On Information Rates Over a Binary-Input Filtered Gaussian Channel

MICHAEL PELEG [1,2] **(Life Senior Member, IEEE), TOMER MICHAELI** [1],
AND SHLOMO SHAMAI [1] **(Life Fellow, IEEE)**

[1]Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel

[2]Rafael, Haifa 3102102, Israel

CORRESPONDING AUTHOR: S. SHAMAI (e-mail: sshlomo@ee.technion.ac.il)

ABSTRACT We study communication systems over band-limited Additive White Gaussian Noise (AWGN) channels in which the transmitter's output is constrained to be symmetric binary (bipolar). We improve the available Ozarow-Wyner-Ziv (OWZ) lower bound on capacity which is based on peak-power constrained pulse-amplitude modulation, by introducing new schemes (achievability) with two advantages over the studied OWZ schemes. Our schemes achieve a moderately improved information rate and they do so with much fewer sign transitions of the binary signal. The gap between the known upper bound, which is based on spectral constrains of bipolar signals, and our new achievable lower bound is reduced to 0.93 bits per Nyquist interval at high SNR.

INDEX TERMS Information capacity, faster-than-Nyquist signaling, filtered Gaussian channel, unit process, bipolar input, peak limited signals.

I. INTRODUCTION AND PROBLEM DEFINITION

WE STUDY communication systems over band-limited Additive White Gaussian Noise (AWGN) channels in which the transmitter's output is constrained to be bipolar, as presented in Figure 1. Such systems arise when the power efficiency must be high or when the transmitter needs to be of very low complexity. Those systems are usually implemented by some form of Pulse Width Modulation (PWM), Pulse Position Modulation (PPM), or similar schemes, operating over Gaussian noise channels [1], [2], [3]. Communication systems with binary transmitted signals are of recent practical interest in millimeter-wave wide-band applications, e.g., [4], [5].

In this work, we examine theoretical limits on communication with binary transmission, not limited to PWM. We are interested in the reliable information rate supported by this system focusing mainly on the region of asymptotically high SNR. This theoretical problem was addressed by Ozarow, Wyner, Ziv (OWZ) [6] using the Pulse Amplitude Modulation (PAM) method. OWZ [6] showed that performance, measured by mutual information, achievable with a signal peak-limited to $\pm\sqrt{P}$ can also be

achieved with a binary-valued $\pm\sqrt{P}$ signal with a very high Sign Transition Rate (STR). They applied this finding to design a PAM scheme with symbols uniformly distributed in $[-\sqrt{P}, +\sqrt{P}]$, which provides an achievable lower bound on the capacity of the system. As implied by [6], peak-limited continuous-time signals such as filtered PAM, can in principle be also band limited [7] and hence represented by sampling at an appropriate rate, while the equivalent (in the sense of [6]) bipolar processes cannot be strictly bandlimited [8]. A lower bound exceeding for low SNR that of [6], was presented in [9], based on improved bounds for intersymbol-interference Gaussian channels. Additional results on capacity of systems with binary inputs, some with additional constraints on average transition rate, minimum inter-transition time and out of band power, are presented in [10] and [11]. Systems with limited minimal transition times were investigated in [12], including systems with mild filtering, that is, not strictly bandlimited as in [6].

The binary channel input carries information in its transition times. Sampling the binary input at a Nyquist rate corresponding to the channel bandwidth would degrade the performance severely, thus the system in Figure 1 falls in
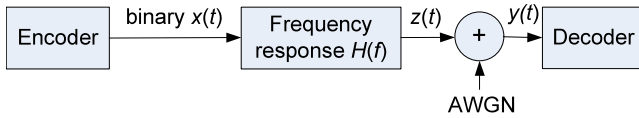
FIGURE 1. Communication system with binary-valued transmitted signal.

general into the category of Faster Than Nyquist (FTN) signaling which is of wide current theoretical and practical interest. In recent years FTN signaling approaches, forms and extensions of classical pulse-amplitude modulation strategies have emerged. See overviews of these relevant domains in [13], [14], [15] and references therein. See also recent examples of advanced theory and techniques in [16], [17], [18], [19]. FTN can provide significant advantages in terms of capacity with prescribed modulation techniques and signaling strategies, though the resultant channel may suffer significant inter-symbol-interference, which demands higher complexity detection procedures. Yet, no peak-power restrictions are imposed on the resultant time-continuous process, which is a central part in our scheme. This is motivated by practical constraints, as was also the case in [6], reflecting the constraints of magnetic storage media.

In this work, we present new schemes with two advantages over [6]. They achieve a moderately improved information rate and do so with much fewer sign transitions of the binary signal. The new schemes require STR of only up to twice the Nyquist rate of the channel, while [6] uses STR many folds higher than the Nyquist rate; if implemented fully, the STR in [6] is infinite. Low STR is easier to implement in systems which are already wide-band and in which each sign transition must pass a power amplifier such as [4], [5]. We extended the technique of analysis in [6] to the new schemes in which the transmitted signal is a non-linear function of the information sequence.

The studied communication system is presented in Figure 1. It comprises an encoder producing a binary-valued $\pm\sqrt{P}$ input $x(t)$ where P is the transmit power, AWGN channel with noise Power Spectral Density (PSD) of $\frac{1}{2}N_0$ watt/Hz (double-sided) and a receiver. The channel has a frequency response $H(f)$, in our case a unity frequency response at frequencies from 0 to B and zero otherwise. The channel output y is

$$y(t) = z(t) + n(t), \tag{1}$$

where $z(t)$ is the filtered desired signal and $n(t)$ is the Gaussian noise.

We denote by B the bandwidth of the low pass brick-wall filter in Hz, $T = 1/(2B)$ is the Nyquist sampling period associated with B, $\rho = \frac{P}{N_0 B}$ is the signal to noise ratio (SNR), log denotes the natural logarithm and bold lower-case letters denote vectors and sequences.

