Time Localization and Faster-Than-Nyquist Signaling

Ather Gattami
Ericsson Research
Stockholm, Sweden
Email: ather.gattami@ericsson.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. We show that the conclusions in [1], that one may transmit symbols carried by sinc pulses at a rate higher than that dictated by Nyquist without loosing in bit error rate don't imply that the bit error rate per time unit decreases. This is demonstrated by showing that if the faster than Nyquist model in [1] gives rise to time limited signals, then non-orthogonal signals can achieve a capacity for the AWGN channel that is higher than the Shannon capacity. We also 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 and compensate for spectral inefficiency 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 over a given bandwidth. We demonstrate the power of the precoding scheme by simulations.

I. INTRODUCTION

A. Background and previous work

In 1949, Shannon [2] presented his famous result on the capacity of the additive white Gaussian noise (AWGN) channel for band limited signals. The result was based on communication using orthogonal pulses (to avoid inter-symbol interference) which corresponds to transmission at the Nyquist sampling rate.

In [1], 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-sinc pulses when the code alphabet is finite.

The original derivation of Shannon’s capacity formula [2] is based on transmission of uncorrelated 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 is itself localized in the vicinity of that interval. In fact, as we will see in Section VI, 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 by signaling at a rate that is higher than the Nyquist rate. We show that the conclusions in [1], 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 is demonstrated by showing that if the model in [1] 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 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 both energy and average power constant) and compensate 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

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 \( \mathbf{X}X^\top = \mathbf{X} \). The empty set is denoted by \( \emptyset \) and \( \text{int}(\Omega) \) denotes the interior of the set \( \Omega \). We define the normalized sinc function as

\[
\text{sinc}(x) = \frac{\sin \pi x}{\pi x}
\]

We also define the function

\[
\text{rect}(x) = \begin{cases} 
1 & |x| \leq \frac{1}{2} \\
0 & |x| > \frac{1}{2}
\end{cases}
\]

The floor function \( \lfloor x \rfloor \) denotes the greatest integer less than or equal to \( x \).

**Definition 1** (Signal Space). Let \( \mathcal{L}_2(\mathbb{R}) \) be the Hilbert space of functions \( f : \mathbb{R} \to \mathbb{C} \) that are square integrable, endowed with the scalar product

\[
\langle f, g \rangle = \int_{-\infty}^{\infty} f(t) g(t) dt
\]

and squared norm

\[
\|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty.
\]

**Definition 2** (Fourier Transform). The Fourier transform of a function \( f \in \mathcal{L}_2(\mathbb{R}) \) is given by

\[
f(t) \overset{F}{\rightarrow} F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt.
\]

**Definition 3.** \( \mathcal{L}_1([-\pi, \pi]) \) is the Banach space of functions \( f : \mathbb{R} \to \mathbb{C} \) that are absolutely integrable, with the norm

\[
\|f\| = \int_{-\pi}^{\pi} |f(t)| dt < \infty.
\]

**Definition 4** (Gaussian White Noise Process [11]). The stochastic process \( Z(t) \) is a white Gaussian noise process if \( Z(t) \) has a constant spectral density over all frequencies.

**Proposition 1.** Let \( Z(t) \) be a zero mean white Gaussian Noise process with spectral density \( \frac{N_0}{2} \). For \( g_k, g_e \in \mathcal{L}_2(\mathbb{R}) \), define

\[
Z_k = \langle Z, g_k \rangle = \int_{-\infty}^{\infty} Z(t) g_k(t) dt
\]

and

\[
Z_e = \langle Z, g_e \rangle = \int_{-\infty}^{\infty} Z(t) g_e(t) dt.
\]

Then, \( Z_k \) and \( Z_e \) are zero mean Gaussian variables with covariance

\[
\mathbb{E}[Z_k Z_e] = \frac{N_0}{2} \langle g_k, g_e \rangle = \frac{N_0}{2} \int_{-\infty}^{\infty} g_k(t) g_e(t) dt.
\]

**Proof.** See [11].

**Proposition 2** (Shannon Capacity of the Discrete Memory-less Gaussian Channel). Let \( \{Z_i\}_{i=1}^{\infty} \) be a set of independent Gaussian variables with \( Z_i \sim \mathcal{N}(0, \frac{N_0}{2}) \) and consider the Gaussian channel with input sequence of samples \( \{X_i\}_{i=1}^{n} \) and output

\[
Y_i = X_i + Z_i, \text{ for } i = 1, \ldots, n.
\]

Suppose that

\[
\sum_{i=1}^{n} \mathbb{E}[|X_i|^2] \leq nE_s.
\]

Then, the capacity of the channel is given by

\[
C = \log_2 \left( 1 + \frac{2E_s}{N_0} \right) \text{ bits/sample}
\]

and is achieved for \( X_i \sim \mathcal{N}(0, E_s) \).

**Proof.** See [12].

**Definition 5** (Toeplitz matrix). A matrix \( T_n \) is called Toeplitz if it has the following form

\[
T_n = \begin{pmatrix}
c_0 & c_{-1} & c_{-2} & \cdots & c_{-(n-1)} \\
c_1 & c_0 & c_{-1} & \cdots & c_{-(n-2)} \\
c_2 & c_1 & c_0 & \cdots & c_{-(n-3)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0
\end{pmatrix}
\]

**Definition 6** (Associate function). Let \( T_n \) be a Toeplitz matrix according to Definition 5. A function \( f \in \mathcal{L}_1([-\pi, \pi]) \) is called an associate function to \( T_n \) if \( f(z) \) is associated with the corresponding Fourier series of the matrix elements i.e.,

\[
f(z) \sim \sum_{k=-\infty}^{\infty} c_k e^{ikz},
\]

where

\[
c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikz} f(z) dz.
\]

The matrix is sometimes denoted \( T_n(f) \), and in the infinite case we write \( T(f) \).

The notation is motivated by the fact that the associated function determines the distribution of the eigenvalues of the matrix. To state this, we need the following definition.

