Time Localization and Capacity of Faster-Than-Nyquist Signaling

Ather Gattami  
Bitynamics Research  
Stockholm, Sweden  
Email: atherg@gmail.com

Emil Ringh  
Ericsson Research  
Stockholm, Sweden  
Email: emil.ringh@ericsson.com

Johan Karlsson  
KTH Royal Institute of Technology  
Stockholm, Sweden  
Email: johan.karlsson@math.kth.se

Abstract—In this paper, we consider communication over the bandwidth limited analog white Gaussian noise channel using non-orthogonal pulses. In particular, we consider non-orthogonal transmission by signaling samples at a rate higher than the Nyquist rate. Using the faster-than-Nyquist (FTN) framework, Mazo showed that one may transmit symbols carried by \( \text{sinc} \) pulses at a higher rate than that dictated by Nyquist without losing bit error rate. However, as we will show in this paper, such pulses are not necessarily well localized in time. In fact, assuming that signals in the FTN framework are well localized in time, one can construct a signaling scheme that violates the Shannon capacity bound. We also show directly that FTN signals are in general not well localized in time.

We also consider FTN signaling in the case of pulses that are different from the \( \text{sinc} \) pulses. We show that one may use a precoding scheme of low complexity, in order to remove the inter-symbol interference. This leads to the possibility of increasing the number of transmitted samples per time unit and compensate for spectral inefficiency due to signaling at the Nyquist rate of the \( \text{sinc} \) pulses. We demonstrate the power of the precoding scheme by simulations.

I. INTRODUCTION

A. Background and previous work

In 1949, Shannon [1] presented his famous result on the capacity of the additive white Gaussian noise (AWGN) channel for band limited signals. The result is based on communication using orthogonal pulses (to avoid inter-symbol interference) which corresponds to transmission at the Nyquist sampling rate.

In [2], Mazo considered the uncoded transmission case for binary symbols, where the sample rate is faster than the one dictated by the Nyquist sampling theorem, so called faster-than-Nyquist (FTN) signaling. The minimum Euclidean distance between pairs of binary signals was used as a performance measure, and it was shown that one can transmit the signal pulses faster than the Nyquist frequency without decreasing the Euclidian distance between any two signals. The limit for such rate increase is known as the Mazo limit: \( \rho \approx 0.802 \), and was derived in [3], [4]. Recently, a series of work has explored FTN signaling, see [5] and the references therein. This work has also been extended to two dimensions, the second dimension being the frequency domain [6]. The signals are packed tighter also in the frequency domain, possibly introducing interference between previously uncorrelated subcarriers. However, this inter-carrier interference does not affect the reliability of the signaling with a certain packing density, and with the use of an optimal detector the error rates would remain unchanged. Also, in [7] it was shown that FTN can achieve the maximum capacity for a given pulse (as opposed to orthogonal transmission). This work also considers the use of FTN for achieving higher capacity for non-\( \text{sinc} \) pulses when the code alphabet is finite.

The original derivation of Shannon’s capacity formula [1] is based on transmission of uncorrelated \( \text{sinc} \) pulses. These signals have infinite support in the time domain and hence, as pointed out by Wyner in [8], “the idea of transmission rate [using band limited pulses] has, at best, a limited meaning.” To treat this problem in a rigorous manner, Wyner proposed several physically consistent models with corresponding coding theorems, thereby justifying Shannon’s capacity formula [8]. Also, in the FTN framework, the concept of transmission rate is problematic and there is no guarantee that a signal consisting of a linear combination of pulses centered at a given time interval has its energy localized in the vicinity of that interval. In fact, as we will see in Section III, one can easily construct examples where this is not the case. Another problem in the FTN framework is that the algorithms used for detection and estimation suffers from high complexity, rendering NP-hard problems in general [9], [10].

B. Contribution

We consider communication over the bandwidth limited analog Gaussian white noise channel using non-orthogonal pulses. In particular, we consider non-orthogonal transmission by signaling samples at a rate higher than the Nyquist rate. Using the faster-than-Nyquist (FTN) framework, Mazo showed that one may transmit symbols carried by \( \text{sinc} \) pulses at a higher rate than that dictated by Nyquist without losing bit error rate. However, we show in this paper that such pulses are not necessarily well localized in time. In fact, assuming that signals in the FTN framework are well localized in time, we show that one can construct a signaling scheme that violates the Shannon capacity bound. We also show directly that FTN signals are in general not well localized in time. Thus, it’s not physically correct to talk about bits per second, as the energy of non-orthogonal signals may not be well localised in time, as opposed to orthogonal sinc pulses, for instance.

