

Modulation Format and Digital Signal Processing for IM-DD Optics at Post-200G Era

Di Che  and Xi Chen 

(Invited Paper)

Abstract—200-Gb/s per lane intensity-modulation (IM) direct-detection (DD) optics are being commercialized to support 800G and 1.6T applications inside datacenters. Though IM-DD remains its cost and power consumption advantages over Coherent at 1.6T for short-reach interconnect below 10 km, its roadmap is not clear towards the next capacity doubling, considering it becomes more challenging to scale the components bandwidth linearly with the capacity demand. This makes both advanced modulation formats and digital signal processing (DSP) indispensable for an IM-DD system aiming at higher speed. From a system perspective, this article reviews candidate modulation formats and DSP for IM-DD optics at post 200G (per lane) era. By taking into account generic constraints for future IM-DD systems like bandwidth limit, peak power constraint, transceiver nonlinearity, fiber dispersion, and so on, it discusses a wide range of techniques including probabilistic constellation shaping (PCS), high symbol rate pulse amplitude modulation (PAM), faster than Nyquist (FTN) signaling, nonlinear equalizations, and multicarrier modulations. Different from prior IM-DD review literature, we mainly focus on if it is meaningful to exploit a DSP technique with respect to constraints in practical systems, rather than just the technique itself. The study is backed with rich simulation and experiment results.

Index Terms—Digital signal processing (DSP), direct detection (DD), faster than nyquist (FTN), intensity modulation (IM), modulation format, multicarrier modulation, nonlinear equalization, optical fiber communication, probabilistic constellation shaping (PCS), pulse amplitude modulation (PAM).

I. INTRODUCTION

DIGITAL signal processing (DSP) has been applied to optical fiber communications for almost two decades. Though the first use of DSP was in 10-Gb/s intensity modulation (IM) and direct detection (DD) systems to combat the fiber dispersion by maximum likelihood sequence estimation (MLSE) [1], [2], it soon became a powerful tool to revive the coherent communication [3], [4]. Nowadays, DSP-enabled coherent transceivers have been widely implemented to optical systems ranging from long-haul transmission over thousands of kilometers (km) to short-reach interconnects below 80 km. On the other hand, despite a variety of DSP techniques being

proposed for IM-DD, it takes a much slower pace to adopt DSP in real-world products. The IM-DD world had been ruled by the antiquated on-off-keying (OOK) modulation over decades till the commercialization of the first 4-ary pulse amplitude modulation (PAM-4) DSP in 2015 [5]. Up to now, PAM-4 remains the dominant IM-DD format which is incorporated with a low-complexity receiver DSP including clock recovery, equalization like a feed-forward equalizer (FFE) and low-latency forward error correction (FEC) decoding like the KP4/KR4 FEC defined in IEEE 802.3 Standard.

Based on 50- and 100-Gb/s per lane IM-DD, optical modules for 400G Ethernet have been shipping in volume since 2020. In the meantime, the 200-Gb/s per lane optical standard is being actively developed in both industry interoperability groups and standard bodies. Such 100-GBd class PAM-4 signals inevitably call for more sophisticated DSP to combat system impairments. For example, the 800G Pluggable MSA group suggests an extra MLSE after the FFE to improve the tolerance to dispersion and bandwidth limit, and a stronger concatenated FEC scheme than KP4 to improve the bit-error ratio (BER) threshold from $2e-4$ to $2e-3$ [6]. Going beyond 200G per lane, it is perceived DSP would play an even more important role for an IM-DD system, considering its component bandwidth growth has begun lagging the next capacity doubling of optical interface.

In European Conference on Optical Communications (ECOC 2022), we discussed whether it is meaningful to pursue higher symbol rate in a bandwidth-limited IM-DD system via DSP [7]. This article extends our ECOC presentation and provides a more comprehensive review on the potential DSP suitable for future IM-DD optics. There have been extensive literatures to review the variety of DSP techniques [8], [9], [10], [11], [12], [13], [14], [15] since they were introduced to IM-DD more than 10 years ago. This work is differentiated from them in the following aspects. First, with Coherent being penetrated to shorter and shorter reach, the next-generation IM-DD will mostly be deployed to unamplified applications within 10 km. The transceiver will be the dominant factor of system impairments than the fiber channel, which changes the role of various DSP techniques consequently. Second, more DSP has been added to the IM-DD portfolio, like various types of faster-than-Nyquist (FTN) precoding and equalization, probabilistic constellation shaping (PCS), multicarrier entropy loading (EL), Volterra nonlinear equalization (VNE) and so on, and it lacks a comprehensive analysis among them. Third, like our ECOC

Manuscript received 3 May 2023; revised 5 August 2023; accepted 20 August 2023. Date of publication 4 September 2023; date of current version 16 January 2024. (Corresponding author: Di Che.)

The authors are with the Nokia Bell Labs, Murray Hill, NJ 07974 USA (e-mail: di.che@nokia-bell-labs.com; xi.v.chen@nokia-bell-labs.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2023.3311716>.

Digital Object Identifier 10.1109/JLT.2023.3311716

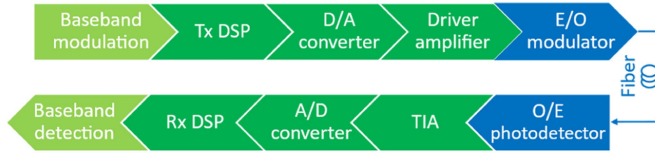


Fig. 1 Schematic of an IM-DD system. Tx/Rx: transmitter/receiver.

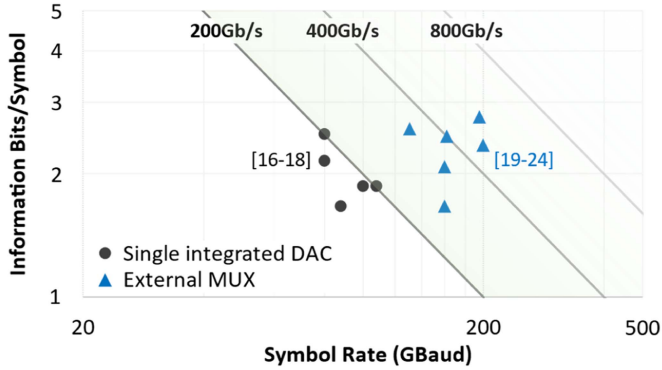


Fig. 2. Recent demonstrations of high-speed IM-DD systems. A reference may contain more than one points.

work [7], we focus on if it is meaningful to exploit each DSP technique rather than just its implementation. Moreover, instead of a study based on a specific experiment setup, we summarize common features for IM-DD systems with different hardware, aiming to generalize some conclusions to help readers tailor the modulation format and DSP based on their own setups.

II. SYSTEM CONSTRAINTS AND IMPAIRMENTS

A generic architecture of the IM-DD system is shown in Fig. 1. We assume DSP is used in the context of this article (200 Gb/s and beyond). An IM transmitter mainly consists of electronic signal generation, radio-frequency (RF) amplification (optional, depending on the required drive swing) and electrical-to-optical (E/O) conversion. The E/O conversion is realized by a directly modulated laser (DML), an electro-absorption modulator (EAM), or a Mach-Zehnder modulator (MZM). The integration of a continuous-wave (CW) laser with an EAM or an MZM is also named an external modulated laser (EML). The signal is detected by a photodiode (PD) after fiber transmissions. There may be a transimpedance amplifier (TIA) or an RF power amplifier before the signal is digitally detected. Because future IM-DD applications will be mostly limited to within 10 km, we do not include optical amplification in the link. In this section, we review the relevant hardware constraints to provide useful guidance on designing the appropriate modulation format and DSP chain for such an IM-DD system.

A. Bandwidth Limit (BWL)

We focus on the transmitter BWL considering it determines the maximum speed of signal generation. In Fig. 2, we illustrate the recent IM-DD experiments with the best effort of pushing forward the envelope of high symbol rate systems [16], [17], [18], [19], [20], [21], [22], [23], [24]. Though the experiments

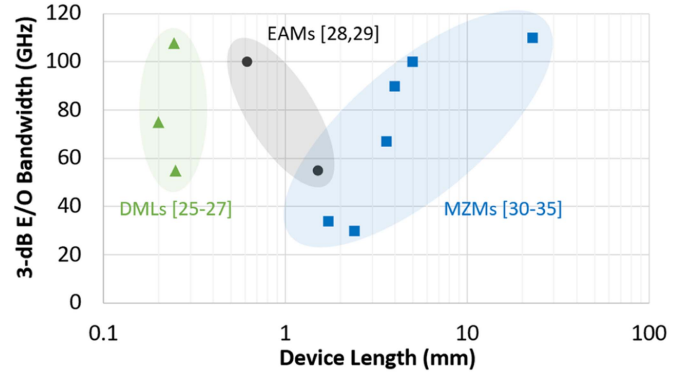


Fig. 3. Bandwidths and device lengths of DMLs, EAMs, and MZMs.

were performed with different types of modulators (making it not possible to make an apple-to-apple comparison), electrical signal generations are commonly their main speed bottleneck. Using a single D/A converter (DAC), the state-of-the-art single-lane IM-DD speed is about 200 Gb/s. The integrated DACs for 200G-class signals in Fig. 2 are mostly based on (down to) 14-nm CMOS (complementary metal-oxide semiconductor) technology or (down to) 55-nm SiGe (Silicon-Germanium) lithography, capable of 100-GBaud class symbol rate. As a main force of application-specific integrated circuits (ASIC), the latest CMOS technologies have evolved to 7- or 5-nm process, enabling 130-GBaud class signal generation in the latest generation coherent transceiver products. Regarding SiGe DACs, though they are believed to offer higher bandwidth than CMOS ones, it is not yet clear how to make them compatible with the fabrication process of CMOS-based DSP ASIC for co-integration. A promising approach to further increase the speed of electronic signals is to exploit external DAC multiplexers (MUX), namely, multiplexing two or more DAC outputs with external analog RF components like in [19], [20] to improve the electrical bandwidth multi-folds. While integrated DACs are commonly part of the CMOS ASIC, external MUXs can use other materials such as SiGe or Indium Phosphide (InP), which are not ideal for DSP but good at handling high-speed analog signals. With external MUXs, an IM-DD system can achieve over 500-Gb/s net bit rate as shown in Fig. 2. The fact that increasing the electronics bandwidth results in higher data rate per lane serves as evidence that the speed of electronics is more limiting than that of the optics at the moment.

In terms of the E/O conversion, lab demonstrations have shown various types of modulators with bandwidth at or higher than 100 GHz. We illustrate the 3-dB bandwidth of three types of E/O modulators in recent demonstrations in Fig. 3 [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35]. With recent advances in device structure design on traditional platforms such as InP, as well as emerging materials such as thin-film lithium niobate (TFLN) for MZM, the state-of-the-art DMLs, EMLs, and MZMs can all achieve the 3-dB bandwidth at 100-GHz class, making them suitable for 200-GBaud signals. Though these modulators are not limited by bandwidth to date, they have ineligible differences in other aspects like extinction ratio (ER) and device nonlinearities. Especially for footprints, DMLs

and EMLs are very compact and typically less than 1-mm long, which are the main force for short-reach interconnect markets. MZMs usually have much larger sizes (anywhere from 3 mm to up to about 20 mm), but their better modulation quality (e.g., little chirp and high linearity) make them popular in recent high-speed IM-DD experiments. As a reference, the footprint of photonic integrated circuits (PIC) in an intensity modulation transmitter is typically $5 \times 5 \text{ mm}^2$ or less.

The BWL can be shown as the spectral roll-off of a signal at higher frequencies. However, rather than looking at the power spectrum, a more relevant way to characterize the BWL is the SNR profile over frequencies. This is because SNR is the only figure of merit on estimating the system capacity. For example, digital pre-equalization can flatten the power spectrum and conceal the effect of BWL, but it enhances the noise in the meantime which can be captured by the SNR profile. Knowing the signal power spectral density (PSD) of $S(f)$, noise PSD of $N(f)$ and the channel response of $H(f)$, the system capacity under the additive Gaussian noise is [36]

$$C = \int \log_2 \left(1 + \frac{S(f)}{N(f)} |H(f)|^2 \right) df \quad (1)$$

Clearly, the capacity is not only determined by the frequency response, but also the frequency-resolved SNR. While a signal achieves higher overall SNR by occupying less bandwidth, its spectrum must be expanded to the entire usable bandwidth of the channel to approach the capacity [37]. This inevitably pushes all IM-DD systems towards the BWL condition when they aim to maximize their data rates.

B. Peak Power Constraint (PPC)

The power constraint of a transmitter limits the output signal power, which then determines the system capacity through SNR as in (1). Unlike amplified coherent systems whose optimum launch power to the fiber is fundamentally limited by the fiber nonlinearity [38], the power constraint of an un-amplified IM-DD system comes from the finite linear driving range of an E/O modulator, commonly understood as the peak power constraint (PPC) of an IM-DD system. Though no E/O conversion has an absolutely linear transfer function, some modulators natively offer more linearity than others. We use EAMs and MZMs for an exemplary comparison. Fig. 4 shows the typical transfer functions of an EAM and an MZM. To characterize the linear driving range, it is crucial to introduce the concept of “dynamic extinction ratio (ER)”. While the static ER (or DC ER) refers to the optical output power ratio between the fully ‘turned on’ and fully ‘turned off’ states by applying DC voltages as shown in Fig. 4, the dynamic ER is the ratio between the maximum and minimal optical power levels with the actual driving RF signal. The DC ER is a direct measure of the design/fabrication quality of an optical device, and the dynamic ER reflects a combination of device and signal quality which is a more relevant metric to understand the PPC in IM-DD systems.

The modulator transfer functions in Fig. 4 are nonlinear for both EAMs and MZMs. However, unlike the EAM where the entire modulation region is nonlinear, the sinusoidal response of MZMs offers a relatively wide region where the response

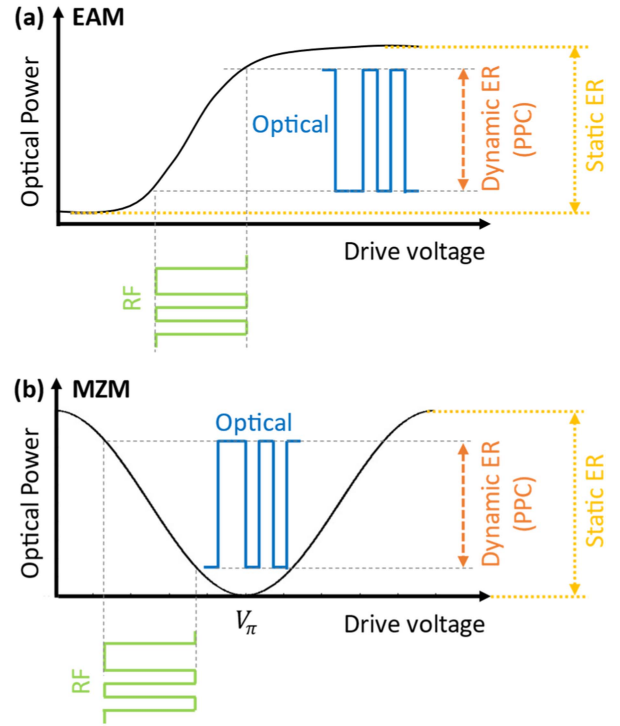


Fig. 4. Exemplary transfer functions for (a) EAM and (b) MZM.

is almost linear. Typically, EMLs offer a dynamic ER between 5 dB to 7 dB, while MZMs are between 10 dB to 15 dB.