**Definition 7** (Equal distribution). Let \( M \) be a constant in \( \mathbb{R} \) and 0 < \( n \in \mathbb{N} \). Also, let \( \{a_{\ell}^{(n)}\}_{\ell=1}^{n} \) and \( \{b_{\ell}^{(n)}\}_{\ell=1}^{n} \) be sets in \( \mathbb{R} \) such that

\[
|a_{\ell}^{(n)}| < M, \quad |b_{\ell}^{(n)}| < M, \quad \text{for all } 1 \leq \ell \leq n
\]
We say that \( \{a^{(n)}_\ell\}_{\ell=1}^n \) and \( \{b^{(n)}_\ell\}_{\ell=1}^n \) are equally distributed on \([-M, M]\) if for any continuous function \( F: [-M, M] \to \mathbb{R} \), we have that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^n F(a^{(n)}_\ell) - F(b^{(n)}_\ell) = 0.
\]

**Proposition 3 (Spectrum of a Toeplitz matrix).** Let \( T_n \) be a Toeplitz matrix with associated function \( f \). Also let \( \lambda^{(n)}_\ell \), \( \ell = 1, 2, \ldots, n \), be the eigenvalues of \( T_n \). Then,
\[
\inf_z f(z) \leq \lambda^{(n)}_\ell \leq \sup_z f(z).
\]
Moreover, the sets \( \{\lambda^{(n)}_\ell\}_{\ell=1}^n \) and \( \{f\left(\frac{2\pi \ell}{n} - \pi\right)\}_{\ell=1}^n \) are equally distributed.

**Proof.** See [13] or [14].

**Proposition 4 (Associate function of the square root inverse).** Let \( T_n \) be a Toeplitz matrix with associate function \( f \) and let \( f(z) > 0 \) for all \(-\pi \leq z \leq \pi\). Moreover, let \( K_n \) be the inverse of the hermitian square root of \( T_n \). \( K_n = (T_n \frac{1}{2})^{-1} \).

Then, \( K_n \) is asymptotically Toeplitz with associate function \( 1/\sqrt{f(z)} \).

**Proof.** The proposition follows from application of Theorem 5.3 (c) and Theorem 5.2 (c) in [14].

**Definition 8 (Circulant matrix).** A matrix \( C_n \) is called Circulant if it has the following form
\[
C_n = \begin{pmatrix}
c_0 & c_1 & c_2 & \cdots & c_{n-1} \\
c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_1 & c_2 & c_3 & \cdots & c_0
\end{pmatrix}.
\]
It is thus a special case of a Toeplitz matrix.

**Proposition 5 (Diagonalizing a Circulant matrix).** Let \( C_n \) be a Circulant matrix. Then, \( C_n = U \Lambda_n U^* \), where \( \Lambda_n = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \) and \( U \) is the Fourier matrix: \( [U]_{m,\ell} = e^{-2\pi im\ell/n} \) for \( m, \ell = 0, 1, \ldots, n - 1 \).

**Proof.** See [14] and [15].

To be concise and to avoid ambiguities regarding the simulations we also define the following words:

**Definition 9 (Payload bits).** A payload bit is a bit realized from some input distribution. This bit represents pure data that we want to communicate and it is on sets of these that we compute block-error rates (BLER) and bit-error rates (BER); these are also used in the measure of throughput.

**Definition 10 (Physical bits).** A physical bit is the bit physically transferred over the channel. A physical bit represents some type of coded realization of a payload bit. The number of physical bits, denoted \#physical bits, is
\[
\#\text{payload bits} = \#\text{physical bits} \cdot \text{code rate}.
\]

When measuring signal-to-noise ratio (SNR), it will be done with respect to physical bits.

**III. PROPERTIES OF NON-ORTHOGONAL PULSES**

**Proposition 6.** 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, \( \exp(\alpha_i \omega), \ldots, \exp(\alpha_n \omega) \) are linearly independent for \( \omega \in \Omega \).

**Proof.** See the Appendix.

**Lemma 1 (Linear independence of non-orthogonal pulses).** Let \( \Omega \subset \mathbb{R} \) with \( \text{int}(\Omega) \neq \emptyset \), \( h(t) \in L_2(\mathbb{R}) \), and
\[
h_k(t) = h(t - \tau_k) \xrightarrow{\mathcal{F}} H(\omega)e^{i\omega \tau_k}, \quad k = 1, \ldots, n,
\]
for arbitrary \( \tau_1, \ldots, \tau_n \in \mathbb{R} \) with \( \tau_i \neq \tau_j \) for \( i \neq j \). Suppose that \( H(\omega) \neq 0 \) for \( \omega \in \text{int}(\Omega) \). Then, the functions \( h_1, \ldots, h_n \) are linearly independent for all \( n \in \mathbb{N} \).

**Proof.** The proof is deferred to the appendix.

Now introduce the scalar products
\[
H_{kl} = \langle h_k, h_l \rangle = \int_{-\infty}^{\infty} h_k(t) \overline{h_l(t)} dt \quad (1)
\]
and define the matrix \( H \) whose elements \( [H]_{kl} = H_{kl} \) for \( k, \ell = 1, \ldots, n \). The matrix \( H \) is the Gramian of the functions \( h_1, \ldots, h_n \).

**Proposition 7 (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 \( H_{kl} = \langle h_k, h_l \rangle \). Then, \( H \succeq 0 \). Furthermore, \( H \) is invertible if and only if \( h_1, \ldots, h_n \) are linearly independent.