II. KNOWN PERFORMANCE BOUNDS

Shamai and Bar-David [20], derived an upper bound on the system capacity, based on the fact that the Power Spectral Density (PSD) of a binary-valued signal is limited by certain constraints presented in [21] and [8]. They analyzed limits on spectral densities of binary signals and then upper bounded

the capacity of the system by Mutual Information (MUI) when the channel input has the capacity achieving Gaussian distribution with the same PSD as the binary-valued signal. For high SNR they proved that relative to the capacity-achieving frequency-flat Gaussian input there is a power loss at least by a factor of $\gamma = 0.9337$, see definition of γ below. The same paper considers Random Telegraph Signal (RTS) as an interesting example rather than a bound and the power factor there is around $\gamma = 0.63$ which is an upper bound on the performance of RTS. The capacity C_G bits/second of the channel with PSD given as $S(f)$ used in [20] is the well-known expression

$$C_G = \int_0^B \log_2\left(1 + \frac{S(f)}{N_0}\right) df,$$

where C_G is achieved with a Gaussian input. In the limit of asymptotically high $\frac{S(f)}{N_0}$ the capacity C_G becomes

$$C_G^h = \int_0^B \log_2 \frac{S(f)}{N_0} df. \tag{2}$$

For a frequency-flat Gaussian signal of bandwidth B this yields

$$C_{G0}^h = B \cdot \log_2 \frac{P}{N_0 B}.$$

Multiplying $S(f)$ in (2) by a factor γ increases C_G^h by $B \cdot \log_2 \gamma$ information bits per second which are $\Delta = \frac{1}{2}\log_2 \gamma$ bits per Nyquist sampling interval. Consequently, the equivalent SNR gain is defined, for a scheme with bandwidth B, as a function of difference Δ in information per Nyquist interval between the scheme and the AWGN channel with the same bandwidth and transmit power as,

$$\gamma = 2^{2\Delta}.$$

OWZ [6] derived the following achievable lower bound using the modulation method [6] described in the introduction.

$$I_{\text{OWZ}} \geq \frac{1}{2}\log_2\left(\frac{2P \cdot e}{\pi^3 N_0 B} + 1\right) = \frac{1}{2}\log_2\left(\rho\frac{2e}{\pi^3} + 1\right) \tag{3}$$

where I_{OWZ} stands for mutual information per Nyquist interval.

This corresponds to

$$\gamma_{\text{OWZ}} = \frac{2e}{\pi^3} = 0.1753.$$

The bipolar signal that achieves the performance of the PAM modulation technique in [6] involves high transition rate of the binary signal. An improved lower bound in the low SNR regime is reported in [9].

III. NEW ACHIEVABLE SCHEMES

The main results of this work are the improved lower bounds on the capacity of the bipolar-input bandlimited AWGN channel, see Proposition 1. The proposition is proved by introducing and analyzing new communication schemes.
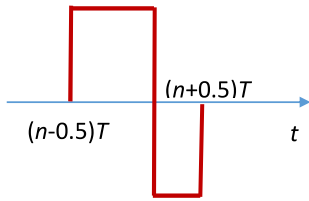
FIGURE 2. Binary signals $x(t)$ of type A. The single symbol with a variable transition time.
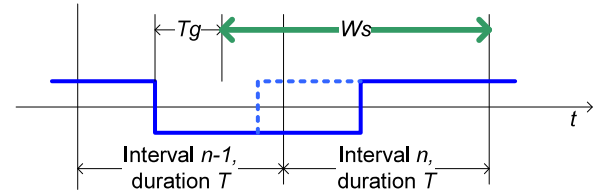

FIGURE 4. Time diagram of scheme D, the solid blue line is the transmitted signal, the dotted blue line is another possible transmitted signal.

TABLE 1. Upper bounds using Gaussian inputs with the same spectra.

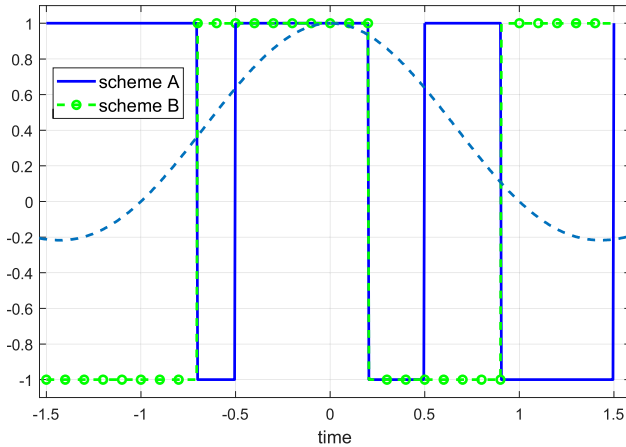| Scheme	Equivalent power factor γ relative to rectangular spectra 0 to B Hz	Information loss Δ relative to rectangular spectra 0 to B Hz in bits/Nyquist interval
Rectangular spectra 0 to B Hz	1	0
Schemes A and B	0.3	0.8686
Scheme C	0.367	0.7232
Upper bound on binary schemes in [20]	0.9337	0.0495
Upper bound on the Random Telegraph signal as in [20]	0.6271	0.3366

FIGURE 3. Binary signals $x(t)$ of type A and B. The dashed line is the impulse response of the channel filter.

We discuss four schemes, denoted by A, B, C and D. In all of them the time axis is partitioned into successive intervals of duration T equal to the Nyquist interval corresponding to B. In scheme A, the binary signal in each interval n of time t spanning $(n-0.5)T \leq t < (n+0.5)T$ is

$$x(t) = \begin{cases} 1 & (n-0.5)T \leq t \leq (n+a_n)T \\ -1 & (n+a_n)T < t < (n+0.5)T, \end{cases} \quad (4)$$

where a_ns are the information-carrying variables, uniformly, independently, and identically distributed (u.i.i.d.) over $[-0.5, +0.5]$. Thus, information is conveyed by the time of sign reversal of the signal, see Figure 2 for an illustration.