We go on and consider FTN signaling in the case of pulses...
that are different from the sinc pulses. We show that one may use a precoding scheme of low complexity in order to remove the inter-symbol interference. This leads to the possibility of increasing the number of transmitted samples per time unit (keeping average energy per time unit bounded by some constant) and compensating for spectral efficiency losses due to signaling at the Nyquist rate, and so, achieve the Shannon capacity when the same energy is spent for transmitting with the ideal sinc pulses. We demonstrate the power of the precoding scheme by simulations.

II. PRELIMINARIES AND PROPERTIES OF NON-ORTHOGONAL PULSES

To start, we introduce some basic notation. Let $\mathbb{N}$ denote the set of natural numbers $\{0, 1, 2, \ldots\}$, $\mathbb{R}$ the set of real numbers, and $\mathbb{C}$ the set of complex numbers. The $n \times n$ identity matrix is denoted by $I_n$. $X \sim \mathcal{N}(0, \mathbf{X})$ denotes that $X$ is a Gaussian variable with $\mathbb{E}\{X\} = 0$ and $\mathbb{E}\{XX^\top\} = \mathbf{X}$. We use $\mathbf{X} \succ 0$ ($\mathbf{X} \succeq 0$) to denote that the matrix $\mathbf{X}$ is positive definite (semidefinite). The empty set is denoted by $\emptyset$ and $\text{int}(\Omega)$ denotes the interior of the set $\Omega$. The floor function $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x$.

**Definition 1 (Signal Space).** $L_2(\mathbb{R})$ denotes the Hilbert space of functions $f : \mathbb{R} \to \mathbb{C}$ that are square integrable.

**Definition 2 (Fourier Transform).** The Fourier transform of a function $f \in L_2(\mathbb{R})$ is given by

$$f(t) \xrightarrow{\mathcal{F}} F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt.$$ 

Next, let $h_k \in L_2(\mathbb{R})$ be square integrable functions for $k = 1, \ldots, n$, introduce the scalar products

$$H_{k\ell} = \langle h_k, h_\ell \rangle = \int_{-\infty}^{\infty} h_k(t)\overline{h_\ell(t)}dt,$$

and define the matrix $H$ with elements $|H|_{k\ell} = H_{k\ell}$ for $k, \ell = 1, \ldots, n$. The matrix $H$ is the Gramian of the functions $h_1, \ldots, h_n$.

**Proposition 1 (Positive Definiteness of the Gramian Matrix).** Let $h_1, \ldots, h_n \in L_2(\mathbb{R})$ and consider the Gram matrix

$$H = \begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1n} \\ H_{21} & H_{22} & \cdots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} & H_{n2} & \cdots & H_{nn} \end{pmatrix},$$

where $H_{k\ell} = \langle h_k, h_\ell \rangle$. Then, $H \succeq 0$. Furthermore, $H$ is invertible if and only if $h_1, \ldots, h_n$ are linearly independent.

**Proof.** See [11, pp. 12-13].

We will consider Mazo’s model [2] consisting of a finite number of modulated sinc pulses transmitted at a rate that is higher than the Nyquist rate. Let $T = \frac{1}{2W}$ and

$$g(t) = \sqrt{2W} \cdot \text{sinc}(2Wt) = \sqrt{T} \cdot \frac{\sin \pi t}{\pi t}.$$ 

Figure 1. Signal of the form (2) where $\rho = 0.81$, $n = 20$, and $T = 1$ with more than 50% of the energy is outside the interval $\Omega = [-15, 34]$.

Introduce

$$h_k(t) = g\left(t - \rho k - \frac{1}{2W}\right)$$

for $\rho \in (0, 1]$. For any real number $\tau$, we have that

$$g(t - \tau) \xrightarrow{\mathcal{F}} G(\omega)e^{-i\omega \tau}$$

and so the spectrum of $g(t - \tau)$ is invariant under (time) shifting since

$$|G(\omega)e^{-i\omega \tau}| = |G(\omega)|.$$ 

Now, let $X(t)$ be a signal consisting of $m = \lfloor n/\rho \rfloor$ samples transmitted with a rate that is faster than Nyquist (that is $\rho < 1$), given by

$$X(t) = \sum_{k=1}^{m} A_k h_k(t).$$

where $\{A_k\}$ are real valued random variables. It’s faster than Nyquist in the sense that the pulses are spaced in-between with a time distance $\rho T$ instead of $T$. Thus, the sample density is $\frac{1}{\rho T} = \frac{2W}{\rho}$ samples/second.