**Proof.** Consult [16], pp. 12–13.

**IV. COMMUNICATION WITH ORTHOGONAL PULSES**

We will start by reviewing the derivation of Shannon for the capacity of an AWGN channel using sinc pulses transmitted with the Nyquist rate. Let \( T = \frac{1}{2W} \) and
\[
g(t) = \sqrt{2W} \cdot \text{sinc}(2Wt) = \sqrt{T} \cdot \frac{\sin 2\pi Wt}{\pi t}.
\]
The Fourier transform of the function \( g(t) \) is given by
\[
g(t) \xrightarrow{\mathcal{F}} G(\omega)
\]
where
\[
G(\omega) = \sqrt{2W} \cdot \text{rect}\left(\frac{\omega}{2W}\right).
\]
Next, introduce
\[ g_k(t) = g \left( t - \frac{k - 1}{2W} \right). \]
A signal \( X(t) \) consisting of \( n \) Nyquist pulses is given by
\[ X(t) = \sum_{k=1}^{n} A_k g_k(t) \tag{2} \]
where \( \{A_k\} \) are real valued random variables. Define the \textit{sample density} as \( 2W = \frac{1}{t} \) samples/second. Since \( g_k \) and \( g_\ell \) are orthogonal for \( k \neq \ell \), we have that \( \langle g_k, g_\ell \rangle = 0 \), and
\[
E[|X|^2] = E \left\| \sum_{k=1}^{n} A_k g_k \right\|^2
= E \left( \sum_{k=1}^{n} A_k g_k, \sum_{\ell=1}^{n} A_{\ell} g_\ell \right)
= E \sum_{k=1}^{n} \sum_{\ell=1}^{n} \langle g_k, g_\ell \rangle A_k A_\ell
= \sum_{k=1}^{n} E|A_k|^2.
\]
Thus, for \( n \) samples, the average energy per sample is given by
\[
\frac{1}{n} \sum_{k=1}^{n} E|A_k|^2.
\]
For a zero mean ergodic process \( \{A_k\} \) the limit of the average above exists as \( n \to \infty \) and is given by
\[
E_s = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} E|A_k|^2.
\]
Furthermore, the average power is given by
\[
P = \lim_{n \to \infty} \frac{1}{nT} \sum_{k=1}^{n} E|A_k|^2 = E_s.
\]
Thus, we have that
\[
E_s = P \cdot T = \frac{P}{2W}. \tag{3}
\]
The received signal over the AWGN channel is given by
\[ Y(t) = X(t) + Z(t) \]
where \( Z(t) \) is a zero mean white noise Gaussian process. We define the measurements
\[
Y_k = \langle g_k, Y \rangle = \langle g_k, X + Z \rangle = \langle g_k, X \rangle + \langle g_k, Z \rangle = A_k + Z_k.
\]
It follows from Proposition 1 that \( \{Z_k\} \) is a sequence of independent identically distributed Gaussian variables since \( \{g_k\} \) is an orthogonal set of functions. The variance of \( Z_k \) is
\[
E(Z_k^2) = \frac{N_0}{2} \int_{-\infty}^{\infty} |g_k(t)|^2 dt = \frac{N_0}{2}.
\]
For a sequence of \( \{A_k\} \) with total energy
\[
\sum_{k=1}^{n} E|A_k|^2 = nE_s
\]
the channel capacity over the AWGN is maximized for \( A_k = X_k \), where \( \{X_k\} \) is a sequence of independent identically distributed Gaussian variables with \( X_k \sim \mathcal{N}(0, E_s) \). Thus, the capacity of \( n \) samples is
\[
C_n = n \cdot \frac{1}{2} \log_2 \left( 1 + \frac{2E_s}{N_0} \right)
= n \cdot \frac{1}{2} \log_2 \left( 1 + \frac{P}{N_0 W} \right) \text{ bits.}
\]
Recall that the sample density is \( 2W \) samples/second. Then, the capacity per sample is \( C_n/n \) and the average capacity in bits per second as time goes to infinity is
\[
C = \lim_{n \to \infty} 2W \cdot \frac{C_n}{n}
= W \log_2 \left( 1 + \frac{P}{N_0 W} \right) \text{ bits/second}
\]
which is the well-known Shannon capacity result of the AWGN channel.

V. COMMUNICATION WITH NON-ORTHOGONAL PULSES

Again, we will use Mazo’s model [1] of considering a finite number of modulated sinc pulses transmitted at a rate that is higher than the Nyquist rate, and let the number of pulses go to infinity.

Let \( T = \frac{1}{\pi W} \) and
\[
g(t) = \sqrt{2W} \cdot \text{sinc} (2Wt) = \sqrt{T} \cdot \frac{\sin 2\pi W t}{\pi t}.
\]
Introduce
\[
h_k(t) = g \left( t - \rho \frac{k - 1}{2W} \right)
\]
for \( \rho \in (0, 1] \). Clearly, \( h_k = g_k \) for \( \rho = 1 \).

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
\[
|G(\omega) e^{-i\omega \tau}| = |G(\omega)|.
\]
Let \( X(t) \) be a signal consisting of \( m = \lfloor n/\rho \rfloor \) samples transmitted with a rate that is \textit{faster than Nyquist} (that is \( \rho < 1 \)), given by
\[
X(t) = \sum_{k=1}^{m} A_k h_k(t) \tag{4}
\]
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 \( \frac{1}{\rho T} = \frac{2W}{n} \) samples/second. Note that we may take \( \rho = \frac{n}{q} \) for \( n < q \in \mathbb{N} \) and obtain \( m = n/\rho = q \) exactly.

**Lemma 2.** The functions

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

\( k = 1, \ldots, m, \) are linearly independent.

**Proof.** Let \( \tau_k = \rho(k-1)/2W \). Then,

\[
g(t - \tau_k) \xrightarrow{\mathcal{F}} G(\omega)e^{-i\omega \tau_k}
\]

where

\[
G(\omega) = \sqrt{2W} \cdot \text{rect} \left( \frac{\omega}{2W} \right).
\]

Clearly, \( G(\omega) \neq 0 \) for \( \omega \in \Omega = [-W, W] \) and \( \text{int}(\Omega) \neq \emptyset \). Hence, it follows from Lemma 1 that \( h_k(t) = g(t - \tau_k) \) are linearly independent for \( k = 1, \ldots, m \).

Now the energy of the signal is

\[
E\|X\|^2 = E \left\| \sum_{k=1}^{m} A_k h_k \right\|^2 = E \left( \sum_{k=1}^{m} A_k^* h_k, \sum_{\ell=1}^{m} A_\ell h_\ell \right)
\]

\[
= E \sum_{k=1}^{m} \sum_{\ell=1}^{m} \langle h_k, h_\ell \rangle A_k A_\ell
\]

\[
= E \sum_{k=1}^{m} \sum_{\ell=1}^{m} H_{k\ell} A_k A_\ell
\]

\[
= E A^T H A.
\]

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:

\[
E A^T H A = m \rho E_s \leq \frac{n}{\rho} \rho E_s = n E_s
\]

which gives an average energy per sample to be

\[
\frac{1}{m} E A^T H A \leq \rho E_s.
\]

Since the time duration between consecutive samples is \( \rho T \), the time spanned of \( m \) samples is \( m \rho T \). Over an infinite time horizon, the average power is given by

\[
\lim_{m \to \infty} \frac{1}{m \rho T} E A^T H A \leq \frac{\rho E_s}{\rho T} = P
\]

Thus, the average power does not exceed that of the Nyquist signaling case. As in the previous section, the received signal over the AWGN channel is given by

\[
Y(t) = X(t) + Z(t)
\]

where \( Z(t) \) is a zero mean white noise Gaussian process. In the same way, we define the measurements

\[
Y_k = \langle h_k, Z \rangle = \langle h_k, X + Z \rangle
\]

\[
= \langle h_k, X \rangle + \langle h_k, Z \rangle
\]

\[
= \frac{1}{2} \left( \langle h_k, \sum_{\ell=1}^{m} A_\ell h_\ell \rangle + Z_k \right)
\]

\[
= \sum_{\ell=1}^{m} \langle h_k, h_\ell \rangle A_\ell + Z_k
\]

\[
= \sum_{\ell=1}^{m} H_{k\ell} A_\ell + Z_k.
\]

