Increased Information Rate by Oversampling

E. N. Gilbert

Abstract—A noiseless ideal low-pass filter, followed by a limiter, is normally used as a binary data channel by sampling the output once per Nyquist interval. Detectors that sample more often encounter intersymbol interference, but can be used in ways that increase the information rate.

Index Terms—Intersymbol interference, band-limited signals, Nyquist rate, oversampling, limiter.

I. INTRODUCTION

Fig. 1 shows a noiseless channel, consisting of an ideal low-pass filter followed by a limiter. If the input function is \( f(t) \), the output is \( y(t) = \text{sgn}(f(t)) \). Channels of this sort are commonly used with binary data serving directly as the Nyquist samples of a low-pass input function \( x(t) \). The detector then samples the limiter output at the Nyquist times to recover the data and achieve a signaling rate of one bit per Nyquist interval.

A detector that merely samples the channel output at Nyquist times receives information no faster than 1 b/Nyquist interval, regardless of the source. One may reasonable ask if, with a suitable source, the detector can achieve a higher information rate by sampling more often (oversampling). The answer is not obvious, as examples in Section II show. Section III will exhibit a signaling system that achieves an information rate 1.072 b/Nyquist interval by allowing the detector to sample at times a half Nyquist interval apart. This rate increase is small, but it shows that, in principle, oversampling does permit faster rates.

Mazo suggested the present problem as a simplification of a more realistic one in Kalet et al. [1]. The simplification omits details like multiple slicing levels (256 versus 2) and a less convenient sample rate (8/7 versus 2) which seem unrelated to the main problem of showing that extra samples can supply new information. Several authors [2]-[4] have derived high bit-error rates for moderately oversampled digital systems.

II. PRELIMINARIES

Time will be measured in units of the Nyquist interval. Then, the cutoff frequency of the filter in Fig. 1 will be \( W = 1/2 \) and the response of the filter to a unit impulse at time 0 will be

\[
F(t) = \sin \frac{\pi t}{\tau}.
\]

The sampling theorem (Shannon [5], [6]) represents an output function \( f(t) \) of the filter in terms of its Nyquist samples \( f(i) \) at integer times

\[
f(t) = \sum f(i)F(t - i).
\]

If the input \( x(t) \) is itself a low-pass signal, \( f(t) = x(t) \) and the output is \( y(t) = \text{sgn}(x(t)) \). Then, taking binary data \( x(n) = +1 \) or \(-1 \) for the input Nyquist samples will make \( y(n) = x(n) \).

Now, suppose the detector samples twice as often, say at times \( t/2 \) with integer \( i \). Times with even \( i \) are again the Nyquist times. If \( i \) is odd, \( t/2 \) has fractional part \( 1/2 \) and will be called a half-Nyquist time. If the detector could sample \( f(t) \) directly, sampling at the half-Nyquist times would be wasted effort; the Nyquist samples in (2) already determine \( f(t) \). But, since the detector can only observe \( y(t) = \text{sgn}(f(t)) \), the half-Nyquist samples of \( y(t) \) may indeed contain extra information. How can a detector use these samples?

A naive approach might transmit two message sequences, say \( a(j) \) at the Nyquist times and \( b(i) \) at the half-Nyquist times, making the filter output

\[
f(t) = \sum a(j)F(t - j) + \sum b(i)F(t - i + 0.5)\quad (3)
\]

Although \( a(j) \) and \( b(i) \) are the coefficients of the biggest contributors to (3) at times \( t = j \) and \( t = i - 0.5 \), the detection is now complicated. For example,

\[
f(n) = a(n) + \sum b(i)F(n - i + 0.5)\quad (4)
\]

The series in (4) represents a kind of noise (intersymbol interference) obscuring the sample \( a(n) \). It is not obvious that intersymbol interferences will not always reduce the information rate of \( y(t) \) to at most 1 b/Nyquist interval. As a special case, let \( a(j) \) and \( b(i) \) be independent Gaussian variates, all with zero mean and the same variance. The two series in (3) become independent Gaussian noise functions with the same power.

The rate of information about both messages in only 1 b/Nyquist interval, even if the detector observes \( f(t) \) instead of \( y(t) \).

