Constrained capacities for faster-than-Nyquist signaling

Rusek, Fredrik; Anderson, John B

Published in:
IEEE Transactions on Information Theory

DOI:
10.1109/TIT.2008.2009832

2009

Constrained Capacities for Faster-than-Nyquist Signaling

FREDRIK RUSEK and JOHN B. ANDERSON
Swedish Strategic Center for High Speed Wireless Communication and
Dept. of Electrical and Information Technology
Lund University
Box 118, SE-221 00 Lund SWEDEN
Email: anderson@it.lth.se; fredrikr@it.lth.se

Abstract

This paper deals with capacity computations of faster-than-Nyquist (FTN) signaling. It shows that the capacity of FTN is higher than the orthogonal pulse linear modulation capacity for all pulse shapes except the sinc. FTN signals can in fact achieve the ultimate capacity for the signal power spectral density (PSD). The paper lower and upper bounds the FTN capacity under the constraint of finite input alphabet. It is often higher than the capacity for comparable orthogonal pulse systems; sometimes it is superior to all forms of orthogonal signaling with the same PSD.

Key Words: *Constrained Capacities, Faster than Nyquist, Intersymbol Interference, Coded Modulation, Bandwidth-Efficient Coding*

1 Introduction

In this paper we consider the problem of transmitting data over a bandlimited additive white Gaussian noise (AWGN) channel by means of a set of signals that are generated by linear modulation. Linear modulation signals have the form $\sum a[n]h(t-nT)$, where we take the $\{a[n]\}$ to be independent and identically distributed (i.i.d.). All such signals can be used, in which case the $\{a[n]\}$ are data, or a codeword set can be formed by using some but not all of the $a[n]$ sequences. These signals have many useful detection properties, especially when $h(t)$ is an orthogonal pulse. Shannon's classical result states that the highest transmission rate over the AWGN channel, its capacity, is

$$W \log_2(1 + P/WN_0) \quad \text{bits/s} \tag{1}$$

when the signals have a rectangular average power spectral density (PSD); here W is the one-sided width of the signal PSD, P is the average power, and $N_0/2$ is the value of the white channel noise PSD. By an application of calculus, the result extends to

$$C = \int_0^W \log_2 \left[1 + \frac{2P}{N_0} |H(f)|^2 \right] df \quad \text{bits/s}, \tag{2}$$

when the signals have PSD $P|H(f)|^2$, where $H(f)$ has unit square integral. The PSD of linear modulation sets is proportional to $|H(f)|^2 = \mathcal{F}\{h(t)\}$. However, such signal sets do not necessarily achieve (2).

On the contrary, it has been known for some time that linear modulation signal sets are usually weaker. In particular, making h orthogonal exacts a penalty. Here are some of the facts. When h is a sinc pulse $(T/\pi t)\sin(\pi t/T)$, the PSD is rectangular, the pulses are orthogonal, and the codeword sets can in principle achieve (1). But the sinc has serious disadvantages in practice. Therefore, smoother orthogonal pulses are used and these introduce excess bandwidth. The detection performance remain unchanged, and so the capacity remains (1), despite the extra bandwidth. That this is ordinarily a loss compared to (2) can be seen as follows. Practical orthogonal pulses obey Nyquist's 1924 symmetry condition, that $|H(f)|^2$ is antisymmetric about the point $(f, |H(f)|^2) = (1/2T, |H(0)|^2/2)$; sinc satisfies this trivially. It is easy to show that the transition from sinc to a non-sinc $H(f)$ with this condition can only *increase* (2). But this higher capacity cannot be achieved by signals based on the non-sinc orthogonal $H(f)$.

The effect becomes worse as power grows—or equivalently, as bits carried per hertz grows. Let $T = 1/2$, which yields a positive sinc bandwidth of 1 Hz. Comparing (1) and (2) for a root raised cosine (root RC) orthogonal pulse with excess bandwidth factor $\alpha = .3$, we get 1 and 1.001 bits/s, respectively, when $P/N_0 = 1$. But when $P/N_0 = 1000$ the capacities are 9.97 and 11.81 bits/s, a difference of 1.84 bits/s. The bit density in this example is roughly that of a good dialup telephone line. To achieve the higher value 11.81 with orthogonal pulses would require 5.5 dB more power. As power diverges, so also does the separation in the two capacities; the ratio of (2) to (1) in fact tends to 1.3, the amount of the excess bandwidth.

The object of this paper is to regain this extra capacity. The method will be faster-than-Nyquist signaling (FTN) and we will explore the information rates associated with it. Originally FTN was not constructed with capacity considerations in mind; the first papers only considered it in conjunction with its minimum Euclidean distance [1], [2], [3]. Recently there has been renewed interest in FTN;

non sinc pulses and receivers are investigated in [4] and non binary FTN is investigated in [5]. All of these papers relate to Euclidean distance. This paper and its precursors [6] and [7] consider FTN capacity. In particular we compare the possible rates of FTN to its main competitor, orthogonal signaling.

An alternate view is that this paper seeks the benefits of excess pulse bandwidth without the capacity shortfall that occurs with orthogonal pulses. A related subject is oversampling. It has been shown that oversampling of certain digital transmission signals can lead to higher information rates [8], [9].

The paper is organized as follows. In Section 2 we give the system model. Capacity comparisons and information rate bounds are found in sections 3 and 4 respectively. Section 5 gives numerical results and comparisons to orthogonal signaling and to the ultimate AWGN capacity for the assumed spectral density.

2 Faster-than-Nyquist Signaling and its Detection

Consider a baseband PAM system based on a real, continuous T-orthogonal pulse $h(t)$. The transmitted *linear modulation* signal is formed by

$$s_{\mathbf{a}}(t) = \sum_{n=0}^{\infty} a[n]h(t - n\tau T), \quad \tau \leq 1 \tag{3}$$

where $a[n]$ is a sequence of zero-mean i.i.d. data symbols drawn from an alphabet $\mathcal{A}$. The signaling rate is $1/\tau T$. We assume $h(t)$ is unit energy, i.e. $\int_{-\infty}^{\infty} |h(t)|^2 dt = 1$. Moreover, T is the smallest time shift such that $h(t)$ is orthogonal; i.e., if $T' < T$, there exists an integer m such that $\int h(t)h(t - mT')dt \neq 0$. Since the pulse is T-orthogonal the signals will not suffer from intersymbol interference (ISI) when $\tau = 1$. This we refer to throughout as *Nyquist* signaling. For $\tau < 1$ we say that we have FTN signaling, and ISI is unavoidable for i.i.d. input symbols.[1]

Signals of the form (3) with zero-mean i.i.d. data symbols have PSD [11]

$$\Phi(f) = \frac{\sigma_a^2}{\tau T}|H(f)|^2 \tag{4}$$

where $\sigma_a^2 = \mathcal{E}\{|a[n]|^2\}$. Note that the FTN with i.i.d. inputs does not change the PSD shape. We consider it fair to compare different systems only if they have the same PSD.

[1] Some type of precoding, for example Tomlinson-Harashima [10], can eliminate the ISI in the receiver, but the i.i.d. assumption on $\{a[n]\}$ is then violated.

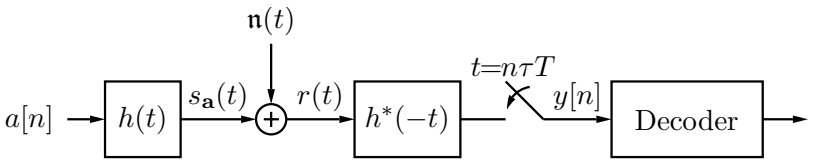

Figure 1: Overall system of faster-than-Nyquist signaling. Consecutive data symbols $a[n]$ are spaced by τT seconds.

The AWGN channel presents to the decoder $r(t) = s_{\mathbf{a}}(t) + \mathfrak{n}(t)$, with $\mathfrak{n}(t)$ real white noise. Forney showed [12] that a set of sufficient statistics to estimate $a[n]$ from the linear modulation signal is the sequence

$$y[n] = \int_{-\infty}^{\infty} r(t) h^*(t - n\tau T) dt \tag{5}$$

Inserting the expression for $r(t)$ into (5) yields the discrete-time model

$$y[n] = \sum_{m=0}^{N} a[m] g[n-m] + \eta[n] \tag{6}$$

where

$$\begin{aligned} g[n-m] &= \int_{-\infty}^{\infty} h(t - n\tau T) h(t - m\tau T) dt \\ \eta[n] &= \int_{-\infty}^{\infty} \mathfrak{n}(t) h(t - n\tau T) dt. \end{aligned}$$

Equation (6) is the so called Ungerboeck observation model [13]. The correlation of the noise sequence $\eta[n]$ is

$$\mathcal{E}\{\eta[n]\eta^*[m]\} = \frac{N_0}{2} g[n-m]. \tag{7}$$

This system model is illustrated in Figure 1. In this paper, $\mathfrak{n}$ and h are real but they need not be in principle.

In matrix notation (6) can be written as $\boldsymbol{y}^N = \boldsymbol{G}^N \boldsymbol{a}^N + \boldsymbol{\eta}^N$. The matrix $\boldsymbol{G}^N$ is a $N \times N$ Toeplitz matrix formed from $\{g[0], g[1], \ldots, g[N]\}$, and $\boldsymbol{a}^N$ is the column vector $\{a[0], \ldots, a[N]\}^{tr}$; later we will use the notation $\boldsymbol{a}_{n_1}^{n_2}$ to denote $\{a[n_1], \ldots, a[n_2]\}^{tr}$.