We may write (5) in the more compact form

\[
Y = HA + Z.
\]

It follows from Proposition 1 that \( \{Z_k\} \) is no longer a sequence of independent Gaussian variables since \( \{h_k\} \) is not an orthogonal set of functions. The covariance of \( Z_k \) and \( Z_\ell \) is given by

\[
E\{Z_k Z_\ell\} = \frac{N_0}{2} \langle h_k, h_\ell \rangle = \frac{N_0}{2} H_{k\ell}.
\]

The functions \( \{h_k\} \) are linearly independent according to Lemma 2, and Proposition 7 implies that \( H \succ 0 \) be the unique matrix such that \( H^{1/2} \cdot H^{1/2} = H \). Then, we may write

\[
Z = H^{1/2} V, \quad V \sim \mathcal{N}(0, \frac{N_0}{2} I_m)
\]

and (6) is equivalent to

\[
Y = HA + H^{1/2} V.
\]

Since \( H \) is positive definite, it’s invertible, and so is \( H^{1/2} \). Multiplying the left and right hand side of Equation (8) by \( H^{-1/2} \) gives

\[
S = H^{-1/2} Y = H^{1/2} A + V.
\]

Now let \( X = H^{1/2} A \). Note that the preceding \( A = H^{-1/2} X \) gives \( H^{1/2} A = X \), so we may always make this substitution. The energy constraint on \( X \) is

\[
\sum_{k=1}^{m} E|X_k|^2 = E\|X\|^2 = E A^T H A \leq pm E_s.
\]

The channel capacity over the discrete time channel

\[
S_k = X_k + V_k, \quad V_k \sim \mathcal{N}(0, \frac{N_0}{2}), \quad k = 1, \ldots, m,
\]

with a sequence \( \{X_k\} \) of total energy

\[
\sum_{k=1}^{m} E|X_k|^2 = pm E_s
\]
is maximized for a sequence \{X_k\} of independent identically distributed Gaussian variables with \(X_k \sim N(0, \rho E_n)\). Thus, the capacity of \(m\) samples is
\[
C_m = m \cdot \frac{1}{2} \log_2 \left( 1 + \frac{2\rho E_n}{N_0} \right) = m \cdot \frac{1}{2} \log_2 \left( 1 + \frac{\rho P}{N_0 W} \right) \text{ bits.}
\]

Recall that the sample density is \(\frac{2W}{T}\) 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
\[
C(\rho) = \lim_{m \to \infty} \frac{2W}{\rho} \cdot \frac{C_m}{m} = \frac{W}{\rho} \log_2 \left( 1 + \frac{\rho P}{N_0 W} \right) \text{ bits/second.}
\]

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 result may be explained by considering the energy localization of signals in time, as will be presented in the next section.

**VI. ENERGY LOCALIZATION IN TIME**

An implicit assumption when considering signals \(X(t)\) of the forms (2) and (4) is that the signal is localized in time to the interval \([0, n-1]\). However, in the faster than Nyquist framework of Mazo, this assumption is violated and as we will see one can create signals with a considerable proportion of its energy content outside this region. This gives insights in why communication in the previous section is not dependent on \(n\), as \(\rho \to 0\), since the class of functions is not restricted to the given interval.

First let the set of pulses we consider be the 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 \(g(t-\rho)\) with 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. In fact, it is possible 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 theorem.

**Proposition 8.** 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 (4) for some positive number \(\rho < 1\). \(\square\)

**Proof.** See the Appendix.

It is possible to construct examples where the energy is not localized to the interval \([0, n-1]\) also for \(\rho\) larger than the Mazo limit. Consider the problem of determining the maximal energy outside the interval \(\Omega_m := [-m, m+n-1]\) for a signal \(X(t)\) of the form (4), i.e.,
\[
\max_{X(t) \text{ s.t. } (4)} \int_{\mathbb{R}\setminus\Omega_m} X(t)^2 dt \text{ subject to } \|X(t)\|_2 \leq 1.
\]

Note that this optimization problem may be written as an eigenvalue problem
\[
\max_{A \in \mathbb{R}^m} A^T(H – H(\Omega_m))A \text{ subject to } A^T HA \leq 1,
\]
where
\[
H(\Omega_m)_{k,\ell} = \int_{\Omega_m} h_k(t)\overline{h_\ell(t)} dt,
\]
and the maximum corresponds to the largest eigenvalue of the matrix \(I – H^{-1}H(\Omega_m)\). Figure (3) shows how the maximal
energy outside $\Omega_m$ depends on $m$ where $\rho \in \{0.7, 0.81, 0.9, 1\}$ for the case $n = 20$. Figure (4) shows an example where $\rho = 0.81$ and more than 50% of the energy is outside the interval $\Omega_{15} = [-15, 34]$.

In the case of constrained alphabets we look at binary signaling, and the example shown in Figure 5. This figure shows sinc pulses with $T = 1$ and $\rho = 0.9$, modulated with the sequence $-1, +1, -1, \ldots$ of length 400 bits with the main peaks in $\rho T, 2\rho T, \ldots 400\rho T$. The darker region of the signal, between the vertical bars, is the time where the signal energy is supposed to be localized. However, we can visually see that a lot of energy falls outside this region. In fact numerical integration shows that about 75% of the energy is confined in the interval $t \in [0, 401\rho T]$, and only about 27% to the interval $t \in [\rho T, \ldots, 400\rho T]$.

VII. FTN FOR NON IDEAL PULSES

A. FTN Gramian and spectrum

We shall see that FTN can be applied to make good use of spectrum leakage, when optimal pulses are difficult to use. We will investigate the situation with the use of Toeplitz operators and connect the spectrum of the matrix, i.e. the eigenvalues, with the spectrum of the pulses. The result is similar to what is presented in [7] but provides additional insight since the associated function is explicitly presented, something we will make use of later. The setup in this proposition is going to be similar as in Lemma 1.