1) Influence of the Nonlinear E/O Transfer Function on PPC:

It is important to point out that the E/O PPC is not an absolute value but rather a tradeoff between the nonlinear distortion and the peak power. Driving into the nonlinear region allows higher peak power at the sacrifice of enhanced nonlinear distortions. Using higher DSP complexity, the nonlinear distortion may be mitigated by various types of nonlinear equalizations, as will be revealed in following sections.

2) *E/O PPC vs. Optical Output Power Limit:* It is also noted that the E/O PPC should not be confused with the *optical* output power limit of an IM transmitter. The optical output power is a constant set by the laser power and the bias condition of the E/O modulator and is independent of modulation formats due to the bipolar nature of the electrical drive signal [39, Section III]. This is different to the MZMs in a coherent modulator which are biased at the null point. The optical output power of a null-point modulator is zero without the electrical drive signal, and consequently, it is determined by both the CW laser power and the electrical drive signal power. In an IM-DD system without optical amplification, optical power is not relevant to finding the optimal modulation format.

C. Peak Enhancement

Due to the PPC from the E/O conversion, the proper metric to characterize the *system* quality for IM-DD would be the peak-signal-noise ratio (PSNR), defined as

$$PSNR = \frac{PPC}{P_{Noise}} \quad (2)$$

where P_{Noise} is the accumulated noise power of the entire IM-DD system, and PPC stands for the peak power of the electrical signal that drives the E/O modulator. Clearly, PSNR is a system metric independent of modulation formats, just like the role of SNR in an average power constraint (APC) system. For all the PSNR simulations in this article, we use the peak-to-peak swing of a signal X to characterize PPC, namely,

$$PPC = |\max X - \min X|^2 \quad (3)$$

While the system metric is different between PPC and APC systems, SNR is always a proper *signal* metric to characterize the signal quality and evaluate its achievable information rate (AIR). To find the optimum modulation format with maximized AIR, the system PSNR should be translated to the signal SNR at the digital baseband where bit-interleaved coded modulations (BICM) perform FEC coding and decoding. The translation is made by the peak-to-average power ratio (PAPR) of the drive signal at the input to the E/O modulator, namely,

$$SNR = \frac{PSNR}{PAPR_{drive_signal}} \quad (4)$$

One should not confuse $PAPR_{drive_signal}$ with the PAPR of the digital modulation format, noted as $PAPR_{constellation}$. Usually, the PAPR is enhanced when we generate the analog drive signal from the digital baseband, which results in $PAPR_{drive_signal} > PAPR_{constellation}$. Such PAPR increment is defined as a PAPR enhancement (PAPRE) factor in [39],

$$PAPRE = \frac{PAPR_{drive_signal}}{PAPR_{constellation}} \quad (5)$$

$$PAPR_{drive_signal} = PAPRE \cdot PAPR_{constellation} \quad (6)$$

In turn, $PAPR_{drive_signal}$ is determined by both the modulation format and the PAPRE factor. PAPRE can happen either in digital domain by the transmitter (Tx) DSP, like pulse shaping and pre-equalization, or in analog domain like the BWL-induced peak distortion from DAC and driver amplifier. Substituting (5) to (4), the signal SNR is expressed as

$$SNR = \frac{PSNR}{PAPR_{constellation} \cdot PAPRE} \quad (7)$$

Clearly, the modulation format itself is not sufficient to tell the SNR from the system PSNR. As a consequence, the modulation design must take into account the PAPRE factors.

Unfortunately, PAPRE is modulation format dependent, more explicitly, it depends on the initial PAPR of the signal. This brings challenge for the modulation format design because such an optimization problem is commonly solved under a fixed signal power constraint. There is a special case of PAPRE named *extreme PAPRE* in [39]. It dominates $PAPR_{drive_signal}$, making it approach a constant regardless of $PAPR_{constellation}$. A unique extreme PAPRE condition is the inverse discrete Fourier transform (IDFT) in a discrete multi-tone (DMT) transmitter, whose time-domain signal PAPR is sensitive to the IDFT size and oversampling ratio but is not the modulation format of each

subcarrier. In this case, the PPC is equivalent to an APC, as the SNR approaches a constant according to (4).

It is important to point out that the PAPRE during (e.g., from the modulator BWL) or after (e.g., from the dispersion induced power fading) the E/O conversion should not be counted in the PAPRE in (7). This is because the PSNR is determined by the PPC of the E/O conversion, and the optical-domain distortion after the E/O conversion is not relevant. To better understand this, we assume two IM transmitters with the same PPC. Their overall Tx BWL are also identical, but one is purely from the electronics and the other is purely from the E/O modulator. It is expected the electronics-BWL transmitter exhibits a lower SNR, because its PAPRE happens before the E/O conversion and degrades the SNR via (7). This will be verified by an experiment in Section V-C.

D. Four-wave Mixing (FWM)

Most field-deployed IM-DD systems are operated at O-band to minimize the chromatic dispersion (CD) penalty. However, the inter-channel four-wave-mixing (FWM) [40], [41] has been found to impose a severe performance limitation on wavelength division multiplexing (WDM) transmissions in state-of-the-art O-band applications [42], [43], [44], [45]. As read from its name, FWM is a nonlinear interaction among three frequencies producing one new frequency. The magnitude of the new frequency f_g is [41]

$$P_g = \frac{D_x^2}{9} \gamma^2 P_i P_j P_k e^{-\alpha L} \eta \quad (8)$$

where D_x is the degeneration factor which equals to 6 for non-degenerate FWM, and 3 for degenerate FWM. $P_{i,j,k}$ is the input power at the frequency of $f_{i,j,k}$, α and L are the loss coefficient and the length of the fiber, respectively, γ is the third-order fiber nonlinearity coefficient, and η is the FWM coefficient. For IM-DD links with a single fiber span, η can be approximated as

$$\eta \approx \frac{1}{(\Delta\beta_{ijk})^2 + \alpha^2} \quad (9)$$

where $\Delta\beta_{ijk} \equiv \beta_i + \beta_j - \beta_k - \beta_g$ is the difference of the propagation constants due to CD and characterizes the phase match among the involved waves. The phase match condition is closely related to (i) the amount of CD, and (ii) the relative frequency spacing between FWM waves. As the WDM grid in O-band is close to the zero-dispersion wavelength of the fiber and the channels are usually evenly spaced, the phase mismatch is very small, leading to high FWM coefficient η . Such inter-channel fiber nonlinearity may be alleviated by unequal channel spacing [43] and polarization interleaving [44], [45], but it is not straightforward to be mitigated by DSP.

E. Fiber Dispersion and Modulation Chirp

Chromatic dispersion (CD) remains a major limiting factor for IM-DD transmissions at O-band. Though O-band covers the zero-dispersion wavelength, the WDM capability is limited by FWM which prevents the deployment of dense WDM (DWDM)

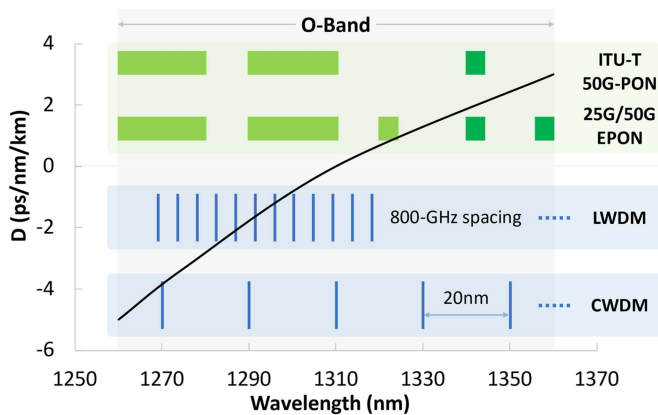


Fig. 5. Chromatic dispersion in standard single mode fiber (SSMF) for typical wavelength plans in O-band. Each horizontal bar (green) in the PON figures indicates a wavelength range.

near the zero-dispersion wavelength. Among typical O-band wavelength plans in Fig. 5, the narrowest channel spacing is 800 GHz for LAN-WDM. This is much wider than the common 50- or 100-GHz DWDM grid at C-band. Moreover, tightly packed channels require precise wavelength stabilization for lasers and wavelength (de-)multiplexers, adding huge cost and power consumption. As a result, most wavelength plans in Fig. 5 spread channels across the O-band, leaving nonnegligible CD on edge channels.

Besides the point-to-point WDM transmission mentioned above, another common application at O-band is passive optical networks (PON). Although the state-of-the-art PON commonly use time division multiplexing (TDM), the “good” wavelength windows are usually reserved for the uplink to tolerate the large disparity of laser operating conditions among end users, leaving tough wavelengths for the downlink. For example, the ITU-T 50G-PON [46] standard chooses the downlink at 1342 nm as shown in Fig. 5, with a CD parameter close to 2 ps/nm/km.

CD introduces frequency-dependent phase variation whose profile follows a parabolic shape symmetric around the carrier frequency. Such symmetry introduces destructive interference at particular frequencies after the square-law intensity detection and leads to frequency-selective spectral fading [47]. A sketch fading spectrum with positive dispersion is shown in Fig. 6(a). The number of notches and their frequencies are determined by a variety of system parameters like accumulated CD and carrier wavelength. Moreover, modulation chirp [48] generated from the E/O conversion also plays a critical role. Chirp refers to the phenomenon of instantaneous frequency or phase variation of the optical signal during intensity modulations. The frequency variation is named adiabatic chirp and is commonly unique to DMLs. The phase variation is named transient chirp and exists in DMLs, EMLs as well as non-push-pull MZMs. DMLs have positive transient chirp, while EMLs can be designed to exhibit either positive or negative transient chirp. The push-pull MZM is the only type of modulator free from both types of chirps.

Because chirp brings about phase/frequency variation on an IM signal, it interacts with the CD-induced phase variation and changes the transfer function of the spectral fading. In Fig. 6,

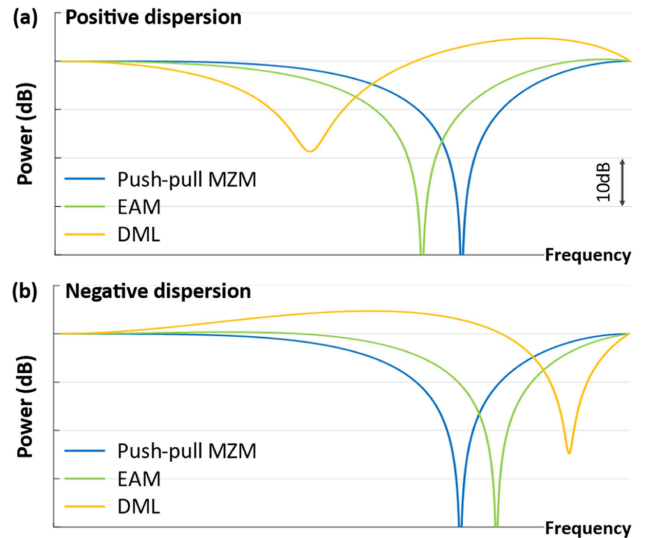


Fig. 6. Sketches of received electrical spectra with frequency-selective fading under (a) positive and (b) negative fiber dispersion for an O-band IM-DD link. The push-pull MZM has no chirp, while both the EAM and DML are assumed to have positive transient chirp; DML is also assumed to have adiabatic chirp.

we show two sets of sketches of the received digital spectra with positive and negative fiber CD. We assume the ideal push-pull MZM has no chirp, while both DML and EML have positive transient chirp. Note that the horizontal axis is left without unit as we intend to show frequency-selective fading illustratively without considering specific symbol rate, fiber type, dispersion parameter, etc. Compared to the chirp-free MZM (blue curves), chirped modulators suffer from severer fading under positive dispersion, whose frequency notches move to lower frequencies in Fig. 6(a). In contrast, if dispersion is negative as in Fig. 6(b), chirp mitigates the fading and pushes the notch towards higher frequencies. Similarly, the spectral fading can also be alleviated given positive dispersion and negative EML chirp. Historically, it is these EMLs designed with negative transient chirp that allow bridging 80-km reach at 10 Gb/s in C-band [14].

Frequency-selective fading can be generalized as a colored-SNR phenomenon, whose channel capacity is determined by (1). The faded spectrum can be estimated by mathematical models accurately [49] given the modulation chirp and fiber dispersion parameters. Though various DSP techniques [50], [51], [52], [53] can alleviate the performance penalty due to fading, it is not a common practice to operate an IM-DD system in a channel with completely faded frequency notches as they induce irreversible information loss at those frequencies. Rather, the accumulated CD is kept below a threshold to push the first notch out from the signal spectrum. In this case, fading behaves like BWL, and consequently, a DSP scheme that combats BWL can potentially be exploited for spectral fading, too.

F. System Nonlinearities

System nonlinearities are more prominent when pushing an IM-DD transceiver towards higher order modulation and higher symbol rate. A common bottleneck on the transmitter integrity is the nonlinear transfer function from various components like E/O modulators [54] and RF amplifiers. As we mentioned in

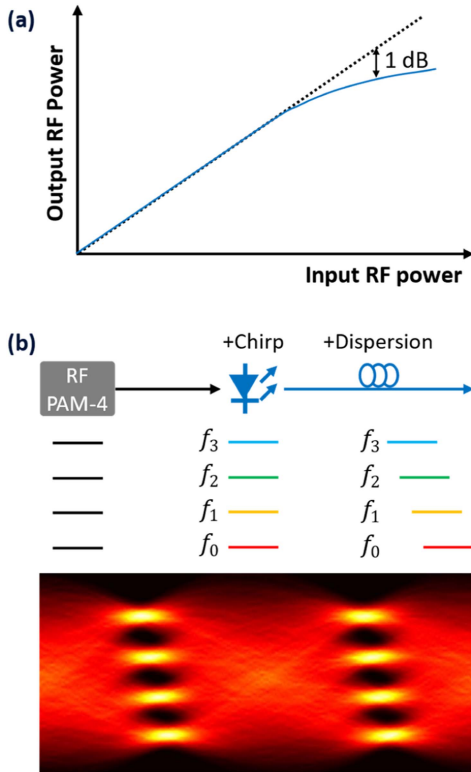


Fig. 7. IM-DD system nonlinearity examples: (a) a nonlinear transfer function of an RF driver (with its P1dB point); (b) a skewed PAM-4 eye diagram due to the interaction between DML adiabatic chirp and fiber dispersion [55].

