Higher-Order Modulation vs Faster-Than-Nyquist PAM-4 for Datacenter IM-DD Optics: An AIR Comparison Under Practical Bandwidth Limits

Di Che and Xi Chen

(Post-Deadline Paper)

***Abstract*—The 4-ary pulse amplitude modulation (PAM-4) has been widely deployed to both electrical and optical interfaces in datacenter applications. Based on 100-GBd PAM-4, the 200G per lane optical interface is being actively developed in both industrial interoperability groups and standard bodies. Going beyond 200G, it is questionable if PAM-4 could support an even faster interface with higher baud-rate, considering the bandwidth growth for both electronics and optical components has started lagging the ever-increasing capacity demand. Under limited bandwidth, one option is to improve the spectral efficiency by higher order formats like PAM-6 and PAM-8, but this induces backward incompatibility as well as potential modulation mismatch between electrical and optical interfaces that requires additional energy for gearboxing. This revives the study of faster than Nyquist (FTN) signaling to pursue higher PAM-4 baud-rate breaking the bandwidth limit. In ECOC 2021, we demonstrated a reduced-complexity maximum a posteriori probability (MAP) decoder in intensity-modulation (IM) direct-detection (DD) systems with superior FTN performance. In this paper, we extend our work by adding more algorithm details and experimental results. Moreover, we perform the achievable information rate (AIR) comparison between the FTN PAM-4 and the higher-order modulation within the Nyquist limit in different channel conditions. Remarkably, in a practical IM-DD channel with gradually decayed frequency response, FTN shows AIR gain over the Nyquist approach even at a low FTN decoding complexity. The finding suggests FTN PAM could be a promising candidate at post 200G era to offer both AIR and implementation advantages.**

***Index Terms*—4-ary pulse amplitude modulation (PAM-4), BCJR algorithm, faster than Nyquist, intensity-modulation direct-detection (IM-DD), intersymbol interference, maximum a posteriori estimation, maximum likelihood sequence estimation, optical fiber communication, optical interconnections.**

I. Introduction

DATACENTER architecture is being paced by the capacity scaling of switching ASIC (application specific integrated circuit) which is doubling approximately every two years. Over the past decade, the capacities of Ethernet switching and optical modules have both increased forty-fold, from 0.64 to 25.6 Tb/s, and from 10 to 400 Gb/s, respectively [1]. The next switch silicon supports 51.2-Tb/s capacity which calls for an optical interface at 800G/1.6T per module or 200 Gb/s per lane. This has stimulated activities for both 800G industry interoperability [2] and MAC-layer standardization being studied by IEEE 802.3 Beyond 400 Gb/s Ethernet Study Group. The 800G pluggable MSA (multi-source agreement) group published its first 200G per lane technical specification recently based on 112.5-GBd 4-ary pulse amplitude modulation (PAM-4) [2]. IEEE Ethernet also suggested a 200G scheme by 106.25-GBd PAM-4 for a smooth upgrade from its 100G per lane standard using 53.125-GBd PAM-4 [3]. The format of PAM-4 aligns well with the electrical interface for the switch SerDes typically defined by the OIF (Optical Internetworking Forum) Common Electrical Interface (CEI) implementation agreements, which have also adopted PAM-4 since the 12.8-Tb/s generation [4].

While PAM-4 has been the unified format for both electrical and optical interfaces at present, it is questionable if it could support an even faster optical interface beyond 200G per lane for the generation after, considering the bandwidth growth for both electronics and optical components has started lagging the ever-increasing capacity demand. A similar situation happens during the transition from 100 GbE (Gigabit Ethernet) to 400 GbE several years ago, when PAM-4 emerged to replace the traditional two-level non-return-to-zero (NRZ) format, trading off the lower signal-to-noise ratio (SNR) requirement for a bit rate doubling at the same baud rate (BR). Such tradeoff won't be efficient beyond PAM-4, because the next bit rate doubling must raise the modulation order to PAM-16 which relies on a much more complicated equalization to maintain the signal integrity. Thus, it requires a combination of higher BR, more advanced modulation format, and more powerful equalization to extend the roadmap for the next speed doubling. Higher-order PAMs have been actively studied in recent intensity-modulation (IM) direct-detection (DD) demonstrations, like PAM-6 [5] and PAM-8 [6], which improve the bit rate by at most 29% and 50%, respectively. For a system with constrained bandwidth, multicarrier modulations like multi-band carrier-less amplitude phase modulation (CAP) and discrete multitone (DMT) [7]–[9] have been considered to maximize the speed. On the other hand, using an advanced modulation other than PAM-4

Manuscript received December 17, 2021; revised March 7, 2022 and April 18, 2022; accepted April 21, 2022. Date of publication May 5, 2022; date of current version May 25, 2022. *(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.2022.3172247.

Digital Object Identifier 10.1109/JLT.2022.3172247

induces a mismatch between optical and electrical interfaces which requires more sophisticated gearboxing between the two for modulation conversions, and consequently, higher power consumption for digital signal processing (DSP). This revives the interest on faster than Nyquist (FTN) technique [10], [11], namely, keeping the format as PAM-4 while increasing its BR beyond the bandwidth limit. Some FTN approaches pre-process the baseband signal for spectral narrowing to fit a bandwidth-limited channel, like partial-response pre-filtering [12]–[15] and Tomlinson-Harashima precoding (THP) [16]–[19]. However, as pre-processing adds gearboxing between the switch SerDes and optical module, it would be more attractive to directly drive the optics by the SerDes output and demodulate the FTN signal by receiver side post-processing, like decision feedback equalizer (DFE), maximum likelihood sequence estimation (MLSE) [20]–[23] and maximum a posteriori probability (MAP) decoding [24]–[27]. In ECOC 2021, we demonstrated ultrahigh baud-rate FTN signaling [27] via a low-complexity MAP decoder in IM-DD systems. As an extended version of our ECOC work, this paper introduces the low-complexity MAP algorithm in more details in Section II, together with more experimental results to better explain the algorithm in Section III.

Many optical FTN literatures use FTN ratio to evaluate the FTN efficiency, defined as the ratio of the FTN BR with respect to the BR limit set by the Nyquist criterion,

$$R_{FTN} = \frac{BR_{FTN}}{BR_{Nyquist}} \tag{1}$$

While FTN ratio provides a fair metric among FTN schemes, its dependence on modulation order and system SNR prevents it from being a general metric for the comparison with other advanced modulation schemes. In this paper, we use achievable information rate (AIR) as the universal metric and evaluate the AIR of FTN PAM-4 in different channel conditions in Section IV. The metric of AIR enables a direct comparison between the FTN PAM-4 and the conventional Nyquist approach, namely, modulating the signal within the Nyquist bandwidth limit using higher-order formats beyond PAM-4. This tells if the enhanced PAM-4 BR via FTN brings real AIR improvements. The latest ultrahigh bandwidth optical transmission systems support the generation of 100-GHz class electronic signals [28]. In the same 100-GHz IM-DD setup, we compare our FTN results in ECOC 2021 with the Nyquist signaling published in ECOC 2020 [28] in Section V, to verify part of the theoretical finding in Section IV. The paper sheds light on the modulation format design for future IM-DD optics beyond 200G per lane.

II. Optimum FTN Decoding and a Simplified Algorithm

In this section, we introduce an optimum decoding structure for the FTN signal, and a simplified algorithm with marginal performance penalty to the capacity limit which will be used throughout the paper. This decoding structure does not rely on pre-processing, namely, the uncontrolled ISI may be added at the transmitter, or be induced arbitrarily by the bandwidth-limited channel, or a combination of both.

A discrete sequence in the presence of ISI is expressed as

$$s_t = \sum_{k=0}^{L-1} h_k x_{t-k} \tag{2}$$

where x_t is the ISI free sequence, h_i is the ISI coefficient and L is the channel memory length. Without known pre-processing parameters, the ISI coefficients are acquired at the receiver for FTN decoding. An optimum receiver structure for s_t cascades a whitened matched filter (WMF) and a MLSE or MAP decoder [29]. We will introduce the two building blocks separately in the sub-sections below.

A. Whitened Matched Filter (WMF)

As indicated by its name, a WMF performs matched filtering and noise whitening to prepare a suitable input sequence for the subsequent MLSE/MAP decoding. The WMF can be realized by a channel shortening filter (CSF) which equalizes the signal with a target impulse response (TIR) [30]–[32]. The CSF is designed to output samples with whitened noise simultaneously, and its TIR provides the ISI model for the MLSE/MAP decoder.