We denote the sequence of all a_n by $\mathbf{a}$, denote the binary transmitted signal in interval n as $x_n(t)$ and the overall transmission by $x(t)$ or $\mathbf{x}$. Scheme A changes sign twice in each Nyquist interval T, thus it's STR is $4B$.

Scheme B is derived from scheme A by inverting the signal in successive intervals of length T to eliminate half of the sign transitions of the binary signal, reducing it's STR to $2B$. See Figure 3.

Scheme C is derived from scheme A by inverting the signal in successive intervals at random where the signs s_n valued as ± 1 are used as additional information inputs. The signs s_n are equi-probable and independent. The signaling in scheme C comprises a_n and s_n, thus the signaling rate is twice the Nyquist rate. The STR of scheme C is $3B$, since the random sign reversal related to s_n occurs in half of the transmission intervals.

Denote by $\mathbf{s}$ the sequence of the sign inversions s_n in schemes A, B and C, so that $s_n = -1$ for the inverted symbols and $s_n = 1$ otherwise.

Scheme D is derived from scheme B. It eliminates pairs of consecutive sign-transitions with exceptionally short inter-transition time, which we found detrimental to our lower bound. Scheme D improves upon scheme B by introducing a minimal inter-transition interval $T_g = 0.2T$ and by extending the range in which each sign-transition time can occur. In particular, as in scheme B, each transmission interval of duration T is associated with one sign transition. However, the transition time specified in (4) as uniformly distributed over the nth transmission interval spanning $(n-0.5)T \leq t < (n+0.5)T$ in scheme B, is, in the new scheme D, distributed uniformly over a window $W_s(n)$, which starts T_g after the previous sign-transition and ends, as in scheme B, at the end of the current interval. This is illustrated in Figure 4. Note that the sign transition associated with the nth transmission interval may occur in the nth interval or in one of the few intervals preceding it. The STR of scheme D is $2B$, identical to scheme B.

Computing the exact capacity of the four schemes, that is, the MUI between the binary input $x(t)$ and the channel output $y(t)$, seems intractable. We therefore resorted to deriving upper and lower bounds. The lower bound on performance of each scheme applies to the scheme itself and it is also a lower bound on the channel capacity. Each upper bound applies only to the specific scheme while the upper bound on the capacity of the system is the upper bound on binary schemes in Table 1.

To compute upper bounds on the communication rates of schemes A, B and C, we first evaluate the PSD of the
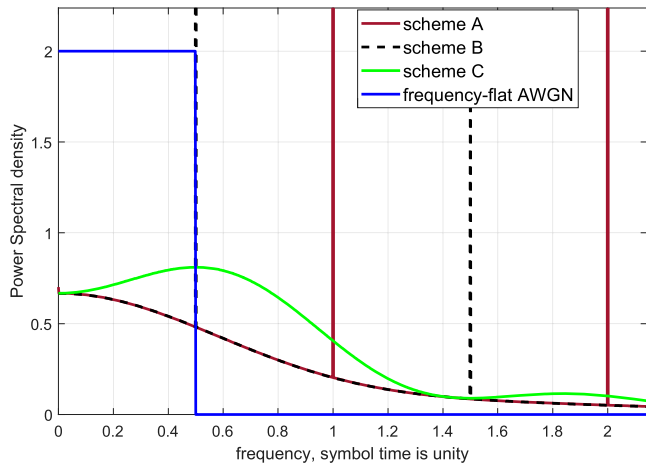
FIGURE 5. Spectrum of the three schemes and of the frequency-flat AWGN. The continuous parts of the scheme A and scheme B curves overlap.

signals. We assume that the signal is randomly shifted as a whole by a delay distributed uniformly over $(0, T)$ to render it stationary. The autocorrelations of the signals in the three schemes are derived in the Appendix and summarized in (5), shown at the bottom of the page.

The PSD was obtained by numerical Fourier transform of the autocorrelations, see Figure 5, and verified by simulation. The PSD is obtainable analytically from the autocorrelations. For example for scheme C with $T = 1$ we have the one-sided PSD:

$$S_c(f) = \frac{3\pi f - 2\sin(2\pi f) + \pi f \cos(2\pi f)}{(\pi f)^3}.$$

The AWGN line in Figure 5 is the PSD of the standard bandlimited capacity-achieving signal without the binary constraint. As well-known from the water-pouring theory, it spreads the available power uniformly over the available bandwidth B. The PSDs of our three schemes suffer the disadvantage of wasting some of the transmitted power out of the channel bandwidth and of not spreading the remaining power uniformly. Scheme C is evidently better than schemes A and B. Indeed the schemes A, B and C are constrained to bipolar transmitted signals and therefore cannot possess a strictly bandlimited spectrum, as we know

from [8]. The spectra of schemes A and B are identical except for the discrete frequency components (tones) which do not influence the outcome of (2). There are no discrete tones in scheme C since it decorrelates the pulses by random sign inversions, limiting the support of the autocorrelation to $[-T, T]$.

Based on the PSD, we compute the upper bounds on performance at high SNR of the three schemes using (2) and compare them to the optimal input which is a Gaussian signal with power P and a flat PSD from 0 to B. The results are presented in Table 1.

We proceed to derive lower bounds on communication rates of the new schemes. As shown in Figure 1, $x(t)$ passes through the channel filter and is then contaminated by AWGN. The receiver filters the signal by the same low pass filter, which is clearly an information-lossless operation. We sample the filtered channel output at the Nyquist rate $1/T$ producing an infinite sequence $\mathbf{y}$ of samples y_n. We denote the signal without the noise component by a sequence $\mathbf{z}$ of samples z_n, see Figure 1.