Section II-B, for EML and MZM, there is a tradeoff between the PPC of drive signals and the nonlinear distortion. An RF amplifier is driven to the nonlinear region beyond an output power threshold, specified as its ‘ x -dB compression point’ like the 1-dB point (P1dB) in Fig. 7(a). Like what we mentioned in Section II-B for the E/O modulator, we can also take advantage of the nonlinear region of an RF amplifier for higher power and utilize DSP to mitigate associated distortions. Nonlinearities may also come from the interaction between components and fiber impairments. For example, a skewed eye diagram is often seen in DML transmissions [25], [55] due to the interplay between adiabatic chirp and CD as shown in Fig. 7(b). Because adiabatic chirp essentially means a frequency shift proportional to the instantaneous signal intensity, different PAM levels travel at different speeds in the dispersive fiber and eventually present themselves as a skewed version upon detection.

G. Phase Noise and Multi-Path Interference (MPI)

Laser phase noise can produce intensity fluctuations in an IM-DD system through various phase-to-intensity conversion mechanisms, like fiber dispersion [56] and optical interference [57]. Among them, MPI has been a non-negligible limiting factor in state-of-the-art IM-DD applications [58]. MPI is an interferometric noise due to the optical reflections during fiber transmission, mostly from the fiber connectors (e.g., an air gap between the two fiber endfaces, or a polluted endface) that induce refractive index discontinuities. It is more severe for a laser

with larger linewidth. MPI can quickly accumulate its power if the number of reflectance (N_R) increases, considering the total number of forward propagated reflections (each is generated by two reflectances) is $(N_R - 1) \cdot N_R/2$ [57].

In a higher-order PAM system, the interferometric noise is greatly enhanced as it mixes more intensity levels. This makes PAM-4 more vulnerable to MPI than the antiquated OOK. As characterized in [58], under 1-dB power penalty, PAM-4 can tolerate about 8-10dB less MPI than OOK. Due to the lack of practical MPI mitigation techniques, standard bodies have tightened the maximum reflectance specification for PAM-4 systems. For instance, IEEE802.3 and 100G Lambda MSA (multi-source agreement) reduced the return loss bar from -26 to -35 dB in 2016 [59]. There emerged a few new proposals to tackle the MPI issue in an analog [59] or a digital [60] manner.

III. PROBABILISTIC CONSTELLATION SHAPING

For an amplified coherent transmission system subjected to an optical APC, probabilistic constellation shaping (PCS) is a powerful tool to approach the Shannon capacity of the optical fiber channel [61], [62]. However, it brings extensive debates on whether the PCS benefits are relevant to an IM-DD system with PPC. Based on the findings in [39], we extend our discussion in this section in an intuitive manner.

A. Role of the MB-Distributed PCS

In APC systems, the Maxwell-Boltzmann (MB) distribution is most widely adopted for PCS, because it maximizes entropy given a fixed system SNR. An MB distribution assigns a higher probability to a symbol with lower power, which enhances the $PAPR_{constellation}$ and reduces the SNR under a PPC according to (7). Due to such reduced SNR, in a PPC system, MB-PCS usually cannot fully achieve the shaping benefits as in an APC system. For higher symbol-rate IM-DD systems, stronger BWL and transmitter pre-equalization induce huge PAPRE, making PAPRE the dominating factor for $PAPR_{drive_signal}$ in (6). This closes the $PAPR_{drive_signal}$ gap of various constellations and converts PPC towards APC via (4). In this case, MB or MB-like PCS can partially achieve PCS benefits [39], [63], [64]. In this case, it becomes crucial to maximize the signal average power under the PPC to get higher SNR. Below we introduce some practical implementations for the MB-PCS to improve the SNR aiming at higher shaping gain in a PPC system.

First, to avoid a huge constellation power loss, the MB-PCS signal should never be strongly shaped. It was revealed [61] that a lightly shaped PCS signal retains a shaping gain in the APC system. Light shaping means the source entropy should be close to (and less than) an integer, namely, the entropy of a uniform PAM- X signal where $X = 2^n$ ($n = 2, 3, 4, \dots$). This restricts the rate adaptation range. A strategy to alleviate the issue is to allow $X \neq 2^n$ for the PAM- X templates. X should remain an even number (i.e., $X = 2n$) to be compatible with the probabilistic amplitude shaping (PAS) architecture [61]. A non- 2^n PAM- X template is simply generated by truncating the 2^n constellation [65]. This is equivalent to setting a probability of zero for the outer PAM levels, in the context of PCS. The strategy has been

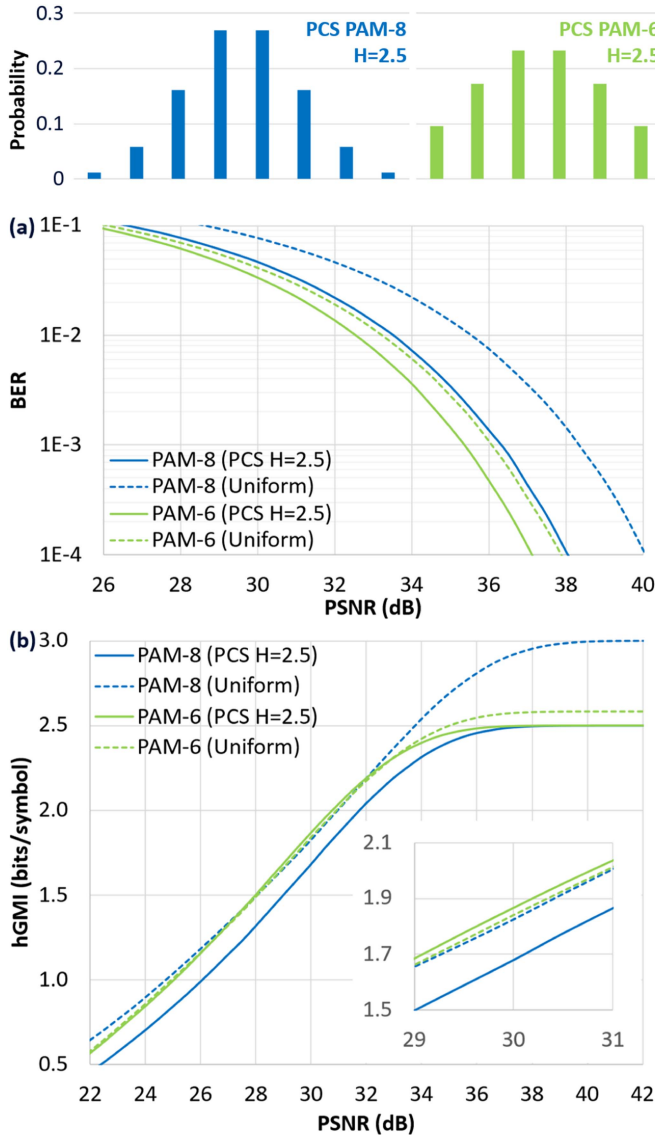


Fig. 8. Comparison of (a) BER and (b) hGMI between PCS PAM-6 and PAM-8 signals with the same entropy (2.5 bits/symbol), with uniform PAM-6 and PAM-8 signals as baselines. The simulation is a PPC system characterized by PSNR defined in (2) and (3), with an RRC (0.01) filter to emulate PAPRE.

popular in latest high-speed short-reach coherent transmissions limited by the transmitter PPC [66], [67]. By simulation, we show how it improves the PCS performance in an IM-DD system, as shown Fig. 8, with root-raised cosine (RRC) filtering (0.01 roll-off) as a PAPRE example. With the same entropy of 2.5 bits/symbol, the PCS signal using a PAM-6 template clearly achieves a lower BER than using a PAM-8 template. We use the generalized mutual information (GMI) under binary hard-decision (HD) decoding [68] (hGMI) as the AIR metric,

$$hGMI = \mathbb{H}(X) - \log_2 |\mathcal{X}| \cdot \mathbb{H}_2(\epsilon) \quad (10)$$

where $\mathbb{H}(X)$ is the entropy of signal X and $\mathbb{H}_2(\cdot)$ is the binary entropy function, ϵ is the bit error probability, and $|\mathcal{X}|$ is the size of modulation alphabet. In the inset of Fig. 8(b), PCS PAM-6 achieves an hGMI gain over uniform PAM-6. In contrast, PCS

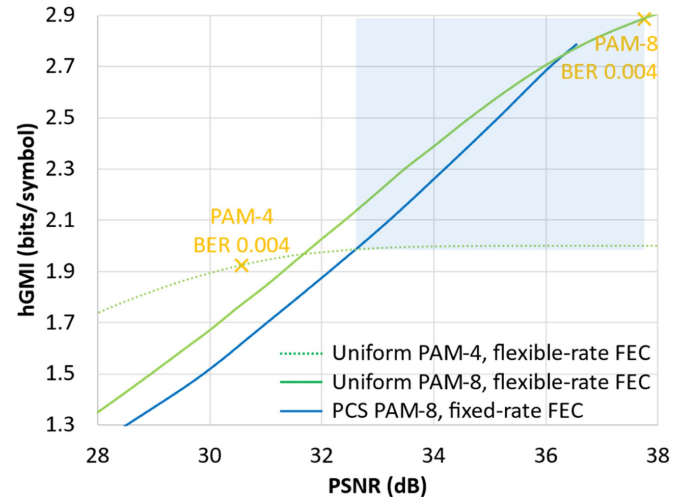


Fig. 9. Rate adaptations using uniform PAM-4/8 with flexible-rate FEC codes, and PCS PAM-8 (entropy up to 2.9 bits/symbol) with a fixed-rate FEC code. The simulation uses a PPC system with a RRC (0.01) filter to emulate PAPRE.

PAM-8, with the same entropy as PCS PAM-6, shows a big penalty with respect to uniform PAM-8.

Second, because the outer PAM levels with higher power occur less frequent than the center levels, the MB-PCS signal has higher tolerance to various transmitter nonlinearities than the uniform signal, like digital clipping, power saturation of the driver amplifier and nonlinear E/O conversion. This means the peak-to-peak swing of the MB-PCS signal can be higher than the uniform signal for a similar nonlinearity penalty. In other words, the PPC is enlarged using MB-PCS. Such nonlinearity tolerance advantage is verified in an EAM-based system [64] with a nonlinear transfer function like in Fig. 4(a).

Together with the light shaping strategy, it is highly possible for an IM transmitter with PAPRE and nonlinearity to achieve shaping gain using the classic MB-PCS.

B. Distinguishing Two Types of PCS Benefits - Shaping Gain & Rate Adaptation

In PCS IM-DD literatures that demonstrate shaping gain, the reference baseline to claim the gain varies from one to another, making it difficult to perform cross comparisons among them. To clarify the gain, we distinguish two types of PCS gains in this section. The simulation example is a PPC system with the PAPRE emulated by a 0.01 roll-off RRC filter. We compare the hGMI as a function of PSNR among uniform PAM-4/8 and PCS PAM-8 signals in Fig. 9. For uniform PAM signals, we assume an ideal flexible-rate FEC for rate adaptation; while for PCS signals, we adjust the entropy up to 2.9 bits/symbol for rate adaptation under a fixed-rate HD-FEC with a BER threshold of 0.004. In Fig. 9, the PCS PAM-8 curve lies below the uniform PAM curves when PSNR < 36.3 dB, indicating no shaping gain for PCS signals. Nevertheless, most low-cost IM-DD systems only allow one FEC code. Given a fixed-rate FEC, if the system PSNR is between the two PSNR thresholds of uniform PAM-4 and PAM-8, PCS may effectively improve the AIR owing to its rate adaptation capability using a fixed-rate FEC. For example,

if $PSNR = 36$ dB in Fig. 9 and the system only allows an HD-FEC with a BER threshold of 0.004, the AIR of uniform PAM-4 is limited to within 2 bits/symbol and uniform PAM-8 cannot meet the BER threshold. In contrast, PCS PAM-8 improves the AIR to 2.69 bits/symbol. The meaningful PCS rate adaptation range with the fixed-rate FEC is indicated by the shaded area in Fig. 9. This phenomenon is experimentally verified in [39].

In general, when a PCS gain over uniform signaling is reported, it should be distinguished between:

- 1) Shaping gain: “absolute” AIR gain when the co-design of modulation format and multi-rate FEC is allowed.
- 2) Rate adaptation: AIR improvement under a variable system condition (e.g., different PSNRs) using only one (or very limited) choice of FEC code.

PCS is guaranteed to achieve both types of shaping benefits in an APC system, and achieve none of the benefits in a strict PPC system without PAPRE. In an IM-DD system with PAPRE, the rate adaptation can be partially achieved, but the shaping gain is not guaranteed, which depends on the transmitter PAPRE (and nonlinearity) condition, as we discussed in Section III-A. Between the two types of benefit, future IM-DD applications may value rate adaptation more than shaping gain, considering shaping gain provides very limited improvement (e.g., <10%) on the absolute AIR even in the best case. On the other hand, rate-adaptive short-reach transceivers have been hotly pursued recently, like the flexible-rate passive optical network (FLCS-PON) aiming at flexible quality of service among end users [69].

The gain of a PCS scheme is closely related to its distribution matcher (DM). The ideal DM can be approximated, e.g., by constant composition distribution matching (CCDM) [70], that achieves marginal rate loss with long block length ($\geq 10^3$). Short block length techniques reduce the DM complexity at a sacrifice of a rate loss [71], [72]. Despite the reduced shaping gain, they usually have little impact on the rate adaptation range. Therefore, if shaping gain is not the primary consideration for an IM-DD application, it would be suitable to utilize the short block length DM for cost saving.