In this work, rather than designing a single CSF directly, we design a WMF by cascading two filters to make the DSP chain compatible with the common IM-DD implementation. The first one is an adaptive feed-forward equalizer (FFE) widely applied in commercial PAM-4 receivers. The full-response FFE aims to minimize ISI which inevitably leads to colored noise due to the partial response of the FTN signal. Consequently, the second filter is a noise whitening filter (NWF) to minimize the colored noise effect, which in turn, induces ISI to the signal. The NWF builds the ISI model for the MLSE/MAP decoder, and a longer NWF can achieve higher degree of whitening at the cost of longer ISI memory. To design an NWF, we first evaluate the noise as $N_t = r_t - x_t$, where r_t is the FFE output and x_t is either a training sequence or the decision of r_t, and then perform power spectrum estimation (PSE) [33] on N_t. Among various PSE solutions, we choose the auto-regressive model with Yule-Walker equations to extract the NWF coefficients.

B. Revisiting MAP Decoding Based on BCJR Algorithm

Both MLSE and MAP decoders can be applied to demodulate an ISI-contaminated sequence. In this work, we use the MAP decoder to evaluate the FTN receiver performance.

The MAP decoder is implemented by the BCJR algorithm [34] commonly characterized by a trellis graph. A toy example is illustrated in Fig. 1. Each trellis stage contains all possible states by tracing back to the past samples with a memory length of L as in (2), and each state is expressed as

$$s = (x_t, x_{t-1}, \ldots, x_{t-L+1}) \tag{3}$$

where $x \in \mathcal{X}$, namely, a symbol in the modulation alphabet $\mathcal{X}$. This leads to the total number of states per trellis stage

$$N_{states} = |\mathcal{X}|^L \tag{4}$$

where $|\mathcal{X}|$ is the alphabet size. In the example of Fig. 1, we use an alphabet $\mathcal{X} = [0\ 1]$ ($|\mathcal{X}| = 2$) and a memory length $L = 3$,

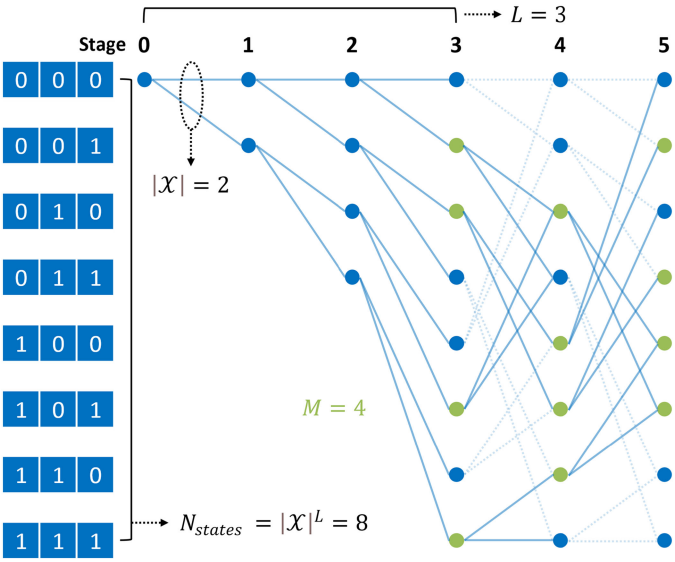


Fig. 1. Trellis graph for the forward $\boldsymbol{\alpha}$-recursion in an M-BCJR decoder. This example assumes an alphabet $\mathcal{X} = [0\ 1]$, an ISI memory length $\boldsymbol{L} = 3$, and a reduced number of states $\boldsymbol{M} = 4$.

leading to $N_{states} = 8$. These states are listed at the left-hand side of Fig. 1. Between two adjacent trellis stages, each state at the last stage generates $|\mathcal{X}|$ transitions to the current stage. By assuming additive white Gaussian noise (AWGN), the transition metric is calculated as

$$\Gamma\ (i,j) = \frac{P\,(x')}{\sqrt{\pi N_0}}\ exp\left[\frac{-(y-l_{i,j})^2}{N_0}\right] \tag{5}$$

where i is the i-th state at the last stage and j is the j-th state at the current stage, $P(x')$ is the a priori probability of the symbol x' that causes the transition, N_0 is the noise variance, y is the ISI channel observation, in other words, the WMF output; the transition (i,j) points to a state at the current stage whose state metric $l_{i,j}$ is calculated by (2). To obtain a metric on a symbol-by-symbol basis, the MAP decoder should calculate the state probability given the transition metric $\Gamma(i,j)$, based on a pair of forward and backward recursions

$$P\ (\ s_n = \ i,\ s_{n+1} = \ j, \boldsymbol{y}) = \alpha_n\ [i]\,\Gamma_n\,(i,j)\,\beta_{n+1}\,[j] \tag{6}$$

where $\boldsymbol{y} = y_1, \ldots, y_N$ are the WMF outputs, $\alpha_n[i]$ is a forward variable for the i-th state at stage n, and $\beta_{n+1}[j]$ is a backward variable for the j-th state at stage $n+1$. Beginning from a randomly chosen initial state at the root of the trellis, α is calculated recursively in a forward trellis pass,

$$\alpha_{n+1}\ [j] = \sum_{i\in\mathcal{S}} \alpha_n\,[i]\,\Gamma_n\,(i,j) \tag{7}$$

where $\mathcal{S}$ is the set of states at stage n that can reach the j-th state at stage $n+1$. Conversely, β is calculated in a backward pass,

$$\beta_n\ [i] = \sum_{j\in\mathcal{S}} \beta_{n+1}\,[j]\,\Gamma_n\,(i,j) \tag{8}$$

where $\mathcal{S}$ is the set of states at stage $n+1$ reached from the i-th state at stage n.

Known the joint probability $P(s_n, s_{n+1}, \boldsymbol{y})$ in (6), the probability of each symbol x' in the modulation alphabet (with the size of $|\mathcal{X}|$) can be calculated as

$$P\ (x'|\boldsymbol{y}) = \frac{P\,(x',\boldsymbol{y})}{P\,(\boldsymbol{y})} = \frac{\sum_{(i,j)\in\Sigma_n^+} P\,(s_n = i, s_{n+1} = j, \boldsymbol{y})}{P\,(\boldsymbol{y})} \tag{9}$$

where Σ_n^+ stands for the set of transitions (i,j) between stages n and $n+1$ caused by the input x'. The actual value of $P(\boldsymbol{y})$ in the denominator can be omitted by normalizing $P(x'|\boldsymbol{y})$ as

$$\sum_{x'\in\mathcal{X}} P\,(x'|\boldsymbol{y}) = \ 1 \tag{10}$$

The bitwise log-likelihood ratio (LLR) can be derived from symbol probabilities by symbol to bit demapping.

The BCJR algorithm above can be computed in the logarithm domain to improve its numerical stability, named the log-MAP algorithm and its suboptimal implementation Max-log-MAP.

C. Simplified MAP Decoding Based on M-BCJR Algorithm

Based on the equations above, the computational complexity of BCJR algorithm is dominated by the calculation of transition metrics. Between two adjacent stages, the algorithm should calculate totally $|\mathcal{X}|^{L+1}$ times of $\Gamma(i,j)$ as in (5). As a result, its complexity is in the order of $\mathcal{O}(L|\mathcal{X}|^{L+1})$. For PAM-4, even a short memory $L = 3$ leads to an unacceptable complexity. To reduce the complexity, it is crucial to reduce N_{states} so that the BCJR algorithm does not need to ergodically traverse the trellis graph. For large L values, Γ-matrices are usually sparse and most non-zero elements are very small. Therefore, it is possible to discard non-significant states with low probabilities at each trellis stage without substantial decoding penalty. This concept is known as M-algorithm [35]–[37], namely, N_{states} is limited to within M. Fig. 1 shows an example of limiting N_{states} at each stage to $M = 4$ for the α-recursion. The algorithm implemented in this work is described as follow [37].