A detector that samples \( y(t) \) many times per Nyquist interval can accurately locate the zeros of \( f(t) \), and perhaps thereby receive information at a rate faster than 1. Zero crossings do not determine a low-pass function, even to within a scale factor. Consider the simple example:

\[
f(t) = \sin (2 \pi ft/3) + c \sin (2 \pi ft) = (1 + c + 2c \cos 4\pi ft/3) \sin (2 \pi ft/3)
\]

where the parameter \( c \) can have any value in \(-1/3 < c < 1\). With \( c \) in that range, the bracketed term remains positive and the zeros of \( f(t) \) are just those of \( \sin (2 \pi ft/3) \). However, at least one low-pass function with prescribed simple zeros does exist if each Nyquist interval contains exactly one of the zeros (see the infinite product for such a function in Logan [7] and Titchmarsh [8]). Using that result, Shamai [9] proposed the following oversampling system, with \( N \) samples per Nyquist interval, to achieve an information rate \( \log(N + 1) \).

Each Nyquist interval \((n, n + 1)\) will contain \( N \) sample points, say at \( n + (2k - 1)/(2N) \), \( k = 1, \ldots, N \), for an evenly spaced
sampling sequence. Sample points cut each Nyquist interval into \( N + 1 \) segments, two of length \( 1/2N \) and the rest of length \( 1/N \). From each Nyquist interval, the source chooses one segment at random, all \( N + 1 \) equally likely. The midpoint of that segment will become a simple zero. The transmitter sends a low-pass signal \( f(t) \) with the chosen zeros. Because a segment has at most one simple zero, a segment with a zero is one having endpoints with \( y(i) \) of opposite sign. Then, zeros in segments between sample points are easily detected. An endpoint of a segment may also be an integer point \( n \), not a sample point, but \( y(n) \) can be deduced from nearby samples because each Nyquist interval contains exactly one zero \( y(n) = -y(n - 1) \). Generating the required signal \( f(t) \) presents practical problems; \( f(t) \) depends on locations of zeros far in the future, and so one must expect long coding delays. The systems to follow are much more elegant.

### III. GAUSSIAN SOURCE

In this section, \( x(t) \) will be a low-pass Gaussian noise signal; the input Nyquist samples \( x(n) \) will be independent Gaussian variables with zero mean and unit variance. The receiver will try to deduce \( x(n) \) from just two output samples, the Nyquist sample \( y(n) \) and half-Nyquist sample \( y(n - 0.5) \). Write \( X = x(n) \) and \( Y = (y(n - 0.5), y(n)) \). The detector reveals information about the input at a rate

\[
H(X) = H(Y) = H(Y|X) \tag{5}
\]

in bits per Nyquist interval. Here, \( H(X), H(Y), H(Y|X), H(Y|X) \) are information and conditional information measures as defined by Shannon [5]. Of course, the detector cannot reconstruct the input \( x(n) \) exactly from a few received binary samples. However, Shannon's coding theorem will guarantee that other sources with rate less than \( R \) can be coded to signal over this channel at low error rate.

Gaussian input samples were chosen purely for analytical convenience. Restricting input samples \( x(n) \) to a few discrete levels would have simplified the practical problem of coding discrete sources for the channel. There is also no reason to expect Gaussian input to maximize the rate \( R \). Using more than two output samples might give a larger rate \( R \), but the objective is to obtain \( R > 1 \) as simply as possible.

It may be intuitively clear that the entire output sequence gives information about the entire input sequence at a rate at least \( R \), the number of two output samples that give about one input sample. A formal proof follows. Write \( X_k \) and \( Y_k \) for the transmitted and received samples in the first \( k \) Nyquist intervals, i.e.,

\[
X_k = (x(1), (2), \ldots, x(k)) \quad \text{and} \quad Y_k = (y(0.5), y(1), \ldots, y(k - 0.5), y(k)).
\]

and write \( kR_k = H(X_k) - H(X_k|Y_k) \) for the information that \( Y_k \) gives about \( X_k \). Since the input samples are independent,

\[
H(X_k) = \sum_{i=1}^{k} H(x(i)).
\]

Also,

\[
H(X_k|Y_k) = H(x(1)|Y_k) + H(x(2)|x(1), Y_k) + \cdots + H(x(i)|x(i-1), Y_k)
\]

\[
\leq \sum_{i=1}^{k} H(x(i))y(i - 0.5), y(i)),
\]

the inequality following because \( X_{i-1} \) and \( Y_k \) determine \( y(i - 0.5) \) and \( y(i) \) exactly. But then,

\[
kR_k \geq \sum_{i=1}^{k} (H(x(i)) - H(x(i)|y(i - 0.5), y(i))) = kR.
\]