Proposition 9 (Toeplitz- and pulse-spectrum). Let $h(t) \in \mathcal{L}_2(\mathbb{R})$, and

$$ h_k(t) = h(t - \tau_k) \frac{e^{-i\omega t}}{H(\omega)} \quad k = 1, \ldots, n, $$

with the restriction that $\tau_k = \tau \cdot (k-1)$ for some constant $\tau \in \mathbb{R}^+ \setminus \{0\}$. Furthermore, let $H_n$ be such that $[H_n]_{k\ell} = \langle h_k, h_\ell \rangle.$

Then,

i) $H_n$ is a Toeplitz matrix $T_n(f)$.

ii) The associated function $f \in \mathcal{L}_1([-\pi, \pi])$ is given by

$$ f(z) = \frac{1}{\tau} \sum_{\ell=-\infty}^{\infty} \left| H \left( \frac{z + 2\pi \ell}{\tau} \right) \right|^2, $$

Figure 3. Maximal energy outside $\Omega_m = [-m, m + n - 1]$ as a function of $m$, of a signal (4) where $\rho \in \{0.7, 0.81, 0.9, 1\}$ and where $n = 20$.

Figure 4. Signal of the form (4) where $\rho = 0.81$ and $n = 20$ with more than 50% of the energy is outside the interval $\Omega_{15} = [-15, 34]$.

Figure 5. This figure shows the resulting signal when a Sinc-pulse is transmitted with a completely alternating sequence of $-1, +1, -1, \ldots$ and length 400 bits. The darker part of the signal, surrounded by vertical bars, is the part $t = 0$ and $t = 401 \cdot \rho T$ respectively; that is one symbol-time before/after the first/last of the main peaks.
Moreover, let \( g(t) \) be a given pulse, and let its Fourier transform be \( G(\omega) \) where \( G(\omega) \neq 0 \) for \( |\omega| \leq \pi W' \) and \( G(\omega) = 0 \) for \( |\omega| > \pi W' \). We want to have communication at the rate \( R = \frac{P}{N_0 W'} \) bits/second, which is exactly the capacity for sinc pulses with average transmit power \( P \) and bandwidth \( W' \).

To be able to rely on the above result, we have to show that the resulting signal is well localized in time as the number of pulses \( n \) becomes large enough. The time localization of a large number of sinc pulses transmitted at the Nyquist rate is well established (as defined and shown in [8]), and thus, we will show that as the number of pulses goes to infinity, the precoded signal approaches a train of sinc pulses transmitted at the Nyquist rate.

**Theorem 1 (Time localization of precoded signals).** Let \( g(t) \) be a given pulse, and let its Fourier transform be \( G(\omega) \) such that \( G(\omega) \neq 0 \) for \( |\omega| \leq \pi W' \) and \( G(\omega) = 0 \) for \( |\omega| > \pi W' \). Moreover, let

\[
T' = \frac{1}{2W'},
\]

and

\[
h_k(t) = g(t - kT').
\]

Let \( H_{2n+1} \) be a \( (2n+1) \times (2n+1) \) matrix where \( [H_{2n+1}]_{k,\ell} = \langle h_k, h_\ell \rangle \), \( k, \ell = -n, \ldots, n \), and \( H_{2n+1}^{-1} \) the inverse of its positive definite square root. Now let \( X \) be a vector of \( 2n + 1 \) symbols, \( A = H_{2n+1}^{-1/2} X \), and \( X(t) \) given by

\[
X(t) = \sum_{k=-n}^{n} A_k h_k(t).
\]

Then \( X(t) \) is well localized in time.

**Proof.** See the Appendix.

\[\square\]

**C. Complexity and implementation**

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

However, the precoding can be implemented very efficiently since we, in most realistic cases, are going to work with Toeplitz matrices according to Proposition 9. This allows us to embed it into a circulant matrix, and since it is diagonalized by the discrete Fourier transform matrix we use FFT to do reduce the computational complexity of the whole algorithm down to \( O(n \log(n)) \). The trick is presented in for example in [15], but for clarity we use the rest of this subsection to describe the algorithm.

The precoding as well as the decoding consists of multiplying a vector with the inverse of a matrix square root, hence both precoding and decoding can described with the same algorithm. The matrix is a Toeplitz matrix as in Definition 5 with elements \( c_0, c_1, \ldots, c_n \). Vectors will be denoted with capital letters, and elements of these are denoted with small letters. Superscripts in parenthesis on vectors are used for clarity in order denote the size. Furthermore [1,17,42,\ldots] also denotes a vector, but where the elements are explicitly given. Also, \( \hat{0}^{(n)} \) is a vector of \( n \) zeros. With this notation the algorithm is described in Algorithm 1.

**Algorithm 1 Multiplication with square-root-inverse**

**Input:** \( X^{(n)} \)

**Output:** \( Y^{(n)} \)

1: \( S^{(2n)} \leftarrow \text{FFT}([c_0, c_1, \ldots, c_n, 0, c_n, \ldots, c_2]) \)
2: \( Z^{(2n)} \leftarrow [s_1^{-1/2}, s_2^{-1/2}, \ldots, s_{2n}^{-1/2}] \)
3: \( U^{(2n)} \leftarrow \text{FFT}([X^{(n)}, \hat{0}^{(n)}]) \)
4: \( V^{(2n)} \leftarrow [u_1 w_1, u_2 w_2, \ldots, u_{2n} w_{2n}] \)
5: \( W^{(2n)} \leftarrow \text{IFFT}(V^{(2n)}) \)
6: \( Y^{(n)} \leftarrow \sqrt{\rho} \cdot [w_1, w_2, w_3, \ldots, w_n] \)

Note that when used for precoding and decoding the two first steps in Algorithm 1, calculating \( Z^{(2n)} \), only need to be performed when the modulation scheme changes; either with the change of pulses or change of FTN-parameter. The last multiplication with \( \sqrt{\rho} \) is for power normalization purposes.

**D. Example - Root-Raised-Cosine**

We exemplify the theory by applying it to the root-raised-cosine pulse, from this it is also easy to see how it works for other roll-off pulses with a spectral leakage.

As before, let \( T = \frac{\pi}{W'} \), and introduce

\[
g_\beta(t) = \frac{4\beta}{\pi \sqrt{T}} \cos \left( \frac{(1 + \beta) \pi}{4} \right) + \frac{\sin((1 - \beta) \pi)}{4(\beta - 1)^2},
\]

(10)
defined for $\beta \in [0, 1]$. Let the Fourier transform of $g_\beta(t)$ be

$$G_\beta(\omega) = \begin{cases} \sqrt{T} & 0 \leq |\omega| \leq (1-\beta)\frac{\pi}{T} \\ \sqrt{T} \sqrt{1 - \sin \left( \frac{T}{2\beta} (|\omega| - \frac{\pi}{2}) \right)} & (1-\beta)\frac{\pi}{T} < |\omega| \leq (1+\beta)\frac{\pi}{T} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

We can see that the pulse is a roll-off pulse that is defined with a spectrum leakage $\beta$, the transform has support in linear frequency on an interval of size $2W(1+\beta)$. It is also a Nyquist pulse, since with $T$ as chosen above the pulses are orthogonal. When applying FTN to the root-raised-cosine we arrive at the following result.