1) Begin the α recursion from the root of trellis (*i.e.*, the initial state) till the stage n whose $N_{states} > M$
2) Store the M largest α values and their corresponding states (*i.e.*, survivors) for the β recursion and discard the others
3) Compute α_{n+1} from the M α_n values, which restricts the Γ calculations to $M|\mathcal{X}|$ times between adjacent stages
4) Repeat steps 2 and 3 for the α recursion till the end of trellis
5) Compute the β recursion from the end of trellis in an opposite direction following similar steps as in 1 to 4
6) At each trellis stage, while the α recursion stores the M largest α values as indicated in step 2, the β recursion stores the β values whose states overlap with those stored in the α recursion to guarantee the survivors overlap between the α and β recursions

After the above α and β recursions, the symbol probabilities are calculated by (6) and (9). Based on step 3, the M-algorithm reduces the BCJR complexity to the order of $\mathcal{O}(LM|\mathcal{X}|)$, which grows linearly rather than exponentially with L. This makes it practical to use a large L to achieve superior ISI-mitigation capability with manageable complexity.

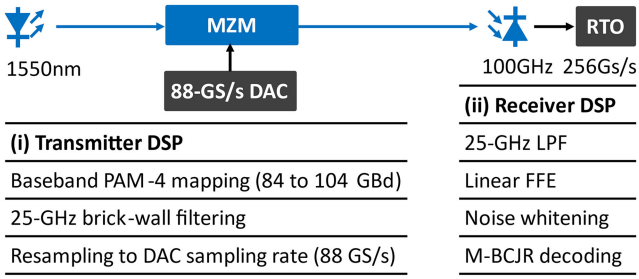


Fig. 2. IM-DD experiment setup with 25-GHz bandwidth limit. MZM: $LiNbO_3$ Mach-Zehnder modulator; DAC: digital-to-analog converter; RTO: real-time oscilloscope; LPF: lowpass filter; FFE: feed-forward equalizer.

D. Discussions on the Forward Error Correction (FEC)

Essentially, a MAP decoder is a soft-input soft-output (SISO) equalizer, whose input and output are usually log likelihood ratios (LLR). It is compatible with hard decision (HD) FEC by taking its output LLR signs as HD bits. If the probabilities of the transmitted symbols are equal as is the case for the uniform PAM, the HD performance of MAP decoding is close to MLSE. For soft decision (SD) FEC, the LLRs are iteratively passing between the SISO MAP equalizer and the SISO FEC decoder, which is commonly referred to as turbo equalization.

Different from the optimum BCJR decoder, M-BCJR faces a severe problem in the SD mode. It frequently produces zero likelihood $\mathcal{L}_{\pm 1}$ leading to infinite LLR without reliable estimate of its magnitude, especially for small M values. This is because a smaller set of states per trellis stage makes it less likely for the β survivors to be overlapped with the α survivors. Since only the β values whose states overlap with the M α survivors are stored, the less overlapping translates to more zero β values and results in more zero state probabilities as in (6). A simple and widely used recovery method is to set a limit for $|LLR|$, in other words, clip the infinite LLRs. It is suggested in [37] to improve the LLR quality for M-BCJR by a 2^{nd}-round α-recursion. These schemes recalculate the infinite LLR using backup values like the preset threshold or the 2^{nd}-round α probabilities and can partially recover the SD performance. A more straightforward way is to increase M at the sacrifice of higher complexity. Thus, M should be properly selected to balance the complexity and the output LLR quality if M-BCJR is operated at the SD mode.

III. Experimental Verification of the Low-Complexity FTN Decoding With M-BCJR Algorithm

In this section, we demonstrate the reduced-complexity FTN decoding structure introduced in Section II via an IM-DD setup shown in Fig. 2. The transmitter consists of an 88-GS/s digital-to-analog converter (DAC) with 35-GHz analog bandwidth and a $LiNbO_3$ Mach-Zehnder modulator (MZM) with 32-GHz 3-dB bandwidth. The 3-dB bandwidth is usually quoted in literatures to describe the bandwidth limit, despite the FTN performance also relies on the channel response beyond the 3-dB frequency. In this section, we precisely define the bandwidth limit by a 25-GHz digital brick-wall filter at the transmitter, in other words, the signal is sinc-pulse shaped. This emulates the worst case of bandwidth limit to reveal the extreme performance of the FTN decoder. The optical signal is generated by a 1550-nm external-cavity laser (ECL) and sent directly to the receiver. The receiver contains a >100-GHz photodiode (PD) and a 113-GHz 256-GS/s real-time oscilloscope (RTO).

For the FTN PAM-4 signal whose BR is higher than the DAC sampling rate, the signal is generated in a sub-sampling manner with the transmitter offline DSP procedures shown in Fig. 2(i). We first generate the signal at 1 sample per symbol (sps) with a nominal BR, and then filter the signal by a frequency-domain rectangular-shaped lowpass filter for the subsequent down-sampling. This enables the high-BR signal generation without a physical clock rate at the same speed. The receiver offline DSP is listed in Fig. 2(ii). We use a 1000-tap linear FFE at 2 sps and then down-convert the signal to 1 sps. The FFE is adapted by least-mean square (LMS) algorithm which equalizes the in-band response without noise enhancement in the out-band (>25 GHz) region. Then the 1-sps signal goes through an NWF followed by an M-BCJR decoder.

As the MAP decoder performs sequential detection, it may suffer from overfitting if the symbol sequence under test is too short with respect to the decoder memory length, in other words, the performance may exhibit a strong pattern dependence which fluctuates dramatically with different sequences under test. This issue is more prominent for M-BCJR as its reduced complexity enables a much longer decoding memory. In this experiment, we set the sequence length as $BR/Sa \cdot 2^{17}$ where Sa stands for the DAC sampling rate. By offline M-BCJR simulation, we find the BER fluctuation is less than 10% using different random sequences even for a large decoding memory $L = 50$, which is sufficiently small for reliable performance evaluations. As a reference, if we try a short sequence length at the order of 2^{10} in the simulation, the BER fluctuation is around 50% when $L = 50$ and $R_{FTN} = 200\%$.

A. Whitened Matched Filter (WMF)

We choose the 100-GBd PAM-4 signal as an example and illustrate its noise power spectral density (PSD) after the LMS equalizer in Fig. 3(a) to show the effect of the WMF structure as introduced in Section II.A. The noise power is low within the Nyquist limit, and slig htly enhanced only when the frequency approaches 25 GHz. In contrast, the out-band noise PSD is levelled with the signal PSD, because the system has no out-band (>25 GHz) response at all. Clearly, the LMS-based FFE results in a colored-noise profile under its minimum mean-square error (MMSE) target, and a 2^{nd}-stage filtering is essential to whiten the noise for optimum MAP decoding. We select the tap numbers of 10 and 50 for NWFs and compare their noise whitening effect in Fig. 3(a). The corresponding coefficients are shown in Fig. 3(b) obtained by the auto-regressive PSE method [24]. Obviously, the 50-tap filter shows a better whitening effect than the 10-tap one and almost exhibits a flat noise PSD, except for the region near 25 GHz where the signal spectrum is sharply cut off by the digital brick-wall filter.

B. M-BCJR Parameter Setting for 100-GBd Signaling

The M-BCJR performance trades off with its complexity. We show the BER of the 100-GBd PAM-4 signal as a function of L

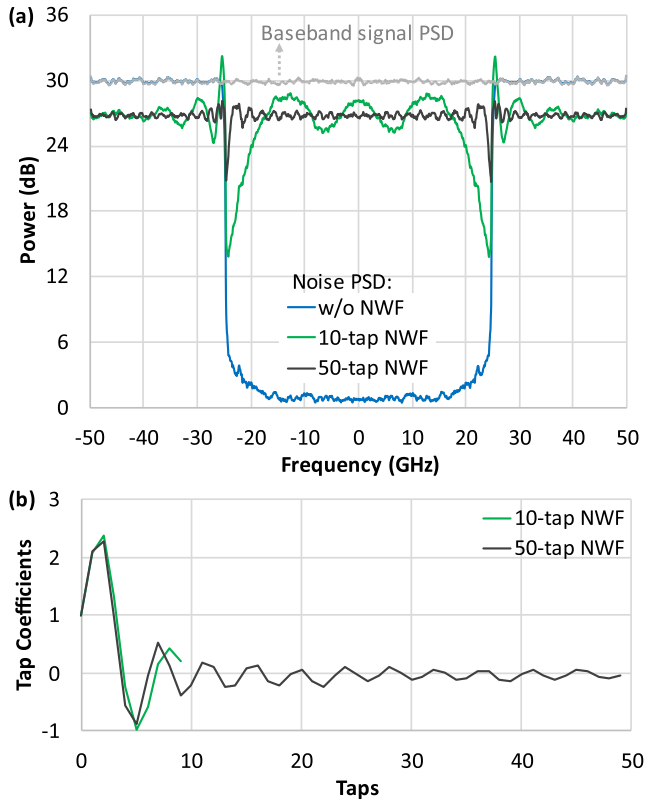


Fig. 3. 100-GBd FTN PAM-4 signal under 25-GHz bandwidth limit. (a) Noise power spectral density (PSD) after FFE with or without (w/o) noise whitening filter (NWF); (b) NWF coefficients extracted from the auto-regressive power spectrum estimation (PSE) by Yule-Walker equations [33].