To calculate \( R \), begin with the joint probability distribution of \( X \) and \( Y \). Define two Gaussian variables \( f = f(n) \) and \( g = f(n - 0.5) \), so that \( X = f \) and \( Y = (\text{sgn}(g), \text{sgn}(f)) \). The correlation coefficient of \( f \) and \( g \) is

\[
\rho = E(fg) = F(0.5) = 2/\pi. \tag{6}
\]

Using (6), one can easily verify that the new Gaussian variables

\[
u = f - g)/\sqrt{2(1 + \rho)}, \quad v = (f - g)/\sqrt{2(1 - \rho)}
\]

have zero mean, unit variance, and are independent. Then, the probability densities determined by \( (f,g) \), or by \( (u,v) \), are integrals of the elements

\[
e^{-r^2/2}dr dv = e^{-r^2/2}\int_{-\infty}^{\infty}df dg / 2\pi \sqrt{1 - \rho^2} \tag{8}
\]

in which \( r^2 \) is the quadratic form

\[
r^2 = u^2 + v^2 = (f^2 - 2pf + g^2)/(1 - \rho^2)
\]

\[
e^{-r^2/2}\int_{-\infty}^{\infty}df dg / 2\pi \sqrt{1 - \rho^2} \tag{9}
\]

i.e., \( r \) is the radial coordinate in the \((u,v)\) plane.

To evaluate \( H(Y) \), find probabilities of the four possible sign pairs \( ++, +-, -+, -- \) for \( Y \). \( Y = ++ \) if \( (f,g) \) is in the first quadrant, i.e., if \((u,v)\) is in an angular sector \( A \) between the lines \( u = \pm \sqrt{(1 + \rho)/(1 - \rho)} \). The probability that \( Y = ++ \) is easily evaluated as an integral in the \((u,v)\) plane. Express \( u \) and \( v \) in polar coordinates to obtain

\[
P(++) = \int_0^1 e^{-r^2/2}rdr \int_0^{\pi/2} d\theta / 2\pi \]

\[
= 1/\pi \sqrt{1 + \rho} / (1 - \rho) = 0.35983396.
\]

The other outcomes for \( Y \) have probabilities

\[
P(--) = P(+) = 0.5, \quad P(+-) = 0.14016604.
\]

Then, \( H(Y) = 1.855903 \) bit/Nyquist interval.

Because \( X \) is Gaussian, the second coordinate of \( Y \) exactly, \( H(Y|X) \) is just the uncertainty of \( \text{sgn}(g) \) when \( f \) is known. For a given \( f \), the conditional probability density for \( g \) is obtained by dividing the joint density in (8) by the marginal density \( (f^2/2)/\sqrt{2\pi} \) for \( f \). Then, (9) shows the conditional distribution for \( g \) to be Gaussian with mean \( \rho f \) and variance \( 1 - \rho^2 \).

An integration over \( 0 < g < \infty \) gives

\[
P(\text{sgn}(g) = + |f) = \Phi(\rho f/\sqrt{1 - \rho^2}) \tag{10}
\]

where

\[
\Phi(x) = \int_x^\infty e^{-t^2/2} dt / \sqrt{2\pi}
\]

is the normal distribution function. Using (10), one can calculate the uncertainty about \( \text{sgn}(g) \) when \( f \) has a known value. That
uncertainty must be multiplied by the Gaussian density function at \( f \) and integrated to get \( H(Y|X) \). The integration was done numerically, with the result \( H(Y|X) = 0.7838 \). The rate in (5) is then \( R = 1.072 \).

IV. DIGITAL SOURCE

The idea in Section III might be used with digital signals by encoding letters into discrete amplitude levels that approximate a Gaussian distribution. A large alphabet should give a good approximation to a Gaussian input and a rate near 1.072. This section modifies that idea to obtain a rate greater than 1 for digital sources having alphabets of only four letters.

