# Maximum-Likelihood Detection and Constrained Coding on Optical Channels

Zeinab Taghavi, *Student Member, IEEE*, Nikola Alic, *Member, IEEE*, and George Papen, *Member, IEEE*

***Abstract*—The most probable error events for a long-haul optical communication system using sequence estimation are characterized, and the performance using a modulation code is analyzed. The modulation code is restricted to be a constrained code that enhances the minimum distance between received signal sequences. We investigate several constrained codes used in combination with 4- and 8-state Viterbi algorithms and apply them to a standard noncoherent NRZ system. The improvement using a rate 0.85 code can be as high as 5 dB relative to an uncoded system operating at the same symbol rate. If the coded and uncoded systems are constrained to have the same information rate, a more modest 1 dB improvement can be obtained for specific fiber distances. We also determine the effect of the code on the number of states of the Viterbi algorithm and the sensitivity of the sampling point. Using these codes, the tolerance of the receiver to sampling errors can be increased by a nearly a factor of two for specific distances. The performance improvements that these codes can provide make them attractive candidates when used in conjunction with sequence estimation.**

***Index Terms*—Constrained codes, distance-enhancing codes, modulation codes, long-haul optical channel, maximum-likelihood sequence estimation, reduced-state Viterbi algorithm.**

## I. Introduction

In an optical communication system that uses a square-law detector, the combination of the square-law characteristic and intersymbol interference (ISI) produces a nonlinear channel model for the output electrical signal [1]. Recently, electrical-domain equalization methods have been investigated as an alternative to all optical dispersion compensation, e.g., dispersion-compensating fibers (DCF). Several compensation methods have been proposed for systems using both the phase and amplitude of the signal, e.g., [2]–[7]. For systems that do not use phase such as a typical noncoherent non-return-to-zero (NRZ) systems, effective compensation methods are more difficult to implement. It can be proved that no linear electrical equalizer can completely compensate this nonlinear channel even if the channel is noiseless [8], [9]. Therefore, the performance achieved by electrical equalization techniques for nonlinear optical channels that do not use phase is inferior to equalization that does use phase such as all optical compensation.

An effective solution to control the ISI is to use maximum-likelihood sequence estimation (MLSE), which is the optimal detection and equalization method for minimizing the word error rate (WER) [8]. It appears that MLSE was first suggested to compensate chromatic dispersion in optical fibers in [10], and [11]. The exact signal statistics for a long-haul link containing a single optical pre-amplifier, where the noise is added in the optical domain, was calculated and applied to maximum likelihood sequence estimation for electronic dispersion compensation in [12]. An approximation of the performance of such a system is analyzed in [13]. In [14] high complexity Vitebi algorithms are used to reach 1 040 km on a standard single-mode fiber at 10 Gbit/s.

In addition to improving detection, the overall system can be enhanced by matching the transmitted signal to the channel[1] using coding and/or modulation. These types of codes are called "modulation codes" or "line codes." Usually, a modulation code is applied in conjunction with an error correction code (ECC), so that the coded data after an ECC is coded again by the modulation code block. Fig. 1 shows a block diagram of a long-haul optical system using both ECC and a modulation code.

Modulation codes are widely implemented in magnetic and optical disc recording and most optical communication systems [15]–[18]. While the ultimate goal of all types of modulation codes is to reduce the BER, different codes use different design criteria to achieve this goal [19]–[22].

Increasing the "distance" between the received points in the signal space is a common design criterion. In the absence of noise, the signal vectors corresponding to a sampled signal sequence form a constellation space. Given the channel statistics, a metric can be defined between each pair of signal vectors and it is commonly referred to as distance[2]. Errors occur predominantly between vectors near in distance [8]. A distance-enhancing modulation code chooses and removes a subset of signal vectors to increase the minimum distance.

A constrained code is a modulation code based on a finite state machine, commonly represented by a state diagram or trellis [19]. Constrained codes restrict transitions in a state diagram. In [23] a constrained code was designed for an optical fiber channel to remove the isolated "one" pattern that caused significant errors in a system using a threshold detector. Here,

Manuscript received January 03, 2008; revised May 03, 2008. Current version published May 13, 2009.

Z. Taghavi and G. Papen are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: ztaghavi@ucsd.edu; gpapen@ucsd.edu).

N. Alic is with the California Institute of Telecommunications and Information Technologies (CALIT2), University of California, San Diego, La Jolla, CA 92093-0436 USA (e-mail: nalic@ece.ucsd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2009.2012728

[1]In this paper, the term channel incorporates the transmitter, the fiber, and the optical sensor.

[2]The metric only corresponds to a Euclidean distance when the noise is a additive Gaussian.

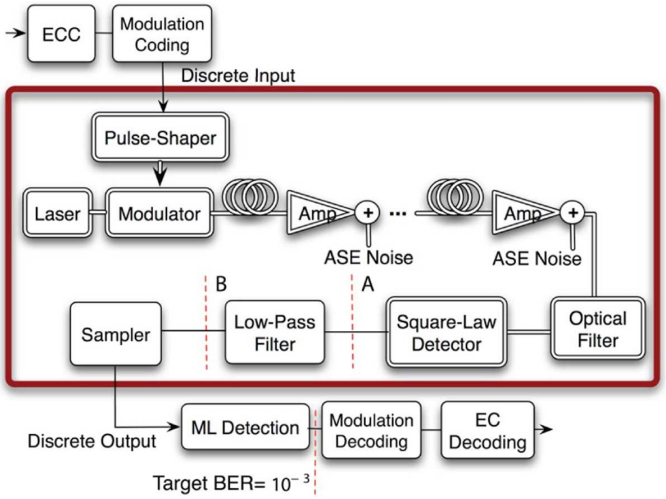
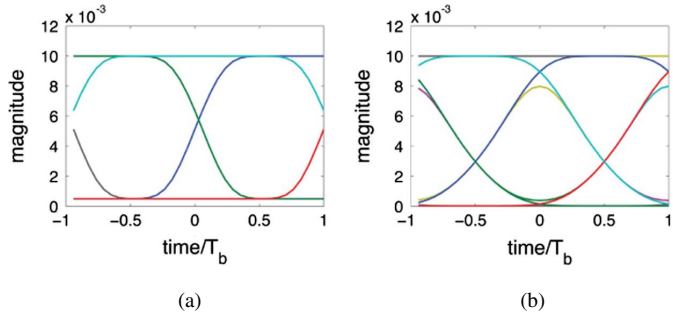



Fig. 1. Long-haul optical channel including the transmitter and the receiver.




(a) (b)

Fig. 2. Eye diagrams of a back-to-back system with $\xi = 0$: (a) before the electrical filter, (b) after the electrical filter.

we extend this result and present a general constrained code design process for a noncoherent NRZ optical channel. In this method, we first characterize the most probable error events and then based on this analysis, we design distance enhancing constrained codes.

Distance enhancing constrained codes increase the distance by specifying a "forbidden list" of code strings whose omission ensures small distance error events do not occur [24], [25]. The technique of generating a constrained code using "forbidden list" can be summarized in three steps [26]. First, the set of most probable input error events, $S$, for the specified channel is determined. Second, a list of forbidden patterns, $Q$, is chosen such that preventing them from being transmitted and received reduces the number of error events in set $S$. Finally, an efficient practical encoder and decoder for this constraint are constructed to avoid the set of patterns in $Q$. For this system, the sequence detector in the receiver is designed to incorporate the channel and code constraints. Therefore, the constraint can reduce the number of sequence detector states relative to the uncoded version.

The rest of the paper is organized as follows. In the next section, a mathematical model for the long-haul optical channel is defined and MLSE at the receiver is characterized without coding. In the Section III, the design of distance-enhancing constrained codes based on a forbidden list of patterns for the nonlinear optical channel is described. In Section IV, the performance of several codes is investigated with respect to key system parameters including effect of sampling shift and the number of states of the MLSE. Finally, the paper concludes with prospects for application of these codes.

## II. Optimal Detection for Optical Channel

In this section, we introduce the mathematical model of the optical channel, the transmitter, and the receiver.