**Lemma 1.** The functions

$$h_k(t) = g\left(t - \rho k - \frac{1}{2W}\right),$$

$k = 1, \ldots, m$, are linearly independent.

**Proof.** See the appendix.
sinc pulses centered in \([0,4)\), for \(\rho=1/4\)
sinc pulse centered at \(t=6\)

Figure 2. Top: sinc pulse centered at \(t=6\). Middle: Basis functions for \(\rho\in\{1/2,1/3,1/4\}\).

III. TIME LOCALIZATION AND FTN

A signal is well localized to a time interval \(\Omega \subset \mathbb{R}\) if the signal energy outside the interval is small compared to the total signal energy, i.e., if

\[
K_D(X, \Omega) = \frac{\int_{\Omega} |X(t)|^2 \, dt}{\|X(t)\|^2} \geq 1 - \mu \tag{3}
\]

where \(\mu\) is close to zero.

An implicit assumption when considering signals \(X(t)\) of the form (2) is that the signal is well localized in time to the interval \(\Omega = [0,(n-1)T]\) (that is, approximately time-limited). However, in the faster than Nyquist framework, this assumption is violated and we will see one can create signals with a considerable proportion of its energy content outside this region. This implies that signals of the form (2) with \(\rho \ll 1\) may not satisfy (3) even for moderate size \(\mu\) and intervals \(\Omega\) that contains \([0,(n-1)T]\), for example \(\Omega = [-M,(n-1)T+M]\), for some \(M > 0\) (e.g., as depicted in Figure 1 with \(M = 15\)).

To see another example of this, consider pulses consisting of the (non-orthogonal) basis functions \(g(t-\rho(k-1))\) for \(k = 0, \ldots, m = \lfloor 4/\rho \rfloor\) for a given \(\rho < 1\) and where \(W = 1/2\). Figure 2 shows the best quadratic approximations of the sinc pulse \(g(t-6)\) for \(\rho \in \{1/2,1/3,1/4\}\). By Lemma 1 such approximation is not exact, but the approximation is very close for \(\rho = 1/4\). This example shows that using the faster than Nyquist framework with small \(\rho\), one can create signals with support outside the designated “time slot” of the signal, for example approximating a sinc pulse centred outside the interval \([0,(n-1)T]\). In fact, it is possible for each fixed \(n > 0\) to approximate any band limited signal as a faster than Nyquist pulse if the rate \(\rho\) is sufficiently small. This is the content of the following proposition.

**Proposition 2.** Let \(n > 0\) be fixed. Any \(L_2(\mathbb{R})\) function whose Fourier Transform is restricted to the frequency band \([-W,W]\) can be approximated arbitrary close (in \(L_2(\mathbb{R})\)) by a function \(X(t)\) of the form (2) for some positive number \(\rho < 1\).

**Proof.** See the Appendix.

This effect may also be seen when the alphabet is constrained to a discrete set of points, e.g., for binary symbols. To illustrate this, let the signal be of the form (2) where \(A_k \in \{-1,+1\}\) with \(T = 1\), \(\rho = 0.9\), and \(m = 400\). The signal corresponding to \(A_k = (-1)^k\) is depicted in Figure 3. The darker region of the signal, between the vertical bars, is the time \([-\rho,\rho m]\) where the signal energy is supposed to be localized. However, as can be seen a considerable proportion of the energy falls outside this region. In fact numerical integration shows that about 75% of the energy is confined in the interval \(t \in [-\rho,\rho m]\), and only about 27% to the interval \(t \in [0,\rho (m-1)]\).