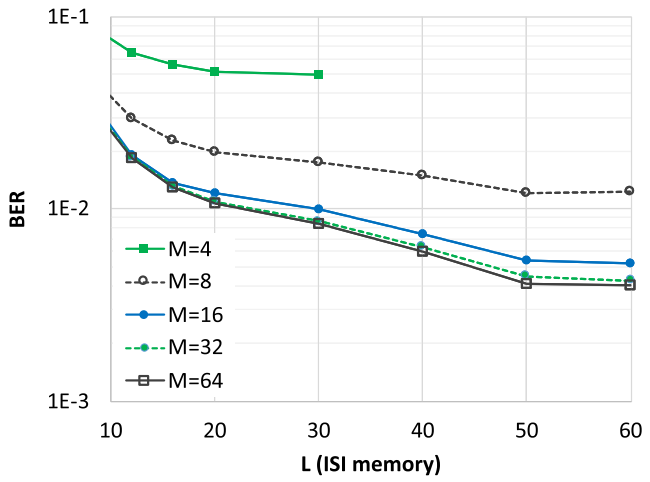


Fig. 4. BER dependence on M-BCJR parameters ($\boldsymbol{L}$ and $\boldsymbol{M}$) for the 100-GBd FTN PAM-4 signal under 25-GHz bandwidth limit.

in Fig. 4, where the BER gradually reduces with the increase of L and saturates around $L = 50$. We also illustrate the BER with variable M to study if the much smaller M than $|\mathcal{X}|^L$ degrades the performance. A big decoding penalty exhibits for $M = 4$, but the marginal benefit of increasing M quickly reduces and almost vanishes for $M > 16$.

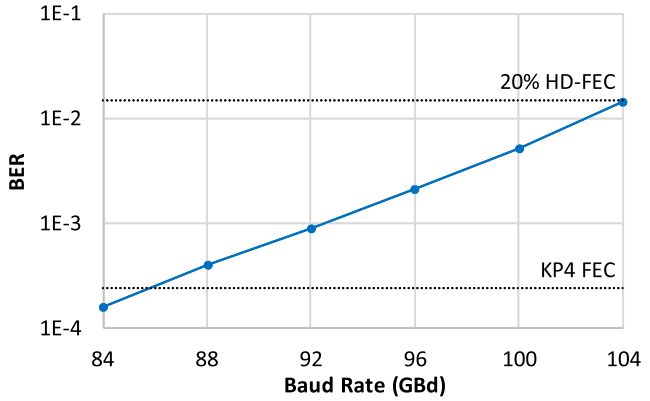


Fig. 5. BER as a function of PAM-4 baud-rate (BR) for the 25-GHz bandwidth setup. 20% HD-FEC [38].

C. BER as a Function of PAM-4 Baud Rate (BR)

We set $M = 16$ and show the BER as a function of PAM-4 BR in Fig. 5. Remarkably, the BER is below the KP4 FEC limit of 2.4e-4 at 84 GBd ($R_{FTN} = 168\%$), and below the 20% HD-FEC limit [38] of 1.5e-2 up to 104 GBd with a record $R_{FTN} = 208\%$ for PAM-4 signaling. The performance saturates when $L \approx 10$ for 84 GBd and saturates when $L \approx 50$ for 104 GBd, which indicates higher R_{FTN} induces longer ISI memory length.

IV. Achievable Information Rate (AIR) Comparison Between Nyquist and FTN Approaches

With respect to the Nyquist BR, an FTN scheme is usually evaluated by R_{FTN} in (1). R_{FTN} trades off with the underlying system SNR given an FTN scheme, just like the AIR trades off with the SNR for the Nyquist signal. More generally, R_{FTN} should be converted to AIR for an FTN signal to be fairly compared with a Nyquist signal. In this section, we conduct a series of AIR comparisons to study if the improved BR of FTN brings actual AIR advantage over the Nyquist counterpart using higher order modulations. Without loss of generality, we fix the format as PAM-4 for FTN signaling.

The AIR of Nyquist signals can be predicted by the Shannon capacity formula. However, it is not easy to find a closed form AIR estimation for FTN signals with tight bounds. Rather, it is suggested in literatures to evaluate the constrained capacity for FTN signaling [39], namely, the maximum AIR under certain restrictions like a fixed modulation format or a chosen decoding method. Such constrained capacity can be evaluated by Monte Carlo simulations [40], [41]. We will follow this idea to evaluate the AIR using the FTN decoding scheme as proposed in Section II. Note that the finding in this section is generalized to other decoding schemes, though their constrained capacities may be different due to their various ISI mitigation capabilities.

A. IM-DD Channel Modeling

The system AIR is closely related to its underlying channel condition. For a practical IM-DD system, its channel response is jointly determined by the transmitter, the fiber channel, and

the receiver. Rather than a sharp cutoff after a certain frequency, the usable bandwidth is commonly extended beyond the 3-dB point with a gradually decayed response for higher frequencies. This leads to a colored-SNR channel (*i.e.*, SNR varies with the frequency), whose capacity [11], [16] is

$$C = \int \log_2 \left(1 + \frac{S(f)}{N(f)} |H(f)|^2\right) df \tag{11}$$

where $S(f)$ is the transmitted signal PSD, $N(f)$ is the noise PSD and $H(f)$ is the channel frequency response. The capacity can be approached by the multicarrier modulation (MCM) with entropy loading that assigns a rate-adaptive format per subcarrier basis to match its channel condition (*e.g.*, SNR) [42]. Although MCM is the optimum Nyquist signaling, we exclude it from the comparison in this paper due to its high complexity for IM-DD optics. Rather, we choose higher-order PAMs for the Nyquist camp to be compared with FTN PAM-4.

For the gradually decayed channel response, it is obscure to define the Nyquist frequency (denoted as f_N) and the Nyquist BR (*i.e.*, $2f_N$). To begin with, we select a strictly band-limited channel whose response is sharply cut-off beyond f_N in Section IV.B. Correspondingly, the signal is sinc-pulse shaped with a rectangular spectrum. Then in Section IV.C, we emulate a more practical bandwidth-limited channel by root raised cosine (RRC) filtering that has a clearly defined f_N but extends its response beyond f_N. The transmitted signal is shaped by the same RRC filter to form a raised cosine (RC) response with zero ISI for the $2f_N$-Baud Nyquist signal. The model also represents a common feature of real-world Nyquist signaling, which usually occupies excess bandwidth beyond f_N. Finally in Section IV.D, we discuss the general FTN gain in a colored-SNR channel where f_N is not well defined. The simulation models are summarized in Fig. 6.