It can be shown that

$$g[k] = \int_{-\infty}^{\infty} |H(f)|^2 e^{\jmath 2\pi k \tau T f} df. \tag{8}$$

Since $e^{\jmath 2\pi \tau T f}$ is periodic with period $1/\tau T$ it follows that the same time-discrete model as in (6) is obtained if $|H(f)|^2$ is replaced by the *folded* pulse spectrum $|H_{\mathrm{fo}}(f)|^2$, where

$$|H_{\mathrm{fo}}(f)|^2 \triangleq \sum_{k=-\infty}^{\infty} \left| H\Big(f + \frac{k}{\tau T}\Big) \right|^2, \quad -1/2\tau T \le f \le 1/2\tau T \tag{9}$$

and zero otherwise. This means that (8) can alternately be expressed as

$$g[k] = \int_{-1/2\tau T}^{1/2\tau T} |H_{\text{fo}}(f)|^2 e^{\jmath 2\pi k\tau T f} df \tag{10}$$

Since $|H_{\text{fo}}(f)|^2$ implies the same time-discrete model as $|H(f)|^2$ it follows that systems based on the folded pulse shape have equivalent detection performance to systems based on $h(t)$.

Since the noise variables are correlated it is convenient to work with the whitened matched filter (WMF) model instead of the Ungerboeck model. By filtering $y[n]$ with a whitening filter the sequence encountered by the decoder becomes

$$x[n] = \sum_{l=0}^{n} b[n-l]a[l] + w[n] \tag{11}$$

where $b[n]$, the new model, is a causal ISI tap sequence such that $b[n] \star b^*[-n] = g[k]$ and $w[n]$ is independent Gaussian noise with variance $\sigma^2 = g[0]N_0/2 = E_h N_0/2$, where E_h is the energy in $h(t)$. Since the whitening filter is invertible, $x[n]$ also forms a set of sufficient statistics. However, in the case of infinite models $b[n]$, the WMF model can be hard to compute. One example is when $\boldsymbol{g}$ does not have a rational $\mathcal{Z}$-transform. In our case the signals may be bandlimited below $1/2\tau T$ Hz, which prohibits the WMF model. But we will nonetheless make use of it later.

The average power P of an FTN transmission, from (4), equals

$$P = \frac{\sigma_a^2}{\tau T} \tag{12}$$

Finally we note that instead of transmitting faster, an equivalent strategy is to transmit a wider pulse in time. Let the pulse be $h_W(t) = \sqrt{\tau}h(\tau t)$. If the widening factor is $1/\tau$ the pulse would be T'-orthogonal if $T' = T/\tau$. Therefore the same time discrete model is obtained and the systems are equivalent in detection and in bits per Hz.

3 FTN Constrained Capacity: Gaussian Alphabet

In this section the capacity of FTN is investigated. First we review the capacity of general ISI signals, and then we apply the results to FTN. Throughout, we reserve the term capacity for AWGN and signals with PSD $P|H(f)|^2$, without further restrictions. Constrained capacity will mean the maximum information rate under some added restriction, such as FTN or an alphabet assumption. Transmitted symbols are always i.i.d. In this section there are no further alphabet restrictions.

3.1 General ISI Signals

We first consider the capacity of discrete-time signals with ISI. Assume either the WMF model, in which $\boldsymbol{x} = \boldsymbol{b} * \boldsymbol{a} + \boldsymbol{w}$, or the Ungerboeck model, in which $\boldsymbol{y} = \boldsymbol{g} * \boldsymbol{a} + \boldsymbol{\eta}$. Let $p_{\boldsymbol{a}}(\boldsymbol{a}) = \prod p_a(a)$ denote the probability distribution of the data symbols. Then the constrained capacity is

$$\begin{aligned} C_{\mathrm{DT}} &\triangleq \sup_{p_a(a)} \lim_{N\to\infty} \frac{1}{N} I(\mathbf{y}^N; \mathbf{a}^N) = \sup_{p_a(a)} \lim_{N\to\infty} \frac{1}{N} I(\mathbf{x}^N; \mathbf{a}^N) \\ &= \sup_{p_a(a)} \lim_{N\to\infty} \frac{1}{N} [H(\mathbf{x}^N) - H(\mathbf{x}^N | \mathbf{a}^N)] \quad \text{bits/channel use} \end{aligned} \tag{13}$$

From [14, Lemma 1], [15] C_{DT} in (13) is obtained when $a[n]$ has a Gaussian distribution and it is

$$C_{\mathrm{DT}} = \frac{1}{2\pi} \int_0^{\pi} \log_2 \Big[1 + \frac{\sigma_a^2}{\sigma^2} G(\lambda) \Big] d\lambda \tag{14}$$

where

$$G(\lambda) = \sum g[k] e^{i\lambda k} = \Big| \sum b[k] e^{i\lambda k} \Big|^2 = |B(\lambda)|^2 \tag{15}$$

and $B(\lambda)$ is the ISI transfer function. The validity of Eq. (14) for infinite ISI is crucial and is addressed in Appendix C.[2] Higher rates than C_{DT} can be supported by allowing correlated $\{a[n]\}$ [14, 15]; water filling then gives the solution.

The special case of no ISI, i.e. $g[n] = b[n] = \delta[n]$, implies that $G(\lambda) = 1$, $0 \le \lambda \le \pi$ and consequently C_{DT} becomes

$$\frac{1}{2} \log_2 \Big[1 + \frac{\sigma_a^2}{\sigma^2} \Big] \tag{16}$$

3.2 Application to FTN

Now consider FTN signals. The constrained capacity is found by first computing the capacity of the resulting discrete-time channel after the matched filter sampler, and then normalizing by the signaling rate, that is

$$C_{\mathrm{FTN}} \triangleq \frac{1}{\tau T} C_{\mathrm{DT}} \quad \text{bits/s} \tag{17}$$

To find the FTN capacity we must find $G(\lambda)$ for the ISI that it creates. $G(\lambda)$ will be a scaled, folded version of $|H(f)|^2$ around $f = 1/2\tau T$; that is,

$$G(\lambda) = \frac{1}{\tau T} \sum_{k=-\infty}^{\infty} \Big| H\Big(\frac{\lambda}{2\pi\tau T} + \frac{k}{\tau T} \Big) \Big|^2 = \frac{1}{\tau T} \Big| H_{\mathrm{fo}}\Big(\frac{\lambda}{2\pi\tau T} \Big) \Big|^2 \tag{18}$$

[2] The appendix makes use of results from forthcoming sections, but those do not assume that (14) is valid for infinite ISI; hence the method of proof is not circular.

The folded spectrum $|H_{\rm fo}(f)|^2$ becomes

$$|H_{\rm fo}(f)|^2 = \tau T G(f 2\pi\tau T) \tag{19}$$

We can now compute $C_{\rm FTN}$. Take $\sigma_a^2 = P\tau T$; as was the case in (14),

$$\begin{aligned} C_{\rm FTN} &= \frac{1}{2\pi\tau T}\int_0^{\pi} \log_2\Big[1+\frac{P\tau T}{\sigma^2}G(\lambda)\Big]d\lambda \\ &= \frac{1}{2\pi\tau T}\int_0^{\pi} \log_2\Big[1+\frac{P}{\sigma^2}|H_{\rm fo}(\frac{\lambda}{2\pi\tau T})|^2\Big]d\lambda = \int_0^{1/2\tau T} \log_2\Big[1+\frac{2P}{N_0}|H_{\rm fo}(f)|^2\Big]df \end{aligned} \tag{20}$$

since $\sigma^2 = N_o/2$.

However, the capacity of a linear modulation signal of form (3) with $\tau = 1$, i.e. orthogonal signaling, does not depend on $h(t)$; it stems only from the T-orthogonality. $|H_{\rm fo}(f)|^2$ is now T. When the transmitter is constrained to power P watts, the capacity is thus

$$C_{\rm N} = \frac{1}{2T}\log_2\Big[1+\frac{2TP}{N_0}\Big] \qquad \text{bits/s} \tag{21}$$

where the subscript denotes Nyquist. The next theorem states that FTN has constrained capacity superior to Nyquist signaling.

Theorem 1 *Unless $h(t)$ is a sinc pulse there exist τ such that*

$$C_{\rm FTN} > C_{\rm N} \tag{22}$$

For $h(t)$ a sinc, $C_{\rm N} = C_{\rm FTN}$ for all τ.

Proof First assume that $h(t)$ is a sinc pulse. Then Liveris [16] shows that FTN signaling based on the sinc pulse has $C_{\rm FTN} = C_{\rm N}$ for all τ.

Now assume that $h(t)$ is not sinc. Since $h(t)$ is T-orthogonal it follows from the Gibby-Smith ISI condition [17] that

$$\sum_{k=-\infty}^{\infty}\Big|H\Big(f+\frac{k}{T}\Big)\Big|^2 = T, \ \ \forall f \tag{23}$$

By taking $\tau = 1/2$, (18) equals

$$G(\lambda) = \frac{2}{T}\sum_{k=-\infty}^{\infty}\Big|H\Big(\frac{\lambda}{\pi T}+\frac{2k}{T}\Big)\Big|^2 \tag{24}$$

It follows that for $\lambda \le \pi/2$

$$G(\pi-\lambda) = \frac{2}{T}\sum_{k=-\infty}^{\infty}\Big|H\Big(\frac{1}{T}-\frac{\lambda}{\pi T}+\frac{2k}{T}\Big)\Big|^2 = \frac{2}{T}\sum_{k=-\infty}^{\infty}\Big|H\Big(\frac{\lambda}{\pi T}-\frac{2k+1}{T}\Big)\Big|^2 = \frac{2}{T}\sum_{k=-\infty}^{\infty}\Big|H\Big(\frac{\lambda}{\pi T}+\frac{2k+1}{T}\Big)\Big|^2 \tag{25}$$

Combining (23), (24) and (25) gives

$$G(\lambda) + G(\pi - \lambda) = \frac{1}{\tau} = 2 \tag{26}$$

To compute $C_{\rm FTN}$ we compute $C_{\rm DT}$ in (14); σ_a^2 is $P\tau T$, which gives

$$\begin{aligned} C_{\rm FTN} &= \frac{1}{\pi T}\int_0^{\pi/2} \log_2\Big[1 + \frac{P\tau T}{\sigma^2}G(\lambda)\Big]d\lambda + \frac{1}{\pi T}\int_{\pi/2}^{\pi} \log_2\Big[+\frac{P\tau T}{\sigma^2}G(\lambda)\Big]d\lambda \\ &= \frac{1}{\pi T}\int_0^{\pi/2} \log_2\Big[1 + \frac{P\tau T}{\sigma^2}G(\lambda)\Big]d\lambda + \frac{1}{\pi T}\int_0^{\pi/2} \log_2\Big[1 + \frac{P\tau T}{\sigma^2}G(\pi-\lambda)\Big]d\lambda \end{aligned}$$

Using $\log_2(1+x) + \log_2(1+y) > \log_2(1+x+y),\ x, y > 0$ gives

$$\begin{aligned} C_{\rm FTN} &> \frac{1}{\pi T}\int_0^{\pi/2} \log_2\Big[1 + \frac{P\tau T}{\sigma^2}[G(\lambda) + G(\pi-\lambda)]\Big]d\lambda = \frac{1}{\pi T}\int_0^{\pi/2} \log_2\Big[1 + \frac{PT}{\sigma^2}\Big]d\lambda \\ &= \frac{1}{2T}\log_2\Big[1 + \frac{2PT}{N_0}\Big] = C_{\rm N} \qquad \blacksquare \end{aligned}$$

Remark In the proof we took $\tau = 1/2$. This can be relaxed to $\tau = 1/q$, q an integer, in which case (26) becomes

$$\sum_{k=0}^{\lceil \frac{q}{2} \rceil - 1} G\left(\lambda + \frac{2\pi k}{q}\right) + \sum_{k=0}^{\lfloor \frac{q}{2} \rfloor - 1} G\left(\frac{2\pi}{q} - \lambda + \frac{2\pi k}{q}\right) = q \tag{27}$$

with $0 \le \lambda \le \pi/q$.