IV. TIME LOCALIZATION AND CAPACITY OF FTN

Next we examine the FTN model from a viewpoint of channel capacity. Assuming that the FTN model (2) only contains signals that are well localized to the interval \([0,(n-1)T]\), we will see that the Shannon capacity can be violated. Again this illustrates that the FTN model contains signals that are not well localized in time.
The energy of a signal given by (2) is
\[ E∥X∥^2 = E \left( \sum_{k=1}^{m} A_k h_k \right)^2 \]
\[ = E \left( \sum_{k=1}^{m} A_k h_k, \sum_{\ell=1}^{m} A_\ell h_\ell \right) \]
\[ = E \sum_{k=1}^{m} \sum_{\ell=1}^{m} H_{k\ell} A_k A_\ell \]
\[ = E A^T H A. \]

Let \( E_s \) be the energy per sample in the case of \( \rho = 1 \). We will restrict the energy to not exceed the energy for the Nyquist signaling case, that is, the same as the energy for signaling \( n \) samples. For \( \rho = n/m \), we have
\[ E A^T H A = m \rho E_s = \frac{n}{\rho} \rho E_s = n E_s \]
which gives an average energy per sample to be
\[ \frac{1}{m} E A^T H A = \rho E_s. \]

Since the time duration between consecutive samples is \( \rho T \), the time spanned of \( m \) samples is \( m \rho T \). Now if the non-orthogonal pulses were well localized in time, we would get over an infinite time horizon the average power
\[ \lim_{m \to \infty} \frac{1}{m \rho T} E A^T H A = \frac{\rho E_s}{\rho T} = P. \]  
(4)

Thus, the average power does not exceed that of the Nyquist signaling case. The received signal over the AWGN channel is thus given by
\[ Y(t) = X(t) + Z(t) \]
where \( Z(t) \) is a zero mean white noise Gaussian process. We define the measurements
\[ Y_k = \langle h_k, Y \rangle = \langle h_k, X + Z \rangle = \langle h_k, \sum_{\ell=1}^{m} A_\ell h_\ell \rangle + Z_k \]
\[ = \sum_{\ell=1}^{m} \langle h_k, h_\ell \rangle A_\ell + Z_k \]
\[ = \sum_{\ell=1}^{m} H_{k\ell} A_\ell + Z_k. \]  
(5)

We may write (5) in the more compact form
\[ Y = HA + Z. \]  
(6)

It’s not hard to check that \( \{ Y_k \} \) is not a sequence of independent variables since \( \{ h_k \} \) is not an orthogonal set of functions. The covariance of \( Z_k \) and \( Z_\ell \) is given by
\[ E \{ Z_k Z_\ell \} = \frac{N_0}{2} \langle h_k, h_\ell \rangle = \frac{N_0}{2} H_{k\ell}. \]  
(7)

The functions \( \{ h_k \} \) are linearly independent according to Lemma 1, and Proposition 1 implies that \( H > 0 \). Let \( H^{\frac{1}{2}} > 0 \) be the unique positive definite matrix such that \( H^{\frac{1}{2}} H^{\frac{1}{2}} = H \). Then, we may write
\[ Z = H^{\frac{1}{2}} V, \quad V \sim N(0, \frac{N_0}{2} I_m) \]
and (6) is equivalent to
\[ Y = HA + H^{\frac{1}{2}} V. \]  
(8)

Since \( H \) is positive definite, it’s invertible, and so is \( H^{\frac{1}{2}} \). Multiplying the left and right hand side of Equation (8) by \( H^{-\frac{1}{2}} \) gives
\[ S = H^{-\frac{1}{2}} Y = H^{-\frac{1}{2}} A + V. \]  
(9)

Now let \( X = H^{-\frac{1}{2}} A \), using the precoding \( A = H^{-\frac{1}{2}} X \). This is always possible since \( H \) is nonsingular. The energy constraint on \( X \) is then
\[ \sum_{k=1}^{m} E \{ X_k \}^2 = E \{ X \}^2 = E A^T H A = \rho m E_s. \]

The channel capacity over the discrete time channel
\[ S_k = X_k + V_k, \quad V_k \sim N(0, \frac{N_0}{2}), \quad k = 1, ..., m, \]  
(10)
with a sequence \( \{ X_k \} \) of total energy
\[ \sum_{k=1}^{m} E \{ X_k \}^2 = \rho m E_s \]
is maximized for a sequence \( \{ X_k \} \) of independent identically distributed Gaussian variables with \( X_k \sim N(0, \rho E_s) \). Thus, the capacity of \( m \) samples is
\[ C_m = m \cdot \frac{1}{2} \log_2 \left( 1 + \frac{2 \rho E_s}{N_0} \right) \]
\[ = m \cdot \frac{1}{2} \log_2 \left( 1 + \frac{\rho P}{N_0 W} \right) \quad \text{bits}. \]