We use generalized mutual information (GMI) to evaluate the AIR. Unless otherwise specified, we assume binary HD-FEC coded modulations with flexible code rates, whose GMI is calculated as [43]

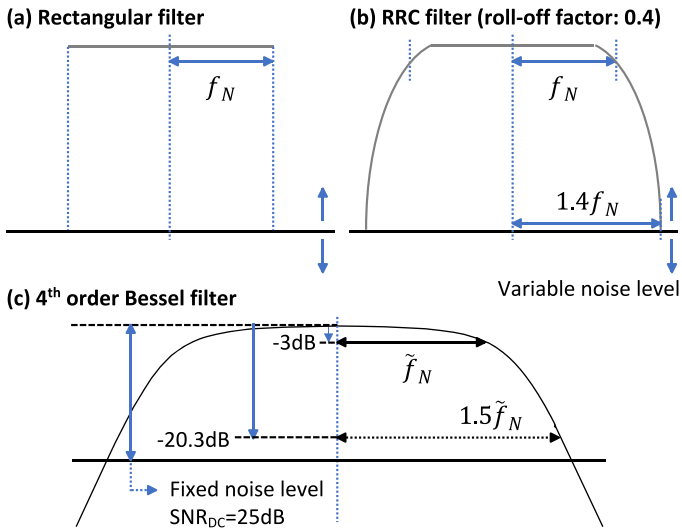
$$GMI_{HD} = H(X) - \log_2 |\mathcal{X}| \cdot \mathbb{H}_2(\epsilon) \tag{12}$$

Where $\mathbb{H}$ stands for the entropy and $\mathbb{H}_2(\cdot)$ is the binary entropy function, ϵ is the bit error probability (*i.e.*, BER) and $|\mathcal{X}|$ is the modulation order. To get the ϵ value for FTN signals, we use Monte Carlo simulations with the decoder structure described in Section II, whose parameters will be listed below accordingly. For a fair AIR comparison between Nyquist and FTN signals, the AIR of an FTN signal is its GMI scaling by R_{FTN},

$$\begin{aligned} AIR_{Nyquist} &= GMI \\ AIR_{FTN} &= R_{FTN} \cdot GMI \end{aligned} \tag{13}$$

namely, the AIR, in the unit of bits/symbol, is always evaluated at the Nyquist BR, and the total data rate is $BR_{Nyquist} \cdot AIR$.

Throughout the section, we fix the format as PAM-4 for FTN signaling and use variable-ary PAM-X ($X = 2^m$, $m = 2, 3, 4 \ldots$) for high-order Nyquist signaling. Note that the AIR may be enhanced for both signaling by using probabilistically shaped (PS) PAM, but we rule out this option due to its implementation controversy in peak-power constrained IM-DD systems [44] to simplify our analysis.



Channel		Rectangular	RRC (ρ=0.4)	4th order Bessel
Nyquist freq.		f_N	f_N	$\tilde{f}_N$ (3-dB BW)
Signal pulse		Sinc	RRC (ρ=0.4)	Sinc
MOD	Nyquist	PAM-X (X=4,8,16), flexible-rate HD-FEC		
	FTN	PAM-4, flexible-rate HD-FEC		
BR	Nyquist	$2f_N$	$2f_N$	$2\tilde{f}_N \cdot [1,1.6]$
	FTN	$2f_N \cdot [1,2]$	$2f_N \cdot [1.4,2]$	$2\tilde{f}_N \cdot [1.8,2.4]$
EQ	Nyquist	-	FFE	FFE
	FTN	NWF+MAP	FFE+NWF+MAP	FFE+NWF+MAP

Fig. 6. Bandwidth-limited channel modeling: (a) Rectangular filter in Section IV.B; (b) RRC filter in Section IV.C; (c) 4th order Bessel filter in Section IV.D. For rectangular and RRC filters, their Nyquist frequency ($\boldsymbol{f_N}$) is well defined; for the Bessel filter with a gradually decayed response, its 3-dB bandwidth (BW) is regarded as the nominal Nyquist frequency $\tilde{\boldsymbol{f}}_{\boldsymbol{N}}$. The notation $[\boldsymbol{a}, \boldsymbol{b}]$ represents a scalar range from $\boldsymbol{a}$ to $\boldsymbol{b}$ with a step size of 0.05 (or 0.1). MOD: modulation format; EQ: equalization; freq.: frequency.

B. AIR Comparisons in a Strictly Band-limited Channel

We first make comparisons in a strictly band-limited channel as shown in Fig. 6(a) with the sinc-pulse shaped signal. In Fig. 7(a), we illustrate the AIR of Nyquist PAM-X signals as the baselines. An FTN PAM-4 system naturally owns the capability of rate adaptation by adjusting R_{FTN}. To obtain an envelope AIR curve for FTN PAM-4, we adjust R_{FTN} from 105 to 200% with a step size of 5%, and find the maximum AIR at each SNR value. We use two sets of M-BCJR parameters: the 1st set ($L = 100$, $M = 32$) aims to approach the MAP performance limit and the 2nd set ($L = 10$, $M = 8$) emulates a more practical receiver. Remarkably, the FTN PAM-4 curve follows the trend of the Nyquist PAM-X curve even when $R_{FTN} \to 200\%$, indicating the superior ISI mitigation capability of the proposed

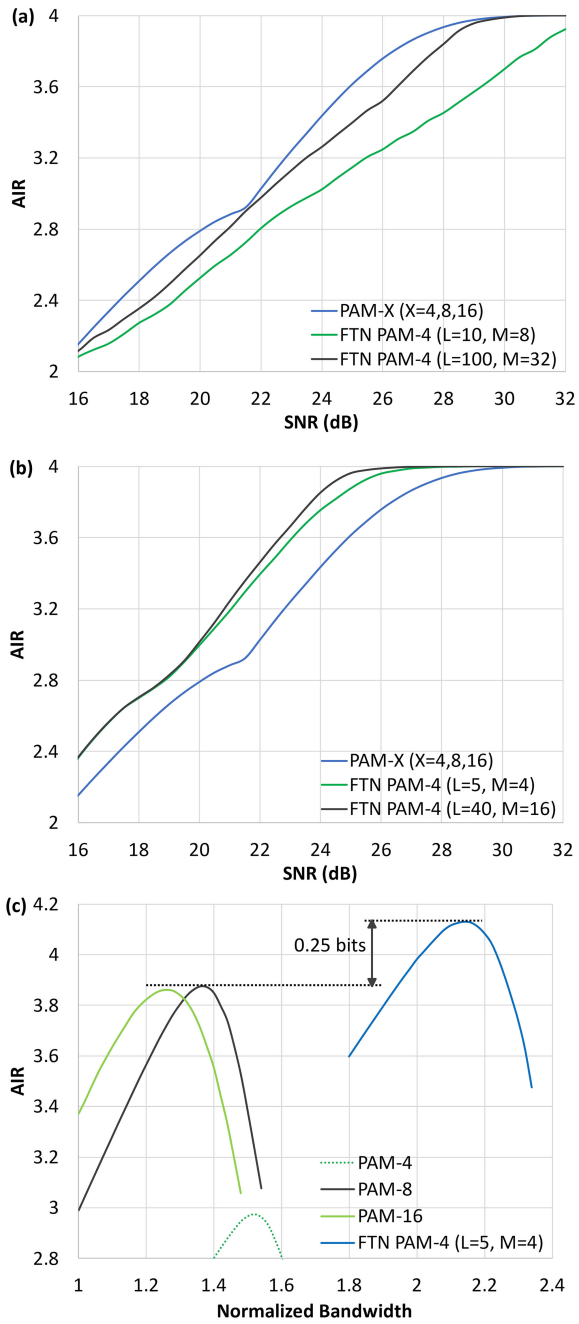


Fig. 7. AIR comparisons between Nyquist and FTN signals under bandwidth-limited channels emulated by: (a) Rectangular filter; (b) RRC filter (roll-off factor 0.4); (c) 4th order Bessel lowpass filter. As indicated in Fig. 8, the noise level is variable in (a-b) and fixed in (c). In (c), the AIR is normalized to the nominal Nyquist BR of $2\tilde{f}_N$ and the bandwidth (horizontal axis) is normalized to the nominal Nyquist frequency of $\tilde{f}_N$.

FTN decoding structure. Despite marginal SNR penalties, FTN can approach the AIR of higher-order PAM-X at any AIR target simply using PAM-4. This means the AIR should not be the primary consideration if we choose between the two schemes, and what really matters is the complexity to achieve the target AIR. The 2nd M-BCJR curve gradually deviates from the 1st one at the high SNR region with higher optimum R_{FTN}, because the more severe ISI needs a longer memory L.