**Lemma 3 (Associated function to the Root-raised cosine Gramian).** Let $h_k(t) = \rho \cdot g_\beta \left( t - \frac{k-1}{2W} \right)$, with $g_\beta$ given by (10), $k = 1, 2, \ldots, n$, and $\rho \leq 1$. Moreover let $H_n$ be the Gramian of these pulses as given by Proposition 7. Then the associated function $f$ to $H_n$ is given by

$$f(z) = \begin{cases} (1 + \beta)\rho & z \in [-\pi, -(1 + \beta)\rho \pi] \\ z \in [-1 + \beta] & z \in [-(1 - \beta)\rho \pi, -1 - \beta)\rho \pi] \\ 1 & z \in [1 - \beta \rho \pi, 1 + \beta \rho \pi] \\ 0 & z \in [(1 + \beta)\rho \pi, \pi] \end{cases} \quad (13)$$

and

$$1 - \sin \left( \frac{2\zeta}{2\beta} \right) \cos \left( \frac{\zeta}{2\rho} \right) z \in [-\pi, -(2 - 1 + \beta)\rho \pi] \quad (14)$$

where $\rho \leq 1$.

**Proof.** See the Appendix.

By inspecting Lemma 3 in the context of Proposition 3, we can see that we only have a well behaved matrix and numerical stability if $(1 + \beta)\rho \geq 1$. This is indeed also true for all roll-off pulses that utilize extra bandwidth $\beta$, which can be concluded from Proposition 9.

We can see that with $\rho = \frac{1}{T_0^2}$ and $T' = \rho T$, the root-raised-cosine fulfills all prerequisites in Theorem 1. Hence, for a fixed average power $P$ the capacity for precoded FTN with root-raised-cosine becomes

$$C = (1 + \beta)W \log_2 \left( 1 + \frac{P}{N_0W(1 + \beta)} \right) \text{ bits/second.}$$

**E. Simulation results**

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

We apply root-raised-cosine as pulse shape and use a roll-off factor $\beta = 0.22$ to mimic existing standards as [19]. 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

$$\sigma = 10^{-\text{SNR}_{db}/20},$$ 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 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 [20]. 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 fist apply FTN decoding, to produce $S$, before using turbo decoding which gives the estimated payload bits, see Figure 1. For every value of $\rho$ and SNR presented, we have looked at code rates from $1/3$ 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 result throughput can be found in Figure 7. It should however be noted that this
Throughput (bits / equivalent time unit)

| Throughput (bits / equivalent time unit) |
|------------------------------------------|
| 1500                                     |
| 3400                                     |
| 2000                                     |
| 3600                                     |
| 2500                                     |
| 3800                                     |
| 4000                                     |
| 4200                                     |
| 4000                                     |
| 4400                                     |
| 4500                                     |

Throughput as given by throughput (1 – block error rate) · #payload bits. One can observe that the FTN scheme seems to need a higher SNR in order to reach any throughput, and the lower the ρ, the higher the SNR. However, once this is reached it starts to outperform the Nyquist case. The stagnation in throughput at high SNR is due to the code rate not going higher than ≈ 0.96.

Figure 6. Throughput based on BLER plotted versus SNR. The equivalent Nyquist block consists of 4000 bits.

Throughput based on BER plotted versus SNR. The equivalent Nyquist block consists of 4000 bits.

Throughput as given by throughput (1 – bit error rate) · #payload bits. This plot shows throughput gains using FTN even at low SNR, although this is less informative from a system perspective. The same stagnation in throughput at high SNR as in Figure 6 can also be seen here, we have reached BER = 0.

VIII. 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 [1], 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 [1] is valued 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.

REFERENCES

[1] J. Mazo, “Faster-than-Nyquist signaling,” Bell System Technical Journal, vol. 54, no. 8, pp. 1451–1462, 1975.
[2] C. E. Shannon, “Communication in the presence of noise,” Proc. Institute of Radio Engineers, vol. 37, no. 1, pp. 10–21, 1949.
[3] J. Mazo and H. Landau, “On the minimum distance problem for faster-than-Nyquist signaling,” IEEE Transactions on Information Theory, vol. 34, no. 6, pp. 1420–1427, 1988.
[4] D. Hajela, “On computing the minimum distance for faster than Nyquist signaling,” IEEE Transactions on Information Theory, vol. 36, no. 2, pp. 289–295, 1990.
[5] J. Anderson, F. Rusek, and V. Öwall, “Faster-than-Nyquist signaling,” Proceedings of the IEEE, vol. 101, no. 8, pp. 1817–1830, 2013.
[6] F. Rusek and J. B. Anderson, “The two dimensional Mazo limit,” in International Symposium on Information Theory, 2005, pp. 970–974.
[7] F. Rusek and J. Anderson, “Constrained capacities for faster-than-Nyquist signaling,” IEEE Transactions on Information Theory, vol. 55, no. 2, pp. 764–775, Feb 2009.
[8] A. Wyner, “The capacity of the band-limited gaussian channel,” Bell System Technical Journal, vol. 45, no. 3, pp. 359–395, 1966.
[9] S. Verdú, “Computational complexity of optimum multuser detection,” Algorithmica, vol. 4, no. 1-4, pp. 303–312, 1989. [Online]. Available: http://www.princeton.edu/~verdu/reprints/Verdu.1989.pdf
[10] J. G. Proakis and M. Salehi, Digital communications, 5th ed. McGraw-Hill, 2008.
[11] R. G. Gallager, Information theory and reliable communication. Wiley, New York, 1968.
[12] C. E. Shannon, “A mathematical theory of communication,” Bell System Tech. J., vol. 27, pp. 379–423 and 623–656, 1948.
[13] U. Grenander and G. Szegö, Toeplitz forms and their applications. American Mathematical Society, 2001.
[14] R. M. Gray, “Toeplitz and circulant matrices: A review,” Department of Electrical Engineering, Stanford, Tech. Rep., 2006. [Online]. Available: http://ee.stanford.edu/~gray/toeplitz.pdf
pendent by the induction hypothesis, we must have 
\[ \omega \]
that our assumption that 
\[ e \]
with 
\[ n \]
any positive integer 
\[ n \]
are linearly dependent. Then, there exists 
\[ c_1, \ldots, c_n \in \mathbb{C}^n \]
such that 
\[ c_1h_1(t) + \cdots + c_nh_n(t) = 0, \quad \forall t \in \mathbb{R}. \]