C. Debate on the Reverse-MB Distribution

In PPC IM-DD systems, a typical competitor of MB-PCS is the reverse-MB (Rev-MB) distribution [73] as illustrated in Fig. 10(a). Opposite to MB, Rev-MB assigns higher probabilities to the PAM levels with higher power. As the outer two PAM levels with only one neighbor are more frequently generated than the inner levels with two neighbors, Rev-MB reduces the wrong decisions to the neighboring levels and improves BER. Such a symbol decision advantage is maximized in a PAPRE-free PPC system, but quickly decreases and even vanishes given higher PAPRE. This is because Rev-MB consumes much higher power than MB for the same entropy, while PAPRE tends to convert a PPC system towards an APC one where the average power becomes “precious”. Using the RRC filtering as an example, we select two roll-off factors of 0.3 and 0.01 to represent different amount of PAPRE and compare the BER of MB and Rev-MB signals in Fig. 10(b). Compared to the case of RRC 0.3, the Rev-MB gain is quickly degraded by the stronger PAPRE (RRC

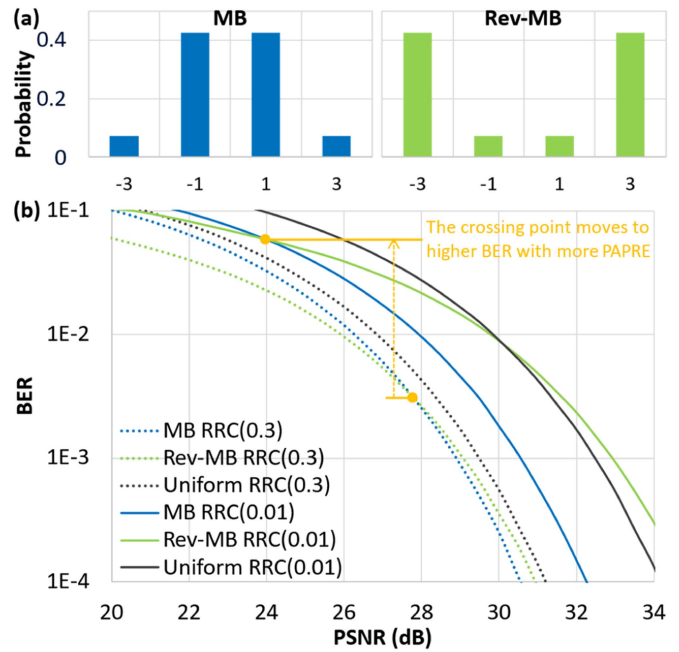


Fig. 10. Comparisons between MB and Rev-MB PCS PAM-4 (with the same entropy of 1.6 bits/symbol): (a) probability mass function, and (b) BER as a function of PSNR. PAPRE is emulated by RRC filters (0.3 and 0.01 roll-off).

TABLE I
INFLUENCE OF DIFFERENT SYSTEM PARAMETERS ON MB & REV-MB PCS

System Parameters		MB	Rev-MB
Tx PAPRE factor*	Higher	Better	Worse
Tx nonlinearities	Higher	Better	Worse
Degree of shaping*	Stronger	Worse	Better
FEC threshold (BER)	Higher	Worse	Better

* The PAPRE factor is defined in (4.1)

* Under the same PAM template, a stronger shaping means a smaller entropy

0.01), as indicated by the move of the crossing point between MB and Rev-MB curves in Fig. 10(b). For the case of RRC 0.01, Rev-MB remains a little gain over MB only at very high BER (>0.05), which is not practical to be corrected by simple FECs in IM-DD systems. Generally, the comparison between MB and Rev-MB depends on various system aspects like the ones shown in Table I, and there is no simple conclusion applicable to all PPC IM-DD systems. Both Rev-MB [69], [73], [74] and MB gain [39], [63], [64] were demonstrated in literatures.

D. An Open Question: The Optimal Distribution for IM-DD

Beyond MB and Rev-MB, a more general task is to find the optimum distribution for a PPC system taking the PAPRE into account. This is not a trivial task, especially for a PAM template with a larger alphabet. There emerge a few attempts to tackle the question, like exhaustively trying a group of distributions [64], and using optimization algorithm or machine learning to find the distribution of a particular system [75], [76], [77]. However, it lacks a generalized method applicable to a system with arbitrary

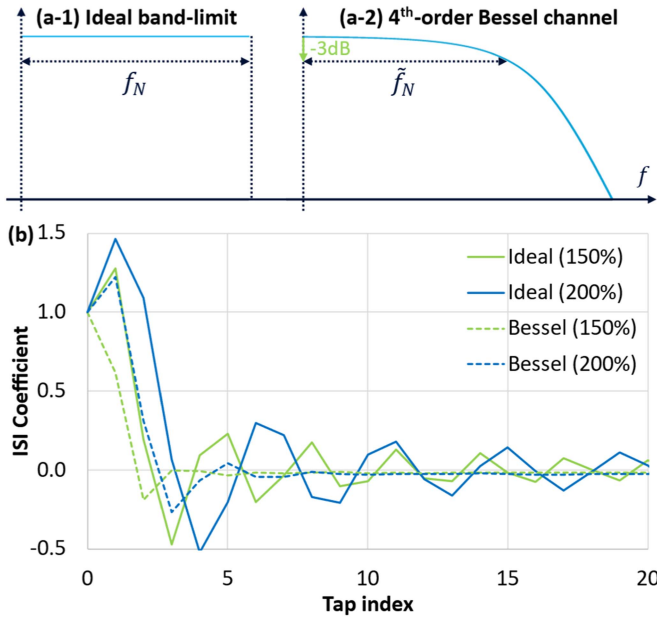


Fig. 11. Examples of BWL channels: (a) frequency-domain model defined in (12), and (b) time-domain model defined in (11). The noise of these channels is assumed to be AWGN.

PAM alphabet, FEC threshold, and transmitter PAPRE. This leaves to be an open question for the near future.

IV. COMBATING THE BANDWIDTH LIMIT (BWL)

BWL introduces inter-symbol interference (ISI) to a signal. Without loss of optimality, a digital receiver may be designed to consist of a whitened matched filter (WMF) and a symbol-rate sampler [36], and the equivalent discrete-time model for the ISI-contaminated received sequence $\{y_t\}$ is

$$y_t = \sum_{k=0}^{L-1} h_k x_{t-k} + n_t \quad (11)$$

where $\{x_k\}$ is the transmitted sequence, $\{h_k\}$ is the sequence of coefficients of a discrete-time channel response and $\{n_k\}$ is the discrete-time white Gaussian noise sequence. In the frequency domain, the equivalent model is

$$Y(f) = H(f) X(f) + N(f) \quad (12)$$

Fig. 11(a) illustrates the spectra for an ideal band-limited channel and a lowpass 4th-order Bessel channel with additive white Gaussian noise (AWGN). The Nyquist bandwidth of the ideal channel is f_N , and the nominal f_N for the Bessel channel is defined as its 3-dB bandwidth (\tilde{f}_N). Given two symbol rates 150% and 200% of the Nyquist limit $2f_N$, respectively, Fig. 11(b) shows their discrete-time models defined in (11) for both channels. The optimum detector for a digital sequence in the presence of ISI is MLSE or maximum a posteriori probability (MAP) decoder [78], [79]. Both are sequential detectors which perform symbol decisions based on sequence observations, and the sequence length is determined by the length (L) of channel memory in (11). Although BWL commonly exists in most

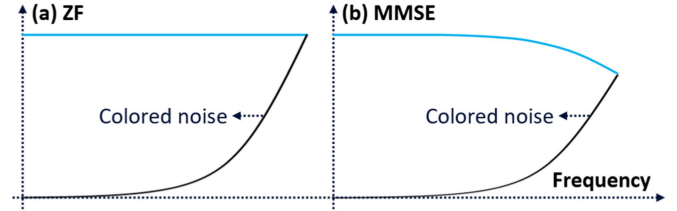


Fig. 12. Illustration of PSDs for signal (blue curves) and noise (black curves) in a colored SNR channel after (a) ZF and (b) MMSE equalizations.

real-world systems, a symbol-by-symbol detector is commonly used for simplicity. The conditions to achieve the optimum symbol-by-symbol detection performance are:

- 1) the signal sequence is ISI free;
- 2) the noise sequence is whitened.

Clearly, a BWL channel cannot simultaneously satisfy the two conditions above. In this section, we review the equalization techniques for ISI mitigation in BWL systems. More crucially, we explain whether it is meaningful to pursue higher-symbol rate beyond the BWL using these techniques. A more detailed study on this topic can be found in [37].

A. Nyquist Signaling and Linear Equalization

The easiest way to avoid a huge amount of ISI is to limit the signal bandwidth to a low-frequency region where the channel response is not severely colored. Besides symbol rate reduction, a useful tool to compress the signal spectrum is pulse shaping [80]. Pulse shaping can shrink the signal spectrum close to a rectangular shape to take full advantage of limited bandwidth and support at most $2B$ -Baud signaling given B -Hz bandwidth according to Nyquist theorem. It usually relies on the Nyquist filter to have zero ISI at the sampling instants. The most popular Nyquist filters are raised cosine (RC) and RRC filters.

For a lightly colored channel response, ISI can be alleviated by linear equalizations. A linear equalizer is commonly a feed-forward equalizer (FFE) that creates a number of delayed copies of its input $\{y_t\}$ and adds them back to $\{y_t\}$ with the proper coefficients $\{c_t\}$. This is equivalent to the convolution between $\{y_t\}$ and $\{c_t\}$ in time domain, namely, the filter output $\{s_t\}$ is

$$s(t) = y(t) * c(t) \quad (13)$$

and the corresponding frequency-domain expression is

$$S(f) = C(f) Y(f) = C(f) H(f) X(f) + C(f) N(f) \quad (14)$$

There are different criteria to obtain the filter coefficients $c(t)$. A typical one is the zero-forcing (ZF) criterion as shown in Fig. 12(a). While a ZF filter minimizes the ISI by forcing $C(f) = 1/H(f)$, it imposes the inverse channel response to the noise as $N(f)/H(f)$, which colors the noise and breaks the optimum condition for symbol-by-symbol detection. An alternative to tradeoff between ISI and noise coloring is the minimum mean-square error (MMSE) criterion as shown in Fig. 12(b). Because the MMSE filter directly minimizes the error between $\{s_t\}$ and the target signal $\{x_t\}$, it usually improves the

symbol decision than a ZF one when the colored response is not ignorable.

B. Why Do We Need Nonlinear Equalization for ISI Mitigation?

A question that usually puzzles a beginner is why a nonlinear equalizer is superior to a linear one for ISI mitigation. In a BWL system, linear equalization is suboptimum, because it can only tradeoff between the two optimum conditions for a symbol-by-symbol detector no matter what criterion is chosen to obtain its coefficients. This is due to its linear nature, namely, it applies the equalization equally to the signal and the noise, as shown in (14). This drawback becomes more severe if we extend the signal bandwidth significantly beyond the “flat” region of the channel response. In this case, the signal can no longer be approximated as a Nyquist one, but goes faster than Nyquist (FTN) [81], [82]. FTN is a generalized terminology to describe a heavy ISI system with a symbol rate higher than its bandwidth can support.

To eliminate ISI without coloring the noise in FTN systems, the equalizer should decouple the ISI mitigation from the noise coloring. This leads to a popular nonlinear symbol-by-symbol detector, i.e., the decision-feedback equalizer (DFE). The basic DFE concept is to perform the symbol decision, a nonlinear operation, during the equalization to remove the noise from the received signal. If all previous decisions \hat{x}_{t-k} ($k \geq 1$) are correct (i.e., the ideal DFE assumption, $\hat{x}_{t-k} = x_{t-k}$), the ISI tail can be simply subtracted from (11) known the channel coefficients $\{h_k\}$ and the equalized sample

$$y'_t = y_t - \sum_{k=1}^{L-1} h_k \hat{x}_{t-k} = x_t + n_t \quad (15)$$

becomes ISI-free with only white noise. This means an ideal DFE can realize the optimum symbol-by-symbol detection.

The DFE performance is degraded if the system SNR is low, because of the increase of wrong symbol decisions. In this case, the subtraction of ISI tail can be moved to the transmitter. This precoding technique is named Tomlinson-Harashima precoding (THP) [36]. THP can narrow the transmitted signal spectrum to accommodate the BWL. After the channel propagation, the pre-coded signal is ISI free and can be detected on a symbol-by-symbol basis. THP has revived recently in a variety of IM-DD applications [83], [84].

C. Type of Nonlinear Equalizations for ISI Mitigation

All nonlinear equalizers aiming at combating the BWL are designed to eliminate the ISI influence while keeping the noise white. Generally, they can be categorized in Fig. 13. DFE and THP belong to the symbol-by-symbol category. As mentioned at the beginning of Section IV, an optimum receiver should observe the entire ISI-contaminated sequence by a sequential detector. The input of a sequential detector should have a white noise as in (11). The noise can be whitened at the receiver by a noise whitening filter (NWF), part of the WMF structure [36]. Alternatively, the transmitter can perform partial-response (PR) pre-filtering to limit the signal bandwidth, aiming to obtain the

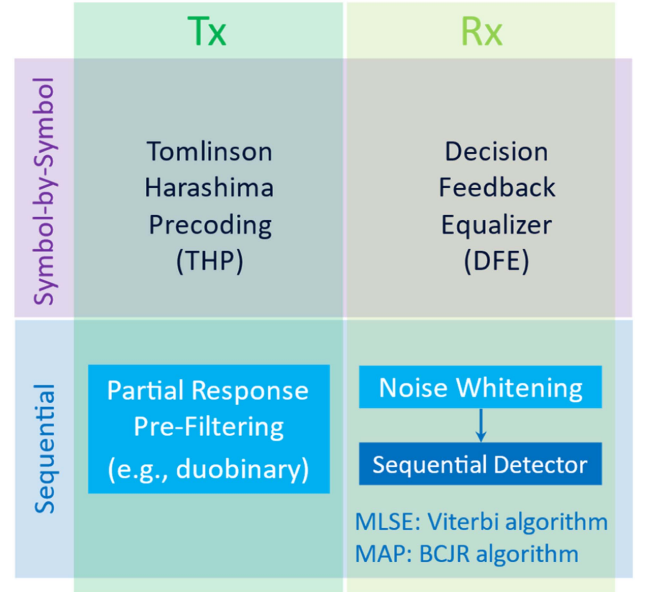


Fig. 13. Types of FTN techniques categorized by their locations (Tx or Rx) and decoding manner (symbol-by-symbol or sequential decoding).