We lower-bound the capacity $I(\mathbf{x}; \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})$ by adapting the approach presented in OWZ [6]. Since $H(\mathbf{y}|\mathbf{x})$ is the known entropy of the noise, the main term to evaluate is $H(\mathbf{y})$. OWZ lower-bounded $H(\mathbf{y})$ as a function of the entropies of its components $H(\mathbf{z})$ and $H(\mathbf{n})$ using the Entropy-Power Inequality (EPI) presented in [22]. OWZ evaluated $H(\mathbf{z})$ using the fact that the channel was an Inter Symbol Interference (ISI) channel representable by a Toeplitz matrix the determinant of which is computable using the Szegö theorem [24].

We begin by determining the entropy of $\mathbf{z}$. The required differential entropy is

$$h_z = \frac{1}{N}h(z_1 \ldots z_N).$$

In schemes A and B, each a_n determines one symbol x_n and those symbols are linearly filtered to produce $\mathbf{z}$. The sequence $\mathbf{a}$, treated as a vector in the next equation, comprises u.i.i.d. components, and therefore its differential entropy is:

$$h_a \triangleq \frac{1}{N}h(\mathbf{a}) = h(a_i) = \log(1). \tag{6}$$

$$R_A(\tau) = \begin{cases} \left(1 - \frac{2|\tau|}{T}\right)\left(1 - \frac{|\tau|}{T}\right) + \left(-\frac{2}{3}\left|\frac{\tau}{T}\right|^3 + 2\left|\frac{\tau}{T}\right|^2 - \left|\frac{\tau}{T}\right|\right); & \frac{|\tau|}{T} < 1 \\ \frac{1}{3} + 2\tau_n^2 - 2\tau_n; & \text{otherwise} \end{cases} \tag{5a}$$

where $\tau_n = \frac{|\tau|_{\text{mod}\, T}}{T}$ and $|\tau|_{\text{mod}\, T}$ denotes the modulo T operation.

$$R_B(\tau) = \begin{cases} \left(1 - \frac{2|\tau|}{T}\right)\left(1 - \frac{|\tau|}{T}\right) - \left(-\frac{2}{3}\left|\frac{\tau}{T}\right|^3 + 2\left|\frac{\tau}{T}\right|^2 - \left|\frac{\tau}{T}\right|\right); & \frac{|\tau|}{T} < 1 \\ \frac{1}{3}\left(4\tau_n^3 - 6\tau_n + 1\right) \cdot (-1)^{\lfloor|\frac{\tau}{T}|\rfloor}; & \text{otherwise} \end{cases} \tag{5b}$$

where $\lfloor \tau \rfloor$ is the largest integer smaller than τ

$$R_C(\tau) = \begin{cases} \left(1 - \frac{2|\tau|}{T}\right)\left(1 - \frac{|\tau|}{T}\right); & \frac{|\tau|}{T} < 1 \\ 0; & \text{otherwise} \end{cases} \tag{5c}$$
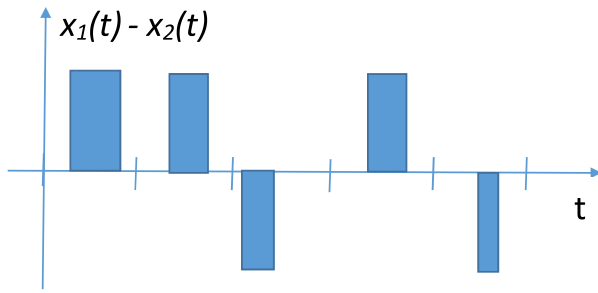
FIGURE 6. Difference signal $x_1 - x_2$.

The noiseless sampled output $\mathbf{z}$ is a function of $\mathbf{a}$, which we denote by $\mathbf{z} = m(\mathbf{a})$. To derive $h(\mathbf{z})$ using the Jacobian formula (7) similarly to [6], we need our transformation $\mathbf{z} = m(\mathbf{a})$ to be a bijection and $\mathbf{a}$ and $\mathbf{z}$ must have identical dimensions.

Lemma 1: For every $\varepsilon > 0$, if the channel's bandwidth is $B = \frac{1}{2T} + \varepsilon$, where T is the signaling period, then the transformation $\mathbf{z} = m(\mathbf{a})$ in schemes A and B is a bijection.

Proof: The modulation scheme in Figure 1 is deterministic, therefore each sequence $\mathbf{a}$ can produce only a single sequence $\mathbf{z}$. It remains to prove that there are no two distinct sequences $\mathbf{a}$ producing the same $\mathbf{z}$. If this would happen, then there would exist a pair of transmitted signals $\mathbf{x}_1 \neq \mathbf{x}_2$ such that $\mathbf{z}(\mathbf{x}_1) = \mathbf{z}(\mathbf{x}_2)$, implying $\mathbf{d} \overset{\text{def}}{=} \mathbf{z}(\mathbf{x}_1) - \mathbf{z}(\mathbf{x}_2) = 0$. Since the low-pass filter is linear, such a $\mathbf{d}$ would be the low-pass filtered signal $\mathbf{x}_1 - \mathbf{x}_2$. By the construction of $\mathbf{x}$, for schemes A and B, not C, the difference $\mathbf{x}_1 - \mathbf{x}_2$ would be a sequence of pulses as depicted in Figure 6 in which each pulse is assigned a symbol interval T during which it has a zero value except for some contiguous duration in which it is ± 2, see Figure 6.