### A. Channel Model

The channel used is a noncoherent long-haul optical channel with a square-law detector. The transmitted signal is NRZ with a finite extinction ratio, $ER$, defined as $20\log_{10}(A_{\max}/A_{\min})$, where $A_{\max}$ is the maximum magnitude of an isolated mark (one) and $A_{\min}$ is the minimum magnitude of an isolated space (zero). For long-haul channels with optical amplifiers, ASE noise caused by amplifiers is the dominant noise. The received signal $r(t)$ is modeled as

$$r(t) = |[s_t(t) \star \ell(t) + n_o(t)] \star h(t)|^2 \star h_e(t)$$

where $\star$ is the convolution operator, $s_t(t)$ is the transmitted signal, $\ell(t)$ is the impulse response of the fiber channel, $n_o(t)$ is the complex additive white Gaussian optical amplifier noise (AWGN), $h(t)$ is the optical filter before detector, and $h_e(t)$ is the electrical filter. A block diagram of the channel model is shown in Fig. 1.

The transmitted NRZ signal is

$$s_t(t) = \max[A_{\min}, \min\{A_{\max}, \mathbf{d} \star p_t(t)\}] \tag{1}$$

where $\mathbf{d}$ is the discrete transmitted bit sequence and $p_t(t)$ is the impulse response of the transmitter. A consequence of using NRZ signals is that, in general, the relationship between the transmitted sequence $\mathbf{d}$ and the transmitted waveform $s_t(t)$ is not linear even with respect to the optical field. A typical transmitted eye diagram is shown in Fig. 2(a).

The impulse response of the fiber channel, $\ell(t)$, which is the source of intersymbol interference, is proportional to $\exp(jt^2/(2\beta_2 L))$, where $L$ is the propagation length or the effective fiber length and $\beta_2$ is the dispersion coefficient [27]. We now define the normalized dispersion index (NDI) as

$$\xi = -2\beta_2 L R_b^2.$$

The impulse response of the channel, $\ell(t)$, can then be written as

$$\ell(t) = \frac{e^{j\pi/4}\mathrm{sign}(\xi)}{T_b\sqrt{\pi|\xi|}} \exp\left[-\frac{j}{\xi}\left(\frac{t}{T_b}\right)^2\right]$$

[28], where $R_b = 1/T_b$ is the symbol or baud rate with typical units of Gsym/s ($10^9$ symbols/second). The normalized dispersion index $\xi$ is dimensionless and characterizes the ISI. The quadratic dependence of $\xi$ on the symbol rate $R_b$, implies that at a fixed fiber distance, the ISI increases rapidly with respect to the symbol rate.

Similar to other communication channels, the memory length of a channel is defined as the number of neighboring samples

TABLE I
RELATIONSHIP BETWEEN $\xi$, THE MEMORY LENGTH, AND THE CORRESPONDING LENGTH OF THE FIBER FOR A CHANNEL WITH THE GROUP VELOCITY DISPERSION $D = 17$ PS/(NM.KM}) AND SYMBOL RATE $R_b = 10.7$ GSYM/S

| $\xi$ | 0.5 | 0.85 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|
| Memory Length | 2 | 3 | 4 | 5 | 6 |
| $L$ (km) | 103 | 175 | 206 | 308 | 411 |

affected by the interference. We define the memory length as the width of the infinite impulse response that contains 95% of the energy of the transmitted signal. The memory length is then a function of $\xi$. Table I provides a relationship between $\xi$, the memory length, and the corresponding length of a standard fiber for a given symbol rate.

In practical systems, $h_e(t)$ is usually chosen to be a lowpass filter. Here, we assume that it is a rectangular filter given by

$$h_e(t) = \begin{cases} 1, & -T_b < t < 0 \\ 0, & \text{Otherwise.} \end{cases}$$

The sensed electrical signal is convolved with this filter function and then sampled to produce

$$\begin{aligned} r_k &= r((k+\Delta)T_b) \\ &= \int_{(k+\Delta)T_b}^{(k+\Delta+1)T_b} |[s_t(t) \star \ell(t) + n_o(t)] \star h(t)|^2 dt \end{aligned} \quad (2)$$

where $k$ is an integer that represents the detected symbol and $-1 < \Delta < 1$ is the shift of sampling point with respect to the start of the symbol. This normalized range corresponds to the width of the convolution of the sensed electrical signal of approximate width $T_b$ with the rectangular filter function that also has a width $T_b$.

To produce a realistic channel model, we extracted the transmitted pulse shape, $(p_t(t))$, and optical filter, $(h(t))$, from the experimental setup in [29]. The pulse shape $p_t(t)$ is assumed to be raised cosine, and $h(t)$ is a Gaussian filter with time-bandwidth product $B_{3\,dB}.T_b = 3$, where $B_{3\,dB}$ is the 3-dB bandwidth of the filter. Fig. 2 shows the simulated back-to-back eye diagram of the signal $r(t)$ for this system. The eye drawn in 2(a) corresponds to what would be seen on an oscilloscope for a typical system without the presence of the electrical filter, i.e., point $\mathbf{A}$ in the Fig. 1. The eye in Fig. 2(b) is the eye after the electrical filter which is used in the detection process, i.e., point $\mathbf{B}$ in the Fig. 1. The sampling point plays a vital role in the performance of this system [30]. The dependence of the performance of a system on the sampling shift, $\Delta$, is investigated in Section IV.

### *B. Optimal Detection: Maximum Likelihood Sequence Estimation*

For an ISI channel, the Viterbi algorithm (VA) is the optimal maximum likelihood sequence estimation (MLSE) receiver when the noise is memoryless and the number of trellis states is large enough to span the memory. The optimal branch metrics of the corresponding trellis are calculated based on the statistics of the channel which is the negative logarithm of the conditional probability distribution function (pdf) of the sampled output of the channel, $r_k$, given a known transmission sequence, $\mathbf{d}$ [31]

$$-\log f_{r_k|\mathbf{d}}(r). \quad (3)$$

For our channel model with an optical noise-reducing filter before the sensor and a rectangular electrical filter at the receiver, (3) does not have a simple closed form. In [12], [32], and the references therein, the channel is modeled as the sum of parallel chi square channels.

The complexity of the VA grows exponentially in the memory length and thus for long-haul channels with long memory lengths, it is prohibitively complex for high data rate transmission. A practical approach is to use a "reduced-state" VA with the memory length of the corresponding trellis set to be less than the memory length of the channel. In contrast to a "full-complexity" VA, the noiseless outputs for this reduced-state trellis corresponding to each outgoing branch from a state are not identical. The optimal detector for this reduced-state trellis uses branch metrics that are the logarithm of the average of the pdf over all the possible noiseless outputs [33]. An example of the implementation of a reduced-state VA is presented in [14]. We implement our reduced state VA including coding using 8 or less states. This choice is based on the fact that 4-state systems are realizable by state-of-the-art chips [34], and 8-state systems represent practical next generation systems.

Fig. 3 shows the result of the simulated channel model described in this section using a 3-bit analog to digital converter (ADC) and a 4-state Viterbi algorithm. The curves are drawn for two choices of electrical filters: a system without electrical filter, and a system with a rectangular filter. The plots are compared to the results presented in [29]. The simulation results are derived by counting at least 1000 errors at the output. In these plots, it can be seen that the outputs of the simulation are typically within 2 dB of the experimental data[3]. The validation of the channel model to the experimental data with no free parameters to within 2 dB gives us confidence that the simulated results of the effect of coding will be accurate. The results presented in this paper are based on this model and an 8-bit ADC to isolate the effect of coding from the effect of quantizer. The effect of the output rate of quantizers on the performance of a system using MLSE at the receiver has been analyzed in [36].

## III. DISTANCE ENHANCING CONSTRAINED CODES

In this section, we characterize error events caused at the output of the VA, analyze them, and based on the dominant error patterns, we design constrained modulation codes. As described in Section I, in order to design these codes, the most probable errors that occur when using a MLSE receiver must be identified. This process is called error event characterization.

In many linear communication channels, a concatenation of a partial-response (PR) equalizer and a sequence detector is used in the receiver. The first block equalizes the channel to a known channel which is not necessarily a channel with all the ISI removed, and thus is called a PR equalizer. The second block is then designed based on the known statistics of the equalized channel. Most of the analysis and code designs in the literature are based on known PR channel models. The main purpose of

[3]The results presented in [29] were derived using a 0.2 nm OSA filter, while the paper stated that it used 0.1 nm filter [35]. Therefore, the curves taken from [29] are shifted by 3 dB to account for this difference.