C. AIR Comparisons in an RRC-Filtered Channel

We use an RRC filtered channel as shown in Fig. 6(b) to emulate a more practical frequency response with a smooth decay. We choose the roll-off factor as 0.4, which corresponds to 40% excess bandwidth beyond f_N, a typical ratio for IM-DD systems. We perform the following simulation procedures for the Nyquist signal:

1) Upsample to 2 sps
2) Apply the f_N-Hz RRC filter (pulse shaping)
3) Add AWGN
4) Apply the same RRC filter (bandwidth limit emulation)
5) Downsample to 1 sps for symbol-by-symbol decision

In terms of the FTN PAM-4 signal with a nominal baud-rate of BR, the simulation procedures are:

1) Upsample to 2 sps
2) Lowpass filter the signal with a rectangular-shaped filter ($2f_N$-Hz bandwidth) to resample it by the ratio $2f_N/BR$
3) Apply the f_N-Hz RRC filter
4) Add AWGN
5) Apply the same RRC filter
6) Resample back to 2 sps by the ratio $BR/(2f_N)$
7) Perform FTN decoding (FFE, NWF, and M-BCJR)

Between the two lists above, we keep the steps 3-5 (FTN) to be the same as the steps 2-4 (Nyquist) to guarantee an identical channel modeling. Though AWGN is added to the upsampled signal, its power is normalized to the frequency range $[0, f_N]$ to align the SNR definitions between the signals with and without excess bandwidth in Fig. 7(a) and (b). We adjust the resampling ratio in step 2 (FTN) to emulate different R_{FTN}. To make full use of the 40% excess bandwidth, we set the minimal R_{FTN} as 140% and increase it up to 200% with a step size of 5% and find the maximum AIR for each SNR value.

Owing to the excess bandwidth and smoother spectrum roll-off, the RRC-filtered FTN channel has a shortened response. In Fig. 7(b), we use the M-BCJR parameter $L = 40, M = 16$ to approach the MAP performance limit and $L = 5, M = 4$ to test an extremely low-complexity decoder. Both curves lie above the Nyquist PAM-X one, indicating the AIR gain of the FTN signal over the Nyquist counterpart. The simpler decoder almost overlaps with the high-complexity one when $SNR < 20$ dB, a practical SNR region for most real-world IM-DD systems.

The gain of FTN is attributed to its capability of salvaging the excess bandwidth. Compared with the results in Fig. 7(a), the excess bandwidth occupied by the RRC filtering shows no influence on Nyquist PAM-X signals, because their optimum demodulation is in a symbol-by-symbol manner after matched filtering and the pulse shaping process is transparent to them. In contrast, an FTN signal is aware of the pulse shaping as the FTN decoder takes sequential observations for demodulation. In essence, the pulse shape leads to a colored $H(f)$, whose excess bandwidth beyond f_N brings extra AIR according to (11).

D. FTN Gain in Colored-SNR Channels: A General View

We turn our focus back to the most practical IM-DD channel with a gradually decayed frequency response. In this case, it is difficult to distinguish between Nyquist and FTN signals via

their BRs. Instead, we use the equalization scheme to define the two in this section. IM-DD receivers usually use the linear FFE only by ignoring the residual ISI after equalization, and we treat this as the Nyquist approach. In contrast, an FTN signal exhibits severe ISI after the linear FFE, making it necessary to apply nonlinear equalizations for ISI mitigation like the MAP decoder introduced in Section II. We treat the demodulation with a MAP decoder as the FTN approach. We use the 4th order Bessel filter to emulate the colored-SNR channel, as shown in Fig. 6(c). This filter is part of the reference receiver in the 400GbE standard [3] that represents a typical bandwidth-limited IM-DD channel. Its 3-dB bandwidth is treated as the nominal Nyquist bandwidth $\tilde{f}_N$. In this subsection, the Nyquist approach can extend the signal bandwidth beyond $\tilde{f}_N$ for a higher BR at a sacrifice of degraded SNR. The tradeoff between BR and SNR results in an optimum BR for the Nyquist approach to maximize the AIR. Without a clearly defined Nyquist BR, we mainly compare the capability of utilizing the excess bandwidth beyond $\tilde{f}_N$ between Nyquist and FTN approaches. Therefore, we fix the noise level and show the AIR as a function of signal bandwidth in Fig. 7(c), where the signal bandwidth is normalized to $\tilde{f}_N$ and the AIR is normalized to the Nyquist BR of $2\tilde{f}_N$. The noise level is chosen to have an SNR of 25 dB at DC (direct current, *i.e.*, the zero frequency), a typical value for state-of-the-art IM-DD systems. With the M-BCJR parameters of $L = 5$, $M = 4$, the FTN approach achieves around 0.25-bit AIR gain over the Nyquist approach, and the gain can be further enhanced if higher M-BCJR complexity is allowed. This clearly indicates FTN can utilize the excess bandwidth in a more effective manner.

Compared with Section IV.C, though the Nyquist approach in this subsection unlocks the capability of extending the signal bandwidth beyond $\tilde{f}_N$, it cannot compete with FTN, because the excess bandwidth with worse frequency response enhances the ISI penalty which quickly overwhelms the BR improvement. Further, in a colored-SNR channel, the FTN gain over Nyquist is a generalized phenomenon, because the FTN decoding allows larger amount of ISI to extend the usable bandwidth range and improve the AIR according to (11). This offers a fundamental incentive to pursue higher BR using FTN signaling with respect to lower-BR Nyquist signaling with higher-order modulations.

V. Record FTN-PAM Baud Rate Demonstrations and Comparisons With Nyquist Signaling

The latest ultrahigh-bandwidth optical transmission systems support the generation of 100-GHz class electronic signals [28]. In this section, we demonstrate both Nyquist and FTN signaling in a 100-GHz IM-DD system, and make comparisons between the two. As shown in Fig. 8, the transmitter consists of a digital-band-interleaved (DBI) DAC [45] with 100-GHz analog bandwidth and a thin-film $LiNbO_3$ MZM [46] with more than 100-GHz bandwidth; and the receiver consists of a >100-GHz PD and a 256-GSa/s 113-GHz RTO. The end-to-end bandwidth of the entire setup is constrained by the DBI-DAC that contains a sharp analog filter at around 100 GHz. Therefore, the system mimics the condition discussed in Section IV.B, namely,

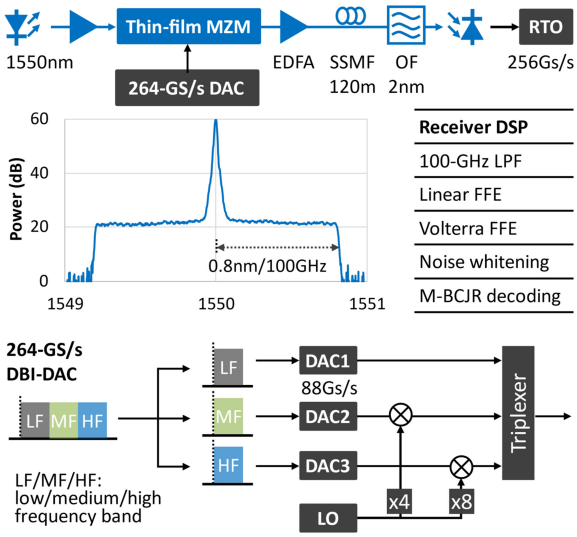


Fig. 8. IM-DD experiment setup with 100-GHz bandwidth limit. OF: optical filter; SSMF: standard single-mode fiber; DBI: digital band interleave [45].

a band-limited channel with rectangular filtering as shown in Fig. 6(a).

The DBI-DAC uses three CMOS DACs, each with around 35-GHz bandwidth, to generate three baseband signals to be up-converted to the low-, medium- and high- frequency (LF, MF, and HF) bands, respectively, as shown in Fig. 8. The triplexer then combines the three bands without inter-band interference by filtering out the undesired frequencies of each band using sharp analog filters. Its output signal has a peak-to-peak voltage of around 1.0 V. Due to the limited DBI-DAC swing, the thin-film MZM is biased away from its quadrature point towards the null point for a higher extinction ratio (ER). The bias condition together with the intrinsic loss of the MZM induces around 30 dB modulation loss. To get a high system optical SNR (OSNR), the 1550-nm ECL output is amplified by an erbium-doped fiber amplifier (EDFA) to inject around 23-dBm light to the MZM. Another EDFA follows the MZM to compensate the modulator loss, and a 250-GHz optical filter rejects the out-band amplified spontaneous emission (ASE) noise ahead of the receiver.

For the FTN signaling, we fix the modulation as either PAM-2 or PAM-4 while changing its BR to realize rate adaptation. The signal is generated in the sub-sampling mode as explained at the beginning of Section III. For the Nyquist signaling as reported in [28], we fix the BR as 200 GBd, the highest BR allowed by the Nyquist sampling theorem, and use PS PAM-16 with variable entropies for rate adaptation. This Nyquist signal is RRC filtered with a roll-off factor of 0.01. The receiver FTN DSP is similar to that in the 25-GHz setup listed in Fig. 2(ii), except for an additional Volterra nonlinear FFE [47] after the linear FFE to compensate for the modulator nonlinearity due to the bias condition near the null point. The Volterra kernel contains 2nd-order polynomials of the 2-sps samples with 40-tap memory, leading to 820 polynomial coefficients. The 200-GBd PS PAM-16 signals, processed in [28], use only the linear FFE with the similar parameter setting as for the FTN signal.