So it is sufficient to prove that such a nonzero signal $\mathbf{x}_1 - \mathbf{x}_2$ cannot have zero spectra in $0 \leq f \leq B + \epsilon$. This follows directly from [23, Th. 1], which proved that signals with zero spectra in $0 \leq f \leq B + \varepsilon$, which are denoted in [23] as high-pass signals or signals with a zero gap, change sign at average rates higher than $1/T = 2B$, which is the highest possible rate of sign changes of the function $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{x}_1(t) - \mathbf{x}_2(t)$ in Figure 6. Thus, such a nonzero $\mathbf{x}_1 - \mathbf{x}_2$ cannot exist and $\mathbf{z} = m(\mathbf{a})$ is a bijection. The asymptotically small change in B is immaterial in this work by the problem definition. ∎

Since $\mathbf{z} = m(\mathbf{a})$ is a bijection, the entropy h_z is

$$h_z \overset{\triangle}{=} \frac{1}{N} h(\mathbf{z}) = \frac{1}{N} h(\mathbf{a}) + \frac{1}{N} \int p(\mathbf{a}) \log \left| \frac{\partial z_i}{\partial a_j} \right| d\mathbf{a}$$
$$= \frac{1}{N} h(\mathbf{a}) + \frac{1}{N} E_{\mathbf{a}} \left(\log \left| \frac{\partial z_i}{\partial a_j} \right| \right), \quad (7)$$

where $\left| \frac{\partial z_i}{\partial a_j} \right|$ denotes the determinant of the Jacobian matrix of $\mathbf{z} = m(\mathbf{a})$, $p(\mathbf{a})$ is the probability density function of $\mathbf{a}$ and $E_{\mathbf{a}}$ denotes expectation with respect to $\mathbf{a}$. The Jacobian matrix is denoted $\left(\frac{\partial z_i}{\partial a_j} \right) \overset{\triangle}{=} J$. Unlike OWZ [6], in our scheme A,

the Jacobian matrix is not Toeplitz since here $\frac{\partial z_i}{\partial a_j}$ depends on each a_j. Therefore, we could not follow OWZ using the Szegö theorem [24]. Instead, we evaluated the expectation in (7) numerically by generating the signals $\mathbf{z}$ with random sequences $\mathbf{a}$, computing J for each $\mathbf{z}$ and averaging h_z. The Jacobian matrix J is evaluated by

$$\frac{\partial z_i}{\partial a_j} = \pm 2 \frac{\sin(\pi t_{ij}/T)}{\pi t_{ij}/T}, \quad (8)$$

where t_{ij} is the time elapsed from the time of transition a_i to the sample z_i. The sign is positive for transitions from 1 to -1 and negative otherwise. The computation was executed on cyclic sequences 500 and 1000 symbols long and verifying identical result in both cases.

Denote

$$h_d = h_z - h_a = \frac{1}{N} \int p(\mathbf{a}) \log \left| \frac{\partial z_i}{\partial a_j} \right| d\mathbf{a}.$$

The result of numerical evaluation of scheme A is $h_d = 0.5197$ nats for $T = 1$ and $P = 1$ and is invariant with T, see (4), (6) and (8).

The entropy (7) is identical in scheme A and in scheme B with its alternate sign inversions. This is because $\frac{\partial z_i}{\partial a_j}$ changes sign when $\mathbf{s}_j = -1$, so for scheme B we can create a new auxiliary vector

$$\hat{\mathbf{a}} = (a_1 s_1, \dots a_i s_i \dots) \quad (9)$$

in which $\left(\frac{\partial z_i}{\partial \hat{a}_j} \right)$ is identical to $\left(\frac{\partial z_i}{\partial a_j} \right)$ in scheme A and $h(\hat{\mathbf{a}}) = h(\mathbf{a})$ yielding the same h_z in schemes A and B.

The true entropy of z is larger by $0.5 \log(P)$ due to multiplication by $\sqrt{P}$ and, a has a unity support, so $h_a = 0$, see (6). Thus,

$$h_z = \frac{1}{N} h(\mathbf{z}) = h_d + 0.5 \log P.$$

The entropy of the sampled noise at filter output is

$$h_n = 0.5 \log (2\pi e B N_0).$$

By EPI [22], the entropy of the sum is upper bounded in terms of entropies of its components:

$$e^{2h_y} \geq e^{2h_z} + e^{2h_n}$$
$$e^{2h_y} \geq e^{2(h_d + 0.5 \log(P))} + 2\pi e B N_0$$
$$h_y \geq 0.5 \log \left(e^{2h_d + \log(P)} + 2\pi e B N_0 \right) \quad (10)$$
$$I_{ya} \overset{\triangle}{=} \frac{1}{N} I(\mathbf{y}; \mathbf{a}) = h_y - h_{y|\mathbf{a}}$$
$$\geq 0.5 \log \left(e^{2h_d + \log(P)} + 2\pi e B N_0 \right) - 0.5 \log(2\pi e B N_0)$$
$$I_{ya} \geq 0.5 \log \left(\frac{e^{2h_d + \log(P)} + 2\pi e B N_0}{2\pi e B N_0} \right)$$
$$I_{ya} \geq 0.5 \log \left(\frac{\frac{1}{2\pi e} P \cdot e^{2h_d} + B N_0}{B N_0} \right) \quad (11)$$

The AWGN capacity is

$$C_{\text{AWGN}} = 0.5 \log \left(\frac{P + B N_0}{B N_0} \right).$$

So the power gain at all SNRs over the AWGN channel is lower-bounded by

$$\gamma = \frac{1}{2\pi e} e^{2h_d} = 0.1656.$$

OWZ [6] reported a better result, $\gamma_{OWZ} = 0.1753$, see (3) above; our software reconstructs this result as a verification.

The same analysis technique used here for the brickwall channel response is applicable to a general channel frequency response $H(f)$. To extend the technique to a more general $H(f)$, the sinc pulse used in (8) to compute the Jacobian matrix and shown in Figure 3 would be replaced by the new channel impulse response. Furthermore, $H(f)$ would need to be non-zero over $0 < f < B$ to fulfill the conditions of Lemma 1.