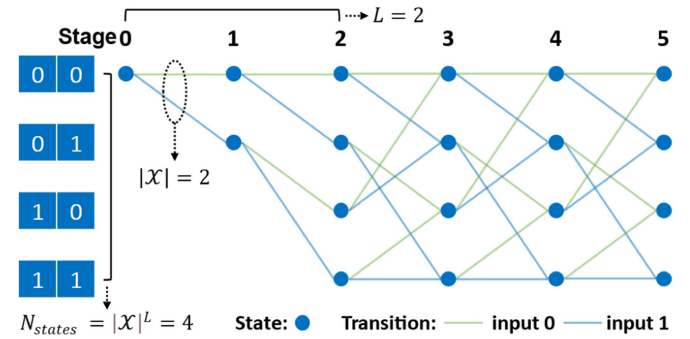


Fig. 14. Trellis graph of a sequential detector with PAM-2 ($\mathcal{X} = \{0, 1\}$) and memory length $L = 2$, leading to $|\mathcal{X}|^L = 4$ states per stage.

white noise after the channel automatically without the receiver NWF. The simplest PR filter has a “ $1 + \alpha$ ” structure, namely, a sample with controlled ISI from the past sample $x_t + \alpha x_{t-1}$. It is called a duobinary filter when $\alpha = 1$, and was considered recently for high-speed 25/50G PON applications [85].

Instead of removing ISI or noise from the received samples, a sequential detector directly compares the received sequence with all possible sequences and finds the one with the maximum likelihood. A sequential detector is usually characterized by a trellis graph. A toy example is illustrated in Fig. 14. Each trellis stage contains all possible sequences (called trellis states),

$$\mathbf{s} = (x_t, x_{t-1}, \dots, x_{t-L+1}), x \in \mathcal{X}$$

by tracing back to the past samples with a memory length of L . The total number of states is $|\mathcal{X}|^L$. The input symbol at the next stage triggers a transition between the states at the current and the next stages, as indicated in Fig. 14. The most popular metric

to characterize the likelihood of each transition is

$$l = \left| y_t - \sum_{k=0}^{L-1} h_k x_{t-k} \right|^2 \quad (16)$$

and the probability is proportional to $\exp(-l/N_0)$ if the noise follows a Gaussian distribution $N \sim (0, N_0)$. The likelihoods of all the sequences for each sample are sent to a Viterbi decoder to generate the symbol-wise hard decisions [78], or to a BCJR decoder to calculate the symbol-wise soft probabilities [78].

Obviously, the total number of sequences $|\mathcal{X}|^L$ determines the complexity of a sequential detector that grows exponentially with L . If the ISI of a BWL system has a long tail, it must be truncated to tradeoff the performance with complexity. This is usually referred to as channel shortening. A channel shortening filter (CSF) [86] can be designed to combine multiple functions like matched filtering and noise whitening to make its output follow a target ISI model as in (11) with a shortened memory. Besides, given the ISI length of L , the complexity can be further reduced by simplifying the trellis search [87], [88]. An easily understood approach is to omit the states (or transitions) with low probabilities. There emerged various proposals to simplify the sequential detectors for IM-DD systems [37], [89], [90], [91], [92].

D. Pre-Processing Vs Post-Processing

In the FTN system, a transmitter can perform pre-processing like pulse shaping, THP and PR filtering which actively shrink the signal bandwidth to accommodate the BWL. Alternatively, the signal can be naturally lowpass filtered by the BWL channel, leaving the entire ISI mitigation at the receiver for post-processing. Pre-processing changes the signal distribution, resulting in more levels after PR filtering and even an arbitrary waveform after pulse shaping or THP. It makes a big difference for a PAM transmitter, because pre-processing (i) aggravates PAPRE and reduces the SNR via (7), and (ii) makes DAC an essential transmitter element. The need of a DAC is problematic, because it is a big portion of power consumption in a PAM DSP ASIC chip [5]. Moreover, it means the electrical signal from any source, like a PAM-4 signal from the switch SerDes, must be regenerated (e.g., retiming, reshaping, reamplifying) to drive an E/O modulator. This calls for extra “gearboxing” between electronics and optics and prevents the “linear” drive optics [93]. As a DAC-based transmitter also support higher-order PAM or even multicarrier modulation (whose capability of combating the BWL will be shown in Section V) to improve the AIR, pre-processing weakens an attractive feature of FTN with respect to these competitors, namely, the modulation simplicity.

E. Is It Meaningful to Use FTN Signaling?

FTN signaling degrades the average system SNR, because it extends the signal spectrum to frequencies with weaker channel response. Such tradeoff between SNR and symbol rate leads to a debate whether it is better to limit the signal bandwidth to the lower frequency region for a higher SNR to enable higher-order modulations. We discussed this in our ECOC presentation [7] and investigated it in detail in [37].

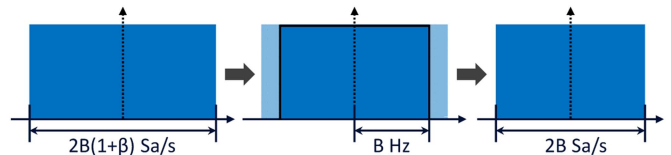


Fig. 15. Generating a $2B(1+\beta)$ Baud signal with a sub-sampling rate of $2B \text{ Sa/s}$. This example uses a rectangular lowpass filter as the anti-aliasing filter.

The answer to this question is closely related to the response of the BWL channel. The channel models in Fig. 11 represent two typical frequency responses: one with a sharp cutoff at f_N and the other with a gradually decayed profile. For the 1st channel, it will be more efficient to approach its capacity by higher-order modulations; while for the 2nd one, FTN signaling can better approach its capacity. The reason has been revealed in (1). The motivation of FTN is to expand the signal spectrum to the entire useable bandwidth to maximize the integral in (1). If the channel provides little response after a cutoff frequency, spectrum expansion beyond the cutoff point no longer improves the bandwidth usage. In this case, though FTN can approach the capacity keeping the low-order modulation by increasing the symbol rate, the computational complexity will be much higher than that of the high-order modulation without FTN. This is clearly shown in Fig. 11(b) by the much longer the ISI tail for the 1st channel with the sharp frequency cutoff.

By FTN signaling, a digital transmitter can generate a symbol rate higher than its physical sampling rate. As shown in Fig. 15, the transmitter first generates a high symbol rate signal and then limits its bandwidth by lowpass-filtering to resample the signal at a lower clock rate without sampling aliasing. This technique is implemented in recent IM-DD experiments to pursue higher symbol rate [17], [94], [95]. A sub-sampling system usually has a channel with a sharp frequency cutoff, because the anti-aliasing filter has a steep roll-off (like the rectangular filter in Fig. 15) to avoid extra power loss within the finite digital bandwidth (i.e., half of the sampling rate). Therefore, sub-sampling signal generation is commonly not helpful to improve the system AIR though it achieves higher symbol rate.

V. MULTICARRIER MODULATIONS

Multicarrier modulation divides the spectrum into multiple subcarriers, and each subcarrier is independently modulated at a lower symbol rate. In essence, it discretizes the integral of the frequency-resolved capacity in (1) as the capacity summation of all subcarriers. It is the most effective way to make full use of the bandwidth resource under BWL owing to its capability of adaptive modulation per subcarrier basis. In this section, we review common multicarrier modulation schemes and adaptive loading algorithms and perform an AIR comparison among a variety of schemes that combat the BWL, including the FTN technique in Section IV. The comparison is based on a 200-Gb/s IM-DD experiment presented in [96].

A. Type of Multicarrier Modulations

Generally, there are two types of multicarrier modulations, namely, subcarrier multiplexing (SCM) [97], [98] and

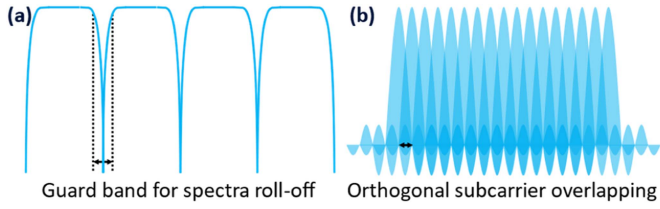


Fig. 16. Frequency-domain illustration of multicarrier signaling using (a) SCM and (b) OFDM (or DMT).

orthogonal frequency division multiplexing (OFDM) [99], [100], as shown in Fig. 16. The SCM signal is simply a combination of multiple single-carrier ones centered at a series of frequencies without spectra overlap. Each single-carrier signal is usually sharply pulse-shaped for a high spectral efficiency. The multiplexing and demultiplexing are commonly performed by a filter bank, which is an array of bandpass filters to combine and separate multiple subcarriers. SCM has been implemented to the state-of-the-art 800G coherent transceivers [101]. For real-valued IM, a complex-valued subcarrier filled with quadrature amplitude modulation (QAM) symbols can be separated into in-phase (I) and quadrature (Q) parts which then go through a pair of filters for frequency upconversion, whose impulse responses form a Hilbert pair. This leads to a double sideband signal whose left sideband is Hermitian symmetric to the right. In IM-DD literatures, it is also referred to as carrierless amplitude phase (CAP) modulation and the multiple-subcarrier version is named multiband CAP [102]. An alternative multiplexing algorithm for SCM is based on the discrete Fourier transform (DFT). It converts multiple single-carrier signals to frequency domain by a small-size DFT, arranges them at different frequencies, and brings the multiband signal back to time domain by a big-size IDFT. The process is named DFT spread (DFTS) OFDM [97], [103], and has been adopted in cellular network standards.

The other multicarrier modulation is OFDM which is usually referred to as discrete multitone (DMT) [104], [105] in IM-DD literatures. DMT allows spectra overlap between *orthogonal* subcarriers, and directly maps the densely spaced subcarriers in frequency domain and converts them to time domain by IDFT. DMT offers much finer frequency granularity to adapt the BWL more precisely than SCM. Each subcarrier has a much lower symbol rate, which simplifies the channel equalization as single tap per subcarrier. On the other hand, the massive number of subcarriers aggravates the PAPRE and reduces the signal SNR according to (7).

B. Bit Loading vs Entropy Loading

A loading algorithm finds an appropriate modulation format for each subcarrier based on its SNR. Its goal is to reserve the same SNR margin to a given FEC threshold for all subcarriers, which enables the same error protection among them by a fixed-rate FEC code. Each subcarrier can be loaded with a uniform QAM format with integer bits using bit loading (BL) [106], [107] or a PCS-QAM format with fine rate granularity using entropy loading (EL) [108], [109], [110], [111]. Because uniform QAMs only provide integer rate granularity, it is almost impossible for

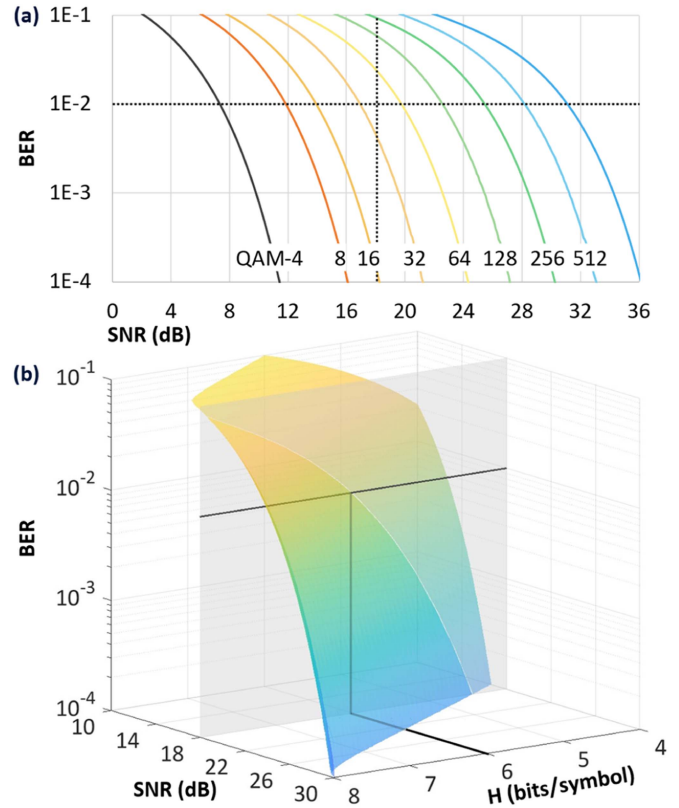


Fig. 17. Lookup tables (BER as a function of SNR and modulation format) for (a) BL and (b) EL (PCS 256-QAM). Both figures show an example of finding the format under the BER target of $1e-2$ given a subcarrier SNR of 18 dB.

BL to find a format exactly matching the loading target. Thus, an additional power loading (PL) usually follows the BL to fine tune the SNR of each subcarrier [106]. In contrast, EL can fractionally adjust the entropy of each subcarrier by PCS, which offers a precise matching without PL [109]. In Fig. 17, we compare BL and EL by illustrating their lookup tables which store the BER of all the available formats given various channel SNRs. As an example, for the channel SNR of 18 dB, BL should choose either 32- or 64-QAM, and then perform an extra PL to match the BER target of $1e-2$. In contrast, EL straightforwardly meets the BER target when the entropy is about 6 bits/symbol (PCS 256-QAM).

EL has two main advantages over BL. First, it avoids the PL per subcarrier. Second, it achieves the PCS gain over BL regardless of the PPC due to the *extreme PAPRE* of IDFT, as explained in Section II-C.

C. Comparison Among Algorithms to Combat the BWL

Both multicarrier and FTN techniques can combat the BWL. The question is who can achieve a higher AIR in the same BWL system. On one hand, a multicarrier signal, especially a DMT one, has a much better frequency resolution to adapt the BWL response owing to its massive number of subcarriers. It also achieves the PCS gain using EL, which vanishes in most PAM systems under the PPC unless strong PAPRE is presented. On the other hand, multicarrier modulations usually come with a significant PAPRE, which translates to bigger SNR penalty

compared to the FTN PAM signal with little transmitter pre-processing. We perform an experimental comparison between PAM and DMT by including the above aspects in a state-of-the-art 200-Gb/s IM-DD system [96].

The experimental setup and DSP details can be found in [96]. In brief, we use an IM transmitter whose BWL is mainly from the E/O modulator. The electronics have sufficient bandwidth to generate a wideband signal without significant distortions up to 128 GBd. Thus, the PAM signal, generated with 1 sample per symbol (sps) with negligible PAPRE, can have a much lower $PAPR_{drive_signal}$ than the DMT signal. Meanwhile, to estimate the lower-bound PAM performance with PAPRE, we include a pre-processed PAM signal with Sinc-pulse shaping for higher $PAPR_{drive_signal}$. The $PAPR_{drive_signal}$ of the Sinc PAM-4 signal is adjusted to be the same as the DMT signal to emulate an identical transmitter signal power penalty for the two. We use MLSE as a typical FTN decoder for PAM-4 signals. The 1-sps and Sinc-pulse shaped PAM-4 signals are compared with both BL- and EL-DMT signals.