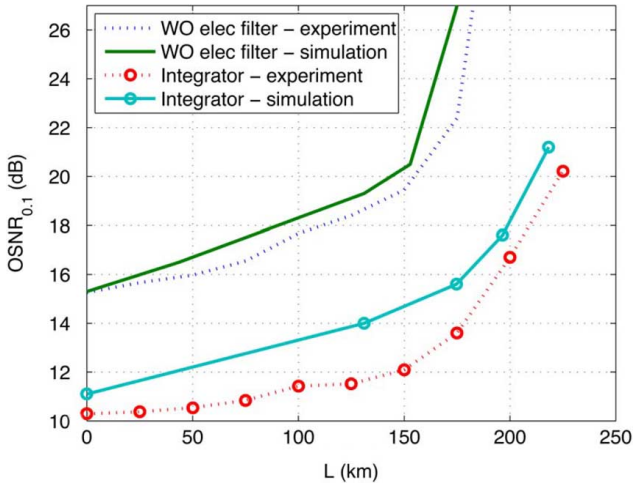



Fig. 3. Comparison of simulation results based on the channel model in (2) and the experimental data presented at [29], using 4-state Viterbi algorithm and 3-bit ADC.

the equalization process is to have a deterministic output regardless of changes within the channel. The MLSE process can then be optimized for this deterministic channel. However, for a nonlinear channel this is not practical because no known equalizer can completely equalize the channel to a known channel model.[4] Therefore, we investigate the properties of MLSE on unequalized channels.

### A. Error Characterization

For linear AWGN channels, there is an algorithm to characterize the error events, i.e., determining the pairs of input sequences resulting in a specific distance at the channel output [37]. This algorithm works in cases where the probability of an error sequence is a function of the absolute difference of the transmitted, $\mathbf{d}$, and detected, $\mathbf{d}'$, sequences, i.e., $|\mathbf{d}-\mathbf{d}'|$. For this class of channels, the Euclidean distance of the output sequences is a suitable metric for determining the most probable error events and is a function of only the difference $|\mathbf{d}-\mathbf{d}'|$. However, for the nonlinear optical channel model, the distance metric defined in (3) does not have a closed-form expression and is dependent on the ordered pair $(\mathbf{d},\mathbf{d}')$ and not the difference $|\mathbf{d}-\mathbf{d}'|$. Specifically, the probability of sending $\mathbf{d}$ and detecting $\mathbf{d}'$ is not equal to the reverse occurrence. Therefore, the standard algorithm that is applicable to linear channels cannot be applied to our channel. The most straightforward approach for finding dominant error events is then a full scale numerical search.

Typical results of a numerical search for the most probable error events for the channel model in (2) are shown in Table II for $\xi = 0.85$ at $\mathrm{BER} = 10^{-3}$. The choice of $\mathrm{BER} = 10^{-3}$ is consistent with the the ECC threshold for fiber optic communication systems according to the ITU guidelines [38]. In this table, each row represents two sequences of bits, their difference, and the probability of sending one sequence and detecting the other. In the column showing the difference "$n$" represents no change, "+" represents transmitting a zero and detecting a 1, and "-" represents transmitting a 1 and detecting a zero. Examining the rows in the table, the probability of error events with equal differences are not necessarily equal. For example, there are 32 error events with the difference sequences $nn{+}nn$ or $nn{-}nn$. Among these, the error event in Row 2 $(\mathbf{d}\rightarrow\mathbf{d}') = (01010 \rightarrow 01110)$ has the probability of $4.3\times10^{-5}(=10^{-4.37})$, while 22 of 32 have probabilities less than $10^{-5}$. The other 10 error events are shown in bold in Table II. In this table, the error events with lengths greater than 11 or with the $\mathrm{BER} < 10^{-5}$ are not listed. The sum of all these unlisted error events yields a $\mathrm{BER} < 4\times10^{-4}(=10^{-3.4})$.

As described in Section II-, the memory length of the channel is monotonically increasing with $\xi$. Therefore, the behavior of the error events changes as $\xi$ changes, which occurs if either the fiber length or data rate change. In Table III, probable error events are listed for four values of $\xi$. To avoid long lists, the error events are grouped based on their difference. Therefore, all error events with differences $\pm(\delta)$ are grouped under one error event class $\delta$, since they contain same pairs of sequences. For example, all error events with the differences $\delta = nn{+}{-}{+}nn$ or $-\delta = -(nn{+}{-}{-}{+}nn) = nn{-}{+}{-}{-}nn$ are grouped under one error event class $\delta = nn+--+nn$. We again note that because the channel is nonlinear, all errors in one class are not equiprobable. The error probability for each class in this table is the total bit error probability caused by the corresponding error events. The classes which have a total probability of less than $10^{-5}$ are also not listed. The grouping of error events into classes implies that some classes listed in Table III are not present in Table II,

TABLE II
LIST OF DOMINANT ERROR EVENTS GENERATED BY NUMERICAL SEARCH FOR $\xi = 0.85, \mathrm{OSNR} = 13.9, \Delta = 0.5$, AND $\mathrm{BER} = 10^{-3}$. THE SYMBOL "$n$" REPRESENTS NO CHANGE, "+" REPRESENTS TRANSMITTING A ZERO AND DETECTING A 1, AND "−" REPRESENTS TRANSMITTING A 1 AND DETECTING A ZERO. $P = -\log_{10}(\mathrm{BER\ of\ the\ Error\ Event})$

| | $\mathbf{d}$ | $\mathbf{d}'$ | Difference | $P$ |
|---|---|---|---|---|
| 1 | 100100 | 101000 | $nn+-nn$ | 4.34 |
| **2** | 01010 | 01110 | $nn+nn$ | 4.37 |
| 3 | 1001001 | 1010101 | $nn+-+nn$ | 4.38 |
| **4** | 01111 | 01011 | $nn-nn$ | 4.45 |
| **5** | 11010 | 11110 | $nn+nn$ | 4.50 |
| **6** | 11111 | 11011 | $nn-nn$ | 4.58 |
| 7 | 1001010 | 1010110 | $nn+-+nn$ | 4.63 |
| **8** | 01001 | 01101 | $nn+nn$ | 4.66 |
| 9 | 100101001 | 101010101 | $nn+-+-+nn$ | 4.66 |
| **10** | 10010 | 10110 | $nn+nn$ | 4.68 |
| 11 | 100111 | 101011 | $nn+-nn$ | 4.69 |
| 12 | 000100 | 001000 | $nn+-nn$ | 4.69 |
| **13** | 01100 | 01000 | $nn-nn$ | 4.71 |
| 14 | 0001001 | 0010101 | $nn+-+nn$ | 4.72 |
| 15 | 0101001 | 0110101 | $nn+-+nn$ | 4.75 |
| 16 | 100101 | 101001 | $nn+-nn$ | 4.76 |
| 17 | 1101001 | 1110101 | $nn+-+nn$ | 4.76 |
| **18** | 01110 | 01010 | $nn-nn$ | 4.82 |
| 19 | 10010100 | 10101000 | $nn+-+-nn$ | 4.83 |
| 20 | 101001 | 100101 | $nn-+nn$ | 4.85 |
| **21** | 11001 | 11101 | $nn+nn$ | 4.91 |
| 22 | 110100 | 111000 | $nn+-nn$ | 4.93 |
| 23 | 110111 | 111011 | $nn+-nn$ | 4.95 |
| **24** | 11110 | 11010 | $nn-nn$ | 4.97 |
| 25 | 1101010 | 1110110 | $nn+-+nn$ | 4.96 |
| | $P>5$ or error length>11 | | Tot. Prob. $< 10^{-3.4}$ | |

[4]An equalizer can be used for a second purpose which is concentrating the energy of the signal spread by ISI to reduce the number of states for sequence estimation. In this case, the output of the equalizer is close to a known channel with a short memory length and the MLSE must tolerate some mismatch.