Interesting pulses from a bandwidth efficiency point of view are the bandlimited pulses. For this class we can prove that values other than $\tau = 1/q$ can be selected while still guaranteeing superior capacity.

Theorem 2 *Assume that the pulse $h(t)$ is bandlimited to W Hz and is not the sinc. Then for $\tau = 1/2WT$*

$$C_{\rm FTN} > C_{\rm N} \tag{28}$$

Moreover this choice of τ is optimal in the sense that $C_{\rm FTN}$ equals the highest possible capacity, C in Eq. (2).

Proof Since $h(t)$ is T-orthogonal it follows that $W > 1/2T$ and $\tau < 1$. Since $\tau = 1/2WT$ there is no aliasing effect and consequently

$$|H_{\rm fo}(f)|^2 = |H(f)|^2, \quad -1/2\tau T \le f \le 1/2\tau T$$

The FTN constrained capacity is computed by (20) and equals

$$C_{\rm FTN} = \int_0^W \log_2\Big[1 + \frac{2P}{N_0}|H(f)|^2\Big]df$$

Because $h(t)$ is T-orthogonal, (23) holds. This implies that for all $f \in [0, 1/2T]$ some power has been removed and distributed to frequencies $f + k/T$, $\forall k$. The computation is illustrated in Figure

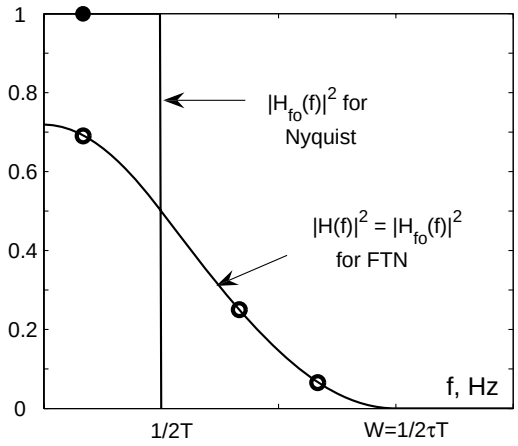

Figure 2: Illustration of Theorem 2. The pulse $h(t)$ is orthogonal under T shifts. When $H(f)$ is folded around $\tau = 1/2WT$, $|H_{\rm fo}(f)|^2$ remains $|H(f)|^2$ in $[-1/2\tau T, 1/2\tau T]$. When $H(f)$ is folded around $1/2T$, $|H_{\rm fo}(f)|^2$ becomes the rectangular block; the three circles sum up to the black dot.

2; orthogonal signaling can only make use of $|H_{\rm fo}(f)|^2$, while FTN makes use of $|H(f)|^2$. The three circles sum to the black dot.

We now use the fact $\sum_k \log_2(1+x_k) > \log_2(1+\sum_k x_k),\ x_k > 0,\ \forall k$. This gives

$$C_{\rm FTN} = \int_0^W \log_2\Big[1+\frac{2P}{N_0}|H(f)|^2\Big]df > \int_0^{1/2T} \log_2\Big[1+\frac{2PT}{N_0}\Big]df = C_{\rm N}$$

But the left side is Eq. (2), implying that $\tau = 1/2WT$ raises the constrained capacity $C_{\rm FTN}$ to the highest possible value for the signal PSD. τ is thus optimal. ■

A reasonable question at this point is whether FTN signals have higher capacity than Nyquist signaling for all τ and all pulses. Write the FTN capacity for a given SNR as a function of τ, $C_{\rm FTN}(\tau)$. For all practical pulses and for many others that we have tested, we find that $dC_{\rm FTN}(\tau)/d\tau < 0,\ 1/2WT < \tau < 1$ for all practical SNRs. But this is not true for every pulse; consider the pulse with

$$|Z(f)|^2 = \begin{cases} 1-a & |f| \le W/2 \\ a & W/2 < |f| \le W \\ 0 & |f| > W \end{cases} \tag{29}$$

where $0 < a < 1$. This pulse is T-orthogonal if T satisfies $T = 1/W$ and there is no $T' < T$ such that the pulse is T'-orthogonal. Take transmission power as P watts. The Nyquist orthogonal capacity equals

$$C_{\rm N} = \frac{W}{2}\log_2\Big(1+\frac{2P}{WN_0}\Big) = \frac{W}{2}\log_2\Big(1+\xi\Big)$$

with $\xi = 2P/WN_0$. Take $\tau = 2/3$; the FTN constrained capacity becomes

$$C_{\rm FTN} = \frac{W}{2}\log_2\Big(1+(1-a)\xi\Big) + \frac{W}{4}\log_2\Big(1+2a\xi\Big)$$

We are interested whether there exist a and ξ such that $C_{\rm FTN} < C_{\rm N}$. By some manipulation it can be shown that if a satisfies

$$a > 1 + \frac{3}{4\xi} - \sqrt{\frac{1}{2\xi} + \frac{9}{16\xi^2}} \tag{30}$$

for any $\xi > 0$ it follows that $C_{\rm FTN} < C_{\rm N}$. Thus FTN is not better than Nyquist signaling for all $\tau < 1$ and all pulses. Observe that the spectrum $|Z(f)|^2$ is rather weird when a satisfies (30); most of the energy lies outside of $W/2$.

From a theoretical point of view it is interesting to characterize pulses whose constrained capacity grows with decreasing τ regardless of the power, but this is a topic for future research. It is not very important from a practical point of view, since in almost all cases the capacity does increase with decreasing τ. For pulses that are not bandlimited, $|H_{\rm fo}(f)|^2 \to |H(f)|^2$ (with respect to the $\|\cdot\|_1$ norm) when $\tau \to 0$. Therefore the constrained capacity $C_{\rm FTN} \to \int_0^\infty \log_2\left(1+\frac{2P}{N_0}|H(f)|^2\right) df > C_{\rm N}$. As in Theorem 2, the full AWGN capacity of signals with PSD shape $|H(f)|^2$ is obtained and the system is optimal.

We close with a statement that as power grows, FTN with nonzero excess bandwidth approaches the sinc bandwidth efficiency. Thus excess bandwidth does not lead to loss.

Theorem 3 *Assume a T-orthogonal pulse $h(t)$ bandlimited to W Hz, with $|H(f)|^2 > 0$, $|f| < W$. $W > 1/2T$, so that there is excess bandwidth. As $P \to \infty$, the bandwidth efficiency of FTN, $C_{\rm FTN}/W$, nonetheless approaches the sinc bandwidth efficiency, i.e.*

$$\frac{C_{\rm FTN}/W}{2TC_{\rm N}} \to 1, \quad P \to \infty \tag{31}$$

Proof Take $\tau \le 1/2WT$, as in Theorem 2, so there is no aliasing and $|H_{\rm fo}(f)|^2 = |H(f)|^2$, $-1/2\tau T \le f \le 1/2\tau T$. Lower bound $C_{\rm FTN}$ in (20) by

$$\begin{aligned} C_{\rm FTN} &\ge \int_0^{W-\epsilon} \log_2\Big[1+\frac{2P}{N_0}|H(f)|^2\Big] df > \int_0^{W-\epsilon} \log_2\Big[\frac{2P}{N_0}|H(f)|^2\Big] df \\ &= \int_0^{W-\epsilon} \log_2 P\, df + \int_0^{W-\epsilon} \log_2\Big[\frac{2}{N_0}|H(f)|^2\Big] df = (W-\epsilon)\log_2 P + Q \end{aligned} \tag{32}$$

for arbitrarily small $\epsilon > 0$. The term Q is finite. Consequently

$$\frac{C_{\rm FTN}/W}{2TC_{\rm N}} > \frac{\frac{Q+(W-\epsilon)\log_2 P}{W}}{2TC_{\rm N}} = \frac{\frac{Q+(W-\epsilon)\log_2 P}{W}}{\log_2(1+2TP/N_0)} \to \frac{W-\epsilon}{W}, \quad \text{as } P \to \infty$$

But ϵ can be arbitrarily small which implies

$$\frac{C_{\rm FTN}/W}{2TC_{\rm N}} \to 1, \quad \text{as } P \to \infty \qquad \blacksquare$$

3.3 Signal Space Representation

Let us investigate why FTN can have rates superior to Nyquist signaling. An orthonormal set of basis functions for orthogonal linear modulation is $\mathcal{B}_{\rm N} = \{h(t-nT),\ -\infty \le n \le \infty\}$. Denote the orthonormal basis for FTN by $\mathcal{B}_{\rm FTN}$. Clearly, $\mathcal{B}_{\rm FTN} \neq \{h(t-n\tau T),\ -\infty \le n \le \infty\}$ because these are not orthogonal. From Theorem 1 we get the following Corollary.