Next we show an improved performance in scheme C. To increase $h(z)$, the polarity of each pulse is inverted at random. As seen in Figure 5 this also removes the wasted discrete tones from the signal spectra. The analysis above cannot be applied directly since now $x(t) \longrightarrow z(t)$ is not a bijection as demonstrated by construction of pairs of signals $x(t)$ the difference of which have period of T and a zero mean, thus zero PSD in the 0 to B frequency band. For scheme C the system mutual information between the modulator inputs $\mathbf{a}$, $\mathbf{s}$ and the channel output is:

$$I(\mathbf{a}, \mathbf{s}; \mathbf{y}) = I(\mathbf{s}; \mathbf{y}) + I(\mathbf{a}; \mathbf{y}|\mathbf{s}). \quad (12)$$

The second term on the r.h.s. is equal to schemes A and B, the signs $\mathbf{s}$ on which this term is conditioned are treated by the auxiliary vector $\hat{\mathbf{a}}$ as defined in (9) for scheme B. The first term on the r.h.s. is the improvement achieved by scheme C relative to schemes A and B. We lower-bound it as follows. Denote the sequence of derivatives of $y(t)$ at times nT by $\dot{\mathbf{y}} = \{\dot{y}_n\}$.

Now

$$
\begin{aligned}
I(\mathbf{s}; \mathbf{y}) &\geq I(\mathbf{s}; \dot{\mathbf{y}}) \\
&= \sum_n I\left(s_n; \dot{\mathbf{y}}|\mathbf{s}_1^{n-1}\right); \quad \mathbf{s}_1^{n-1} = s_1 \ldots, s_{n-1} \\
&\geq \sum_n I(s_n; \dot{\mathbf{y}}) \\
&\geq \sum_n I(s_n; \dot{y}_i). \quad (13)
\end{aligned}
$$

The first line is since $\dot{\mathbf{y}}$ is a function of $\mathbf{y}$. The second line is by the standard mutual information decomposition [22]. The third line is since s_n is independent of $\mathbf{s}_1^{n-1}$, the fourth line is since $\dot{y}_i$ is a subset of the sequence $\dot{\mathbf{y}}$. The last term was evaluated by simulation of scheme C while estimating the symbol-wise probability densities $P(\dot{y}_n|s_n = 1)$, $P(\dot{y}_n|s_n = -1)$ and $P(\dot{y}_n)$ as plotted in Figure 7. It adds 0.136 bits per symbol at asymptotically high SNR which is equivalent to a power gain of $\gamma = 1.207$. Scheme C achieves $\gamma = 0.20$, moderately better than OWZ. We expect that better detectors would improve upon this lower bound.
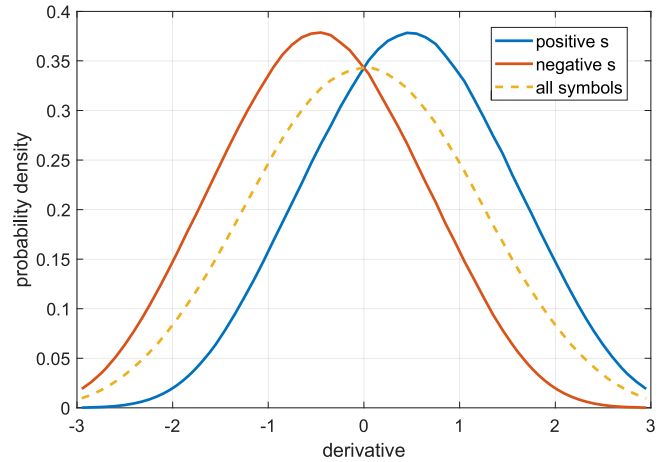

FIGURE 7. Probability density functions of signal derivatives conditioned on signs s_n.

Next we evaluate lower bound on performance of scheme D. Lemma 1 holds for scheme D in which the difference signal is as in Figure 6 with the same average number of pulses except for not confining each pulse to its own T-interval. That is, in both the schemes B and D, the total number of all the negative and positive pulses in the difference signal is half of the total number of sign transitions in x_1 and x_2.

The entropy of $\mathbf{a}$ in (6) is now calculated numerically as

$$
\begin{aligned}
h_a &\triangleq \frac{1}{N} \sum_n h\left(a_n|a_1^{n-1}\right) \\
&= \frac{1}{N} \sum_n \log(W_s(n)). \quad (14)
\end{aligned}
$$

It is larger by 0.4095 nats than that of scheme B, contributing to the performance. Scheme D achieved the best performance among the four schemes, see Table 2.

With scheme D, the achievable lower bound has an advantage of a power factor of 1.47 at all SNRs and of 0.28 bits per Nyquist interval T at high SNR over the scheme reported in [6]. Comparing to Table 1, the gap between the upper and the lower bounds specific to the schemes is 0.43 and 0.438 bits per Nyquist interval for schemes A and C respectively. The gap between the upper bound in [20], entry 6 in Table 2, and the best achievable lower bound, scheme D in the table, is 0.93 bits per Nyquist interval.

The four schemes have distinct attributes. Schemes A and B serve to build up the theoretical base and they provide a lower bound on capacity valid for all SNRs. Scheme C is an extension providing an improved lower bound at high SNR and an improved spectra. Scheme D provides the best lower bound at all SNRs.

Proposition 1: The capacity loss incurred by imposing a constraint of a symmetric binary bipolar input on the AWGN frequency-flat low-pass channel with a given average input power is limited to 0.976 bits per Nyquist interval at high SNR. The equivalent power loss ratio is no less than 0.2586 at all SNR.

TABLE 2. Comparison of different approaches.

Scheme	Equivalent power factor γ relative to Gaussian signal with rectangular spectra 0 to B Hz	Information loss at high SNR relative to Gaussian signal with rectangular spectra 0 to B Hz in bits/Nyquist interval
Gaussian signal with rectangular spectra 0 to B Hz	1	0
OWZ , [6], achievable lower bound	0.1753	1.256
Schemes A and B, achievable lower bound	0.1656	1.2973
Scheme C, achievable lower bound	0.20	1.1613
Scheme D, achievable lower bound	0.2586	0.976
Upper bound in [20]	0.9337	0.0495
Upper bound on RTS as in [20]	0.6271	0.3366

*Proof:* Compare scheme D in Table 2 to the first entry in the table. ∎

IV. CONCLUSION AND OUTLOOK

We studied communication systems over the band-limited AWGN channel in which the transmitter output is constrained to be binary bipolar. We presented new schemes which provide an improved lower bound on the capacity of this channel. The gap between the known upper bound and our new achievable lower bound is reduced to 0.93 bits per Nyquist interval at high SNR. Furthermore, the schemes operate at a much lower rate of sign transitions than the bipolar signaling that achieves the PAM based bounds in [6].