Taking the Fourier transform of the equation above, we get 
\[ c_1H(\omega)e^{-i\omega\tau_1} + \cdots + c_nH(\omega)e^{-i\omega\tau_n} = 0, \quad \forall \omega \in \mathbb{R}. \]

In particular, we have that 
\[ c_1H(\omega)e^{-i\omega\tau_1} + \cdots + c_nH(\omega)e^{-i\omega\tau_n} = 0, \quad \forall \omega \in \text{int}(\Omega) \]
which implies that 
\[ c_1e^{-i\omega\tau_1} + \cdots + c_ne^{-i\omega\tau_n} = 0, \quad \forall \omega \in \text{int}(\Omega). \]

Now letting \( \alpha_k = -i\tau_k \), for \( k = 1, \ldots, n \), and using Proposition 6 together with the assumption that \( \text{int}(\Omega) \neq \emptyset \), we see that \( e^{-i\omega\tau_k} \) are linearly independent so we must have 
\[ (c_1, \ldots, c_n) = 0, \quad \text{a contradiction.} \]
We conclude that \( h_1, \ldots, h_n \) must be linearly independent, and we are done.

**Proof of Proposition 8.** First, 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 4.** The closure of the range of \( BD \) is equal to the range of \( B \).

**Proof of Lemma 4.** 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 \( X(t) \in \text{Range}(B) \circ \text{Range}(BD) \). 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 (16) 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.

By Lemma 4 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 $\sum_{k=1}^{n} A_k \delta(t - \rho(k-1)) \to Y$ weakly as $\rho \to 0$. Since the total variation norm is uniformly bounded by $\sum_{k=1}^{n} |A_k| \leq \|Y\|_{L_1}$ we have that $X(t) = \sum_{k=1}^{n} A_k h_k(t) \to X$ as $\rho \to 0$, which proves the theorem.

**Proof of Proposition 9**

The first statement is obvious by using $\tau_k = \tau \cdot (k - 1)$ and direct calculation of (1).

To prove the second statement we look at

$$[H_n]_k = c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(z)e^{-ikz} dz,$$

but also we know that

$$[H_n]_k = \langle h_{k+1}, h_1 \rangle = \int_{-\infty}^{\infty} h_{k+1}(t) h_1(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega)e^{-ik\tau \omega} H(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\tau} \left| H \left( \frac{z}{\tau} \right) \right|^2 e^{-ikz} dz.$$

The integral is absolutely convergent since $h \in L_2(\mathbb{R})$, which allows us to do some further manipulations of that expression.

$$\int_{-\infty}^{\infty} \frac{1}{\tau} \left| H \left( \frac{z}{\tau} \right) \right|^2 e^{-ikz} dz$$

$$= \sum_{\ell=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\tau} \left| H \left( \frac{z}{\tau} \right) \right|^2 e^{-ikz} dz$$

$$= \sum_{\ell=-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\tau} \left| H \left( \frac{z + 2\pi\ell}{\tau} \right) \right|^2 e^{-ikz-2\pi\ell} dz$$

$$= \sum_{\ell=-\infty}^{\infty} e^{-ikz} \int_{-\infty}^{\infty} \left| H \left( \frac{z + 2\pi\ell}{\tau} \right) \right|^2 dz.$$  

This shows that $c_k$ is the $k$th Fourier series coefficient of the function $\frac{1}{\tau} \int_{-\infty}^{\infty} \left| H \left( \frac{z + 2\pi\ell}{\tau} \right) \right|^2$, and the result now follows from the uniqueness of Fourier series.

Finally to show that $f \in L_1([-\pi, \pi])$ we use Proposition 3 and Proposition 7 to get that $f(z) \geq 0$. Hence $|f(z)| = f(z)$, and

$$\int_{-\pi}^{\pi} |f(z)| dz = c_0 = \int_{-\infty}^{\infty} \frac{1}{\tau} \left| H \left( \frac{z}{\tau} \right) \right|^2 dz < \infty,$$

since $h$ is in $L_2(\mathbb{R})$.

**Proof of Theorem 1**

For ease of notation, although the arguments holds in general, we will consider a symmetric transmission, that is

$$X(t) = \sum_{k=-n}^{n} A_k g(t - kT'),$$

where $A$ is the vector of precoded symbols $A = H^{-\frac{1}{2}}_{2n+1} X$. Propositions 3, 4, and 3, together with the definition of $G(\omega)$ and $T'$ assures us of the existence of $H^{-\frac{1}{2}}_{2n+1}$. For the matrix we use the following notation

$$H^{-\frac{1}{2}}_{2n+1} = \begin{pmatrix} c_{1,0} & \cdots & c_{1,n} & \cdots & c_{1,2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{n+1,-n} & \cdots & c_{n+1,0} & \cdots & c_{n+1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{2n+1,-2n} & \cdots & c_{2n+1,-n} & \cdots & c_{2n+1,0} \end{pmatrix}.$$  

Although the notation might indicate otherwise, the matrix is actually symmetric and the reasoning for numbering its transpose will be apparent later. By Proposition 4 the matrix is asymptotically Toeplitz, and tends towards

$$H^{-\frac{1}{2}} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & c_{-1} & 0 & c_1 & \cdots \\ \cdots & c_{-2} & c_{-1} & 0 & c_2 & \cdots \\ \cdots & \cdots & c_{-3} & c_{-2} & c_{-1} & 0 & c_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

with associate function $\hat{f}$, as $n \to \infty$. The function $\hat{f}$ can in turn be deduced from Propositions 4 and 9. The signal $X(t)$ can now be written as

$$X(t) = X^T H^{-\frac{1}{2}}_{2n+1} \begin{pmatrix} g(t + nT') \\ g(t + (n - 1)T') \\ \vdots \\ g((n-1)T') \\ g(t - (n - 1)T') \\ g(t - nT') \end{pmatrix}.$$  