Corollary 1 *Assume a pulse $h(t)$, and let T be the smallest time such that $h(t)$ is T-orthogonal. Let $\mathcal{B}_{\rm N} = \{h(t-nT),\ -\infty \le n \le \infty\}$ and let $\mathcal{B}_{\rm FTN}$ be the set of orthonormal basis functions for FTN signals with $\tau = 1/q$, i.e. signals of the form $s(t) = \sum_{n=-\infty}^{\infty} a[n]h(t-nT/q)$. Then*

$$\mathcal{B}_{\rm N} \subseteq \mathcal{B}_{\rm FTN} \tag{33}$$

with equality in (33) if and only if $h(t)$ is a sinc pulse.

Proof Assume that $h(t)$ is not a sinc pulse; then the Remark to Theorem 1 shows that $C_{\rm FTN} > C_{\rm N}$. Although the FTN and Nyquist signals are constrained to the same transmission power, and Nyquist signaling is free to place power in any way such that the average constrained power requirement is fulfilled, Nyquist signaling can evidently not create the same signals that FTN can and it follows that $\mathcal{B}_{\rm N} \subset \mathcal{B}_{\rm FTN}$. Now assume that $h(t)$ is a sinc pulse. Since the spectrum of $s(t)$ is still limited to $[-1/2T, 1/2T]$ Hz, it follows from the Sampling Theorem that $\mathcal{B}_{\rm N} = \mathcal{B}_{\rm FTN}$. ■

The Corollary states that for other pulse shapes than sinc, some of the energy in an FTN signal is put outside of $\mathcal{B}_{\rm N}$. In other words, the FTN signal cannot be generated by Nyquist signaling with $h(t)$. One set of orthonormal basis functions called offset QAM that can be used for FTN with certain pulses (including root RC) is given in [18]. Another can be found via a standard Gram-Schmidt orthogonalization. Because each G-S basis function produced is a weighted superposition of previous ones, each has spectrum band limited to W, if the original $H(f)$ is so limited. Asymptotically, then, the size of the basis must be less than $2W\mathfrak{T}$, where $\mathfrak{T}$ is the length of the transmission in seconds. For a root RC h with excess bandwidth α for example, the number of dimensions lies asymptotically in the range $[2W\mathfrak{T},\ 2(1+\alpha)W\mathfrak{T}]$, where W is $1/2T$ and the FTN symbol time is τT. This can be expressed as $[\tau N,\ (1+\alpha)\tau N]$, where N is the number of symbols. Since non-sinc FTN spans more dimensions than than Nyquist signaling, it is natural that is should have a higher capacity.

4 FTN Constrained Capacity: Restricted Alphabets

Until now we have studied the capacity of linear modulation and FTN, without restriction on the alphabet $\mathcal{A}$ that drives the codewords. Primarily, comparison has been made to classic capacity

expressions. We now consider a finite $\mathcal{A}$ and a fixed $a[n]$ distribution $p_a(a) = 1/|\mathcal{A}|,\ \forall a \in \mathcal{A}$. The analytical tools will change. There is no maximization over $p_a(a)$ and we will call this type of constrained capacity an information rate. I_h will denote such a rate, in bits/s; the notation thus implies the uniform alphabet (unlike Section 3), a certain FTN τ, and the pulse h. I_h is defined to be $I_{\rm DT}/\tau T$, where

$$I_{\rm DT} \triangleq \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{y}^N; \boldsymbol{a}^N) \ = \ \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{x}^N; \boldsymbol{a}^N) \quad \text{bits/channel use} \tag{34}$$

Here DT again denotes a time-discrete channel. Currently there exists no closed-form expression for $I_{\rm DT}$. Instead, lower and upper bounds or accurate Monter Carlo based methods must be used. This paper does not improve much on these methods; instead they are applied to FTN. One improvement does appear in Appendix B.

The fact that the FTN system has a time discrete model of infinite length is a problem. If one wishes to base the constrained capacity computations on the WMF model, it can be difficult to find the model for an arbitrary $\boldsymbol{g}$. So the first step is to find a signal system with simple WMF model that has information rate related to the FTN system, and this is done in section 4.1. One must then estimate the information rate of the new system. This is done in two ways; in section 4.2 we lower bound the rate and in 4.3 we describe a simulation-based method to find the bound.

4.1 Finite Systems Related to FTN

Here we find two different pulses $h_L(t)$ and $h_U(t)$ that have finite, preferably short, WMF models, and that have information rates $I_{h_L} \le I_h \le I_{h_U}$. The latter is accomplished if $|H_L(f)|^2 \le |H(f)|^2 \le |H_U(f)|^2,\ \forall f$. To show that, we write $|H(f)|^2 = |H_L(f)|^2 + |H_1(f)|^2$. The optimal decoder works with the autocorrelation $\boldsymbol{g}_h$, where the subscript means that $\boldsymbol{g}$ is obtained from the sampled autocorrelation of $h(t)$. But this can be expressed as $\boldsymbol{g}_h = \boldsymbol{g}_{h_L} + \boldsymbol{g}_{h_1}$ due to (8). Now, construct a pulse $h_2(t)$ such that

$$|H_2(f)|^2 = \frac{1}{2}|H_1(f-\delta)|^2 + \frac{1}{2}|H_1(f+\delta)|^2 \tag{35}$$

where $\delta = s/\tau T$ and s is an integer chosen large enough so that $|H_2(f)|^2$ and $|H(f)|^2$ are non overlapping. If $|H(f)|^2$ is not bandlimited, then replace $|H(f)|^2$ with $|H_{\rm fo}(f)|^2$, which is bandlimited to $1/2\tau T$ Hz and gives the same performance. The construction of $|H_2(f)|^2$ in (35) implies that $\boldsymbol{g}_{h_2} = \boldsymbol{g}_{h_1}$. This means that systems based on pulse $h_L(t)+h_2(t)$ have detection performance identical to systems based on $h(t)$. But a suboptimal strategy for the decoder is to match the receiver filter to

$h_L(t)$ instead of $h_L(t) + h_2(t)$. Since $h_L(t)$ and $h_2(t)$ are non overlapping in frequency, the resulting time-discrete model of this suboptimal receiver strategy is the same as for an optimal strategy for systems based on $h_L(t)$. This proves that $I_{h_L} \leq I_h$; the same argument on the pair I_h,I_{h_U} gives $I_h \leq I_{h_U}$.

To control the tightness of the bound we solve an optimization problem. We express $h_L(t)$ as

$$h_L(t) = \sum_{n=0}^{K-1} b_L[n]\psi(t - n\tau T) \tag{36}$$

and similarly for $h_U(t)$, where the pulse $\psi(t)$ is a τT-orthogonal sinc pulse. Let $\tilde{h}(t;\boldsymbol{b})$ be a pulse constructed from $\boldsymbol{b}$ by $\sum b[n]\psi(t - n\tau T)$. The optimization is

$$\begin{aligned} \boldsymbol{b}_L &= \arg\min_{\boldsymbol{b}} \int_0^{1/2\tau T} |H_{\text{fo}}(f)|^2 - |\tilde{H}(f;\boldsymbol{b})|^2 df \\ \text{s.t.} \quad & |H_{\text{fo}}(f)|^2 \geq |\tilde{H}(f;\boldsymbol{b})|^2, \quad \forall f \end{aligned} \tag{37}$$

Thus $h_L(t)$ has less information rate than $h(t)$ has, but it has finite WMF model (with K taps). Clearly $h_L(t)$ has energy E_{h_L} lower than the unit-energy $h(t)$; the resulting noise in model (11) has lower variance than the noise for pulse $h(t)$. The optimization (37) is linear in the autocorrelation of $\boldsymbol{b}$ and can thus easily be solved by a linear program (LP); the formulation is given in Appendix A. The tightness of the bounds can be controlled by optimizing over enough taps so that E_{h_L} is close to unity.

Slightly poorer bounds can be designed by much simpler methods than the LP; a $\boldsymbol{b}_U$ can be obtained as follows. Define the sequence

$$e[n] = \begin{cases} g[n], & |n| < K \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

Unfortunately, the sequence $\boldsymbol{e}$ is not a valid autocorrelation sequence and in fact its transfer function $E(\lambda)$ is not always larger than $G(\lambda)$. To fix this, define

$$\theta \triangleq \min\left(\min_{0\leq\lambda\leq 1}[E(\lambda) - G(\lambda)], 0\right)$$

and also

$$b_U[n] = \begin{cases} e[n] + |\theta|, & |n| = 0 \\ e[n], & 1 \leq |n| < K \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

The $\boldsymbol{b}_U$ so constructed is a valid autocorrelation sequence (a necessary and sufficient condition for a sequence to be a valid autocorrelation is Eq. (58) in Appendix A) and it holds that $|H_U(f)| \geq |H(f)|$.

The above LP method (37) to obtain finite pulses that give tight lower and upper bounds works well in general but the lower bound unfortunately fails for some important cases. Assume a pulse $h(t)$ is bandlimited to W Hz and let $\tau < 1/2WT$. Then $H_{\rm fo}(f) = 0,\ 1/2T \le f \le 1/2\tau T$. Consequently there must be equality in the constraint in (37) and $\boldsymbol{b}_L$ has to synthesize an ideal low pass filter. But this is not possible for K taps and the method fails.