To compare the AIR under different BER targets (i.e., HD-FEC thresholds), we evaluate the $hGMI$ of the received signals as a function of measured BER, as shown in Fig. 18. Between the DMT-BL and Sinc-shaped PAM-4 (w/o MLSE) signals, BL improves the maximum AIR by $>10\%$ owing to the capability of combating the BWL. In contrast, with FTN decoding, Sinc PAM-4 achieves the maximum AIR only 2.3% away from that of BL, and even outperforms BL at lower BER. Furthermore, by minimizing the PAPRE, the 1-sps PAM-4 signal obtains a higher transmitter signal power that greatly improves its AIR with respect to the Sinc PAM-4 signal. On the other hand, with PCS, EL achieves 7.9% AIR gain over BL in the capacity-approaching region with a BER of around 0.02, and $>10\%$ gain when the BER is less than 0.01. The AIR is similar between 1-sps PAM-4 and DMT-EL signals. While the DMT-EL achieves a maximum AIR 2.5% higher than that of the 1-sps PAM-4, the 1-sps PAM-4 slightly outperforms DMT-EL in lower BER region.

In brief, if the BWL is mainly from the E/O modulator rather than the electronics, the EL advantage over PAM may be marginal, in which case FTN-PAM is more appealing due to its simplicity. In contrast, if the electronics induce huge PAPRE for PAM signals (e.g., by digital pre-processing or strong BWL), the EL gain may be increased to more than 10%, making DMT a meaningful option with higher rate flexibility and AIR.

VI. ROLE OF UNIVERSAL NONLINEAR EQUALIZATIONS

A nonlinear equalizer can be designed to accommodate a particular type of nonlinearity, like the group of nonlinear equalizers customized for BWL in Section IV. On the other hand, it may be designed to universally compensate for any type of nonlinear impairments based on the universal approximation theorem in mathematics. For IM-DD systems, the most widely studied universal nonlinear equalizers are the Volterra series [112], [113], [114], [115], [116], [117], [118], [119], [120] and the artificial neural networks (ANN) [121], [122], [123], [124]. In this section, we analyze the Volterra nonlinear equalizer

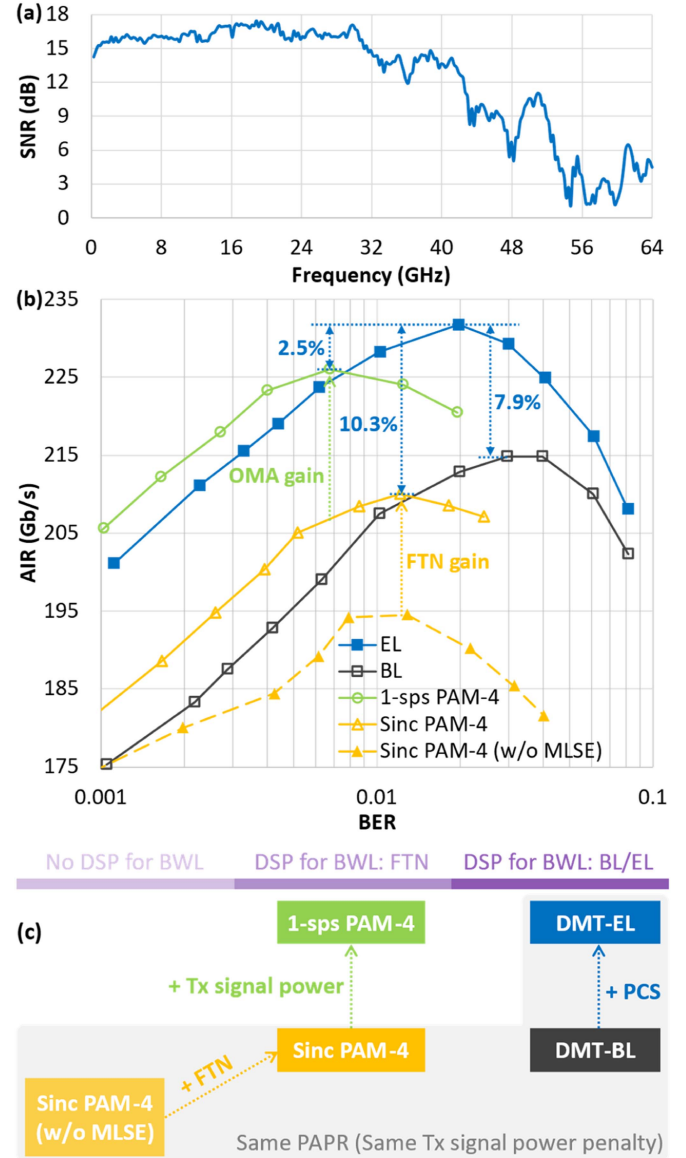


Fig. 18. Experimental comparison between PAM and DMT signals in a BWL IM-DD system [96]. (a) End-to-end system SNR as a function of frequency; (b) AIR (evaluated by $hGMI$) as a function of received BER; (c) illustration of the key factors that influence the comparison results in (b).

(VNE) as a typical universal equalizer, and then briefly extend our discussion to ANN.

A. VNE Structure

The Volterra series is originated from the theorem of Taylor expansion [112]. For a nonlinear system with no memory, its output $y(t)$ can be expressed in a Taylor series of its input $x(t)$

$$y(t) = \sum c_n x^n(t) \quad (17)$$

namely, the weighted summation of polynomials. The Volterra series is an extension of (17) in a system with memory. Below

is an example of 3rd-order Volterra series

$$\begin{aligned}
 y(t) &= c_0 + \sum_{t_1=0}^{L_1-1} c_1(t_1) x(t-t_1) \\
 &+ \sum_{t_1=0}^{L_2-1} \sum_{t_2=0}^{L_2-1} c_2(t_1, t_2) x(t-t_1) x(t-t_2) \\
 &+ \sum_{t_1=0}^{L_3-1} \sum_{t_2=0}^{L_3-1} \sum_{t_3=0}^{L_3-1} c_3(t_1, t_2, t_3) x(t-t_1) x(t-t_2) x(t-t_3)
 \end{aligned} \tag{18}$$

The multiplication terms $\prod_i x(t-t_i)$ are the Volterra kernels. The total number of the r -th order Volterra kernels is

$$\binom{L+r-1}{r}$$

where L is the memory length. The total number of coefficients is in the order of $O(L^R)$, where R is the highest order of kernel.

Besides the polynomial kernels, VNE can extend its kernel with absolute terms $|x(t-t_i)|^r$ [115], which was found to be effective to combat the nonlinearity from power amplifiers.

B. VNE-based Components Nonlinearity Compensation

A common application of Volterra series is the components nonlinearity compensation, like the RF driver nonlinearity as shown in Fig. 7(a) and the nonlinear E/O transfer function in Fig. 4. Using VNE to enhance the nonlinear tolerance, the drive signal swing can be extended to the nonlinear region of those components to achieve higher transmitter signal power. Since VNE effectively improves the SNR, it is popular in recent short-reach transmission experiments with transceiver-limited performance [67].

VNE can be applied both to the receiver for post equalization, and to the transmitter as digital pre-distortion (DPD). As a pre-processing at transmitter, DPD brings about similar concerns as mentioned in Section IV-D. On the other hand, DPD is applied to a clean signal free from channel noises that avoids potential equalization enhanced noises at receiver. There are two well-known learning architectures to obtain the VNE coefficients for DPD, namely, the indirect learning architecture (ILA) [125] and the direct learning architecture (DLA) [126], as illustrated in Fig. 19. In brief, ILA first trains the VNE as a post-equalizer at the receiver, and then moves it to the transmitter for DPD. In contrast, DLA first models the channel as a Volterra series (named an auxiliary channel) and trains this model to minimize the output difference between the auxiliary and the real-world channels. Then it puts another Volterra series in front of the auxiliary channel for DPD, which is trained to minimize the difference between the DPD input and the auxiliary channel output. The auxiliary channel provides a differentiable model to back propagate the gradients of the loss function. DLA doubles the complexity of ILA due to the use of two Volterra series, but often achieves a better nonlinearity mitigation [54].

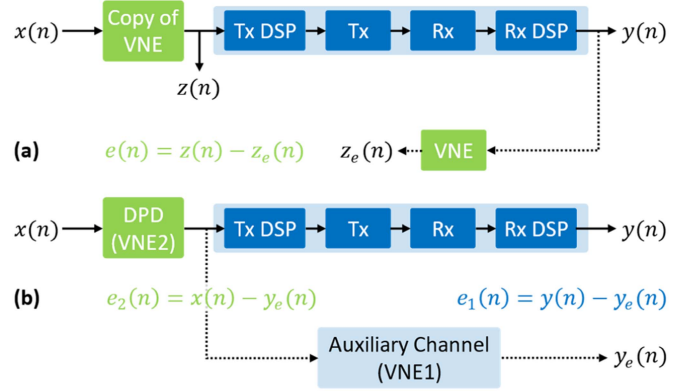


Fig. 19. Two DPD architectures using VNE: (a) indirect learning architecture (ILA) and (b) direct learning architecture (DLA).

C. Is VNE Suitable for Combating the BWL?

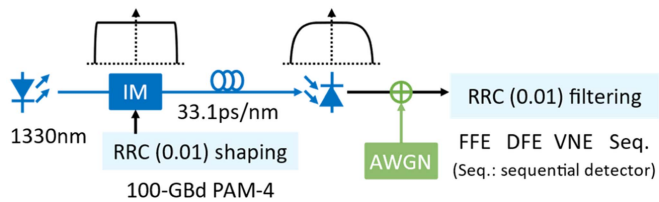
VNE has been considered in many IM-DD works to combat the BWL [119], [123]. A question is whether VNE could compete with the equalizers specifically designed for the BWL channel as summarized in Section IV. We make a comparison between them and dig into the reason in this section.

In the simulation, we emulate a BWL channel by transmitting an IM signal over the dispersive fiber. The combination of CD and square-law detection induces the frequency selective fading, as explained in Section II-D. A 100-GBd PAM-4 signal is pulse shaped by an RRC filter with 0.01 roll-off. It is modulated on a 1330-nm light using a chirp-free E/O modulator. We choose the CD value of 33.1 ps/nm to make the first fading notch around 50 GHz, i.e., the Nyquist frequency of the 100-GBd signal. The received signal spectrum is illustrated in Fig. 20, with a clear spectrum narrowing effect approaching the Nyquist frequency. We evaluate the performance for linear FFE, VNE, DFE and MLSE, with the equalizer parameters listed in Fig. 20.

Among all the equalization schemes, the sequential detector with MLSE achieves the best performance as expected. For the remaining symbol-by-symbol detectors, DFE exhibits the best performance in most cases. As explained in Section IV-B, DFE is the optimum symbol-by-symbol detector given the ideal DFE assumption in (15). In contrast, VNE only slightly improves the BER over the linear FFE. Despite the higher complexity, it has an SNR gap to DFE, especially at the lower BER region. Hence, VNE does not seem to be cost-effective to combat the BWL.

D. Implementation Considerations on VNE

As a close-to-optimum symbol-by-symbol detector in the presence of ISI, DFE relies on HD, a nonlinear operation to calculate the feedback ISI. If a VNE targets a performance close to DFE, it should approximate the HD using polynomials. The HD on a PAM signal is a step function with discontinuity at the decision boundary. A necessary condition for Taylor expansion is that the function should be continuously differentiable, and this is inherited by the Volterra series [112]. Therefore, Volterra Series is not efficient to approximate DFE. This reveals an application limitation of VNE, namely, it only approximates



FFE	32 taps @2sps
VNE	(32,16,16) @2sps, (32,136,816) taps
DFE	32 feedforward taps @2sps + 16 feedback taps @1sps
Sequential Detector	= FFE @2sps + NWF @1sps + MLSE @1sps 32 taps + 3 taps + 64 states

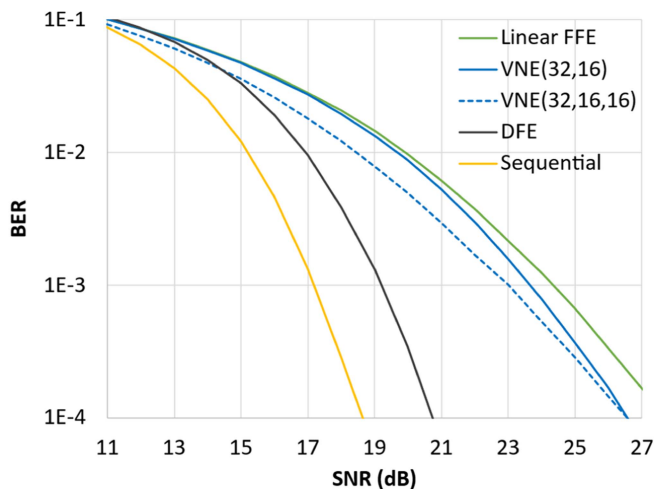


Fig. 20. Performance comparison among various linear/nonlinear equalization schemes under BWL. The BWL is emulated by transmitting a 1330-nm 100-GBd PAM signal over a span of SSMF with 33.1-ps/nm CD, leading to the 1st frequency-selective fading notch at 50 GHz. FF: feedforward; FB: feedback.

a nonlinear system well if it is continuously differentiable. Non-polynomial kernels may help alleviate this limitation.

Another challenge of VNE is how to measure the coefficients. Because the input of VNE contains multi-order of polynomials, the autocorrelation matrix of its input vector is not diagonalized with a huge eigenvalue spread. For common adaptive equalizers in IM-DD systems based on least mean square (LMS) algorithm, their convergence may be degraded and even stuck at local minima [127]. To put it simple, among all the kernels, both their significance and convergence speed can be drastically different, making it difficult for the adaptive filter to reach its optimum. A well-known approach to enhance the LMS performance is to orthogonalize the input of VNE by Wiener solution [112], which has been applied to recent IM-DD experiments [113], [114]. The reduction of VNE kernels is also found to improve the convergence of LMS [113], [114], [115], [116]. The recursive least square (RLS) algorithm [127] or the Kalman filter [128] can greatly improve the convergence, but they usually come with an unacceptable complexity for a simple IM-DD system.

E. A Brief Discussion on ANN

Compared to VNE, ANN is a much more powerful universal function approximator. First, while the Volterra series only has

polynomial kernels, ANN offers much richer types of nonlinear activation function (e.g., sigmoid, rectified linear unit) together with a variety of network structures (e.g., feedforward NN, convolutional NN, recurrent NN) to emulate the nonlinearity in a versatile manner [121]. Second, besides the classic stochastic gradient descent method, ANN is equipped with a mature group of optimization algorithms to find its coefficients [121]. Same as VNE, ANN can be applied for both post equalization and DPD [129]. As ANN is not originally intended to adapt a time-variant system, the coefficients are commonly trained offline as a static nonlinear model, despite some preliminary practice on real-time training of time-variant systems [130]. ANN faces a few practical limitations like VNE. It may be inefficient to approximate a non-continuous-differentiable nonlinear system, and its better convergence performance may be backed with a huge complexity. Readers may refer to [122], [123], [124] for more details on the application of ANN to IM-DD systems.