TABLE III
LIST OF DOMINANT ERROR EVENTS GENERATED BY NUMERICAL SEARCH FOR: $\xi = 0.43, \mathrm{OSNR} = 12.8, \Delta = 0.875; \xi = 0.85, \mathrm{OSNR} = 13.9, \Delta = 0.5; \xi = 1.49, \mathrm{OSNR} = 16.8, \Delta = 0; \xi = 1.76, \mathrm{OSNR} = 28.2, \Delta = 0$. THE TARGET $\mathrm{BER} = 10^{-3}$ FOR ALL CHANNELS AND $P = -\log_{10}(\mathrm{BER\ of\ the\ Error\ Event\ Class})$

| $\xi$ | | Error Event Class | $P$ |
|---|---|---|---|
| 0.43 | 1 | $nn+nn$ | 3.19 |
| | 2 | $nn+-nn$ | 3.75 |
| | 3 | $nn+-+nn$ | 4.16 |
| | 4 | $nn+-+-nn$ | 4.55 |
| | 5 | $nn+-+-+nn$ | 4.95 |
| | | $P$>5 or error length>11 | Tot. Prob. = $10^{-4.75}$ |
| 0.85 | 1 | $nn+nn$ | 3.52 |
| | 2 | $nn+-nn$ | 3.65 |
| | 3 | $nn+-+nn$ | 3.75 |
| | 4 | $nn+-+-nn$ | 4.06 |
| | 5 | $nn+-+-+nn$ | 4.14 |
| | 6 | $nn+-+-+-nn$ | 4.42 |
| | 7 | $nn+-+-+-+nn$ | 4.54 |
| | | $P$>5 or error length>11 | Tot. Prob. = $10^{-4.45}$ |
| 1.49 | 1 | $nn+nn$ | 3.64 |
| | 2 | $nn+-nn$ | 4.01 |
| | 3 | $nn+-n+-nn$ | 4.07 |
| | 4 | $nn+n-+nn$ | 4.17 |
| | 5 | $nn+-n+nn$ | 4.26 |
| | 6 | $nn+-n+-n+nn$ | 4.42 |
| | 7 | $nn+-n+-n+-nn$ | 4.48 |
| | 8 | $nn+n-+n-+nn$ | 4.54 |
| | 9 | $nn+-+n-+nn$ | 4.6 |
| | 10 | $nn+-n+n-+nn$ | 4.74 |
| | 11 | $nn+-n+-+nn$ | 4.75 |
| | 12 | $nn+-+nn$ | 4.91 |
| | | $P$>5 or error length>12 | Tot. Prob. = $10^{-3.49}$ |
| 1.76 | 1 | $nn+-nn$ | 3.07 |
| | 2 | $nn+-n+-nn$ | 3.93 |
| | 3 | $nn+-n-+nn$ | 4.65 |
| | | $P$>5 or error length>11 | Tot. Prob. = $10^{-5.6}$ |

because the BER of the total class is greater than $10^{-5}$, while none of the individual error events within the class is greater than $10^{-5}$.

For small values of $\xi$, shorter error events have a higher probability of occurrence. For these cases, the general error pattern is alternating plus and minus signs. This indicates that most of the sequences causing errors contain at least one of the patterns 101 or 010 corresponding to an isolated mark or space. These patterns are similar to the error patterns seen in the linear $(1+D)$ PR channel [37].

For higher values of $\xi$, the next most common patterns, besides alternating plus and minus are $-+n-$, $+-n+$, $-n+-$, or $+n-+$. These patterns correspond to sequences containing the patterns 1001 or 0110 which are isolated double-mark or double-space.

### B. Forbidden List of Patterns

There is no unique set of forbidden patterns to reduce the BER to a desired level. However, some constraints will limit the selection process. The main issue is the capacity of the constraint, $\mathcal{C}$, which is an upper limit for the rate of the code, $\rho$, designed for the constraint [39]. Code rate is the amount of information sent for each transmitted binary symbol. The capacity of a finite state constraint can be calculated by the corresponding state diagram [39], [19]. Assume that the state diagram of a constraint has $N$ states. The adjacency matrix of this diagram is $A = [a_{ij}]$, $i, j = 1 \cdots N$, where $a_{ij}$ is the number of edges (transitions) starting at state $i$ and ending at state $j$. The capacity of the constraint is equal to the $\mathcal{C} = \log_2 \lambda$, where $\lambda$ is the maximum eigenvalue of $A$.

TABLE IV
FORBIDDEN LISTS OF PATTERNS AND THE CORRESPONDING CODING SCHEMES USED FOR OPTICAL CHANNEL. NOTATION—SUPERSCRIPTS $e$ AND $o$: EVEN AND ODD POSITIONS IN THE SEQUENCE, RESPECTIVELY; BS: BIT-STUFFING; LT: LOOK-UP TABLE

| Code Name | Memory Length | Forbidden Patterns | Coding Rule | Constraint Capacity $\mathcal{C}$ | Code Rate $\rho$ |
|---|---|---|---|---|---|
| *A* | 2 | $\begin{matrix}101\\010\end{matrix}$ | BS | 0.69 | 0.66 |
| *B* | 2 | $\begin{matrix}010^e\\101^e\end{matrix}$ | LT | 0.792 | 0.75 |
| *C* | 2 | $\begin{matrix}101^e\\010^o\end{matrix}$ | BS | 0.879 | 0.856 |
| *D* | 3 | $\begin{matrix}0101^e\\1010^e\end{matrix}$ | LT | 0.916 | 0.89 |

Another concern in the selection of a forbidden list is the size of the corresponding state diagram. To avoid receiving forbidden patterns, the VA at the receiver should use a trellis based on the state diagram of the constraint. Therefore, restrictions in the number of states are translated to restrictions on the number and length of the forbidden patterns. For example, for a receiver with memory length 2, i.e., a 4-state trellis, the receiver can not process a received signal sequence longer than 3 symbols. Therefore, for this trellis, the maximum length of forbidden patterns can not exceed 3.

In this paper, we use four different forbidden lists, shown in Table IV, for the optical channel in (2). Constraints $A$, $B$, and $C$ are based on a 4-state VA and support a channel with memory length 2. These constraints have capacities of 0.69, 0.79, and 0.879, respectively. Constraint $A$ forbids patterns 101 and 010 from the whole sequence, while constraint $B$ forbids these patterns starting at even positions in the sequence. For constraint $C$, the pattern 101 starting at even positions and 010 starting at odd positions are forbidden. Constraint $D$ with capacity 0.916 is designed for VA trellises supporting a memory length of 3, i.e., using eight states, and forbids the patterns 1010 and 0101 at the even positions of the sequence.

### C. Method of Coding

For a given constraint, there are different encoding methods that map the information sequences into coded sequences. The coding rate is upper-bounded by the capacity imposed by the constraint. For systems working at high rates, e.g., optical systems, simplicity is a critical issue in designing the encoders and decoders. Block codes using look-up tables for encoding and decoding are relatively simple codes. However, finding high rate block codes based on a constraint can be difficult. Bit stuffing [40], [41] is another method of code mapping, which produces a coding rate close to the capacity of the constraint. The main drawback of this method is that it produces variable-length coded sequences, which causes error propagation at the output of the decoder and some practical problems in the implementation of the code. Solutions exist to overcome these problems by producing fixed rate bit stuffing encoders [42], or

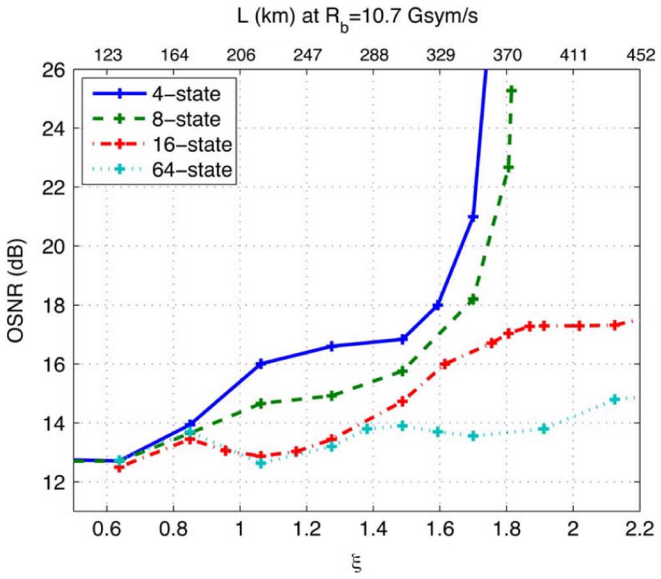



Fig. 4. Required OSNR for 4-, 8-, 16-, and 64-state Viterbi algorithm at the receiver to reach a BER of $10^{-3}$. The x axis is relabeled at the top of the plot by the length of the corresponding fiber at the rate of $R_b = 10.7$ Gsym/s using $D = 17$ ps/(nm.km). The simulations are run for the points and the lines connecting the points are only drawn to aid in viewing.