Instead of assigning a specific level to each of the four letters of the digital signal alphabet, the encoding will choose levels at random. The random choices will be designed to make the channel input samples Gaussian so that the methods of Section IV again apply. First, classify the values \( f \) of inputs to the channel input samples Gaussian so that the methods of Section IV again apply. First, classify the values \( f \) of inputs to the

\[
\begin{align*}
-B &= (-\infty, -L), \\
-S &= (-L, 0), \\
0 &= (0, L), \\
+B &= (L, \infty)
\end{align*}
\]

with \( L \) any convenient positive number (\( B \) and \( S \) stand for “big” and “small”). Also, use \(-B, -S, S, B\) as names for the four letters of the digital alphabet. Whenever letter \( Z \) is to be transmitted \((Z = -B, -S, S, B)\), the input Nyquist sample to the channel will be chosen at random from the range \( Z \) in (11). To make the input Gaussian, the letter probabilities must be

\[
P(B) = P(-B) = 1 - \Phi(L),
\]

\[
P(S) = P(-S) = \Phi(L) - 0.5.
\]

(12)

Thus, a digital source with four equally likely letters will require \( L = 0.67449 \). Also, for each letter \( Z = -B, -S, S, B \), the random Nyquist sample \( f \) must be chosen from range \( Z \) with a conditional probability density

\[
p(f|Z) = \begin{cases} 
\frac{e^{-f^2/2}}{P(Z)\sqrt{2\pi}} & \text{if } f \in Z \text{ in (11)}, \\
0 & \text{otherwise}.
\end{cases}
\]

(13)

The signaling rate is now

\[
R = H(Z) - H(Z|Y) = H(Y) - H(Y|Z).
\]

(14)

Input letters \( Z \) will be assumed to arrive independently with the probabilities (12). The outputs \( Y \) are binary pairs \((\text{sgn}(g), \text{sgn}(f))\) of one half-Nyquist sample and one Nyquist sample, as in Section III. The rate \( R \) is then what would be achieved by a memoryless source signaling over a noisy memoryless discrete channel with four inputs and four outputs. The transition probabilities of this channel can be found from (8) and (13) in order to find \( R \).

The second equation of (14) will be convenient; Section III has already found \( H(Y) = 1.855903 \). Since the output sample \( y(n) \) always has the same sign as \( Z \), there are only eight nonzero transition probabilities \( P(Y|Z) \). These are simply related to one joint probability \( Q = P(S, +) = P(-S, -) \) and letter probabilities (12) by

\[
P(+ + | S) = P(- - | - S) = Q / P(S)
\]

\[
P(+ + | B) = P(- - | - B) = (P(+ +) - Q) / P(B)
\]

\[
P(- - | S) = P(+) + S) = 1 - P(+ + | S)
\]

\[
P(- + | B) = P(+) + B) = 1 - P(+ + | B)
\]

(15)

The value \( P(+ +) = 0.35983396 \), from Section III, appears in (15). \( Q \) is found by integrating (8) over the region \( 0 < g < \infty \), \( 0 < f < L \). Use the last equation in (9) to obtain

\[
Q = \int_{1}^{\infty} \phi \left( \sqrt{1 - \rho^2} \right) \frac{e^{-f^2/2}}{\sqrt{2\pi}} df
\]

and compute \( Q \) numerically. The letter probabilities (12) and transition probabilities (15) now determine the conditional information \( H(Y|Z) \) and the information rate (14). These numbers all depend on the value \( L \). At \( L = 0.67449 \), which makes the four letters equally likely, \( Q = 0.1511 \), \( P(+) = 0.6644 \), \( P(+ | B) = 0.8349 \), \( P(+) = 0.8073 \), and \( R = 1.049 \).

The choice \( L = 0.67449 \) maximizes \( H(Z) \), but not \( R(H(Z|Y) \) also depends on \( L \). As a function of \( L \), \( R \) has a broad maximum, attaining \( R = 1.0592 \) at \( L = 0.83 \).

This signaling system might be used to multiplex two binary sources. Let a sign source have alphabet \(+, -\) and rate 1. Let an amplitude source have alphabet \((S, B)\) and rate less than \( R = 1 \). The letter pairs \(-B, -S, +S, +B\) of the combined source are then suitable channel inputs. The receiver simply decodes \(-, +, +, +, +, -B, -S, +S, +B\), respectively. That reconstructs the sign message exactly. Errors in the decoded letters \( S, B \) may be corrected if the amplitude message was suitably encoded.