F. Application Consideration for Nonlinear Equalizers

Like the common practice for other DSP techniques, we must understand the system impairments before we choose a proper DSP to combat it. Regarding nonlinear equalization, it is crucial to recognize where nonlinearity comes from. If the nonlinearity has a well-defined model, like the ISI model in (11) for BWL, it may be more efficient to use an application specific equalizer than a universal equalizer to mitigate it. If the system contains multiple nonlinear elements, they can be separated into the ones with known models and the ones without, to be mitigated by the application-specific and the universal equalizers, respectively. This has been verified in recent IM-DD experiments impaired by both BWL and components nonlinearities. The combination of ISI mitigation algorithms (like THP and MLSE) and VNE have been found to outperform the VNE-only scheme [115], [116].

VII. CONCLUSION

IM-DD is facing big challenges for the next capacity doubling at post 200G era. The symbol rate of 100 GBd for the 200G per lane optics is expected to be doubled, putting high pressure on both electronics and E/O components. Powerful equalizations, especially the nonlinear ones to combat the BWL like MLSE, will be in demand to enable FTN signaling. The goal of FTN is to make full use of the spectrum resource beyond the traditional 3-dB bandwidth while keeping the mature PAM technique. The other potential option to achieve such a goal is to replace PAM with multicarrier modulation formats like SCM or DMT, which accommodate the BWL by adaptive BL/EL across the usable spectrum. Since SCM has been implemented in commercial coherent DSP ASIC, it is feasible to be leveraged by future IM-DD to avoid the sophisticated nonlinear equalizer in the FTN decoding. A secondary-level method to improve the speed of an IM-DD system is the nonlinearity compensation, to mitigate the nonlinearities from components (e.g., RF amplifiers and E/O modulators), fiber channel (e.g., dispersion) or the interaction between the two. The nonlinear equalizers for such purposes (like VNE) are commonly significantly more sophisticated than

the linear FFE widely implemented in commercial IM-DD products. Therefore, the trade-off between AIR improvement and complexity must be considered to design a practical DSP chain with manageable power consumption.

The analysis on the PCS benefit for IM-DD systems is more complicated than for amplified coherent systems due to the PPC nature of the IM transmitter. In general, the AIR improvement of PCS may be marginal over uniform signaling, but the simple rate adaptation via a fixed-rate FEC code is appealing for future IM-DD applications with higher demand on rate flexibility. The optimum PCS distribution for a PPC IM-DD system is related to a variety of system parameters like PAPRE, FEC threshold, and transmitter nonlinearities, and there is no unique conclusion applicable to all IM-DD systems.

Post-200G IM-DD systems rely on closer co-optimization between software (modulation format and DSP in the context of this article) and hardware. For a real-world system, we should recognize its constraints and find the most efficient tool for each constraint. For instance, although both VNE and MLSE can combat the channel BWL, MLSE is more efficient and closer to optimum. More crucially, when using a DSP technique to optimize a specific system metric, we must evaluate if it brings an overall cost-per-bit reduction. A common effort of recent IM-DD works is to improve the symbol rate via FTN techniques, but the symbol rate improvement does not always translate to an AIR gain, which is closely related to the underlying system condition like the channel response. In short, cost-efficient DSP designs are indispensable for IM-DD to meet the next capacity target and retain its competitiveness to the coherent counterpart.

REFERENCES

- [1] F. Buchali and H. Bulow, "Adaptive PMD compensation by electrical and optical techniques," *J. Lightw. Technol.*, vol. 22, no. 4, pp. 1116–1126, Apr. 2004.
- [2] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 749–763, Feb. 2005.
- [3] S. Tsukamoto, D.-S. Ly-Gagnon, K. Katoh, and K. Kikuchi, "Coherent demodulation of 40-gbit/s polarization-multiplexed QPSK signals with 16-GHz spacing after 200-km transmission," in *Proc. Opt. Fiber Commun. Conf.*, 2005, Paper PDP29.
- [4] W. Shieh and C. Athaudage, "Coherent optical orthogonal frequency division multiplexing," *Electron. Lett.*, vol. 42, no. 10, pp. 587–589, 2006.
- [5] R. Nagarajan, I. Lyubomirsky, and O. Agazzi, "Low power DSP-based transceivers for data center optical fiber communications," *J. Lightw. Technol.*, vol. 39, no. 16, pp. 5221–5231, Aug. 2021.
- [6] "800G FR4 2km specification v1.0," 800G Pluggable MSA, 2021. [Online]. Available: <https://www.800gmsa.com/documents>
- [7] D. Che, "Is it meaningful to pursue higher symbol rate beyond bandwidth constraint for short-reach interconnects?," in *Proc. Eur. Conf. Opt. Commun.*, 2022, Paper Tu1A.3.
- [8] S. Randel, F. Breyer, S. C. J. Lee, and J. W. Walewski, "Advanced modulation schemes for short-range optical communications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1280–1289, Sep./Oct. 2010.
- [9] K. Zhong et al., "Experimental study of PAM-4, CAP-16, and DMT for 100 gb/s short reach optical transmission systems," *Opt. Exp.*, vol. 23, no. 2, pp. 1176–1189, 2015.
- [10] X. Xu et al., "Advanced modulation formats for 400-gbps short-reach optical inter-connection," *Opt. Exp.*, vol. 23, no. 1, pp. 492–500, 2015.
- [11] J. Shi, J. Zhang, Y. Zhou, Y. Wang, N. Chi, and J. Yu, "Transmission performance comparison for 100-gb/s PAM-4, CAP-16, and DFT-S OFDM with direct detection," *J. Lightw. Technol.*, vol. 35, no. 23, pp. 5127–5133, Dec. 2017.
- [12] N. Eiselt et al., "Performance comparison of 112-gb/s DMT, Nyquist PAM4, and partial-response PAM4 for future 5G ethernet-based fronthaul architecture," *J. Lightw. Technol.*, vol. 36, no. 10, pp. 1807–1814, May 2018.
- [13] X. Pang et al., "200 Gbps/lane IM/DD technologies for short reach optical interconnects," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 492–503, Jan. 2020.
- [14] M. Chagnon, "Optical communications for short reach," *J. Lightw. Technol.*, vol. 37, no. 8, pp. 1779–1797, Apr. 2019.
- [15] T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, and S. Pachnicke, "DSP for high-speed short-reach IM/DD systems using PAM," *J. Lightw. Technol.*, vol. 38, no. 24, pp. 6771–6778, Dec. 2020.
- [16] D. Che et al., "Direct modulation of a 54-GHz distributed Bragg reflector laser with 100-GBaud PAM-4 and 80-GBaud PAM-8," in *Proc. Opt. Fiber Commun. Conf.*, 2020, Paper Th3C.1.
- [17] D. Che et al., "Long-term reliable >200-gb/s directly modulated lasers with 800GbE-compliant DSP," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2021, Paper F3A.3.
- [18] F. Zhang et al., "Up to single lane 200G optical interconnects with silicon photonic modulator," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2019, Paper Th4A.6.
- [19] X. Chen et al., "Single-wavelength and single-photodiode 700 Gb/s entropy-loaded PS-256-QAM and 200-GBaud PS-PAM-16 transmission over 10-km SMF," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2020, Paper Th3A-2.
- [20] H. Yamazaki et al., "160-GBd (320-Gb/s) PAM4 transmission using 97-GHz bandwidth analog multiplexer," *IEEE Photon. Technol. Lett.*, vol. 30, no. 20, pp. 1749–1751, Oct. 2018.
- [21] H. Yamazaki et al., "Transmission of 400-Gbps discrete multi-tone signal using >100-GHz-bandwidth analog multiplexer and InP Mach-Zehnder modulator," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2018, Paper Th3B.2.
- [22] H. Yamazaki et al., "Net-400-Gbps PS-PAM transmission using integrated AMUX-MZM," *Opt. Exp.*, vol. 27, no. 18, pp. 25544–25550, 2019.
- [23] D. Che, Y. Matsui, X. Chen, R. Schatz, and P. Iannone, "400-Gb/s direct modulation using a DFB+R laser," *Opt. Lett.*, vol. 45, no. 12, pp. 3337–3339, 2020.
- [24] W. Heni et al., "Ultra-high-speed 2:1 digital selector and plasmonic modulator IM/DD transmitter operating at 222 GBaud for intra-datacenter applications," *J. Lightw. Technol.*, vol. 38, no. 9, pp. 2734–2739, May 2020.
- [25] Y. Matsui et al., "55-GHz bandwidth short-cavity distributed reflector laser and its application to 112-Gb/s PAM-4," in *Proc. Opt. Fiber Commun. Conf.*, 2016, Paper Th5B.4.
- [26] S. Yamaoka et al., "Directly modulated membrane lasers with 108 GHz bandwidth on a high-thermal-conductivity silicon carbide substrate," *Nature Photon.*, vol. 15, no. 1, pp. 28–35, 2021.
- [27] Y. Matsui, R. Schatz, D. Che, F. Khan, M. Kwakernaak, and T. Sudo, "Low-chirp isolator-free 65-GHz-bandwidth directly modulated lasers," *Nature Photon.*, vol. 15, no. 1, pp. 59–63, 2021.
- [28] M. Chacinski, U. Westergren, B. Stoltz, L. Thylen, R. Schatz, and S. Hammerfeldt, "Monolithically integrated 100 GHz DFB-TWEAM," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3410–3415, Aug. 2009.
- [29] O. Ozolins et al., "100 GHz externally modulated laser for optical interconnects," *J. Lightw. Technol.*, vol. 35, no. 6, pp. 1174–1179, Mar. 2017.
- [30] P. Dong, X. Liu, S. Chandrasekhar, L. L. Buhl, R. Aroca, and Y.-K. Chen, "Monolithic silicon photonic integrated circuits for compact 100⁺ Gb/s coherent optical receivers and transmitters," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 4, pp. 150–157, Jul./Aug. 2014.
- [31] S. S. Azadeh et al., "Low $v\pi$ silicon photonics modulators with highly linear epitaxially grown phase shifters," *Opt. Exp.*, vol. 23, no. 18, pp. 23526–23550, 2015.
- [32] Y. Ogiso et al., "Over 67 GHz bandwidth and 1.5 V $v\pi$ InP-based optical IQ modulator with n-i-p-n heterostructure," *J. Lightw. Technol.*, vol. 35, no. 8, pp. 1450–1455, Apr. 2017.
- [33] J. Ozaki et al., "Over-85-GHz-bandwidth InP-based coherent driver modulator capable of 1-Tb/s/λ-class operation," *J. Lightw. Technol.*, vol. 41, no. 11, pp. 3290–3296, Jun. 2023, doi: [10.1109/JLT.2023.3236962](https://doi.org/10.1109/JLT.2023.3236962).
- [34] C. Wang et al., "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature*, vol. 562, pp. 101–104, 2018.
- [35] M. Xu et al., "Dual-polarization thin-film lithium niobate in-phase quadrature modulators for terabit-per-second transmission," *Optica*, vol. 9, no. 1 pp. 61–62, 2022.
- [36] G. D. Forney and M. V. Eyuboglu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, vol. 29, no. 12, pp. 25–34, Dec. 1991.