This we solve by using an $h_U(t)$ that always exists. First express $|H_U(f)|^2$ as $|H_U(f)|^2 = |H(f)|^2 + |H_e(f)|^2$. As above, use a frequency shifted $|H_e(f)|^2$ such that $|H_e(f)|^2$ and $|H(f)|^2$ are non overlapping. If a signaling scheme is based on pulse $h_U(t)$, then this generates, after sampling, the vector $\boldsymbol{y}_U^N$. However, $\boldsymbol{y}_U^N$ could be obtained by having two matched filters $h(t)$ and $h_e(t)$; after sampling, the decoder then sees $\boldsymbol{y}^N$ and $\boldsymbol{y}_e^N$. Since $h(t)$ and $h_e(t)$ are mutually orthogonal it is clear that $\boldsymbol{y}_U^N = \boldsymbol{y}^N + \boldsymbol{y}_e^N$ and thus $\boldsymbol{y}_e^N$ and $\boldsymbol{y}^N$ also form a set of sufficient statistics for detection of $\boldsymbol{a}^N$. Furthermore, since optimal decoding can be done either with $\boldsymbol{y}_U^N$ or with the pair $(\boldsymbol{y}^N, \boldsymbol{y}_e^N)$ we have

$$I(\boldsymbol{y}_U^N; \boldsymbol{a}^N) = I((\boldsymbol{y}^N, \boldsymbol{y}_e^N); \boldsymbol{a}^N) \tag{40}$$

I_{h_U} and I_h can be related by the following inequality:

$$\begin{aligned} I((\boldsymbol{y}^N, \boldsymbol{y}_e^N); \boldsymbol{a}^N) &= H(\boldsymbol{y}^N, \boldsymbol{y}_e^N) - H(\boldsymbol{y}^N, \boldsymbol{y}_e^N \big| \boldsymbol{a}^N) \\ &\le H(\boldsymbol{y}^N) + H(\boldsymbol{y}_e^N) - H(\boldsymbol{\eta}^N) - H(\boldsymbol{\eta}_e^N) \\ &= I(\boldsymbol{y}^N; \boldsymbol{a}^N) + I(\boldsymbol{y}_e^N; \boldsymbol{a}^N). \end{aligned} \tag{41}$$

Inserting (40) in (41) and rearranging we obtain the following information rate inequality

$$I_h \ge I_{h_U} - I_{h_e}. \tag{42}$$

In total we have obtained the following inequalies for the information rates:

$$I_{h_U} - I_{h_e} \le I_h \le I_{h_U}, \quad \tau < 1/2WT \tag{43}$$

and

$$I_{h_L} \le I_h \le I_{h_U}, \quad \tau \ge 1/2WT. \tag{44}$$

We will next show how to use these bounds.

4.2 Bounding Techniques

In this section we try to lower bound the lower bounds of (43) and (44). In order to do that we need a lower bound for I_{h_U} and I_{h_L} in (43) and (44), and an upper bound for I_{h_e} in (43). The

upper bounds of (43) and (44) will not be calculated by bounding techniques; they will be simulated according to the next section. Lower and upper bounding the information rate of an ISI channel has a long history. Bounds appear in for example [14]. In [19] a lower bound that is remarkably tight is proposed, but there is no proof of it. Bounds on the cutoff rate, which is a lower bound to the information rate, appear in [20].

We start with the two lower bounds and assume the Forney model $\boldsymbol{x} = \boldsymbol{b} \star \boldsymbol{a} + \boldsymbol{w}$. We use an extended version of a lower bound from [14].[3] We need to compute I_{DT}, which is

$$\lim_{N\to\infty} \frac{1}{N}[H(\boldsymbol{x}^N) - H(\boldsymbol{x}^N\big|\boldsymbol{a}^N)]. \tag{45}$$

The second term of can be directly computed as $H(\boldsymbol{x}^N\big|\boldsymbol{a}^N) = H(\boldsymbol{w}^N) = (N/2)\log_2(2\pi e\sigma^2)$. By the chain rule,

$$H(\boldsymbol{x}^N) = \sum_{n=0}^{N} H(x[n]\big|\boldsymbol{x}^{n-1}), \tag{46}$$

where $\boldsymbol{x}^{-1} \triangleq \emptyset$. Since conditioning does not increase entropy we get

$$H(x[n]\big|\boldsymbol{x}^{n-1}) \geq H(x[n]\big|\boldsymbol{x}^{n-1}, \boldsymbol{a}^{n-\kappa}) \tag{47}$$

where $1 \geq \kappa \geq n$ is an integer. This implies that

$$I_{\mathrm{DT}} \geq \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} H(x[n]\big|\boldsymbol{x}^{n-1}, \boldsymbol{a}^{n-\kappa}) - \frac{N}{2}\log_2(2\pi e\sigma^2). \tag{48}$$

In [14] with $\kappa = 1$, it was proved that $I_{\mathrm{DT}} \geq I(\rho a + w; x)$, where a is distributed as $a[n]$, w is zero mean Gaussian with variance σ^2 and

$$\rho = \exp\frac{1}{2\pi}\int_0^{\pi} \log|B(\lambda)|^2 d\lambda$$

with $B(\lambda)$ defined in (15). Now we take $\kappa > 1$ to strengthen the bound. Calculate (48) by direct computation. By the chain rule

$$H(x[n]\big|\boldsymbol{x}^{n-1}, \boldsymbol{a}^{n-\kappa}) = H(\boldsymbol{x}_{n-\kappa+1}^{n}\big|\boldsymbol{x}^{n-\kappa}, \boldsymbol{a}^{n-\kappa}) - H(\boldsymbol{x}_{n-\kappa+1}^{n-1}\big|\boldsymbol{x}^{n-\kappa}, \boldsymbol{a}^{n-\kappa}) \tag{49}$$

and this is the entropy sought in (47). Since $\boldsymbol{w}^N$ are independent and the ISI response is causal, the conditioning on $\boldsymbol{x}^{n-\kappa}$ in (49) can be removed, i.e.

$$H(\boldsymbol{x}_{n-\kappa+1}^{n}\big|\boldsymbol{x}^{n-\kappa}, \boldsymbol{a}^{n-\kappa}) = H(\boldsymbol{x}_{n-\kappa+1}^{n}\big|\boldsymbol{a}^{n-\kappa}).$$

[3]The possibility of the extension is obvious and was mentioned in [14].

Now, since the data sequence $\boldsymbol{a}$ is i.i.d.,

$$H(\boldsymbol{x}_{n-\kappa+1}^{n}\big|\boldsymbol{a}^{n-\kappa}) = H(\boldsymbol{z}_{n-\kappa+1}^{n}) \tag{50}$$

where

$$z[l] = \sum_{m=0}^{l-n+\kappa-1} a[l-m]b[m] + w[l], \quad n-\kappa+1 \le l \le n. \tag{51}$$

That is, $z[l]$ is formed by subtracting the interference from $\boldsymbol{a}^{n-\kappa}$ on $\boldsymbol{x}_{n-\kappa+1}^{n}$. To summarize, the entropy in (50) is equal to the entropy of a variable $\boldsymbol{z}^{\kappa-1}$, where

$$\boldsymbol{z}^{\kappa-1} = \boldsymbol{B}^{\kappa-1}\boldsymbol{a}^{\kappa-1} + \boldsymbol{w}^{\kappa-1}, \tag{52}$$

and $\boldsymbol{B}^{\kappa-1}$ is the $\kappa \times \kappa$ matrix

$$\boldsymbol{B}^{\kappa-1} = \begin{bmatrix} b[0] & 0 & 0 & \cdots & 0 \\ b[1] & b[0] & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ b[\kappa-2] & \cdots & b[1] & b[0] & 0 \\ b[\kappa-1] & \cdots & \cdots & b[1] & b[0] \end{bmatrix}.$$

For l larger than the ISI length K the terms to be summed up in (48) are all identical, thus the limit can be removed and I_{DT} can be computed by computation of a single term in (48) with any $n > K$. In bits per channel use, we have shown that

$$I_{\mathrm{DT}} \ge I_{\mathrm{LB}}^{\kappa} \triangleq I(\boldsymbol{z}^{\kappa-1}; \boldsymbol{a}^{\kappa-1}) - I(\boldsymbol{z}^{\kappa-2}; \boldsymbol{a}^{\kappa-2}). \tag{53}$$

An upper bound for I_{h_e} in (43) is also needed. If E_{h_e} is small it suffices with a rather loose bound. We make use of the following bound from [14]:

$$I_{\mathrm{DT}} \le I_{\mathrm{UB}} \triangleq I(\|\boldsymbol{b}\|a + w; a) \tag{54}$$

where a is distributed as $a[n]$, w is zero-mean Gaussian with σ^2 and $\|\boldsymbol{b}\|^2$ is

$$\|\boldsymbol{b}\|^2 = \frac{1}{\pi}\int_0^{\pi} |B(\lambda)|^2 d\lambda.$$

Let $I_{\mathrm{LB},h_U}^{\kappa}$ denote the bound I_{LB}^{κ} in (53) computed from the time-discrete models stemming from $h_U(t)$, and similarily for I_{UB,h_e}. Insert (53) and (54) into (43) and (44); we then obtain the lower bounds

$$I_h \ge \begin{cases} (I_{\mathrm{LB},h_U}^{\kappa} - I_{\mathrm{UB},h_e})/\tau T, & \tau < 1/2WT \\ I_{\mathrm{LB},h_L}^{\kappa}/\tau T & \tau \ge 1/2WT \end{cases} \tag{55}$$

where W is the bandwidth of bandlimited pulses. We evaluate (55) by numerically establishing the distribution of $\boldsymbol{z}^{\kappa-1}$ and then numerically computing the entropy $H(\boldsymbol{z}^{\kappa-1})$.

4.3 Monte Carlo Based Methods

Recently, methods to find information rates have been proposed [21], [22]. A brief review of the technique is given here.

We need to compute $\frac{1}{N}H(\boldsymbol{x}^N)$ in (45), and this can be accomplished as follows. First generate a long sequence $\boldsymbol{x}^N$ and run a forward pass of the BCJR algorithm. The output of this algorithm is $p(\boldsymbol{x}^N)$. Due to the Shannon-Breiman-McMillan Theorem, $-\frac{1}{N}\log_2 p(\boldsymbol{x}^N)$ converges with probability 1 to a well-defined entropy rate which is the $\lim_{N\to\infty}\frac{1}{N}H(\boldsymbol{x}^N)$ in (45). When actually performing the pass, one computes $-\frac{1}{N}\log_2 p(\boldsymbol{x}^N)$ as $\frac{1}{N}\sum\log_2\gamma_k$, where γ_k are the scale factors that bring the sum of state metrics to 1 at each trellis stage. Applying the method, we find the information rates I_{h_U} and I_{h_L} in (43) and (44). The upper bound in (54) is still used for I_{h_e} in (43).