By inspecting (17) we can see that each uncoded symbol $X_k$ will be carried by an effective pulse $\xi_{\ell}^{(2n+1)}(t)$ such that

$$X(t) = \sum_{\ell=-n}^{n} X_\ell \xi_{\ell}^{(2n+1)}(t),$$

Without loss of generality we consider the uncoded symbol $X_0$. It will then be carried by the effective pulse $\xi_0^{(2n+1)}(t)$ given by

$$\xi_0^{(2n+1)}(t) = \sum_{k=-n}^{n} c_{n+1,k} g(t - k\rho T'),$$

and which has a Fourier transform given by

$$\xi_0^{(2n+1)}(\omega) = G(\omega) \sum_{k=-n}^{n} c_{n+1,k} e^{-ik\omega \rho T}.$$
Examining the power spectral density as \( n \to \infty \) we get
\[
|\Xi_{0}^{n+1}(\omega)|^2 = \lim_{n \to \infty} |\Xi_{\omega}(\omega)|^2
\]
\[
= |G(\omega)|^2 \sum_{k=-\infty}^{\infty} c_{k} e^{-ik\omega T} = |G(\omega)|^2 \left| \tilde{f}(\omega T) \right|^2
\]
\[
= T' \sum_{k=-\infty}^{\infty} |G(\omega)|^2.
\]
Now with the definition of \( T' \) we get
\[
|\Xi_{\omega}(\omega)|^2 = \begin{cases} T' & |\omega| \leq \pi W' \\ 0 & |\omega| > \pi W' \end{cases}
\]
which is effectively the spectrum of a sinc pulse with bandwidth \( 2W' \). Since the pulses are transmitted with a time spacing of \( T' = \frac{1}{\pi W'} \) seconds, we conclude that the signal \( X(t) \) is well localized in time.

**Proof of Lemma 3**

The lemma is proved by direct calculations. The assumption \( \rho \leq 1 \) is a technical one which ensures \( (1-\beta)p \geq 2 -(1+\beta)p \), and with \( \beta \in (0,1) \) it also implies \( (1+\beta)p \leq 2 \).

We have that the Gramian for the root-raised-cosine pulses is simply given by a raised cosine with roll-off \( \beta \) and that is sampled at integer instants and with \( T = 1/\rho \); namely
\[
[H_{\omega}]_{mn} = \rho \sin(\rho(m-n)) \cos(\pi \rho(\rho(m-n)))
\]
This can in turn be rewritten as
\[
[H_{\omega}]_{mn} = \rho \cdot \frac{1}{1-4\rho^2} \cdot \sin((1-\beta)p(m-n))
\]
\[
+ \rho \cdot \frac{1}{1-4\rho^2} \cdot \sin((1+\beta)p(m-n))
\]
We will use one of Parseval’s identities, regarding periodic convolution, namely:
\[
f(z) = \sum_{k=-\infty}^{\infty} c_{k} e^{ikz} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_{1}(t) g_{2}(z-t) dt.
\]
Thus we identify the sought associate function \( f \) as the convolution of the two functions \( g_{1}(z) \) and \( g_{2}(z) \) with the corresponding Fourier series coefficients:
\[
c_{k}^1 = \frac{(1+\beta)p}{2} \sin((1+\beta)p) k + \frac{(1-\beta)p}{2} \sin((1-\beta)p)
\]
and
\[
c_{k}^2 = -1 \frac{1}{1-4\rho^2} k^2 - 1
\]
We can then based on (19) conclude that \( g_{1} \) is a sum of rectangular functions, periodic on the interval \([-\pi, \pi]\) and looks like:
\[
g_{1}(z) = \begin{cases} \frac{1}{2} \text{rect} \left( \frac{z}{2(1+\beta)p} \right) + \frac{1}{2} \text{rect} \left( \frac{z}{2(1-\beta)p} \right) & (22) \\
\frac{1}{2} \text{rect} \left( \frac{z}{2(1-\beta)p} \right) & (23) \end{cases}
\]
with \( \frac{1}{2} \text{rect}(z) = \begin{cases} z & z \geq b \\
\frac{1}{2} & a \leq z < b \\
0 & z \leq a \end{cases} \)
\[
1 - \frac{1}{2} \text{rect} \left( \frac{z}{2(1-\beta)p} \right) \right.
\]
\[
+ \frac{1}{2} \text{rect} \left( \frac{z}{2(1-\beta)p} \right).
\]
The function \( g_{1} \) is hence different depending on the value of the parameter combination \( (1+\beta)p \).

Evaluating \( g_{2}(z) \) we try to find the Fourier series of the following function
\[
g_{2}(z) = a \cdot \sin \left( \frac{z-d}{2b} \right) \quad z \in [-\pi, \pi],
\]
and find that it has the following coefficients:
\[
eg_{k}^2 = \frac{1}{4\beta^2 k^2 - 1} \cdot \left( \frac{2ab}{\pi} \cos \left( k \frac{d}{2b} \right) + 2ab(1)^n \right)
\]
Comparing \( c_{k}^2 \) with \( c_{k}^2 \), that is (25) with (20), we can identify that in our case we have \( b = \beta p, d = (1-\beta)p \), and \( a = \frac{\pi}{2\beta} \cdot \frac{1}{\cos((1-\beta)p \pi/(2b))} \).

Making some rearrangements we arrive at
\[
g_{2}(z) = \frac{\pi}{2\beta} \cdot \frac{1}{\sin \left( \frac{\pi}{2\beta} \right)} \cdot \cos \left( \frac{z-\pi}{2\beta} \right).
\]
Now we can use Parseval’s identity, (18), to retrieve \( f \). Getting a closed form expression for the convolution is made easier since \( g_{1} \) is constant over intervals. However since \( g_{1} \) changes depending on the value of \( (1+\beta)p \) and \( g_{2} \) contains and absolute value \( |z-t| \) where \( t \) is the integration variable the problem grows combinatorially. As an intermediate step we solve
\[
\int_{a}^{b} \cos \left( \frac{z-t}{2\beta} \right) dt = \begin{cases} \frac{\pi}{2\beta} & z \geq b \\
\frac{\pi}{2\beta} & a < z < b \\
0 & z \leq a \end{cases}
\]
With
\[
2\beta \left( (z-\frac{\pi-b}{\beta}) + \sin \left( \frac{z-\pi-a}{2\beta} \right) \right)
\]
\[
2\beta \left( \sin \left( \frac{z-b}{\beta} \right) + \sin \left( \frac{z-\pi-a}{2\beta} \right) \right)
\]
\[
+ 2 \sin \left( \frac{\pi}{2\beta} \right)
\]
\[
2\beta \left( \sin \left( \frac{z-b}{\beta} \right) - \sin \left( \frac{z-\pi+a}{2\beta} \right) \right).
\]
Combining the results (21), (26) and (27), and doing some simplifications to the expressions gives the sought associate function found in (12).