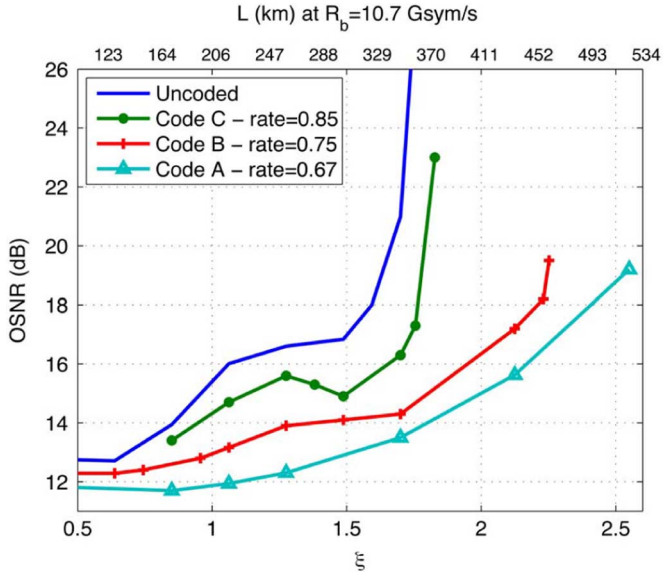



Fig. 5. Comparison of uncoded and coded 4-state Viterbi detectors. The **symbol rate** of all of the systems is equal to 10.7 Gsym/s.

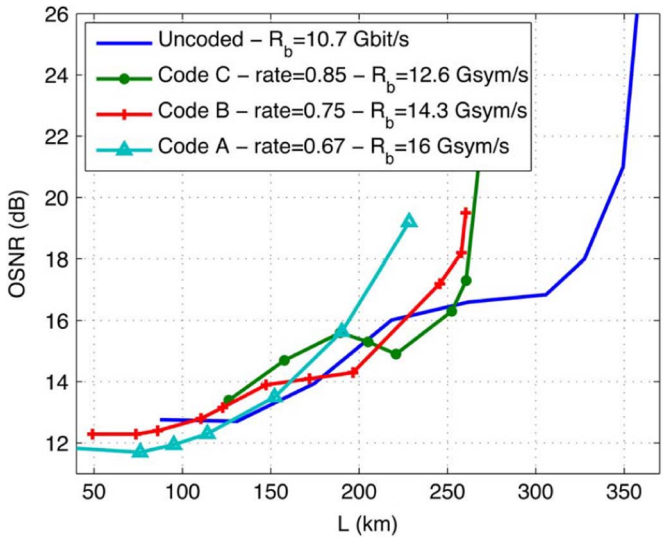



Fig. 6. Comparison of uncoded and coded 4-state Viterbi detectors. The **information rate** of all of the systems is equal to 10.7 Gbit/s.

changing the order of the inner and outer codes for the encoders and decoders, i.e., the modulation code and the error correction code [43]–[45]. However, the cost of all of these solutions is the coding rate. To remove these issues from the analysis, we restrict our results to the error rate before the decoder, as is a standard practice [19]. Noting that not including the decoder can bias the error rates. As an example, if the target is set to be $10^{-3}$ after the decoder, for code $B$ the OSNR increases less than 1 dB. For codes that use bit-stuffing, the decoder causes insertion-deletion errors.

Each of the forbidden lists of patterns illustrated in Table IV used a different coding method. For constraints $B$ and $D$, simple block codes of rates 3/4 and 8/9, respectively, designed in [46], [47], were used. These codes assume a non-return-to-zero inverted (NRZI) format while the transmission in our channel model is NRZ. Therefore, after the encoder, the format of the transmitted signal changes. The NRZ format discrete data sequence $\mathbf{d}^{\mathrm{NRZ}}$can be generated from the NRZI format sequence $\mathbf{d}^{\mathrm{NRZI}}$ using the rule $d_{k+1}^{\mathrm{NRZ}} = d_k^{\mathrm{NRZ}} + d_k^{\mathrm{NRZI}} (\mathrm{mod}\, 2)$. Block codes $B$ and $D$ are designed such that, in addition to distance-enhancing characteristics, they forbid transmitted runs of ones or zeros longer than 7 and 12 for NRZ data for the codes $B$ and $D$, respectively. For codes $A$ and $C$, bit-stuffing is used. The encoders for bit-stuffing add one extra 0 after the pattern 10 where 101 is unwanted and an extra 1 after the pattern 01 where 010 is unwanted. The codes designed for constraints $A$ and $C$ have the average rates of 0.66 and 0.856. For all codes listed in Table IV the probabilities of the bits 0 and 1 at the output of the decoder are equal given they are equiprobable at the input.

## IV. Results

The dominant error events depend on several system parameters including $\xi$, the sampling time $\Delta$, and the number of states used in the VA. In this section, we investigate the dependence of error events on these three parameters based on the channel model described at (Section II-A).

### A. Normalized Dispersion Index, $\xi$

Figs. 4–7 show the behavior of the performance of the VA and codes with respect to the changes in the parameter $\xi$. The graphs show the required optical signal-to-noise ratio (OSNR) to reach a BER of $10^{-3}$ for different values of $\xi$. Fig. 4 is the performance of the uncoded system using either a 4-, 8-, 16-, or 64-state VA. For uncoded systems, the transmitted symbols are information bits and the units Gbit/s and Gsym/s are synonymous. In this figure, the performance of the reduced-state VAs is compared to the full-complexity VA. For $\xi < 2.2$, 95% of the energy of the transmitted symbol is spread within the 6 neighboring symbols. Therefore, the VA with 64 state can be considered as full-complexity VA. As it can be seen, the 8-state

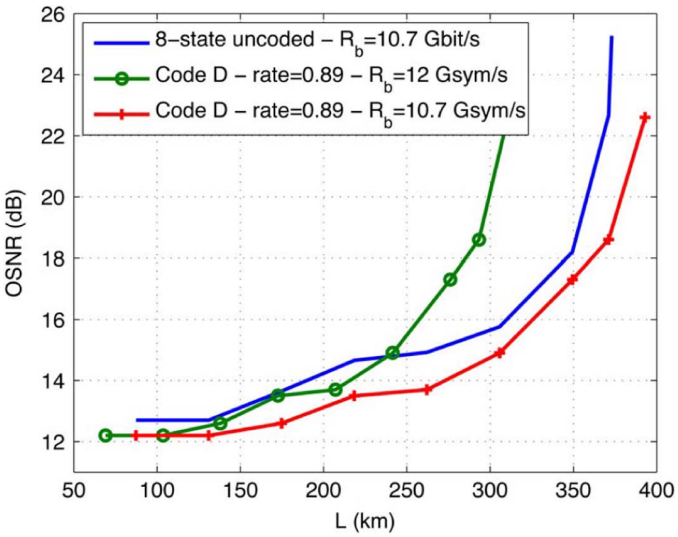



Fig. 7. Comparison of uncoded and coded 8-state Viterbi detectors. The coded systems with $R_b = 10.7$ Gsym/s and $R_b = 12$ Gsym/s have the same symbol rate and information rate, respectively, as the uncoded system.

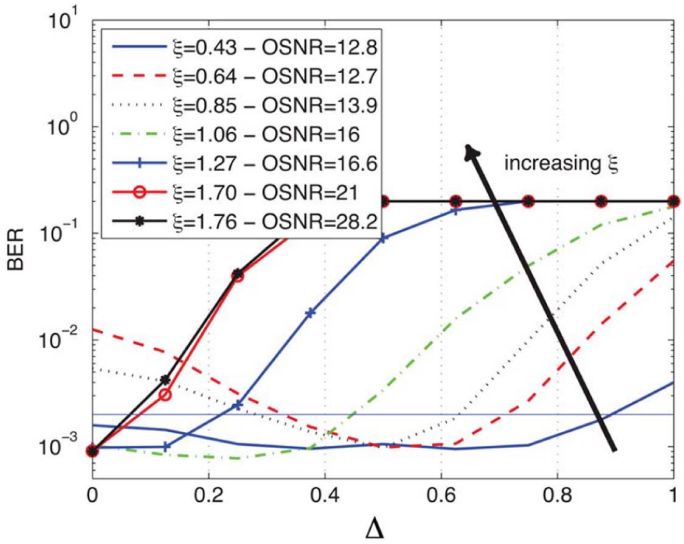



Fig. 8. BER versus sampling shift, $\Delta$, using 4-state Viterbi detector for different values of $\xi$. The thin horizontal line is at BER=$2 \times 10^{-3}$.

receiver does not perform significantly better than the 4-state, while the performance of 16- and 64-state VA at values of $\xi$ higher than 1.8 shows significant improvement. The difference in the performance can be explained by examining the memory length as a function of $\xi$ relative to the memory span for the VA. For $\xi = 1.8$, 87% of the energy is within the 4 neighboring symbol periods while only 77% is within the 3 neighboring symbols. The additional 10% of energy collected by the 16-state VA causes a significant improvement because this energy is not collected by the 8-state system and thus acts as additional interference producing a degradation in performance. For comparison, when $\xi = 0.9$, there is less than 2 dB difference in performance between a 4-, 8-, 16-, and 64-state VA. In this case, the percentage of the collected energy in the neighboring 2, 3, 4, and 6 symbols are 86%, 94%, 98%, and greater than 99% and all four systems operate with approximately the same performance.