The complexity of this method is clearly $|\mathcal{A}|^{K-1}$, which can be very large. In [21] reduced complexity methods were given that bound the information rates; a non optimal decoding algorithm, such as the M-BCJR, was used. Here we propose a new type of bound. It is based on the inequality (43) followed by an application of the full-complexity approach. An example and comparison to earlier simulation methods are found in Appendix B.

We conclude by mentioning that the WMF model is not required. In [23] a MAP equalizer was derived for the Ungerboeck model, which can be used in place of the BCJR computation.

5 Numerical Results and Comparisons

Throughout we take $h(t)$ as root RC with excess bandwidth α ($0\leq\alpha\leq 1$).

The first case is $\alpha = .3$ and binary signaling. We used the simulation method from Section 4.3 to find the bounds (43) and (44) and the result is shown in Figure 3. Symbol sequences $\boldsymbol{a}$ of length 2×10^6 are used and $\tau\in\{.5,.6,.7,.8,.9\}$. For $\tau = .5$ the number of ISI taps K is $K = 9$ (i.e. 256 states); for $\tau = .6$, $K = 8$ (128 states); for $\tau\in\{.7,.8,.9\}$, $K = 6$ (32 states). The upper bounds coincide (to the accuracy of the picture) with the lower bounds for $\tau = .8$ and .9. For the other cases the upper bounds are dotted and the lower bounds are solid. The bounds are rather tight even for small τ. Also included is the ultimate capacity C of the PSD (this capacity coincides with the FTN capacity $C_{\rm FTN}$ if $\tau = 1/2WT$) as well as the capacity of Nyquist signaling with the same $h(t)$, $C_{\rm N}$ in (21). Capacity $C_{\rm N}$ is the ultimate capacity for signals with $\tau = 1$. It is seen that both $\tau = .5$ and $\tau = .6$ have lower bounds that are larger than $C_{\rm N}$, which implies that these systems are superior to any form of Nyquist signaling, and that they are not very far from C at low E_b/N_0; thus not much

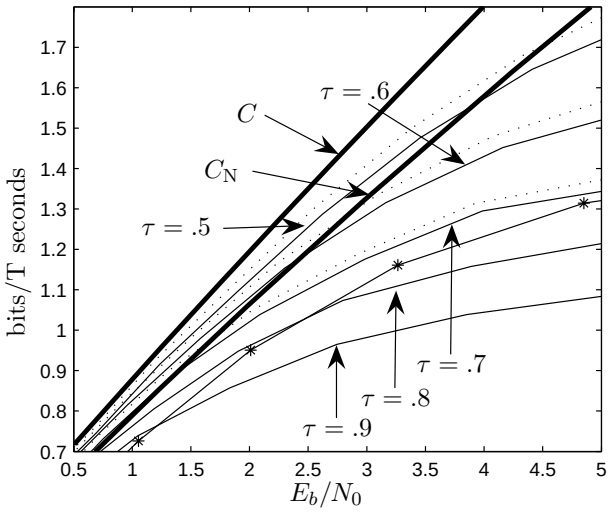

Figure 3: Information rates for binary signaling with with $\alpha = .3$ root RC pulse and various τ. Solid heavy line marked by C is the ultimate capacity for the PSD and the one marked by C_N is capacity of Nyquist signaling. Lighter solid lines are lower bounds to the information rate; dotted lines are upper bounds. The line with markers is the lower bound for the case $\tau = .3$ but computed with the bounding method with $\kappa = 2$.

further can be gained even though the alphabet is binary. The line with markers is a lower bound for the case $\tau = .7$ but computed with the bounding method from Section 4.2 with $\kappa = 2$. This is the only place where the bounding method is used for $\tau < 1/2WT$. In general the bounding method is not very tight for these τ and in the sequel we use the Monte Carlo based method for $\tau < 1/2WT$; for $\tau > 1/2WT$ the bounding method will sometimes be used. For verification purposes we list some of the $\boldsymbol{b}_U$ used. For $\tau = .5$, $\boldsymbol{b}_U = .4374, .7493, .4387, -.1030, -.2011, .0483, .0973, -.0518, -.0023$ and for $\tau = .7$, $\boldsymbol{b}_U = .7281, .6584, -.1975, -.0352, .0737, -.0331$.

Information rates for binary signaling with $\alpha = .2$ are shown in Figure 4. We set $\tau \in \{.7, .8, .9\}$ and K is 11 (1024 states) for all τ. The system with $\tau = .5$ has an information rate superior to $C_{\rm N}$. The bounds are found with the Monte Carlo based method and they are rather tight.

Next we consider is $\alpha = .5$. This choice of α is not very bandwidth efficient, but is worthwhile to discuss because FTN will significantly benefit from the excess bandwidth. We investigate binary signaling with $\tau = 1/3$ (system 1) and quaternary with $\tau = .8$ (system 2). These choices result in maximum information rates 3 and 2.5 bits per T seconds respectively. We compare these systems with Nyquist systems that can achieve the same bits/Hz; this is a fair comparison in our opinion. We consider only lower bounds to the information rates; these are shown in Figure 5. For the binary system we found Monte Carlo bounds; for the quaternary we used the lower bounding technique,

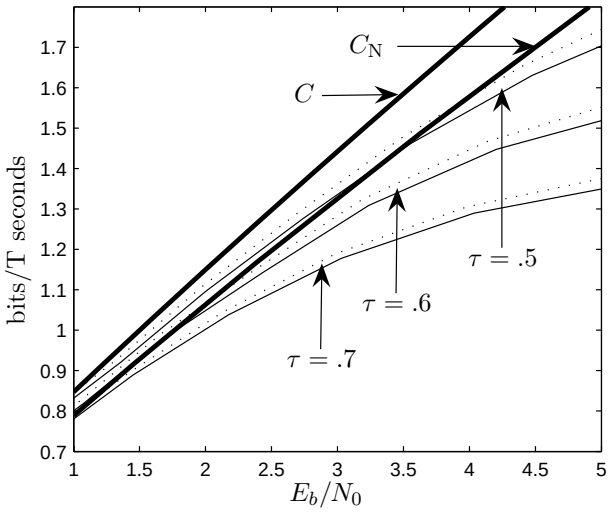

Figure 4: Information rates for binary signaling with an $\alpha = .2$ root RC pulse and various τ. Solid heavy line marked by C is the ultimate capacity for the PSD and the one marked by C_N is capacity of Nyquist signaling. Lighter solid lines are lower bounds to the information rates; dotted lines are upper bounds.

with $\kappa = 2$, from Section 4.2. To obtain a maximum throughput of 2.5 bits/T with Nyquist signaling, we use a 32-cross (32CR) constellation [11] (and measure the information rate per dimension). To obtain 3 bits/T we use 64QAM.

It is clear from Figure 5 that both FTN systems outperform their Nyquist counterparts. Moreover, they significantly outperform $C_{\rm N}$. If a Nyquist system is to be equal to an FTN system in both power and bandwidth efficiency, a smaller, more difficult to realize value for α is required. Consider system 1; for $E_b/N_0 \lessapprox 7$ dB it has an information rate I_h which is $\approx 1.17 C_{\rm N}$ or $\approx 1.27 I_{64}$. Thus a Nyquist system achieving $C_{\rm N}$ needs $\alpha \approx 1.5/1.17 - 1 = .28$, and 64QAM needs $\alpha \approx 1.5/1.27 - 1 = .18$. So the binary FTN system with $\alpha = .5$ is in terms of power and bandwidth efficiency equivalent to a Nyquist system with much more constrained excess bandwidth. System 2 has information rate $I_h \approx 1.15 I_{32}$ in the range $E_b/N_0 \lessapprox 5.5$ dB. Thus the Nyquist 32CR system needs $\alpha \approx 1.5/1.15 - 1 = .3$. Also seen in Figure 5 is that system 1 has information rate very close to C, so not much more rate can be gained.

In Figure 6 a lower bound with $\kappa = 2$ is given for octal FTN with $\alpha = .2$ and $\tau = .8517$; this is system 3. The τ value is chosen to give an FTN system with 3.5 bits/T maximum information rate, which is the maximum rate of a 128CR Nyquist system. Even for α as low as .2 the 128CR system has lower information rate than the FTN system. In fact the FTN beats even the Nyquist system with Gaussian symbol distribution. This we find remarkable for such a small α. Equalizing both the

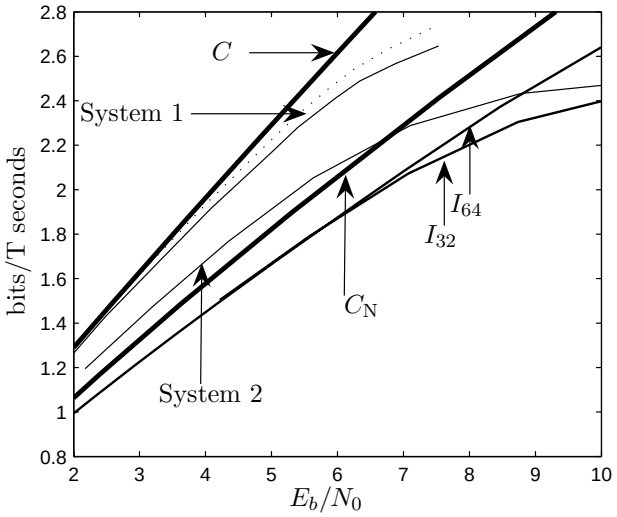

Figure 5: Information rates for signaling with an $\alpha = .5$ root RC pulse. System 1 is binary with $\tau = 1/3$. System 2 is quaternary with $\tau = .8$. The heavy line marked $C_{\rm N}$ is capacity of Nyquist signaling; the heavy line marked C is the ultimate capacity for the PSD. The dotted line is an upper bound. I_{32} and I_{64} mark the information rates of Nyquist 32CR and 64QAM.

bandwidth and power efficiencies of the systems, we find that the 128CR Nyquist requires $\alpha \approx .11$ for $E_b/N_0 \lessapprox 12$ dB. Also shown is a lower bound with $\kappa = 2$ for quaternary FTN with $\alpha = .2$ and $\tau = .9$; this is system 4. The bound lies above the information rates for both 16QAM and 32CR Nyquist systems at low SNR.