Recall that the sample density is \( \frac{2W}{\rho} \) samples/second. Then, the capacity per sample is \( C_m/m \) and the average capacity in bits per second as time goes to infinity is
\[ C(\rho) = \lim_{m \to \infty} \frac{2W}{\rho} \cdot \frac{C_m}{m} \]
\[ = W \log_2 \left( 1 + \frac{\rho P}{N_0 W} \right) \quad \text{bits/second}. \]

From the derivations above, we see that Mazo’s model implies that the channel capacity in fact increases as we pack signals tighter in time, that is when \( \rho \) decreases. This rather unexpected (and false) result may be explained away by considering the energy localization of signals in time. The signals are not well localized in time and the limit in the expression of the average power in (4) is not equal to \( P \) but is equal to a smaller number since the energy could be arbitrarily small over a time period \( m \rho T \).
be approximated by a circulant matrix and the algorithm can be well approximated by an implementation using FFT with complexity $O(n \log(n))$. Note that if $\rho$ is selected to be too small, then the eigenvalues of the Toeplitz matrix $H$ will also be small and this approximate approach will break down. The implementation used in following subsection is done using both this method and an implementation based on full size matrices, and the results are identical for all the simulations.

V. FTN FOR NON IDEAL PULSES

Consider a pulse shape $g(t)$ with Fourier transform $G(\omega)$ where $G(\omega) \neq 0$ for $|\omega| \leq W'$ and $G(\omega) = 0$ for $|\omega| > W'$. If the transmitted time-shifted pulses $g(t-kT)$ are to be orthogonal, the time shift $T$ must be larger than $T^* = \frac{1}{2W'}$, unless $g(t)$ is a sinc pulse [12]. However, by using the precoding described in Section IV, we can transmit the non-orthogonal pulses $h_k(t) = g(t-kT')$ and get an easy detection algorithm (see Section IV). Then, the channel capacity becomes

$$C = W' \log_2 \left(1 + \frac{P}{N_0 W} \right) \text{ bits/second,}$$

which is exactly the capacity for sinc pulses with average transmit power $P$ and bandwidth $W'$. One can show that the signal $X(t) = \sum_{k} A_k h_k(t)$ is well localized in time when $n$ is large enough. For more details, see [13].

A. Complexity and implementation

In general the problem of estimating the transmitted sequence is NP-hard [9] in the case of inter-symbol interference and usually, one has to rely on the Viterbi algorithm which has exponential complexity [10].

However, the precoding scheme that has been suggested in Section IV reduces the problem to an independent estimation for each sample, which is a much easier problem. A block diagram of the process can be found in Figure 4. Starting from the transmitter side we take the discrete data $X$ as input. This is precoded using $A = H^{-1/2} X$ and sent through a pulse shaping filter (2) over the AWGN-channel. On the receiver side this is sampled using a matched filter, these samples are decoded to give the discrete output samples $S$, which according to (10) are decoupled such that $S_k$ only depends on $X_k$.

For many cases the algorithm can be implemented efficiently for large $n$ utilizing the fact that the Toeplitz matrix $H$ is asymptotically circulant [14]. For such cases, $H$ can be approximated by a circulant matrix and the algorithm can be implemented using FFT with complexity $O(n \log(n))$. Note that if $\rho$ is selected to be too small, then the eigenvalues of the Toeplitz matrix $H$ will also be small and this approximate approach will break down. The implementation used in following subsection is done using both this method and an implementation based on full size matrices, and the results are identical for all the simulations.

B. Example - Root-Raised-Cosine

We exemplify the theory by applying it to the root-raised-cosine pulse, which also demonstrates how it works for other pulses.

As before, let $T = \frac{1}{2W'}$, and introduce

$$g_\beta(t) = \frac{4\beta}{\pi \sqrt{T}} \cdot \cos \left(\left(1 + \beta\right)\frac{\pi}{T} t \right) + \frac{\sin \left(\left(1 - \beta\right)\frac{\pi}{T} t \right)}{4\beta T} ,$$

defined for $\beta \in [0, 1]$.