There is a room for future work attempting to improve the achievable lower bound. For this purpose signals with spectra more concentrated in the lower frequency regions than our scheme C, see Figure 5, should be investigated. Interestingly, the maximal power factor γ of the Random Telegraph Signal (RTS) is achieved with average transition rate of about 0.67 per Nyquist interval, less than the 1.5 average transition rate of our scheme C leading to a narrower PSD, thus a future analysis of performance of the RTS signaling might reduce the gap between the upper and lower bounds further.

The lower bound on performance presented here might be improved in future work based on techniques that consider PWM and also RTS in terms of lower bounding the filtered minimum mean square error, and incorporating the Information Estimation relations [27]. Further interesting useful techniques developed for ISI channels [25], [26] should also be considered.

In this paper the signals are designed for good performance in the high SNR regime while the results for schemes A, B and D hold for all SNRs, see (11). Future work may address the non-asymptotic low and intermediate SNR region based on new schemes adapted to SNR and on advanced FTN techniques listed in the introduction for which the Shamai-Ozarow-Wyner [9] bound is of direct relevance.
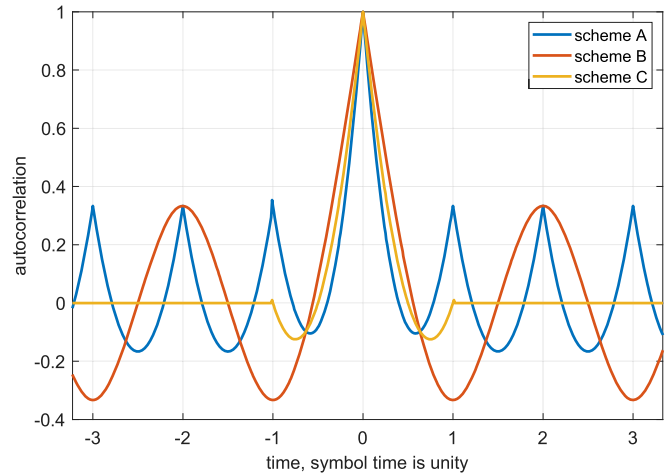
FIGURE 8. Autocorrelation functions for schemes A, B and C.

APPENDIX—AUTOCORRELATIONS

Denote the autocorrelation of $x(t)$ as

$$R(\tau) = E_{x,t}[x(t) \cdot x(t+\tau)],$$

where $E_{x,t}$ denotes expectation over x and over $-T < t < T$.

For scheme C we have

$$R_c = \begin{cases} \left(1 - \frac{2|\tau|}{T}\right)\left(1 - \frac{|\tau|}{T}\right); & \frac{|\tau|}{T} < 1 \\ 0; & \text{otherwise.} \end{cases} \quad (15)$$

The first parenthesis is the correlation given that t and $+t$ are in the same symbol interval, the second parenthesis is the probability of this occurrence. The expression for cases A and B is a little more involved. For $\frac{|\tau|}{T} > 1$, $x(t)$, and $x(t+\tau)$ are independent, thus

$$R(\tau) = E_t\{E_x[x(t)] \cdot E_x[x(t+\tau)]\}; \quad \frac{|\tau|}{T} > 1; \quad (16)$$

For scheme A, $E[x(t)] = 1 - 2/T$. It follows by a straightforward integration for scheme A:

$$R_A(\tau) = \frac{1}{3} + 2\tau_n^2 - 2\tau_n; \quad \frac{|\tau|}{T} > 1,$$

where $\tau_n = \frac{|\tau|_{\text{mod}\,T}}{T}$ and $|\tau|_{\text{mod}\,T}$ denotes the modulo T operation. For scheme B:

$$R_B(\tau) = \frac{1}{3}(4\tau_n^3 - 6\tau_n^2 + 1) \cdot (-1)^{\lfloor|\frac{\tau}{T}|\rfloor}; \quad \frac{|\tau|}{T} > 1,$$

where $\lfloor \tau \rfloor$ is the largest integer smaller than τ.

For $\frac{|\tau|}{T} < 1$, the expectation over t is the sum over the events in which t and $\tau + t$ are in the same symbol interval which yields (15) and of a term contributed by the events where t and $\tau + t$ fall into successive symbol intervals where (16) applies. The result is:

$$R_{AB}(\tau) = \left(1 - \frac{2|\tau|}{T}\right)\left(1 - \frac{|\tau|}{T}\right)$$
$$\pm \left(-\frac{2}{3}\left|\frac{\tau}{T}\right|^3 + 2\left|\frac{\tau}{T}\right|^2 - \left|\frac{\tau}{T}\right|\right); \quad \frac{|\tau|}{T} < 1 \quad (17)$$

where the sign is positive for A and negative for B. Collecting the equations above yields (5) and Figure 8.

ACKNOWLEDGMENT

REFERENCES

[1] J. Huang, K. Padmanabhan, and O. M. Collins, "The sampling theorem with constant amplitude variable width pulses," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 6, pp. 1178–1190, Jun. 2011.

[2] Y. Fan and R. J. Green, "Comparison of pulse position modulation and pulse width modulation for application in optical communications," *Opt. Eng.*, vol. 46, no. 6, Jun. 2007, Art. no. 065001.

[3] A. Pradana, N. Ahmadi, and T. Adionos, "Design and implementation of visible light communication system using pulse width modulation," in *Proc. Int. Conf. Elect. Eng. Informat. (ICEEI)*, Denpasar, Indonesia, 2015, pp. 25–30, doi: 10.1109/ICEEI.2015.7352464.

