et al.*, "Silicon photonic circuits: on-CMOS integration, fiber optical coupling, and packaging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 3, pp. 498–509, May/Jun. 2011.

[40] C. Li, H. Zhang, M. Yu, and G. Lo, "CMOS-compatible high efficiency double-etched apodized waveguide grating coupler," *Opt. Exp.*, vol. 21, no. 7, pp. 7868–7874, Apr. 2013.

[41] A. Mekis *et al.*, "A grating-coupler-enabled CMOS photonics platform," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 3, pp. 597–608, May/Jun. 2011.

[42] Y. Ding, C. Peucheret, H. Ou, and K. Yvind, "Fully etched apodized grating coupler on the SOI platform with -0.58 dB coupling efficiency," *Opt. Lett.*, vol. 39, no. 18, pp. 5348–5350, Sep. 2014.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd ed.* Hoboken, NJ, USA: Wiley, 2006.

[44] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.
[45] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding. ii. Coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.
[46] J. M. Cioffi, Advanced Digital Communication-EE397C Class Notes, 2017. [Online]. Available: http://web.stanford.edu/group/cioffi/
[47] *Media Access Control Parameters, Physical Layers and Management Parameters for 400 Gb/s Operation*, IEEE Standard P802.3bs/D1.3, Apr. 2016.
[48] J. D'Ambrosia, "IEEE P802.3bs baseline summary," 2015. [Online]. Available: http://www.ieee802.org/3/bs/
[49] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," Sep. 2013. [Online]. Available: http://cvxr.com/cvx
[50] D. Marcuse, *Theory of Dielectric Optical Waveguides*. New York, NY, USA: Academic, 1974.
[51] K. Balemarthy, A. Polley, and S. E. Ralph, "Electronic equalization of multikilometer 10-Gb/s multimode fiber links: Mode-coupling effects," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4885–4894, Dec. 2006.
[52] P. Pepeljugoski, S. E. Golowich, A. J. Ritger, P. Kolesar, and A. Risteski, "Modeling and simulation of next-generation multimode fiber links," *J. Lightw. Technol.*, vol. 21, no. 5, pp. 1242–1255, May 2003.
[53] S. Kawakami and H. Tanji, "Evolution of power distribution in graded-index fibres," *Electron. Lett.*, vol. 19, no. 3, pp. 100–102, 1983.
[54] R. Olshansky, "Mode coupling effects in graded-index optical fibers," *Appl. Opt.*, vol. 14, no. 4, pp. 935–945, 1975.
[55] *108 fiber model*, 2004. [Online]. Available: http://www.ieee802.org/3/aq/public/tools/108fiberModel
[56] J. Ding and A. Zhou, "Eigenvalues of rank-one updated matrices with some applications," *Appl. Math. Lett.*, vol. 20, no. 12, pp. 1223–1226, 2007.
[57] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1990.