6 Conclusion

We have investigated constrained capacities that apply to faster-than Nyquist signaling. In particular we have compared them with the capacities for signals based on orthogonal linear modulation. FTN has an infinite-state structure, and in order to analyze it we develop short finite models and bound FTN with these; we also derive a linear program formulation that leads to bounds. These can also be used for standard ISI channels and they are tighter than other, comparable, bounds for the cases considered here.

Our conclusion is that when signals have excess bandwidth it is, in general, beneficial to signal faster-than-Nyquist, that is, to give up orthogonality between the data pulses. For suitable signaling rates the ultimate capacity for signals with a given power spectral density can be achieved by FTN, despite the linear modulation in its design. More specifically, the capacity of FTN signals exceeds that of Nyquist signals for some FTN symbol time τT and for all $\tau T < 1/2W$ for signals bandlimited to W Hz. It often comes close to or equals the ultimate C for the PSD. As well, the capacity of binary and

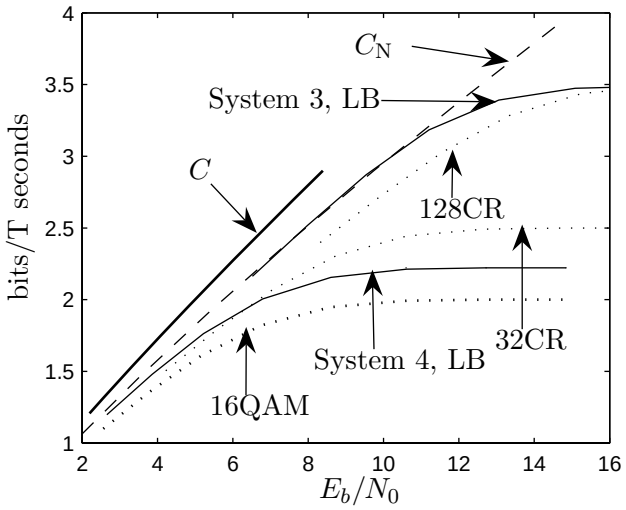

Figure 6: Lower bounds on information rates for FTN. System 3 is octal signaling, $\tau = .8517$ and $\alpha = .2$. System 4 is quaternary signaling, $\tau = .9$ and $\alpha = .2$. The dashed line is Nyquist signaling with an unconstrained alphabet; heavy solid line is ultimate capacity C for the $\alpha = .2$ root RC PSD.

quaternary FTN at many SNRs exceeds that of signals based on such higher-alphabet constellations as 32CR and 64QAM. The capacity gains can also be converted into more generous excess bandwidth: FTN can have more than Nyquist can, at the same power and bandwidth efficiency. As power grows, the bandwidth efficiency of FTN signals in fact approaches that of sinc pulse signaling.

Overall we can conclude that using orthogonality in the synthesis of codewords has a significant cost, but the linear modulation formulation of itself is much less costly.

Acknowledgements

The authors wish to acknowledge an anonymous reviewer who pointed out the infinite duration ISI problem that was solved in Appendix C. Parts of this work were supported by the Swedish Science Board through grant 621-2003-3210, and by the Swedish Foundation for Strategic Research through its Center for High Speed Wireless Communication at Lund University.

Appendix A: Linear Programming Formulation

An LP formulation for optimal ISI signals is well described in [24], or more compactly in [25]; it has appeared earlier in [26]. Since $E_h = 1$ by assumption, we can express the objective function in (37) simply as

$$\boldsymbol{b}_L = \arg\max_{\boldsymbol{b}} E_{h_L} = \arg\max_{\boldsymbol{b}} g[0] \tag{56}$$

where $\boldsymbol{g}$ is the autocorrelation of $\boldsymbol{b}$. Moreover, $|\tilde{H}(f;\boldsymbol{b})|^2$ can be expressed as

$$|\tilde{H}(f;\boldsymbol{b})|^2 = \begin{cases} \sum_{|k|<K} g[k]e^{-\jmath 2\pi k f \tau T}, & -1/2\tau T \le f \le 1/2\tau T \\ 0, & |f| > 1/2\tau T \end{cases} \tag{57}$$

Since the optimization is performed over $\boldsymbol{g}$ we must guarantee that $\boldsymbol{g}$ is a valid autocorrelation, i.e. that there exists a tap sequence $\boldsymbol{b}$ such that $g[k] = b[n] \star b[-n]$. This is true if and only if $\boldsymbol{g}$ satisfies

$$\sum_{|k|<K} g[k]e^{-\jmath 2\pi k f} \ge 0, \quad -1/2 \le f \le 1/2 \tag{58}$$

In practice the frequency axis must be sampled. Let the sample points be the set Ω. We use 1000 points throughout, but after the optimization we use 10^5 to verify that the constraints are fullfilled even for frequencies outside Ω. The optimization for $h_L(t)$ can now be formulated as

$$\begin{aligned} \boldsymbol{b}_L &= \arg\max_{\boldsymbol{b}} g[0] \\ s.t. \qquad & 0 \le \sum_{|k|<K} g[k]e^{-\jmath 2\pi k f \tau T} \le |H_{\text{fo}}(f)|^2, \quad f \in \Omega \end{aligned}$$

For $h_U(t)$ the optimization becomes

$$\begin{aligned} \boldsymbol{b}_U &= \arg\min_{\boldsymbol{b}} g[0] \\ s.t. \qquad & \sum_{|k|<K} g[k]e^{-\jmath 2\pi k f \tau T} \ge |H_{\text{fo}}(f)|^2, \quad f \in \Omega \end{aligned}$$

Appendix B: New Types of Bounds

This appendix compares applies our bounding technique to the one presented in [21]; the outcome is that our method is superior in the low SNR regime. We only consider the upper bound. Applying a reduced state decoder in Section 4.3 yields an upper bound to the information rate [21]. But we have already lowered the complexity with the use of the $h_L(t)$ and $h_U(t)$ that gave inequality (44). A natural question arises: Is reduced complexity computation of strong ISI better than full complexity computation of weak ISI. In the latter case the true information rate I_{h_U} will be found, and this is known to be larger than I_h.

In [21] the following 11-tap impulse response was considered: $b[n] = 1/[1+(n-5)^2],\ 0 \le n \le 11$. An equivalent unit-energy minimum phase filter was found and a reduced state decoder with 100 states was applied, which results in the upper bound.

Here we construct a pulse $h_U(t)$ first with 7 taps, i.e. $K = 7$, and then 4; this results in 64 and 8 states respectively. Full complexity forward BCJR recursions are then performed. The results are

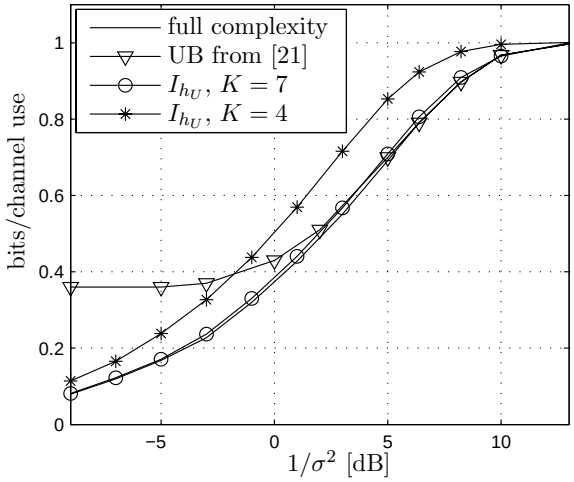

Figure 7: Comparisons between the Monte Carlo based upper bound from [21] and the bounds proposed here. The solid line is the true information rate of an 11-tap filter. Triangles mark the upper bound from [21] computed with 100 states; asterixes and circles mark our upper bounds computed with 8 and 64 states.

given in Figure 7. The solid curve is the true information rate, i.e. the one found by full complexity (1024 states). The curve UB marked by triangles is the upper bound from [21]. The curves I_{h_U} marked by asterixes and squares are the information rates of the filters I_{h_U} with $K = 4$ and 7; these are known to be upper bounds. The filter taps we used were $\boldsymbol{b}_U = 0.8044, 0.7586, 0.4453, 0.2888$ and $\boldsymbol{b}_U = 0.6223, 0.6281, 0.4008, 0.2369, 0.1442, 0.1097, 0.0608$.

It is clear from Figure 7 that the new bounds are better than the bound from [21] for low SNR; this is true even for a state complexity as low as 8 ($K = 4$). The curve computed with $K = 7$ is tight for all SNR; however, the upper bound from [21] is still tighter for SNR above 3–4 dB. Why the new bound works so well for low SNR can be explained as follows. The power spectrum of $h_U(t)$, $|H_U(f)|^2$, has roughly the same shape as $|H(f)|^2$, but it has more energy. Since the shapes are similar $h_U(t)$ will simply act as $h(t)$ but at a higher SNR. Since the energy E_{h_U} is rather close to E_h for large K and the slope of the information rate curve is small at low SNR, this increase will not be very significant, and the bound becomes tight.

Appendix C: Capacity of Infinite Duration ISI

The validity of the constrained capacity formula (14) under infinite ISI is crucial. Shamai et. al. [14] use a result of [15], but [15] only considers finite ISI. Here it will be shown that (14) holds for infinite ISI as well. The idea in what follows is to construct a finite ISI response that has capacity arbitrarily close to that of the infinite one.

If the spectrum $G(\lambda)$ in Eq. (15) stems from finite ISI, then there is no problem, (14) is the capacity as proved in [15]. Now assume that $G(\lambda)$ stems from infinite ISI and has a finite number d of discontinuities. Denote by λ_ℓ the ℓth discontinuity point with corresponding function values $c_\ell^- \triangleq \lim_{\lambda\to\lambda_\ell^-} G(\lambda)$ and $c_\ell^+ \triangleq \lim_{\lambda\to\lambda_\ell^+} G(\lambda)$. Define $c_\ell \triangleq |c_\ell^+ - c_\ell^-|$. Now create a continuous spectrum, if $c_\ell^- < c_\ell^+$ then connect the point $G(\lambda_\ell - \epsilon)$ and $G(\lambda_\ell)$ with a smooth arc; if $c_\ell^- > c_\ell^+$, connect $G(\lambda_\ell)$ and $G(\lambda_\ell + \epsilon)$. Let $G(\lambda) + Z(\lambda)$ denote the continuous spectra and construct it such that $0 \le Z(\lambda) \le \max(c_1, \ldots, c_d),\ 0 \le \lambda \le \pi$.