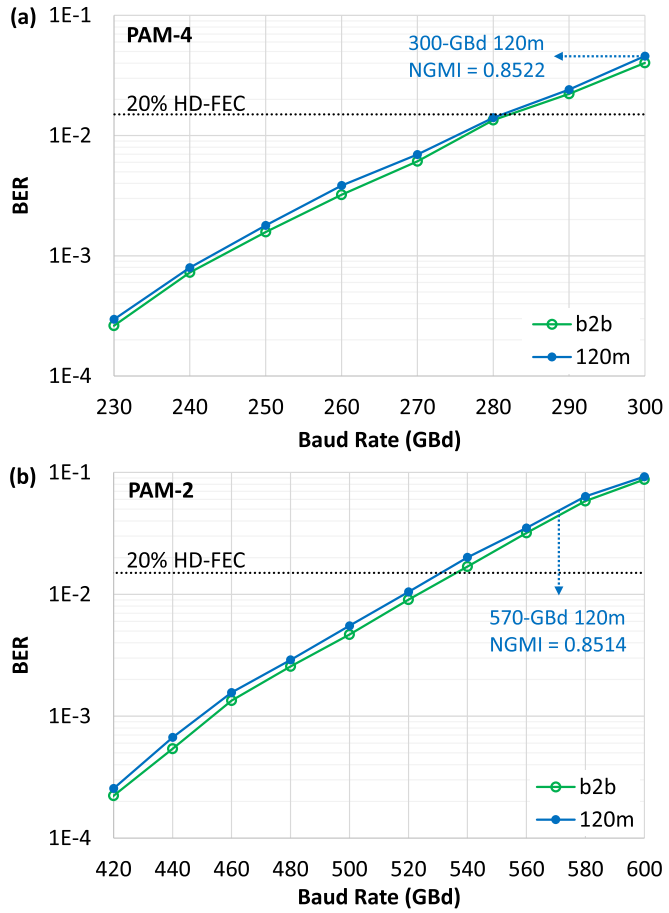
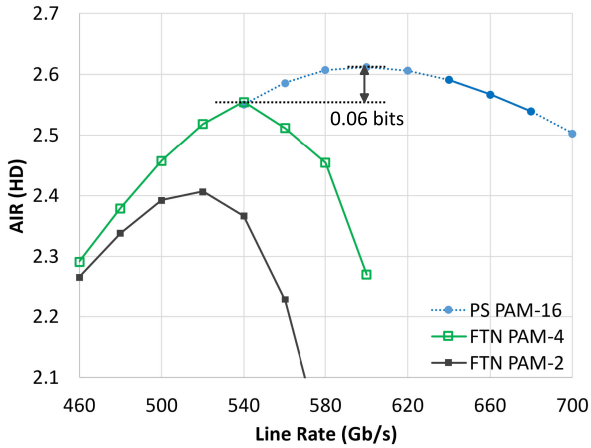


Fig. 9. BER as a function of PAM baud-rate for the 100-GHz bandwidth setup: (a) PAM-4; (b) PAM-2. 20% HD-FEC. The figure also shows the NGMI of 300-GBd PAM-4 and 570-GBd PAM-2 after 120-m SSMF, both are above the 26% SD-FEC threshold. The M-BCJR decoder uses $\boldsymbol{L} = 40,\ \boldsymbol{M} = 16$ for HD, and $\boldsymbol{L} = 200,\ \boldsymbol{M} = 512$ for SD.



	PS PAM-16	FTN PAM-4	FTN PAM-2
AIR (SD)	2.741	2.557	2.426
NBR (26% SD)	494.5	476.2	452.4

Fig. 10. AIR comparisons between Nyquist and FTN signals in a strictly band-limited 100-GHz IM-DD experiment. The Nyquist signal is a 200-GBd RRC-filtered (0.01 roll-off) PS PAM-16 signal, and the FTN signal is either uniform PAM-4 or PAM-2. The upper figure assumes HD-FEC, and the lower table assumes SD-FEC and compares AIR and NBR when the system performance approaches the NGMI threshold (0.8456) of a 26% SD-FEC [48].

A. FTN Performance

By choosing $L = 40$ and $M = 16$ for the M-BCJR decoder, we evaluate the PAM-4 BER with variable BRs up to 300 GBd, as shown in Fig. 9(a). Using a 20% HD-FEC [38], at most 280-GBd BR is achieved for PAM-4. We also show the BER vs BR for PAM-2 in Fig. 9(b), where >520-GBd BR is achieved at 20% HD-FEC threshold. The AIR of PAM-2 is slightly lower than PAM-4, because the receiver DSP exhibits more implementation penalty with the much faster PAM-2 signal.

The transmission performance of datacenter links is usually characterized by their tolerance to chromatic dispersion (CD). Besides the back-to-back (BtB) test, we transmit the FTN signal over 120-m standard single mode fiber (SSMF) accumulating 2-ps/nm CD at 1550 nm. This is equivalent to the CD of 1-km SSMF at the longest LAN-WDM wavelength of 1330 nm in O-band. Compared to BtB, the penalty induced by the 2-ps/nm CD is trivial though the optical bandwidth is as high as 200 GHz. As CD essentially induces ISI penalty to an IM-DD system, the MAP decoder helps alleviate its influence in a similar manner as the FTN-induced ISI.

To achieve record PAM BRs using the more powerful SD-FEC, we dramatically increase M to 512 to alleviate the zero $\mathcal{L}_{\pm 1}$ issue as mentioned in Section II.D. Moreover, we increase L to 200 to boost the performance at high R_{FTN}. We calculate normalized generalized mutual information (NGMI) from the LLR to be compared with the NGMI threshold of the SD-FEC code. We use a 26% SD-FEC scheme with an NGMI threshold of 0.8456 [48]. The FTN PAM-4 achieves 300-GBd BR with an NGMI of 0.8522 after 120-m SSMF, and the PAM-2 achieves 570 GBd with an NGMI of 0.8514 after 120-m SSMF.

B. Comparison With the Nyquist Signaling

We compare the FTN signaling with the Nyquist signaling as reported in [28] in the BtB condition. Though the two share the same experimental setup, the comparison is not strictly apple-to-apple because: (i) the FTN signal is processed with an extra VNE, while (ii) the PS PAM-16 achieves constellation shaping gain over the uniform PAM signaling [44]. Nevertheless, it offers an intuitive insight from an experimental perspective.

For the cross reference with the simulation result in Section IV, we first assume HD-FEC for the comparison and show the AIR (HD) as a function of line rate in Fig. 10. Note that only three line rates were reported in [28]. As the effective SNR was similar (17.6 dB) among different PS PAM-16 entropies in the experiment, we extend the line rate range in Fig. 10 for PS PAM-16 (as indicated by the dotted blue line) by evaluating the theoretical AIR under the measured SNR. For both Nyquist and FTN signaling, there exists an optimum line rate to achieve the maximum AIR. The FTN PAM-4 exhibits around 0.06-bit AIR gap to the PS PAM-16. This aligns well with what we observed

in Fig. 7(a), namely, FTN can approach the AIR of Nyquist signaling in a rectangular-filtered channel given sufficient ISI mitigation capability. Using the same 26% SD-FEC [48], we also compare the AIR (SD) and net bit rate (NBR) at the specific line rate that leads to the NGMI approaching FEC threshold, as listed in the table below Fig. 10. The AIR (SD) is evaluated by GMI_{SD} [43], and the NBR is calculated by subtracting the FEC overhead from the line rate. The gap between Nyquist PS PAM-16 and FTN PAM-4 signals is slightly wider using SD-FEC than HD-FEC, indicating the M-BCJR algorithm inside the FTN decoder needs even higher computational complexity to get reliable SD performance as we explained in Section II.D. This implies HD-FEC may be more cost efficient than SD-FEC in the FTN system with a high FTN ratio; and fortunately, HD-FEC is the primary choice for IM-DD systems deployed in datacenter links.

It is noted again that the comparison here is made in a strictly band-limited channel. In the system with a gradually decayed channel response, FTN can potentially outperform the Nyquist approach as we discussed in Section IV.C and IV.D.