V. OTHER SPACING

An alternative to sampling at evenly spaced times is to displace the half-Nyquist times slightly. Let \( y(n - T) \) and \( y(n) \), for some \( T \) different from 0.5, be used for \( g \) and \( f \) in Sections III and IV. The equations of Sections III and IV remain unchanged, with \( \rho \) in (6) replaced by a new value \( \rho = F(T) \). In fact, other values of \( T \) can improve the rate \( R \).

This section will take \( T = 0.44294647 \), which makes \( \rho = 1/\sqrt{2} \). Then, in (10) and (16), \( \rho / \sqrt{1 - \rho^2} = 1/\sqrt{2} \). That simplification allows all the integrals of Sections III and IV to be done analytically. \( P(+ +) = P(- -) = 3/8, P(+ -) = P(- +) \) is 1/8, so that \( H(Y) = 1.811278 \). In (10), \( P(\text{sgn}(g) = +1|f) = \Phi(f) \), and now

\[
H(Y|X) = 2 \int_{0}^{1} \{ - \Phi \log \Phi - (1 - \Phi) \log (1 - \Phi) \} d\Phi,
\]

\[
= 1/(2 \ln 2) = 0.72135.
\]

The rate in Section III increases to \( R = 1.08993 \).

In Section IV, (16) simplifies to

\[
Q = \int_{\Phi(0)}^{\Phi(L)} \frac{d\Phi}{\Phi(0)} \Phi d\Phi \frac{1}{2} \left( \Phi^2(L) - \frac{1}{4} \right)
\]

which, together with \( P(+ +) = 3/8 \), reduces the transition probabilities (15) to expressions involving \( \Phi(L) \) only. With \( L = 0.67449 \) for equally likely letters, the rate is \( R = 1.06228 \). Changing \( L \) to 0.8 increases the rate to \( R = 1.06344 \).

REFERENCES


Cascading Runlength-Limited Sequences

Jos H. Weber, Member, IEEE, and Khaled A. S. Abdel-Ghaffar

Abstract—In magnetic or optical storage devices, it is often required to map the data into runlength-limited sequences. To ensure that cascading such sequences does not violate the runlength constraints, a number of merging bits are inserted between two consecutive sequences. A theory is developed, in which the minimum number of merging bits is determined and the efficiency of a runlength-limited fixed-length coding scheme is considered.

Index Terms—Fixed-length codes, modulation systems, runlength-limited sequences.

I. INTRODUCTION

Many modulation systems used in magnetic and optical recording are based on binary runlength-limited codes. A string of bits is said to be runlength-limited if the number of zeros between two consecutive ones is bounded by a certain minimum value \( d \) and a certain maximum value \( k \). Such sequences are also called \((d,k)\) constrained. The lower runlength constraint is imposed to reduce intersymbol interference, while the upper runlength constraint is imposed to maintain synchronization. For a general introduction to the theory of runlength-limited sequences and its applications, we refer to the excellent overview papers by Siegel [7] and Immink [3].

In order to prevent the cascading of runlength-limited sequences of fixed length from violating the runlength constraints, some merging techniques have been given by Tang and Bahl [8] and Beekker and Immink [1]. In this paper, we will develop a general theory on cascading runlength-limited sequences, in which the above-mentioned methods appear as special cases. The following notation will be used. Let \( d,k,l,r \), and \( n \) be integers satisfying \( 0 \leq d < k \), \( 0 \leq l \leq r \leq k \), and \( n \geq 1 \), unless explicitly stated otherwise. Let \( M_{d,k,l,r}(n) \) denote the set of binary sequences of length \( n \), starting with at most \( l \) zeros, ending with at most \( r \) zeros, and having at least \( d \) and at most \( k \) zeros.

Manuscript received May 23, 1992; revised March 3, 1993. This work was supported by the National Science Foundation under Grants NCR 89-08105 and NCR 91-15423. The work of K. A. S. Abdel-Ghaffar was supported by an IBM Faculty Development Award. This paper was presented in part at the 12th Symposium on Information Theory in the Benelux, Eindhoven, The Netherlands, May 1990, and at the 1991 IEEE International Symposium on Information Theory, Budapest, Hungary, June 1991.

J. H. Weber is with the Information Theory Group, Department of Electrical Engineering, Delft University of Technology, 2600 GA Delft, The Netherlands.

K. A. S. Abdel-Ghaffar is with the Department of Electrical Engineering and Computer Science, University of California, Davis, CA 95616.

IEEE Log Number 9213023.