For each channel model with given parameters, there is an optimum sampling position, $\Delta$. In Section IV-C, it is shown that the optimum sampling positions vary with changes in the value of $\xi$. Although, increasing $\xi$ increases the memory length and typically would increase the required OSNR, for some ranges of $\xi$ the required OSNR does not increase significantly, or even decreases. This behavior is the result of small shifts in the optimum sampling positions while $\xi$ increases and is consistent with other results presented in the literature (cf. [48, Fig. 15]). Similar effects also occur for coded systems presented next.

Figs. 5 and 6 show the performance of the uncoded and coded systems with a 4-state VA using codes $A$, $B$, and $C$ before the decoder. Fig. 5 shows the comparison based on the assumption that the *symbol rates* are equal for all systems. It can be seen that the codes can produce significant improvement in the required OSNR at a fixed value of $\xi$. For example, this improvement can be as high as 5 dB for code $C$ at $\xi = 1.7$. The improvement in the performance can also be measured as the increase in the achievable fiber distance for a fixed OSNR. For example, using code $C$ at OSNR=16 dB, the fiber distance increases by 53% or 120 km at $R_b = 10.7$ Gsym/s.

While all the curves in Fig. 5 have the same symbol rate, their transmission information rates vary because the code rates are different. To compare systems with equal information rates, the symbol rates of each of the systems are rescaled by the code rate. For example, for the system using code $C$, with a code rate $\rho = 0.85$, the symbol rate is increased to $R_b = 10.7/0.85 = 12.6$ Gsym/s. Fig. 6 rescales each of the curves in Fig. 5 by the corresponding code rate so that the *information rate* is equal for all systems as a function of fiber distance. However, since the symbol rate is now different for each code and $\xi$ depends on the symbol rate, there is no longer a fixed relationship between the fiber distance, $L$, and $\xi$ as there was in Fig. 5. It can be seen that if the system can tolerate the reduction in information rate by using codes, then for a fixed OSNR, significantly longer fiber lengths can be reached. However, if the reduction in information rate can not be tolerated and the system uses a higher symbol rate to achieve the same information rate, then for different fiber lengths, there is a significant reduction in performance. For the best cases, the coded systems show about a 1 dB improvement at specific distances. Code $A$ shows an improvement in the range of fiber lengths less than 123 km, code $B$ between 180 km and 230 km, and code $C$ between 205 km and 250 km.

To study a higher rate constraint, code $D$, in Table IV, with rate 0.89 and longer forbidden patterns was tested. Fig. 7 compares the performance of uncoded and coded 8-state trellis for both fixed symbol and information rates using code $D$. It can be seen that the coded system with the symbol rate of 12 Gsym/s enhances the performance of 8-state trellis up to 0.5 dB for fiber distances less than 245 km. Compared to the best performance of the three 4-state codes, code $D$ performs at most 1 dB better, reaching the best improvement at the fiber distance of 210 km.

### B. Number of States

Another parameter which has a significant effect on the distribution of errors and the effectiveness of the code is the number

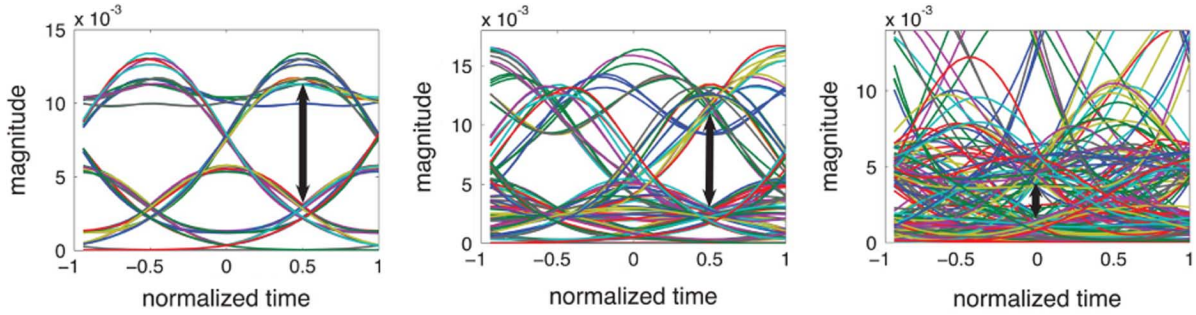



Fig. 9. Eye diagrams of the received signals after electrical filter (point $\mathbf{B}$ in Fig. 1) for systems with: (a) $\xi = 0.43$, (b) $\xi = 0.85$, (c) $\xi = 1.70$. The arrows show widest opening of the eye.

TABLE V
LIST OF DOMINANT ERROR EVENTS FOR $\xi = 1.49$, OSNR $= 14.5$, $\Delta = 0$, AND BER $= 10^{-3}$ USING A 16-STATE VITERBI DETECTOR. $P = -\log_{10}$ (BER OF THE ERROR EVENT CLASS)

| | Error Event Class | $P$ |
|---|---|---|
| 1 | $nnnn + -nnnn$ | 3.39 |
| 2 | $nnnn + -n + -nnnn$ | 3.72 |
| 3 | $nnnn + nnnn$ | 3.8 |
| 4 | $nnnn + n - +nnnn$ | 4.55 |
| 5 | $nnnn + -n + nnnn$ | 4.57 |
| 6 | $nnnn + -n + - + nnnn$ | 4.85 |
| | $P$>5 or error length>14 | Tot. Prob. = $10^{-3.7}$ |

of states of the trellis. Table V lists the dominant error events of a system with $\xi = 1.49$ using a 16-state VA at the receiver. Comparison with the corresponding list in Table III shows that the error events can be grouped in fewer classes for the 16-state VA relative to the 4-state VA. Consider that for $\xi = 1.49$, only 72% of the energy of a symbol can be captured from the two neighboring symbols. Therefore, a 4-state VA treats the remaining 28% of the energy, which is spread over other symbols, as interference. Consequently, because the dominant error events lists are different for systems using different Viterbi receivers, the improvement in the performance of the systems is different. For example, the improvement of the system with a 16-state VA using code $C$ can be seen in the range of the optical distances of 140 km to 205 km with about 1 dB at the best case while the Fig. 6 implies that the same code with the 4-state VA improves the system performance for the fiber distances of 200 km to 250 km.

### C. Optimum Sampling Point

As mentioned in Section II.A, an optimum sampling position, $\Delta_{\text{opt}}$, exists at the receiver to achieve the best performance. The performance demonstrated in Sections III-A and IV are based on the optimum value of $\Delta$. In this section, we determine the sensitivity of the performance as a function of $\Delta$. Fig. 8 plots the BER versus $\Delta$ for several values of $\xi$. The OSNR for each value of $\xi$ is chosen such that the BER at the optimum sampling point $(\Delta_{\text{opt}})$ is equal to $10^{-3}$. In this graph, $\Delta = 0$ represents the sample $r_k$ generated by the integration of the signal over the time period corresponding to the desired symbol, while $\Delta = 0.5$ represents the sample $r_k$ generated by an integration interval that corresponds to half of the desired symbol and half of the next symbol. The performance is, therefore, symmetric about $\Delta = 0$ and the graph is shown for only $0 < \Delta < 1$.

The graph indicates that $\Delta_{\text{opt}}$ is a strong function of $\xi$. The optimum sampling point for lower values of $\xi$ is at the cross-over point of the eye, $\Delta_{\text{opt}} = 0.5$, as shown in 9(a). The fact that the optimal sampling point is at the cross-over point is a direct consequence of the electrical filter which shifts the unfiltered eye shown in Fig. 2 by a half symbol period for small values of $\xi$. As $\xi$ increases, $\Delta_{\text{opt}}$ shifts to the center of the eye, $\Delta_{\text{opt}} = 1$. This effect can be intuitively justified by looking at Fig. 9 which illustrates the eye diagrams for different values of $\xi$ after electrical filtering, i.e., point $\mathbf{B}$ in the Fig. 1. In these eye diagrams, arrows show the widest opening in the eye. Comparing to the results shown Fig. 8, these arrows are exactly located at the optimum sampling points of the corresponding systems. For example, the maximum opening shown in Fig. 9(c), for $\xi = 1.70$ occurs at the normalized time corresponding to $\Delta = 0$, which in Fig. 8, is the sampling position that produce the minimum BER for the system.