VI. Conclusion

200G per lane optics is being developed as the next-gen optical interface for datacenter applications based on 100-GBd PAM-4. Beyond 100 GBd, it is commonly believed that higher-order formats like PAM-6/8 may be an inevitable choice to improve the data rate, as the bandwidth growth of opto-electronics may start lagging the increasing capacity demand. In this paper, we reveal that FTN PAM-4 can approach the capacity of a practical IM-DD channel with a gradually decayed response and may achieve AIR gain over higher-order modulations. This offers a fundamental incentive to consider FTN technique at post 200G era. The FTN advantage relies on an effective decoding scheme. In this paper, we introduce a reduced-complexity MAP decoder by limiting the total number of states per trellis stage, namely, the M-BCJR decoder, which enables much longer ISI memory for a superior ISI mitigation capability. Based on the proposed FTN receiver structure cascading FFE, NWF and M-BCJR, we experimentally demonstrate record FTN baud-rates of 300 GBd for PAM-4 and 570 GBd for PAM-2 in a band-limited 100-GHz IM-DD system.

While this paper focuses on the AIR, the implementation of FTN technique relies on additional factors like complexity and latency. Encouragingly, the FTN advantage as revealed in this paper under practical bandwidth limits reserves the room for the tradeoff between AIR and complexity/latency, making it promising to design a practical FTN scheme to be competitive with the conventional Nyquist approach.

Acknowledgment

We would like to thank HyperLight for their support on the thin-film modulator.

References

[1] 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.

[2] "Enabling the next generation of cloud and AI using 800Gb/s optical modules," 800G Pluggable Multi-Source Agreement (MSA). [Online]. Available: https://www.800gmsa.com/documents/white-paper

[3] *IEEE 802.3bs–IEEE Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation*, IEEE Standard 802.3bs, 2017. [Online]. Available: https://standards.ieee.org/standard/802_3bs-2017.html

[4] "Common Electrical I/O (CEI)-112G," *Opt. Internetworking Forum (OIF)*. [Online]. Available: https://www.oiforum.com/technical-work/hot-topics/common-electrical-interface-cei-112g-2/

[5] N. Stojanovic, C. Prodaniuc, L. Zhang, and J. Wei, "210/225 Gbit/s PAM-6 transmission with BER below KP4-FEC/EFEC and at least 14 dB link budget," in *Proc. Eur. Conf. Opt. Commun.*, Rome, Italy, 2018, Paper We1H-4.

[6] M. A. Mestre *et al.*, "Direct detection transceiver at 150-Gbit/s net data rate using PAM 8 for optical interconnects," in *Proc. Eur. Conf. Opt. Commun.*, Valencia, Spain, 2015, Paper PDP2.4.

[7] 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, 2014.

[8] W. Yan *et al.*, "100 Gb/s optical IM-DD transmission with 10G-Class devices enabled by 65G Samples/s CMOS DAC core," in *Proc. Opt. Fiber Commun. Conf.*, Anaheim, CA, USA, 2013, Paper OM3H.1.

[9] T. Tanaka *et al.*, "Experimental demonstration of 448-Gbps+ DMT transmission over 30-km SMF," in *Proc. Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, 2014, Paper M2I.5.

[10] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1451–1462, 1975.

[11] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[12] K. Igarashi *et al.*, "1.03-Exabit/skm super-Nyquist-WDM transmission over 7,326-km seven-core fiber," in *Proc. Eur. Conf. Opt. Commun.*, London, UK, 2013, Paper PD3.E.3.

[13] V. Houtsma, 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.*, Los Angeles, CA, USA, 2015, Paper Th4H.1.

[14] 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.*, Los Angeles, CA, USA, 2017, Paper Th3D.1.

[15] S. Yamamoto, A. Masuda, H. Taniguchi, and Y. Kisaka, "92-Gbaud PAM4 transmission using spectral-shaping trellis-coded-modulation with 20-GHz bandwidth limitation," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2019, Paper W4I.5.

[16] 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.

[17] 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.

[18] Q. Hu, K. Schuh, M. Chagnon, F. Buchali, and H. Bülow, "Up to 94 GBd THP PAM-4 transmission with 33 GHz bandwidth limitation," in *Proc. Eur. Conf. Opt. Commun.*, Rome, Italy, 2018, Paper Th3F.6.

[19] G. Huang, H. Nakashima, and T. Hoshida, "PDM-16QAM transmission of 135 GBaud enabled by 120 GSa/s DAC and Tomlinson–Harashima precoding," *Opt. Lett.*, vol. 46, no. 18, pp. 4518–4521, 2021.

[20] J. X. Cai *et al.*, "20 Tbit/s transmission over 6860 km with sub-Nyquist channel spacing," *J. Lightw. Technol.*, vol. 30, no. 4, pp. 651–657, Feb. 2012.

[21] J. Zhang *et al.*, "Transmission of 20×440-Gb/s super-nyquist-filtered signals over 3600 km based on single-carrier 110-GBaud PDM QPSK with 100-GHz grid," in *Proc. Opt. Fiber Commun. Conf.*, San Francisco, CA, USA, 2014, Paper Th5B.3.

[22] Y. Zhu, L. Zhang, X. Fang, F. Zhang, and W. Hu, "Sub-sampling generation and transmission of 205Gbaud OOK signal with 120GSa/s DAC based on high-order partial response narrowing," in *Proc. Asian Commun. Photon. Conf.*, Beijing, China, 2020, Paper M4D.4.

[23] D. Che *et al.*, "Long-term reliable >200-Gb/s directly modulated lasers with 800GbE-compliant DSP," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2021, Paper F3A.3.

[24] L. Liu, L. Li, and Y. Lu, "Detection of 56GBaud PDM-QPSK generated by commercial CMOS DAC with 11GHz analog bandwidth," in *Proc. Eur. Conf. Opt. Commun.*, Cannes, France, 2014, Paper P.3.1.

[25] M. Secondini *et al.*, "Optical time–frequency packing: Principles, design, implementation, and experimental demonstration," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3558–3570, Sep. 2015.

[26] 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.

[27] D. Che and X. Chen, "Faster-than-Nyquist signaling up to 300-GBd PAM-4 and 570-GBd OOK suitable for co-packaged optics," in *Proc. Eur. Conf. Opt. Commun.*, Bordeaux, France, 2021, Paper Th3C2-PD1.6.

[28] 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. Eur. Conf. Opt. Commun.*, Brussels, Belgium, 2020, Paper Th3A-2.

[29] G. 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.

[30] F. Rusek and A. Prlja, "Optimal channel shortening for MIMO and ISI channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 810–818, Feb. 2012.

[31] J. Fan, Y. Ren, Y. Zhang, and X. Luo, "MLSE equalizer with channel shortening for faster-than-Nyquist signaling," *IEEE Photon. Technol. Lett.*, vol. 30, no. 9, pp. 793–796, May 2018.

[32] A. Masuda, H. Taniguchi, S. Yamamoto, Y. Ogiso, and M. Fukutoku, "96-Gbaud PAM-4 transmission with 1 sample/symbol under 22-GHz bandwidth limitation using NL-MLSE based on third-order Volterra filter," in *Proc. Eur. Conf. Opt. Commun.*, Rome, Italy, 2018, Paper We4G.5.

[33] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.

[34] 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.

[35] 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.

[36] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.

[37] 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.

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

[39] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 764–775, Feb. 2009.

[40] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[41] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. IEEE Glob. Telecommun. Conf.*, San Antonio, TX, USA, 2001, pp. 2992–2996.

[42] 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, 2018.

[43] G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv:1707.01134*.

[44] 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.

[45] X. Chen *et al.*, "All-electronic 100-GHz bandwidth digital-to-analog converter generating PAM signals up to 190 GBaud," *J. Lightw. Technol.*, vol. 35, no. 3, pp. 411–417, 2017.

[46] C. Wang *et al.*, "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature*, vol. 562, pp. 101–104, 2018.

[47] N. Stojanovic, F. Karinou, Z. Qiang, and C. Prodaniuc, "Volterra and Wiener equalizers for short-reach 100G PAM-4 applications," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4583–4594, 2017.

[48] J. M. Gené, X. Chen, J. Cho, S. Chandrasekhar, and P. J. Winzer, "Experimental demonstration of widely tunable rate/reach adaptation from 80 km to 12,000 km using probabilistic constellation shaping," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2020, Paper M3G.3.