We have the Fourier transform pair

$$\begin{aligned} G(\lambda) + Z(\lambda) &= \sum_{n=-\infty}^{\infty} (g[n] + z[n]) e^{\jmath 2\pi n\lambda} \\ g[n] + z[n] &= \int_{-1/2}^{1/2} [G(\lambda) + Z(\lambda)]\, e^{-\jmath 2\pi n\lambda} d\lambda \end{aligned}$$

Now apply Fejer summation:

$$S_N(\lambda) \triangleq \sum_{n=-N}^{N} (g[n] + z[n]) \left(1 - \frac{|n|}{N}\right) e^{\jmath 2\pi n\lambda}.$$

A well known result is that if $G(\lambda) + Z(\lambda)$ is continuous, then $S_N(\lambda)$ converges uniformly to $G(\lambda) + Z(\lambda)$ as $N \to \infty$. That is, there exists N such that $\max_\lambda |S_N(\lambda) - G(\lambda) - Z(\lambda)| < \varsigma$, with arbitrarily small ς. The finite sequence $s_N[n] \triangleq (g[n] + z[n]) \left(1 - \frac{|n|}{N}\right),\ -N \le n \le N$ is, from (58), an autocorrelation sequence only if $S_N(\lambda) > 0,\ 0 \le \lambda \le \pi$. If we define $\tilde{s}_N[n] = s_N[n] + \varsigma\delta[n]$ then it follows by the uniform convergence that $\tilde{s}_N[n]$ is a valid autocorrelation and that $\tilde{S}_N(\lambda) \ge G(\lambda) + Z(\lambda),\ 0 \le \lambda \le \pi$. This we write as $\tilde{S}_N(\lambda) = G(\lambda) + Z(\lambda) + E(\lambda)$, with $E(\lambda) \ge 0,\ 0 \le \lambda \le \pi$. The implication of $\tilde{s}_N[n]$ is that there exists a finite ISI sequence, with $\tilde{s}_N[n]$ as its autocorrelation, that has capacity given by (14) and that can be related to the capacity of the infinite ISI spectrum $G(\lambda)$.

Denote the capacity of an arbitrary spectrum $X(\lambda)$ as $C_{\mathrm{DT}}(X)$. If $X(\lambda)$ stems from finite ISI, then we have $C_{\mathrm{DT}}(X) = C(X) \triangleq \int \log_2(1 + \xi X(\lambda))\, d\lambda$, where ξ is the SNR. For infinite ISI, $C_{\mathrm{DT}}(X)$ is unknown; it will next be shown that $C_{\mathrm{DT}}(X)$ is the form $C(X)$ also in this case. Because $C(X)$ is the ultimate capacity for $X(\lambda)$ it is clear that we must have

$$C_{\mathrm{DT}}(X) \le C(X).$$

From Section 4.1 we have that

$$C_{\mathrm{DT}}(G) \ge C_{\mathrm{DT}}(G + Z + E) - C_{\mathrm{DT}}(Z + E) = C(G + Z + E) - C_{\mathrm{DT}}(Z + E). \tag{59}$$

where the last equality holds because $G(\lambda) + Z(\lambda) + E(\lambda)$ represents finite ISI. Moreover, because $Z(\lambda) \geq 0$ all λ and $E(\lambda) \geq 0$ all λ we have from the form of $C(X)$ that $C(G) \leq C(G + Z + E)$. Using the inequality chains $C_{\mathrm{DT}}(G) \leq C(G) \leq C(G + E + Z)$ and $C_{\mathrm{DT}}(Z + E) \leq C(Z + E)$ in (59) gives

$$C_{\mathrm{DT}}(G) \geq C(G) - C_{\mathrm{DT}}(Z + E) \geq C(G) - C(Z + E). \tag{60}$$

We now proceed by upper bounding the term $C(Z + E)$. Split the integral into two parts

$$C(Z + E) = \int_{W(\epsilon)} \log_2 \left(1 + \xi E(\lambda)\right) d\lambda + \int_{\bar{W}(\epsilon)} \log_2 \left(1 + \xi \left[Z(\lambda) + E(\lambda)\right]\right) d\lambda, \tag{61}$$

where $W(\epsilon)$ is the interval where $Z(\lambda) = 0$ and $\bar{W}(\epsilon)$ is the interval where $Z(\lambda) > 0$. The two right hand terms of (61) can be bounded as

$$C(Z + E) < \left(\frac{1}{2} - d\epsilon\right) \log_2 \left(1 + \xi\varsigma\right) + d\epsilon \log_2 \left(1 + \xi \left[\varsigma + \max(c_1, \ldots, c_d)\right]\right). \tag{62}$$

The right hand side can be made arbitrarily small by taking small enough ϵ and a large enough N such that ς is sufficiently small. This implies that $C_{\mathrm{DT}}(G)$ is bounded by

$$C(G) - \epsilon' \leq C_{\mathrm{DT}}(G) \leq C(G)$$

for all $\epsilon' > 0$ and it thus follows that $C_{\mathrm{DT}}(G) = C(G)$. This proves that (14) holds also for infinite duration ISI signals, so long as $G(\lambda)$ has finitely many disconinuities.

References

[1] J.E. Mazo, "Faster-than-Nyquist Signaling," *Bell Syst. Tech. J.*, vol. 54, pp. 1451–1462, Oct. 1975.

[2] D. Hajela, "On Computing the Minimum Distance for Faster-than-Nyquist Signaling," *IEEE Trans. Information Theory*, vol. 36, pp. 289–295, Mar. 1990.

[3] J.E. Mazo and H.J. Landau, "On the Minimum Distance Problem for Faster-than-Nyquist Signaling," *IEEE Trans. Information Theory*, vol. 34, pp. 1420–1427, Nov. 1988.

[4] A.D. Liveris and C.N. Georghiades, "Exploiting Faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 51, pp. 1502–1511, Sep. 2003.

[5] F. Rusek and J. B. Anderson, "Non Binary and Precoded Faster than Nyquist Signaling," accepted for publication, *IEEE Trans. Commun.*, July 2007.

[6] F. Rusek and J. B. Anderson, "Serial and Parallel Concatenations Based on Faster-than-Nyquist Signaling," *Proc.* IEEE Int. Symp. Information Theory, Seattle, pp. 970–974, July 2006.

[7] F. Rusek and J. B. Anderson, "On Information Rates of Faster-than-Nyquist Signaling," *Proc.*, IEEE Global Commun. Conf., San Francisco, Nov. 2006.

[8] E.N. Gilbert, "Increased Information Rate by Oversampling," *IEEE Trans. Information Theory*, vol. 39, pp. 1973–1976, Nov. 1993.

[9] S. Shamai, "Information Rates by Oversampling the Sign of a Bandlimited Process," *IEEE Trans. Information Theory*, vol. 40, pp. 1230–1236, July 1994.

[10] J.G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.

[11] J.B. Anderson and A. Svensson, *Coded Modulation Systems*, Plenum, New York, 2003.

[12] G. D. Forney, Jr., "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Information Theory*, vol. 18, pp. 363–378, May 1972.

[13] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," *IEEE Trans. Commun.*, vol. 22, pp. 624–636, May 1974.

[14] S. Shamai, L.H. Ozarow and A.D. Wyner, "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. Information Theory*, vol. 37, pp. 1527–1539, Nov. 1991.

[15] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," Ph.D Thesis, no. ETH 8671, Inst. Signal and Information Processing, Swiss Federal Inst. Technol., Zurich, 1988.

[16] A.D. Liveris, *On Distributed Coding, Quantization of Channel Measurements and Faster-than-Nyquist Signaling*, Ph.D. thesis, Dept. Elec. Eng., Texas A&M Univ., Apr. 2006.

[17] R.A. Gibby and J.W. Smith, "Some Extensions of Nyquist's Telegraph Transmission Theory," *Bell Syst. Tech. J.*, vol. 44, pp. 1487-1510, Sept. 1965.

[18] B. Le Floch, M. Alard and C. Berrou, "Coded Orthogonal Frequency Division Multiplex," *Proc. IEEE*, vol. 83, pp. 982–996, June 1995.

[19] S. Shamai and R. Laroia "The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel Precoding Loss," *IEEE Trans. Information Theory*, vol. 42, pp. 1388–1404, Sept. 1996.

[20] S. Shamai and A. Dembo, "Bounds on the Symmetric Binary Cutoff Rate for Dispersive Gaussian Channels," *IEEE Trans. Commun.*, vol. 42, pp. 39–53, Jan. 1994.

[21] 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. Information Theory*, vol. 52, pp. 3498–3508, Aug. 2006.

[22] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the Achievable Information Rates of Finite State ISI Channels," *Proc.* IEEE Global Commun. Conf., San Antonio, TX, Nov. 2001, pp. 2992–2996.

[23] G. Colavolpe and A. Barbieri, "On MAP Symbol Detection for ISI Channels Using the Ungerboeck Observation Model", *IEEE Commun. Letters*, vol. 9, Aug. 2005.

[24] A. Said, "Design of Optimal Signals for Bandwidth-Efficient Linear Coded Modulation," Ph.D. thesis, Dept. Elec., Computer and Systems Eng., Rensselaer Poly. Inst., Troy, N.Y., Feb. 1994.

[25] A. Said and J.B. Anderson, "Design of Optimal Signals for Bandwidth-Efficient Linear Coded Modulation", *IEEE Trans. Information Theory*, vol. 44, pp. 701–713, Mar. 1998.

[26] G.J. Foschini, "Contrasting Performance of Faster Binary Signaling with QAM," *Bell Syst. Tech. J.*, vol. 63, pp. 1419–1445, Oct. 1984.