We now define the sampling range tolerance, TR, $0 < \text{TR} < 2$, as the continuous range of $\Delta$ for which BER is less than $2 \times 10^{-3}$. Large values of TR indicate more tolerance to sampling errors. For example, for $\text{TR} = 1.5$, the $\text{BER} < 2 \times 10^{-3}$, if the sampling occurs anywhere within 1.5 symbol intervals. The reason the sampling range tolerance can be larger than a symbol interval is again a consequence of the fact that the width of the convolution of the sensed electrical signal over $T_b$ with the rectangular filter function that also has a width $T_b$ is at least $2T_b$. Fig. 8 shows the largest value of the sampling range, $\text{TR} = 1.8$, occurs when the dispersion is low and $\xi = 0.43$. For larger values of $\xi$, TR decreases, as expected, and the BER becomes more sensitive to the sampling point. When $\xi < 1.5$, there are only small changes in the ordering of the most probable error events when the sampling point shifts within the sampling tolerance range defined for $\text{BER} < 2 \times 10^{-3}$. Therefore, for this range of parameters, the constrained codes studied are reasonably robust with respect to the sampling point.

Fig. 10 shows TR for 4-state uncoded and coded systems operating at the same information rate. For these systems, the symbol rates are not equal and thus the timing tolerance is expressed in absolute time units using $\text{TR} \times T_b$. The sampling tolerance improves from approximately 40 ps for the uncoded system to at most 100 ps for coded systems at specific ranges. For example, at $L = 125$ km, using code B produces a two-fold

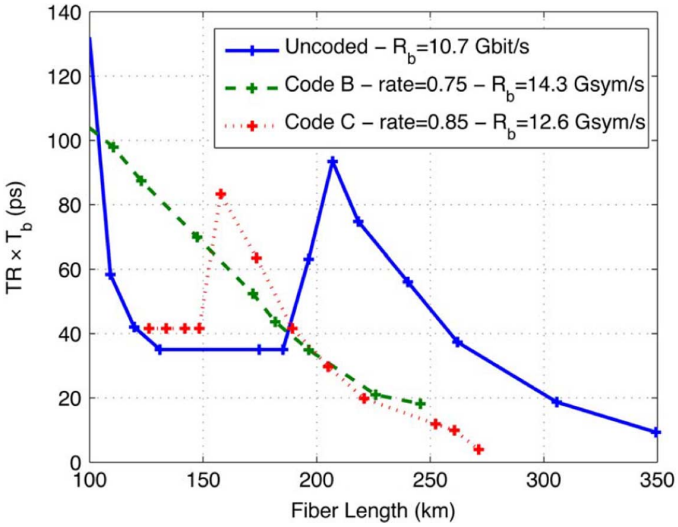



Fig. 10. Comparison of sampling shift time tolerance for 4-state uncoded and coded systems.

improvement increasing $\mathrm{TR} \times T_b$ from 35 ps to 80 ps. If the systems are operated at the same symbol rate, then the improvement is more dramatic. For example, code $B$ increases TR to 2, 0.9, and 0.5 at $\xi = 0.64, 1.27$, and 1.70 respectively. For the uncoded system, the equivalent values determined from Fig. 8 are 0.375, 0.4, and 0.1, respectively. For $\xi = 1.70$, this represents over a five-fold improvement in the sampling tolerance. These results show that, in addition to the distance-enhancing characteristic, the constrained codes can improve the system performance by increasing the robustness of the detector with respect to the sampling errors. Intuitively, the use of these codes opens the eye more horizontally than vertically. This improves the tolerance of the system to sampling error and timing jitter. Timing recovery is an important issue which adds complexity to the receiver [49]–[51]. Our results indicate that using a constrained code can reduce the complexity of the timing recovery process for specific distances.

## V. Conclusion

Detection based on a combination of a reduced-state sequence estimation algorithm and constrained coding can improve both the bit error rate, as well as the robustness of the system to sampling errors for a realistic long-haul fiber channel that has nonlinear ISI. This is a low cost approach to combat ISI for systems of several hundred kilometers. Other approaches may be more applicable at longer distances. The generation of a forbidden list of patterns to incorporate into the code as well as several specific codes was presented. These codes can also incorporate other features such as run-length limits. The improvement depends on the code, the fiber distance, and the symbol rate. For systems with a fixed symbol rate, the coding gain can approach 5 dB relative to the uncoded system for a rate 0.85 code. This improvement can be useful for systems with a fixed symbol rate when the loss of 15% in the information rate is acceptable. In this case, these simple codes can provide longer reach. For fixed information rate systems with varying symbol rates, the quadratic dependence of normalized system memory $\xi$ with respect to the symbol rate reduces the gain to the point where there is only a modest improvement in the error performance over specific distances. However, this improvement is achieved using a very simple encoder and decoder. The improvement with respect to sampling errors is more pronounced. The tolerance to sampling errors can be improved by a factor of 2 for systems operating at the same information rate and over 5 times for systems operating at the same symbol rate. This improves the timing recovery process and reduces the effect of jitter on the BER. The two attributes of improving the BER and increasing the tolerance to sampling errors make these codes attractive candidates for applications when one of these improvements is required. For a specific distance simultaneously achieving both a reduction in error rate and improving timing performance requires further investigation. These codes are used in conjunction with sequence estimation and are compatible with existing electronic compensation technology. We believe this is the first error characterization and analysis of constrained codes for a nonlinear optical ISI channels. This study, also, provides a framework for further investigations in the area of modulation coding for optical channels.

## Acknowledgment

The authors would like to thank Dr. R. I. Killey for useful discussions and his insight on the course of the manuscript preparation.

## References

[1] B. E. A. Saleh and M. I. Irshid, "Coherence and intersymbol interference in digital fiber optic communication systems," *IEEE J. Quantum Electron.*, vol. QE-18, no. 6, pp. 944–51, Jun. 1982.

[2] C. R. S. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, and T. Wuth, "$10 \times 111$ Gb/s, 50 GHz spaced, POLMUX-RZ-DQPSK transmission Over 2375 km employing coherent equalization," in *Proc. C/NFOEC 2007*, Anaheim, CA, p. PDP22.

[3] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel, "Electronic compensation of chromatic dispersion using a digital coherent receiver," *Opt. Express*, vol. 15, no. 5, pp. 2120–6, Mar. 2007.

[4] A. J. Lowery and J. Armstrong, "Orthogonal-frequency-division multiplexing for dispersion compensation of long-haul optical systems," *Opt. Express*, vol. 14, no. 6, pp. 2079–84, Mar. 2006.

[5] I. B. Djordjevic and B. Vasic, "Orthogonal frequency division multiplexing for high-speed optical transmission," *Opt. Express*, vol. 14, no. 9, pp. 3767–75, May 2006.

[6] W. Shieh and C. Athaudage, "Coherent optical orthogonal frequency division multiplexing," *Electron. Lett.*, vol. 42, no. 10, pp. 587–9, May 2006.

[7] S. L. Jansen, I. Morita, N. Takeda, and H. Tanaka, "20-Gb/s OFDM transmission over 4 160-km SSMF enabled by RF-pilot tone phase noise compensation," in *Proc. C/NFOEC 2007*, Anaheim, CA, p. PDP15.

[8] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[9] N. Alic, G. C. Papen, L. B. Milstein, P. H. Siegel, and Y. Fainman, "Performance bounds of maximum likelihood sequence estimation in intensity modulated fiber optic links," presented at the TIWC 2004, Pisa, Italy.

[10] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1439–53, Sep. 1990.

[11] J. H. Winters, R. D. Gitlin, and S. Kasturia, "Reducing the effects of transmission impairments in digital fiber optic systems," *IEEE Commun. Mag.*, vol. 31, no. 6, pp. 68–76, Jun. 1993.

[12] N. Alic, G. C. Papen, R. E. Saperstein, L. B. Milstein, and Y. Fainman, "Signal statistics and maximum likelihood sequence estimation in intensity modulated fiber optic links containing a single optical pre-amplifier," *Opt. Express*, vol. 13, no. 12, pp. 4568–79, Jun. 2005.

[13] O. E. Agazzi, M. R. Hueda, D. E. Crivelli, and H. S. Carrer, "Maximum-likelihood sequence estimation in dispersive optical channels," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 749–763, Feb. 2005.
[14] S. J. Savory, Y. Benlachtar, R. I. Killey, P. Bayvel, G. Bosco, P. Poggiolini, J. Prat, and M. Omella, "IMDD transmission over 1 040 km of standard single-mode fiber at 10 Gbit/s using a one-sample-per-bit reduced-complexity MLSE receiver," in *Proc. C/NFOEC 2007*, Anaheim, CA, p. OThK2.
[15] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*, IEEE 802.3, 2005.
[16] *Fibre Channel Framing and Signaling—3 (FC-FS-3)*, , 2007, Project T11/1861-D Rev 0.30.
[17] *Data Interchange on Read-Only 120 mm Optical Data Disks (CD-ROM)*, Standard ECMA-130, 1996.
[18] *120 mm DVD - Read-Only Disk*, Standard ECMA-267, 2001.
[19] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260–99, Oct. 1998.
[20] N. Kashyap, P. H. Siegel, and A. Vardy, "Coding for the optical channel: The ghost-pulse constraint," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 64–77, Jan. 2006.
[21] I. B. Djordjevic and B. Vasic, "Constrained coding techniques for the suppression of intrachannel nonlinear effects in high-speed optical transmission," *J. Lightw. Technol.*, vol. 24, no. 1, pp. 411–9, Jan. 2006.
[22] V. Pechenkin and F. R. Kschischang, "Constrained coding for quasi-linear optical data transmission systems," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4895–902, Dec. 2006.
[23] N. L. Swenson and J. M. Cioffi, "Sliding-block line codes to increase dispersion-limited distance of optical fiber channels," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 3, pp. 485–498, Apr. 1995.
[24] R. Karabed and P. H. Siegel, "Coding for higher order partial response channels," in *Proc. 1992 SPIE Int. Symp. Voice, Video, and Data Commun.*, Philadelphia, PA, vol. 2605, pp. 115–126.
[25] E. Soljanin, "On-track and off-track distance properties of class 4 partial response channels," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Commun.*, Philadelphia, PA, vol. 2605, pp. 92–102.
[26] R. Karabed, P. H. Siegel, and E. Soljanin, "Constrained coding for binary channels with high intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1777–97, Sep. 1999.
[27] G. P. Agrawal, *Nonlinear Fiber Optics.*, 4th ed. San Diego, CA: Academic Press, 2006.
[28] R. Gaudino, , E. Forestieri, Ed., "Theoretical limits for the dispersion limited optical channel," in *Optical Communication Theory and Techniques*. Berlin, Germany: Springer, 2004, pp. 29–36.
[29] N. Alic, G. C. Papen, S. Radic, and Y. Fainman, "Receiver structure trade-offs in equalized high-speed fiber-optic links," *IEEE Photo. Tech. Let.*, vol. 18, no. 3, pp. 1810–2, Sep. 2006.
[30] A. Faerbert, S. Langenbach, N. Stojanovic, C. Dorschky, T. Kupfer, C. Schulien, J. P. Elbers, H. Wernz, H. Griesser, and C. Glingener, "Performance of a 10.7 Gb/s receiver with digital equalizer using maximum likelihood sequence estimation," in *ECOC 2004*, Stockholm, Sweden, p. Th4.
[31] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 362–378, May 1972.
[32] M. Keskinoz and B. V. K. V. Kumar, "Discrete magnitude-squared channel modeling, equalization, and detection for volume holographic storage channels," *Appl. Optics.*, vol. 43, no. 6, pp. 1368–78, Feb. 2004.
[33] D. E. Crivelli, H. S. Carrer, and M. R. Hueda, "On the performance of reduced-state viterbi receivers in IM/DD optical transmission systems," in *Proc. ECOC 2004*, Stockholm, Sweden, p. We4.P.083.
[34] H. Bae, J. B. Ashbrook, J. Park, N. R. Shanbhag, A. C. Singer, and S. Chopra, "An MLSE receiver for electronic dispersion compensation of OC-192 fiber links," *IEEE J. Solid-Sate Circuits*, vol. 41, no. 11, pp. 2541–2554, Nov. 2006.
[35] N. Alic, Private Communication.
[36] N. Alic, "Information processing for improved performance of optical networks," Ph.D. dissertation, Dept. Elect. and Comp. Eng., University of California, San Diego, CA, 2006.
[37] S. A. Altekar, M. Berggren, B. E. Moision, P. H. Siegel, and J. K. Wolf, "Error-event characterization on partial-response channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 241–247, Jan. 1999.
[38] *Series G: Transmission Systems and Media, Digital Systems and Networks—Digital Sections and Digital Line System—Optical Fibre Submarine Cable Systems—Forward Error Correction for High Bit-Rate DWDM submarine systems*, International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Recommendation G.975.1, 2004.
[39] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379-423–623-656, Jul./Oct. 1948.
[40] P. Lee, "Combined error-correcting/modulation recording codes," Ph.D. dissertation, Dept. Elect. and Comp. Eng., Univ. California, San Diego, CA, 1988.
[41] P. Bender and J. K. Wolf, "A universal algorithm for generating optimal and nearly optimal run-length-limited, charge constrained binary sequences," in *Proc. 1993 IEEE Int. Symp. on Inf. Theory*, San Antonio, TX, p. 6.
[42] Y. Sankarasubramaniam, "New Capacity-approaching codes for run-length-limited channels," Ph.D. dissertation, School of Elec. and Comp. Eng., Georgia Inst. of Tech., Atlanta, CA, 2006.
[43] W. G. Bliss, "Circuitry for performing error correction calculations on baseband encoded data to eliminate error propagation," *IBM Tech. Discl. Bulletin*, vol. 23, pp. 4633–34, 1981.
[44] K. A. S. Immink, "A practical method for approaching the channel capacity of constrained channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1389–99, Sep. 1997.
[45] J. L. Fan and A. R. Calderbank, "A modified concatenated coding scheme with applications to magnetic data storage," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1565–74, Jul. 1998.
[46] B. E. Moision, P. H. Siegel, and E. Soljanin, "Distance-enhancing codes for digital recording," *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 69–74, Jan. 1998.
[47] W. G. Bliss, "An 8/9 rate time-varying trellis code for high density magnetic recording," *IEEE Trans. Magn.*, vol. 33, no. 5, pp. 2746–8, Sep. 1997.
[48] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, "Maximum-likelihood sequence detection with closed-form metrics in OOK optical systems impaired by GVD and PMD," *J. Lightw. Technol.*, vol. 24, no. 8, pp. 3073–87, Aug. 2006.
[49] J. R. Barry, A. Kavcic, S. W. McLaughlin, A. Nayak, and W. Zeng, "Iterative timing recovery," *IEEE Sig. Processing Mag.*, vol. 21, pp. 89–102, Jan. 2004.
[50] W. Chung, W. A. Sethares, and C. R. Johnson, Jr., "Timing phase offset recovery based on dispersion minimization," *IEEE Trans. Sig. Processing*, vol. 53, no. 3, Mar. 2005.
[51] S. U. H. Qureshi, "Timing recovery for equalized partial-response systems," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1326–30, Dec. 1976.

**Zeinab Taghavi** (S'09) received the B.S. and M.S. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran in 2001 and 2003, respectively. She is currently working toward the Ph.D. degree at the University of California at San Diego.

Her research interest includes coding and information theory and optical fiber communication.

**Nikola Alic** (S'00–M'06) received the B.S. degree in optoelectronics from the Belgrade School of Electrical Engineering, Belgrade, Serbia in 1998 and the M.S. and Ph.D. degrees in photonics from the University of California, San Diego, in 2001 and 2006, respectively, for his pioneering work on the role of electronic equalization in optical communications.

In 1998, he served as a Junior Scientist at the "Vinca" Institute of Nuclear Sciences in Belgrade, Serbia, where he worked on solitonic pulse compression in nonlinear fiber arrays. He is currently a Research Scientist at the California Institute of Telecommunications and Information Technologies (CALIT2), University of California San Diego. His research interests include equalization and coding theory in optical communications; high-speed transmission, detection theory, fiber optic parametric amplifiers and all-optical signal processing.

**George Papen** (M'09) received the M.S. degree and the Ph.D. in electrical engineering from the University of Wisconsin, Madison, in 1987 and 1989, respectively.

He joined the Faculty of the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, as an Assistant Professor in 1989, an Associate Professor in 1996, and a Full Professor in May 2001. In July 2002, he accepted a position as a Professor at the University of California,San Diego. His research is in systems applications of optics in computing, communication, and remote sensing. Current research topics include the development of coding techniques to mitigate signal impairments in optical communication systems and for remote sensing applications.