- [37] D. Che and X. Chen, "Higher-order modulation vs faster-than-Nyquist PAM-4 for datacenter IM-DD optics: An AIR comparison under practical bandwidth limits," *J. Lightw. Technol.*, vol. 40, no. 10, pp. 3347–3357, May 2022.
- [38] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
- [39] D. Che, J. Cho, and X. Chen, "Does probabilistic constellation shaping benefit IM-DD systems without optical amplifiers?," *J. Lightw. Technol.*, vol. 39, no. 15, pp. 4997–5007, Aug. 2021.
- [40] K. Inoue, "Phase-mismatching characteristic of four-wave mixing in fiber lines with multistage optical amplifiers," *Opt. Lett.*, vol. 17, no. 11, pp. 801–803, 1992.
- [41] R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, and R. M. Derosier, "Four-photon mixing and high-speed WDM systems," *J. Lightw. Technol.*, vol. 13, no. 5, pp. 841–849, May 1995.
- [42] X. Zhou and C. Lam, "Four-wave mixing penalty for WDM-based ethernet PMDs in O-band," IEEE P802.3df 400 Gb/s and 800 Gb/s Ethernet Task Force, May Session, 2022.
- [43] R. Rodes, V. Bhatt, and C. Cole, "On technical feasibility of 800G-LR4 with direct-detection," IEEE P802.3df 400 Gb/s and 800 Gb/s Ethernet Task Force, March Session, 2022.
- [44] J. Johnson, "FWM analysis of PAM-4 LR/ER PMDs," IEEE P802.3df 400 Gb/s and 800 Gb/s Ethernet Task Force, April Session, 2022.
- [45] X. Liu and Q. Fan, "Inter-channel FWM mitigation techniques for Oband WDM based 800G/1.6T LR and 5G fronthaul applications," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2023, Paper W4F.1.
- [46] R. Bonk et al., "50G-PON: The first ITU-T higher-speed PON system," *IEEE Commun. Mag.*, vol. 60, no. 3, pp. 48–54, Mar. 2022.
- [47] D. Che, Q. Hu, and W. Shieh, "Linearization of direct detection optical channels using self-coherent subsystems," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 516–524, Jan. 2016.
- [48] R. S. Tucker, "High-speed modulation of semiconductor lasers," *IEEE Trans. Electron. Devices*, vol. 32, no. 12, pp. 2572–2584, Dec. 1985.
- [49] L. A. Neto et al., "Simple estimation of fiber dispersion and laser chirp parameters using the downhill simplex fitting algorithm," *J. Lightw. Technol.*, vol. 31, no. 2, pp. 334–342, Jan. 2013.
- [50] C. Xie et al., "Single-VCSEL 100-Gb/s short-reach system using discrete multi-tone modulation and direct detection," in *Proc. Opt. Fiber Commun. Conf.*, 2015, Paper Tu2H.2.
- [51] H. Wang et al., "Adaptive channel-matched detection for C-band 64-gbit/s optical OOK system over 100-km dispersion-uncompensated link," *J. Lightw. Technol.*, vol. 38, no. 18, pp. 5048–5055, Sep. 2020.
- [52] S. Liu et al., "Fixed-state log-MAP detection for intensity-modulation and direct-detection systems over dispersion-uncompensated links," *IEEE Photon. J.*, vol. 13, no. 3, Jun. 2021, Art. no. 7900110.
- [53] X. Wu, J. Zhang, A. P. T. Lau, and C. Lu, "C-band 120-Gb/s PAM-4 transmissions over a 100-km dispersion-uncompensated SSMF using joint combined pulse shaping and low-complexity nonlinear equalization," *Opt. Lett.*, vol. 47, no. 19, pp. 5144–5147, 2022.
- [54] P. W. Berenguer et al., "Nonlinear digital pre-distortion of transmitter components," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1739–1745, Apr. 2016.
- [55] K. Zhang et al., "Intensity directed equalizer for the mitigation of DML chirp induced distortion in dispersion-unmanaged C-band PAM transmission," *Opt. Exp.*, vol. 25, no. 23, pp. 28123–28135, 2017.
- [56] J. Wang and K. Petermann, "Small signal analysis for dispersive optical fiber communication systems," *J. Lightw. Technol.*, vol. 10, no. 1, pp. 96–100, Jan. 1992.
- [57] J. L. Gimlett and N. K. Cheung, "Effects of phase-to-intensity noise conversion by multiple reflections on gigabit-per-second DFB laser transmission systems," *J. Lightw. Technol.*, vol. 7, no. 6, pp. 888–895, Jun. 1989.
- [58] C. R. S. Fludger, M. Mazzini, T. Kupfer, and M. Traverso, "Experimental measurements of the impact of multi-path interference on PAM signals," in *Proc. Opt. Fiber Commun. Conf.*, 2014, Paper W1F.6.
- [59] Y. J. Wen, Y. Cui, and Y. Bai, "Mitigation of optical multipath interference impact for directly detected PAMn system," *Opt. Exp.*, vol. 28, no. 25, pp. 38317–38333, 2020.
- [60] C. Huang et al., "Optical multipath interference mitigation for high-speed PAM4 IMDD transmission system," *J. Lightw. Technol.*, vol. 40, no. 16, pp. 5490–5501, Aug. 2022.
- [61] F. Buchali, F. Steiner, G. Bocherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.
- [62] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.
- [63] J. Zhang et al., "Demonstration of 260-Gb/s single-lane EML-based PS-PAM-8 IM/DD for datacenter interconnects," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper W4I.4.
- [64] M. Sabbir-Bin Hossain et al., "Probabilistic shaping for high-speed unamplified IM/DD systems with an O-band EML," *J. Lightw. Technol.*, vol. 41, no. 16, pp. 5373–5382, Aug. 2023.
- [65] F. Buchali, W. Idler, R. Dischler, T. Eriksson, and L. Schmalen, "Spectrally efficient probabilistically shaped square 64QAM to 256 QAM," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2017, Paper Tu.2.D.1.
- [66] F. Buchali et al., "DCI field trial demonstrating 1.3-Tb/s single-channel and 50.8-Tb/s WDM transmission capacity," *J. Lightw. Technol.*, vol. 38, no. 9, pp. 2710–2718, May 2020.
- [67] M. Nakamura et al., "Over 2-Tb/s net bitrate single-carrier transmission based on >130-GHz-bandwidth InP-DHBT baseband amplifier module," in *Proc. Eur. Conf. Opt. Commun.*, 2022, Paper Th3C.1.
- [68] G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv:1707.01134*.
- [69] R. Borkowski et al., "FLCS-PON—An opportunistic 100Gbit/s flexible PON prototype with probabilistic shaping and soft-input FEC: Operator trial and ODN case studies," *J. Opt. Commun. Netw.*, vol. 14, no. 6, pp. C82–C91, 2022.
- [70] P. Schulte and G. Bocherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [71] J. Cho, S. Chandrasekhar, R. Dar, and P. J. Winzer, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. Eur. Conf. Opt. Commun.*, 2016, Paper W1C.2.
- [72] A. Amari et al., "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5926–5936, Dec. 2019.
- [73] T. Wiegart, F. Da Ros, M. P. Yankov, F. Steiner, S. Gaiarin, and R. D. Wesel, "Probabilistically shaped 4-PAM for short-reach IM/DD links with a peak power constraint," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 400–405, Jan. 2021.
- [74] B. M. de Oliveira, J. Silva, M. Neves, F. Guiomar, M. do Carmo Medeiros, and P. Monteiro, "Revisiting probabilistic constellation shaping in unamplified coherent optical links," in *Proc. Opt. Fiber Commun. Conf.*, 2023, Paper Th3E.1.
- [75] S. Yao, A. Mahadevan, Y. Lefevre, N. Kaneda, V. Houtsmas, and D. van Veen, "ANN-based optimization of probabilistic and geometric shaping for flexible rate 50G and beyond PON," in *Proc. Opt. Fiber Commun. Conf.*, 2022, Paper W3G.2.
- [76] D. Kim and H. Kim, "Capacity-achieving symbol distributions for directly modulated laser and direct detection systems," *Opt. Exp.*, vol. 31, no. 8, pp. 12609–12623, 2023.
- [77] D. Zou, W. Wang, S. Qi, F. Li, and Z. Li, "Theoretical evaluation of the capacity-achieving distribution for IM-DD fiber-optic channels," *arXiv:2302.11851*.
- [78] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.
- [79] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.
- [80] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. Amer. Inst. Elect. Engineers*, vol. 47, no. 2, pp. 617–644, Apr. 1928.
- [81] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1451–1462, Oct. 1975.
- [82] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, Aug. 2013.
- [83] R. Rath, D. Clausen, S. Ohlendorf, S. Pachnicke, and W. Rosenkranz, "Tomlinson-Harashima precoding for dispersion uncompensated PAM-4 transmission with direct-detection," *J. Lightw. Technol.*, vol. 35, no. 18, pp. 3909–3917, Sep. 2017.
- [84] Q. Hu, M. Chagnon, K. Schuh, F. Buchali, and H. Bülow, "IM/DD beyond bandwidth limitation for data center optical interconnects," *J. Lightw. Technol.*, vol. 37, no. 19, pp. 4940–4946, Oct. 2019.
- [85] V. Houtsmas, D. van Veen, A. Gnauck, and P. Iannone, "APD-based duobinary direct detection receivers for 40 Gbps TDM-PON," in *Proc. Opt. Fiber Commun. Conf.*, 2015, Paper Th4H.1.
- [86] N. Al-Dhahir, "FIR channel-shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

- [87] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.
- [88] A. Prlja and J. B. Anderson, "Reduced-complexity receivers for strongly narrowband intersymbol interference introduced by faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2591–2601, Sep. 2012.
- [89] S. V. D. Heide, N. Eiselt, H. Griesser, J. J. V. Olmos, I. T. Monroy, and C. Okonkwo, "Experimental investigation of impulse response shortening for low-complexity MLSE of a 112-gbit/s PAM-4 transceiver," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2016, Paper M.2.B.5.
- [90] T. Xu et al., "Decoding of 10-G optics-based 50-gb/s PAM-4 signal using simplified MLSE," *IEEE Photon. J.*, vol. 10, no. 4, Aug. 2018, Art. no. 7905508.
- [91] W. Wang et al., "Demonstration of 214 Gbps per lane IM/DD PAM-4 transmission using O-band 35GHz-class EML with advanced MLSE and KP4-FEC," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2020, Paper M4F.4.
- [92] Y. Yu, Y. Che, T. Bo, D. Kim, and H. Kim, "Reduced-state MLSE for an IM/DD system using PAM modulation," *Opt. Exp.*, vol. 28, no. 26, pp. 38505–38515, 2020.
- [93] C. Minkenberg, R. Krishnaswamy, A. Zilkie, and D. Nelson, "Co-packaged datacenter optics: Opportunities and challenges," *IET Optoelectron.*, vol. 15, no. 2, pp. 77–91, 2021.
- [94] Q. Zhang, N. Stojanovic, L. Zhang, T. Zuo, C. Xie, and E. Zhou, "Up to 190-Gb/s OOK signal generation using a coding and cutting technique with a 92 GSa/s DAC," in *Proc. Opt. Fiber Commun. Conf.*, 2017, Paper Th3D.1.
- [95] Y. Zhu, X. Fang, L. Zhang, F. Zhang, and W. Hu, "Sub-sampling generation of ultra-high baud rate PAM/QAM signals via high-order partial response narrowing," *Opt. Exp.*, vol. 29, no. 26, pp. 44063–44079, 2021.
- [96] D. Che, "Comparison between PAM and DMT in a 200-Gb/s IM-DD system considering the interaction between bandwidth limit and peak power constraint at transmitter," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2023, Paper W4E.4.
- [97] Y. Tang, W. Shieh, and B. S. Krongold, "DFT-spread OFDM for fiber nonlinearity mitigation," *IEEE Photon. Technol. Lett.*, vol. 22, no. 16, pp. 1250–1252, Aug. 2010.
- [98] M. Qiu et al., "Digital subcarrier multiplexing for fiber nonlinearity mitigation in coherent optical communication systems," *Opt. Exp.*, vol. 22, no. 15, pp. 18770–18777, 2014.
- [99] A. J. Lowery and J. Armstrong, "Orthogonal-frequency-division multiplexing for dispersion compensation of long-haul optical systems," *Opt. Exp.*, vol. 14, no. 6, pp. 2079–2084, 2006.
- [100] W. Shieh, H. Bao, and Y. Tang, "Coherent optical OFDM: Theory and design," *Opt. Exp.*, vol. 16, no. 2, pp. 841–859, 2008.
- [101] H. Sun et al., "800G DSP ASIC design using probabilistic shaping and digital sub-carrier multiplexing," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4744–4756, Sep. 2020.
- [102] M. I. Olmedo et al., "Multiband carrierless amplitude phase modulation for high capacity optical data links," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 798–804, Feb. 2014.
- [103] Y. Wang, J. Yu, and N. Chi, "Demonstration of 4×128 -Gb/s DFT-S OFDM signal transmission over 320-km SMF with IM/DD," *IEEE Photon. J.*, vol. 8, no. 2, Apr. 2016, Art. no. 7903209.
- [104] S. C. J. Lee, F. Breyer, S. Randel, H. P. A. van den Boom, and A. M. J. Koonen, "High-speed transmission over multimode fiber using discrete multitone modulation," *J. Opt. Netw.*, vol. 7, no. 2, pp. 183–196, 2008.
- [105] W. Yan et al., "100 Gb/s optical IM-DD transmission with 10G-class devices enabled by 65 GSamples/s CMOS DAC core," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2013, Paper OM3H.1.
- [106] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 773–775, Feb.–Apr. 1995.
- [107] J. Campello, "Practical bit loading for DMT," in *Proc. IEEE Int. Conf. Commun.*, 1999, pp. 801–805.
- [108] D. Che and W. Shieh, "Approaching the capacity of colored-SNR optical channels by multicarrier entropy loading," *J. Lightw. Technol.*, vol. 36, no. 1, pp. 68–78, Jan. 2018.
- [109] D. Che and W. Shieh, "Squeezing out the last few bits from band-limited channels with entropy loading," *Opt. Exp.*, vol. 27, no. 7, pp. 9321–9329, 2019.
- [110] C. Xie et al., "Achievable information rate enhancement of visible light communication using probabilistically shaped OFDM modulation," *Opt. Exp.*, vol. 26, no. 1, pp. 367–375, 2018.
- [111] X. Chen, S. Chandrasekhar, J. Cho, and P. Winzer, "Single-wavelength and single-photon entropy-loaded 554-Gb/s transmission over 22-km SMF," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper Th4B.5.
- [112] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, vol. 69, no. 12, pp. 1557–1573, Dec. 1981.
- [113] N. Stojanovic, F. Karinou, Z. Qiang, and C. Prodanic, "Volterra and Wiener equalizers for short-reach 100G PAM-4 applications," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4583–4594, Nov. 2017.
- [114] W.-J. Huang et al., "93% complexity reduction of Volterra nonlinear equalizer by L1-regularization for 112-Gbps PAM-4 850-nm VCSEL optical interconnect," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2018, Paper M2D.7.
- [115] N.-P. Diamantopoulos, H. Nishi, W. Kobayashi, K. Takeda, T. Kakitsuka, and S. Matsuo, "On the complexity reduction of the second-order Volterra nonlinear equalizer for IM/DD systems," *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1214–1224, Feb. 2019.
- [116] M. Xiang, Z. Xing, E. El-Fiky, M. Morsy-Osman, Q. Zhuge, and D. V. Plant, "Single-lane 145 Gbit/s IM/DD transmission with faster-than-Nyquist PAM4 signaling," *IEEE Photon. Technol. Lett.*, vol. 30, no. 13, pp. 1238–1241, Jul. 2018.
- [117] G. sharan Yadav, C.-Y. Chuang, K.-M. Feng, J.-H. Yan, J. Chen, and Y.-K. Chen, "Reducing computation complexity by using elastic net regularization based pruned Volterra equalization in a 80 Gbps PAM-4 signal for inter-data center interconnects," *Opt. Exp.*, vol. 28, no. 26, pp. 38539–38552, 2020.
- [118] Q. Zhang, N. Stojanovic, J. Wei, and C. Xie, "Single-lane 180 Gb/s DB-PAM-4-signal transmission over an 80km DCF-free SSMF link," *Opt. Lett.*, vol. 42, no. 4, pp. 883–886, 2017.
- [119] J. Jignesh et al., "Transmitter-side Volterra filtering for increased dispersion tolerance in 56 Gbaud PAM-4 systems," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper M2C.6.
- [120] A. Masuda, S. Yamamoto, H. Taniguchi, M. Nakamura, and Y. Kisaka, "255-Gbps PAM-8 transmission under 20-GHz bandwidth limitation using NL-MLSE based on Volterra filter," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper W4L.6.
- [121] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 493–516, Jan. 2019.
- [122] B. Karanov et al., "End-to-end deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, Oct. 2018.
- [123] L. Yi, T. Liao, L. Huang, L. Xue, P. Li, and W. Hu, "Machine learning for 100 Gb/s/λ passive optical network," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1621–1630, Mar. 2019.
- [124] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Feedforward and recurrent neural network-based transfer learning for nonlinear equalization in short-reach optical links," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 475–480, Jan. 2021.
- [125] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, Jan. 1997.
- [126] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, 2001.
- [127] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2937–2944, Nov. 1994.
- [128] G. Welch and G. Bishop, "An introduction to the Kalman filter," Dept. Comput. Sci., Univ. North Carolina, Chapel Hill, Tech. Report TR95041, 1995.
- [129] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, "Deep neural network-based digital pre-distortion for high baudrate optical coherent transmission," *J. Lightw. Technol.*, vol. 40, no. 3, pp. 597–606, Feb. 2022.
- [130] K. Liu, C. Häger, E. Börjeson, and P. Larsson-Edefors, "FPGA implementation of multi-layer machine learning equalizer with on-chip training," in *Proc. Opt. Fiber Commun. Conf.*, 2023, Paper M1F.4.