One can show that with $\rho = \frac{1}{2W'}$, the capacity for precoded FTN with root-raised-cosine becomes

$$C = \left(1 + \beta\right) W \log_2 \left(1 + \frac{P}{N_0 W (1 + \beta)} \right) \text{ bits/second,}$$

which is also the capacity for transmitting with orthogonal sinc pulses over the bandwidth $(1 + \beta)W$. It can also be shown that the resulting signal is well localized in time but this is omitted because of space constraints, see [13].

C. Simulation results

As a proof of concept, we have conducted simulations where the suggested precoding is applied and implemented, using FFT approximation as described above. These simulations are based on the model presented in (8).

We apply the root-raised-cosine function as a pulse shape and use a roll-off factor $\beta = 0.22$ to mimic existing standards as in [15]. We are also only looking at binary input-output and don’t use any higher order modulation.

These simulations are such that we keep the time for a block constant, with the reference that in the Nyquist case ($\rho = 1$) one block should be 4000 physical bits. The Nyquist transmission also serves as a reference in the sense that these pulses are of unit energy. We are keeping the power constant for all the schemes, meaning that FTN must use a lower energy per physical bit. This in turn relates to the SNR given in the plots, as this is a Nyquist reference type of SNR in order to be able to make a fair comparison between the different schemes. The SNR is given in dB and relates to the standard deviation of the sampled noise as

$$\sigma = 10^{-\text{SNR}_{\text{dB}}/10},$$

since Nyquist transmissions applied unit energy per physical bit. This standard deviation is also used for FTN transmissions regardless of the fact that FTN is using lower energy per physical bit, so the experienced SNR per physical bit will be worse than what is actually written on the axes in the FTN case. This comparison is motivated since it allows us to judge,
for a given channel state, if it is worth changing a Nyquist scheme for FTN.

In addition, this simulation also applies (WCDMA) turbo codes according to [16]. The coding is applied on the payload bits to create the physical bits, \( X \), on which we then apply the precoding. Similarly at the receiver we first apply FTN decoding, to produce \( S \), before using turbo decoding which gives the estimated payload bits. For every value of \( \rho \) and SNR presented, we have looked at code rates from \( \frac{1}{3} \) up to \( \approx 0.96 \) in a grid containing 18 points (exact code rates are rounded due to finite block length). From this, we have then computed the throughput as:

\[
\text{throughput} = (1 - \text{block-error rate}) \cdot \text{#payload bits},
\]

and selecting the code rate giving highest throughput. The throughput is given as bits per equivalent time unit and the reference is, as mentioned, the Nyquist scheme transmits 4000 physical bits, the exact bandwidth is then just a scaling from this.

Another way of calculating the throughput, although less attractive for practical purposes, is to rely on the bit-error rate instead of the block-error rate. The resulting throughput can be found in Figure 6. It should however be noted that this optimization resulted in all schemes using the highest possible code rate for all SNR.

VI. CONCLUSIONS

We considered communication over the analog white Gaussian noise channel using a finite bandwidth \([-W, W]\) and non-orthogonal pulses by signaling at a rate that is higher than the Nyquist rate. We showed that the conclusions in [2], that one may transmit symbols carried by sinc pulses at a higher rate than that dictated by Nyquist without loosing in bit error rate don’t imply that the bit error rate per time unit decreases. This was demonstrated by showing that if the model in [2] is valid to consider bit error rates per time unit, then it means that non-orthogonal signals may achieve a capacity for the AWGN channel that is higher than the Shannon capacity. We explain this phenomenon by means of an example where we show that non-orthogonal signals do not give rise to well localized energy in time. Thus, it’s not physically correct to talk about bits per second, as the energy of non-orthogonal signals may be more spread over time.

We also considered FTN signaling in the case of pulses that are different from the sinc pulses. We showed that one may use a precoding scheme of low complexity, in order to remove the inter-symbol interference. This leads to the possibility of increasing the number of transmitted samples per time unit and compensate for spectral efficiency losses due to signaling at the Nyquist rate of the non sinc pulses. Thus we can achieve the Shannon capacity when the same energy is spent on transmitting with the ideal sinc pulses.

VII. ACKNOWLEDGEMENTS

This work was in part supported by Bitynamics Research, Ericsson Research, and the ACCESS Linnaeus Centre at KTH.
Appendix

A. Proof of Lemma 1

To start with, consider the following proposition. It may be shown using induction over $n$. See [13] for a detailed proof.

**Proposition 3.** Let $\Omega \subset \mathbb{R}$ with $\text{int}(\Omega) \neq \emptyset$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ with $\alpha_i \neq \alpha_j$ for $i \neq j$. Then, $e^{\alpha_1 \omega}, \ldots, e^{\alpha_n \omega}$ are linearly independent for $\omega \in \Omega$.

Next, we will prove Lemma 1 by contradiction. Suppose that $h_1, \ldots, h_n$ are linearly dependent. Then, there exists $0 \neq (c_1, \ldots, c_n) \in \mathbb{C}^n$ such that

$$c_1 h_1(t) + \cdots + c_n h_n(t) = 0, \quad \forall t \in \mathbb{R}. $$

Taking the Fourier transform of the equation above, we get

$$c_1 H(\omega) e^{-i\omega \tau_1} + \cdots + c_n H(\omega) e^{-i\omega \tau_n} = 0, \quad \forall \omega \in \mathbb{R}. $$

In particular, we have that

$$c_1 e^{-i\omega \tau_1} + \cdots + c_n e^{-i\omega \tau_n} = 0, \quad \forall \omega \in (-W, W),$$

which implies that

$$c_1 e^{-i\omega \tau_1} + \cdots + c_n e^{-i\omega \tau_n} = 0, \quad \forall \omega \in (-W, W).$$

Now letting $\alpha_k = -i\tau_k$, for $k = 1, \ldots, n$, and using Proposition 3 with $\Omega = [-W, W]$, we see that $e^{-i\omega \tau_k}$ are linearly independent so we must have $(c_1, \ldots, c_n) = 0$, a contradiction. Therefore $h_1, \ldots, h_n$ must be linearly independent, which concludes the proof. \qed

B. Proof of Proposition 2

First, note that we may take $T = 1$ without loss of generality. Next, introduce the following two linear projection operators. Let $D : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ be the orthogonal projection onto the set of time limited function

$$(DX)(t) = \begin{cases} X(t) & t \in [0, n] \\ 0 & \text{otherwise,} \end{cases}$$

and let $B : L_2(\mathbb{R}) \to L_2(\mathbb{R})$ be the orthogonal projection onto the set of band limited functions with frequency support $[-W, W]$

$$(BX)(t) = \int_{-\infty}^{\infty} X(\tau) \text{sinc}(2W(t - \tau)) d\tau.$$ 

We now have the following lemma.

**Lemma 2.** The closure of the range of $BD$ is equal to the range of $B$, i.e., $\text{Range}(BD) = \text{Range}(B)$.

**Proof of Lemma 2.** Note that $\text{Range}(BD) \subseteq \text{Range}(B)$ since $\text{Range}(B)$ is closed. To show that $\text{Range}(BD) \supseteq \text{Range}(B)$, assume that $X(t) = \text{Range}(B) \cap \text{Range}(BD)^\perp$. By construction, we have that

$$\langle X, BDY \rangle = 0 \quad \text{for all } Y \in L_2(\mathbb{R}),$$

and hence

$$\langle BX, DY \rangle = \langle X, DY \rangle = 0 \quad \text{for all } Y \in L_2(\mathbb{R}),$$

since $B$ is an orthogonal projection and is thus self-adjoint. Now (12) holds only if $X(t) = 0$ on $t \in [0, n]$, which is only possible if $X \equiv 0$, since $X(t)$ is band limited. Therefore $\text{Range}(BD) \supseteq \text{Range}(B)$, and the lemma is complete. \qed

By Lemma 2 we may pick $Y \in L_2(\mathbb{R})$ such that $\hat{X} = BDY$ is arbitrary close to the desired band limited function. Without loss of generality such $Y$ may be chosen with support in $[0, n]$. Next, let

$$A_k = \int_{\rho(k-1)/(2W)}^{\rho k/(2W)} Y(t) dt, \quad \text{for } k = 1, \ldots, m = \lfloor n/\rho \rfloor.$$ 

With this construction $\sum_{k=1}^{m} A_k \delta(t - \rho \frac{k-1}{2W}) \to Y$ weakly as $\rho \to 0$. Since the total variation norm is uniformly bounded $\sum_{k=1}^{m} |A_k| \leq \|Y\|_{L_2}$, we have that $X(t) = \sum_{k=1}^{m} A_k h_k(t) \to X$ as $\rho \to 0$, and the proof is complete. \qed

**REFERENCES**