[4] A. Li, C. Masouros, A. L. Swindlehurst, and W. Yu, "1-bit massive MIMO transmission: Embracing interference with symbol-level precoding," *IEEE Commun. Mag.*, vol. 59, no. 5, pp. 121–127, May 2021.

[5] P. Neuhaus, M. Dörpinghaus, and G. Fettweis, "Zero-crossing modulation for wideband systems employing 1-bit quantization and temporal oversampling: Transceiver design and performance evaluation," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1915–1934, 2021, doi: 10.1109/OJCOMS.2021.3094927.

[6] L. H. Ozarow, A. D. Wyner, and J. Ziv, "Achievable rates for a constrained Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 365–370, May 1988.

[7] S. Shamai, "On the capacity of a Gaussian channel with peak power constrained and bandlimited input signals," *AEÜ Electron. Commun.*, vol. 42, no. 6, pp. 340–346, Nov./Dec. 1988.

[8] L. A. Shepp, "Covariances of unit processes," in *Proc. Working Conf. Stochastic Process.*, Santa Barbara, CA, USA, 1967, pp. 205–218.

[9] 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. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov. 1991.

[10] S. Shamai, "On the information capacity of the continuous-time Gaussian channels with constrained bipolar inputs," in *Proc. Inf. Theory Workshop*, Susono, Japan, Jun. 1993.

[11] S. Shamai and I. Bar-David, "On the capacity penalty due to input-bandwidth restrictions with an application to rate-limited binary signaling," *IEEE Trans. Inf. Theory*, vol. 36, no. 3, pp. 623–626, May 1990.

[12] N. Chayat and S. Shamai, "Bounds on the information rate of intertransition-time-restricted binary signaling over an AWGN channel," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1992–2006, Sep. 1999.

[13] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen, "Faster-than-Nyquist signaling: An overview," *IEEE Access*, vol. 5, pp. 1925–1940, 2017.

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

[15] J. Zhou *et al.*, "Capacity limit for faster-than-Nyquist non-orthogonal frequency-division multiplexing signaling," *Sci. Rep.*, vol. 7, p. 3380, Jun. 2017.

[16] W. Kang, H. Che, and K. Zhu, "Precoding optimization for faster-than-Nyquist signaling with probabilistic shaping," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11470–11478, Oct. 2020.

[17] S. Li, J. Yuan, B. Bai, and N. Benvenuto, "Code-based channel shortening for faster-than-Nyquist signaling: Reduced-complexity detection and code design," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 3996–4011, Jul. 2020.

[18] M. Ganji, X. Zou, and H. Jafarkhani, "On the capacity of faster than Nyquist signaling," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1197–1201, Jun. 2020.

[19] Y. J. D. Kim, J. Bajcsy, and D. Vargas, "Faster-than-Nyquist broadcasting in Gaussian channels: Achievable rate regions and coding," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1016–1030, Mar. 2016.

[20] S. Shamai and I. Bar-David, "Upper bounds on capacity for a constrained Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1079–1084, Sep. 1989.

[21] B. McMillan, "History of a problem," *J. Soc. Indust. Appl. Math.*, vol. 3, no. 3, pp. 114–128, 1955.

[22] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA, USA: Addison-Wesley, 1987.

[23] A. Eremenko and D. Novikov, "Oscillation of functions with a spectral gap," in *Proc. Nat. Acad. Sci. United States Amer.*, vol. 101, no. 16, pp. 5872–5873, 2004. [Online]. Available: https://doi.org/10.1073/pnas.0302874101

[24] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. New York, NY, USA: Chelsea, 1984.

[25] S. Shamai and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1388–1404, Sep. 1996, doi: 10.1109/18.532881.

[26] Y. Carmon and S. S. Shitz, "Lower bounds and approximations for the information rate of the ISI channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5417–5431, Oct. 2015, doi: 10.1109/TIT.2015.2460252.

[27] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

MICHAEL PELEG (Life Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Technion–Israel Institute of Technology, Haifa, Israel, in 1978 and 1986, respectively. From 1980 to 2000, he was with the Communication Research Facilities, Israel Ministry of Defense. Since 2000, he has been with Rafael Ltd. He is associated with the EE Department, Technion–Israel Institute of Technology, where he is collaborating in research in communications and information theory. His research interests include wireless communications, iterative decoding, multiantenna systems, and radiation safety.

TOMER MICHAELI received the B.Sc. and Ph.D. degrees from the ECE Faculty, Technion–Israel Institute of Technology in 2005 and 2012, respectively, where he is currently an Associate Professor. From 2012 to 2015, he was a Postdoctoral Fellow with the CS and Applied Math Department, Weizmann Institute of Science. His research lies in computer vision and machine learning. In particular, he studies problems in image restoration, generation, and manipulation. He won several awards, including the Marr Prize in 2019, and the Krill Prize of the Wolf Foundation in 2020.
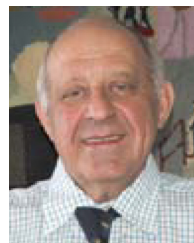
SHLOMO SHAMAI (SHITZ) (Life Fellow, IEEE) is with the Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, where he is also a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering, and the 2017 IEEE Richard W. Hamming Medal. He was a co-recipient of the 2018 Third Bell Labs Prize for Shaping the Future of Information and Communications Technology. He was also a recipient of numerous technical and paper awards and recognitions of the IEEE (Donald G. Fink Prize Paper Award), Information Theory, Communications and Signal Processing Societies, and EURASIP. He is listed as a Highly Cited Researcher (Computer Science) for the years 2004, 2005, 2006, 2007, 2008, and 2013. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY. He has also served twice on the Board of Governors for the Information Theory Society. He has also served on the Executive Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE Information Theory Society Nominations and Appointments Committee, and the IEEE Information Theory Society, Shannon Award Committee. He is also an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering.