This paper appears in *Code Division Multiple Access Communications* (Eds. S. G. Glisic and P. A. Leppänen). Boston, Dordrecht and London: Kluwer, 1995, pp. 29-46.

Towards an Information Theory of Spread-Spectrum Systems

James L. Massey

Signal & Info. Proc. Lab., Swiss Federal Inst. of Tech., Ch-8092 Zürich

Abstract–A novel definition of a spread-spectrum signal as a signal whose Fourier bandwidth is much greater than its Shannon bandwith (one-half the number of dimensions of signal space required per second) is proposed. Six different communication systems are analyzed in terms of this definition. It is shown that there is a fundamental difference between the bandwidth expansion due to coding and that due to "spectrum spreading". It is further shown that spectrum spreading plays no role in increasing channel capacity, but can perform other useful roles such as providing low probability of interception of the signal, good electromagnetic compatibility, and a multiple-access capability. The effects of linear and nonlinear filtering on bandwidth are considered and seen to be quite different for Fourier bandwidth and for Shannon bandwidth. The concepts developed are used to resolve two paradoxes in spread-spectrum communications: the apparent increase in capacity when users become unsynchronized in a code-division multiple-access (CDMA) system and the fact that a heavily loaded CDMA system is as energy-efficient for transmitting information as a single-user system with the same (total) average power constraint. Areas of spread-spectrum communications where further information-theoretic development is needed are indicated.

1. INTRODUCTION

The main purpose of this paper is to consider, from the fundamental viewpoint of Shannon's information theory [1], systems that employ spread-spectrum signals . To do this requires that we carefully define what we mean by a spread-spectrum signal. This is done in Section 2 in which we give a rather unconventional definition of a spread-spectrum signal, but the only one that we were able to formulate that we ourselves found to be satisfactory. To illustrate the implications of this definition, we consider the transmitted signals in six different communication systems in Section 3 to see which qualify (under our definition) to be called spread-spectrum signals. In Section 4, we consider various reasons why one might wish to use a spread-spectrum signal. In Section 5, we make a more strictly information-theoretic investigation of single-sender systems where we show that spreading the spectrum of the transmitted signal can never increase capacity but also that such spreading need not decrease capacity significantly. In Section 6, we consider the quite different effects of linear and nonlinear filtering on the Shannon bandwidth and the Fourier bandwidth of a signal. In Section 7, we use the theory that has been developed in the previous sections to resolve two paradoxes that arise in spread-spectrum communications, namely the apparent increase in capacity when users become unsynchronized in a code-division multiple-access (CDMA) system and the fact that a heavily loaded CDMA system is as efficient for transmitting information as a single-user system with the same (total) average power constraint. In Section 8, we conclude with some remarks as to what more must be done to reach an information theory of spread-spectrum systems that can be used as a basis for making sound practical judgements and choices.

Throughout this paper, we have limited ourselves for simplicity to baseband signals, but the reader should have no difficulty in adapting our approach to passband signals.

2. WHAT IS A SPREAD-SPECTRUM SIGNAL?

In his brilliant treatise [1] that established the field, Shannon called information theory the "mathematical theory of communication". We have often maintained that, in a very real sense, mathematics is definitions. Once the definitions are in place, all the lemmas, theorems and corollaries are determined; one has only to find them and prove them. If we wish to say something about the information theory of spread-spectrum systems, it follows that our unavoidable first task must be to define such systems. Of course, it is "signals" rather than "systems" that have spectra so that our task, more precisely formulated, is to define spread-spectrum signals. This task may well strike the reader as either superfluous or quixotic. Like the U.S. supreme court justice who admitted the difficulty of defining pornography but claimed that he knew it when he saw it, many communication engineers might maintain that a definition is not needed; they know a spread-spectrum signal when they see it. One such friend described a spread-spectrum communication signal to us as "one that uses much more bandwidth than it needs". There seems to be a certain coarse truth in this description, but it will hardly do for mathematical purposes. After some futile attempts to make this description more precise, our friend concluded that a satisfactory general definition of a spread-spectrum signal is not possible, which whetted our appetite to take a stab at formulating one.

Every communication engineer is familiar with the ordinary notion of bandwidth, which we will call *Fourier bandwidth* both to honor the French pioneer in this field and to distinguish it from a less familiar but no less important type of bandwidth. The "sinc pulse" $m(t) = \text{sinc}(2Wt)$, where $\text{sinc}(x) = \sin(\pi \cdot x)/(\pi \cdot x)$, has a Fourier Bandwidth of W Hz, as one sees immediately from its Fourier transform M(f) shown in Fig. 1. For less dichotomous spectra, there are many options for calculating the precise Fourier bandwidth (rms bandwidth, 3 dB bandwidth, 99% energy bandwidth, etc.), but they are all roughly equivalent and any is good enough for our purposes. The notion of Fourier bandwidth extends easily from deterministic signals to stochastic processes (such as modulated signals) in a way familiar to all communication engineers.

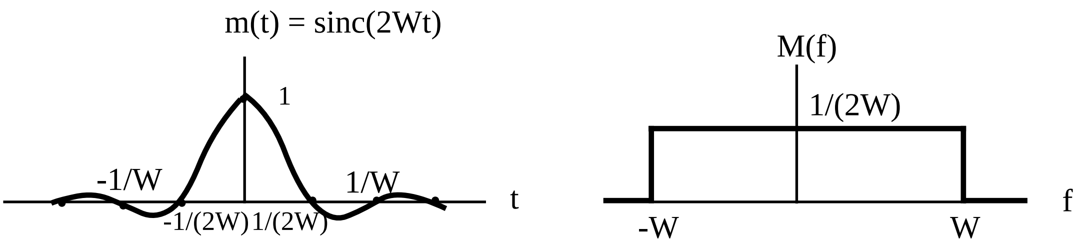


Fig. 1: The Sinc pulse m(t) = sinc(2Wt)
and its Fourier Transform M(f).

The second type of bandwidth, which we will call *Shannon bandwidth* because Shannon [2] was the first to appreciate its importance, makes no real sense for a deterministic signal since it always zero for a single time function. Non-zero Shannon bandwidth implies a "variable" signal (or a stochastic process) such as a modulated signal s(t) that can take on any of a multiplicity of time functions as its value. To determine the Shannon bandwidth of such a signal, one must in principle consider a signal-space representation of s(t) over some very long time interval, say the interval $0 \le t < T$. By this we mean that one must find orthonormal functions, $\phi_i(t)$, i = 1, 2, .., N, so that one can represent (or very well approximate) every possible realization of s(t) by some choice of the coefficients s_1, s_2,..., and s_N in the linear combination

$$s(t) = \sum_{i=1}^{N} s_i \phi_i(t) \qquad (1)$$

for $0 \le t < T$. One says then that one has a signal-space representation of s(t) as a vector $\mathbf{s} = (s_1, s_2, ..., s_N)$ in N-dimensional Euclidean space. When one does this in such a way as to minimize the dimensionality N of the signal space, i. e., to minimize the number of orthonormal functions used, then one has arrived at the *Shannon bandwidth* B, which we now define as

$$B = \frac{1}{2}\frac{N}{T} \text{ (dim/sec).} \qquad (2)$$

Equivalently, the Shannon bandwidth is *one-half the minimum number of dimensions per second required to represent the modulated signal in a signal space.* [In earlier papers [3], [4] where we used the notion of Shannon bandwidth, we omitted the division by 2 in (2). Emboldened by Emerson's dictum that "a foolish consistency is the hobgoblin of small minds" [5], we have now opted for the the factor 2 in the denominator of (2) in order to avoid many such factors elsewhere.]

We now state what might be called *the fundamental theorem of bandwidth:*

> The Shannon bandwidth B of a modulated signal is at most equal to its Fourier bandwidth W; [Rough] equality holds when the orthonormal functions are $\phi_i(t) = \sqrt{2W}$ sinc(2Wt - i) [or any orthonormal functions whose spectra are nearly flat in magnitutde for -W < f < W and nearly zero elsewhere].

There are many proofs of this theorem; Shannon [2] gives a conceptually simple proof whose essence he credits to Nyquist [6] and Gabor [7]; essentially one shows that one can construct 2WT orthonormal functions of Fourier bandwidth W or less that are confined within the time interval $0 \le t < T$ when WT >> 1, but that one can construct no more than this. See the insightful book of Wozencraft and Jacobs [8] and the penetrating paper of Slepian [9] for further discussion of this theorem.

We are now ready to offer our *definition of a spread-spectrum signal as a signal whose Fourier bandwidth is substantially greater than its Shannon bandwidth.* If one considers the Shannon bandwidth to be the amount of bandwidth that the signal *needs* (and we will offer arguments to this effect later) and the Fourier bandwidth to be the amount of bandwidth that the signal *uses*, then we are back at our friend's pithy characterization of a spread-spectrum signal as "one that uses much more bandwidth than it needs".

It is an obvious next step to define the *spreading factor*, γ, of a modulated signal as the ratio of its Fourier bandwidth to its Shannon bandwidth, i.e.,

$$\gamma = W/B. \tag{3}$$

For every modulated signal, $\gamma \le 1$. A spread-spectrum signal is a modulated signal with "large" γ, say $\gamma \le 5$, but of course the precise dividing line between a spread-spectrum signal and an unspread signal is rather arbitrary.

We now have our definitions. It remains to show that they make sense, i.e., that they lead to interesting and useful conclusions.

3. SOME EXAMPLES AND THEIR LESSONS

In digital communication systems (to which we will mostly confine our discussion), the modulated signal in an interval $0 \le t < T$ can assume only finitely many values. In this case, one can in principle always apply the familiar Gram-Schmidt orthogonalization technique (see, e.g., [10, p. 277]) to these signals to obtain an orthonormal basis $\phi_1(t), \phi_2(t), ..., \phi_N(t)$ for the signal space of smallest dimension N containing these signals and can thus determine the Shannon bandwidth according to (2). In most cases of practical interest (as the following examples will illustrate), however, one can find such a basis (and hence find N) by inspection. For an analog system, the modulated signal must generally be treated as a stochastic process. In this case, one can use the Karhunen-Loéve expansion [9, pp. 96-99] to determine N, which will be the number of orthonormal functions in the expansion that have coefficients of non-negligible energy.

We now give several examples of modulated signals, whose analysis will give insight into our definition of a spread-spectrum signal.

Example 1: The modulated signal corresponds to that for one of K users in a time-division multiple-access system in which each user sends L data symbols during each TDMA frame of duration T seconds. Choosing sinc pulses to make the Fourier bandwidth unambiguous, we can write the selected user's modulated signal as

$$s(t) = \sum_{i=1}^{n} b_i \,\mathrm{sinc}\left(\frac{KL}{T} t - i\right), \quad 0 \le t < T \tag{4}$$

where $b_1, b_2, ..., b_L$ are the data symbols. We see that the Fourier bandwidth W must satisfy 2W = KL/T and hence that

$$W = \frac{KL}{2T} . \tag{5}$$

Example 2: The modulated signal corresponds to that for one user in a code-division multiple-access (CDMA) system in which each user modulates a user-specific binary (±1) code sequence of length L with one data symbol in each symbol period of duration T. We can write the selected user's modulated signal as

$$s(t) = b_1 \sum_{i=1}^{L} a_i \operatorname{sinc}(\frac{L}{T} t - i), \quad 0 \le t < T \tag{6}$$

where b_1 is the data symbol and where $(a_1, a_2, ..., a_L)$ is the binary (±1) code sequence. The Fourier bandwidth W satisfies 2W = L/T and hence

$$W = \frac{L}{2T} . \tag{7}$$

Example 3: A user sends an M-ary pulse-position modulated signal in each T second interval, i.e.,

$$s(t) = A \operatorname{sinc}(\frac{M}{T} t - b_1), \quad 0 \le t < T \tag{8}$$

where $b_1 \in \{1, 2, ..., M\}$ is the single data symbol and where A is some fixed amplitude. The Fourier bandwidth satisfies 2W = M/T and hence

$$W = \frac{M}{2T} . \tag{9}$$

Example 4: A user employs binary antipodal signalling to transmit random binary (±1) data symbols in such a manner as to send n such symbols in each T second interval, i.e.,

$$s(t) = \sum_{i=1}^{n} b_i \operatorname{sinc}(\frac{n}{T} t - i), \quad 0 \le t < T. \tag{10}$$

The Fourier bandwidth satisfies 2W = n/T and hence

$$W = \frac{n}{2T} . \tag{11}$$

Example 5: Same as example 4 except that now the "data symbols" are the output of a powerful rate r = 1/n (bits/symbol) trellis encoder fed by random "information bits". Equations (10) and (11) apply unchanged.

Example 6: Same as example 5 except that the code is a trivial r = 1/n code with two binary (±1) codewords, $(a_1, a_2, ..., a_n)$ and its negative. Then

$$s(t) = b_1 \sum_{i=1}^{n} a_i \operatorname{sinc}(\frac{n}{T} t - i), \quad 0 \le t < T \tag{12}$$

where b_1 is the information bit encoded. The Fourier bandwidth is

$$W = \frac{n}{2T}, \tag{13}$$

except for a few pathological choices such as $a_1 = a_2 = ... = a_n$ that cause the Fourier bandwidth to collapse from its typical value as given by (13).

We now consider in which of the above six systems the transmitted signal is in fact a spread-spectrum signal (by our definition). The task reduces essentially to finding the Shannon bandwidth B of each signal.

Example 1 (concluded): By inspection of (4), we see that the signal space is minimally spanned by the orthonormal functions $\phi_i(t) = \sqrt{2W}\,\mathrm{sinc}(2Wt - i)$, i = 1, 2, ..., L, where W is given by (5). Thus, its dimension is N = L so that (2) gives the Shannon bandwidth as B = L/(2T). The spreading factor according to (3) is then just $\gamma = K$. The transmitted signal in this TDMA system is a spread-spectrum signal when the number K of users is large.

Example 2 (concluded): By inspection of (6) and because $a_1, a_2, ..., a_L$ are fixed, we see that the signal space is one-dimensional, i.e., N = 1. Thus the Shannon bandwidth is only B = 1/(2T) and the spreading factor is $\gamma = L$. The transmitted signal in this CDMA system is indeed a spread-spectrum signal when L is large (and our definition of a spread-spectrum signal would be totally indefensible if it were not!).

Example 3 (concluded): From (8), we see that the signal space is minimally spanned by the orthonormal signals $\sqrt{2W}\,\mathrm{sinc}(2Wt - i)$ for i = 1, 2, ..., M, where W is given by (9). Thus N = M and the Shannon bandwidth is B = M/(2T). The spreading factor is $\gamma = 1$, the minimum possible. It follows that such an M-ary PPM signal is *never a spread-spectrum ssignal*, even though according to (9) it requires *Fourier bandwidth very much larger than $\log_2(M)/T$ Hz,* when M is large, *to carry at most $\log_2(M)$ bits of information in each T second interval*. This s(t) is a very wideband signal when M is large, bit it is not a spread signal.

Example 4 (concluded): From (19), we see that the signal space is minimally spanned by the orthonormal signals $\sqrt{2W}\,\mathrm{sinc}(2Wt - i)$ for i = 1, 2, ..., n, where W is given by (11). Thus, N = n, B = n/(2T) and $\gamma = 1$. Binary antipodal modulation, not surprisingly, never produces a spread-spectrum signal.

Example 5 (concluded): For virtually any nontrivial trellis coding system, the encoded symbols b_1, b_2, ..., b_n that appear in (10) will take on such a variety of different possible binary (±1) patterns that one cannot imbed the set of possible s(t) in a signal space of smaller dimension than that required when all n binary symbols can be independently chosen [even though the code, whose binary {0, 1}codewords have been mapped to {+1, -1} in the manner $0 \to +1$ and $1 \to -1$, may be linear over the finite field GF(2)]. Thus both the Fourier and Shannon bandwidths are the same as for the uncoded system of example 4 and $\gamma = 1$. The transmitted signal resulting from binary antipodal modulation of a non-trivially coded information sequence is *never a spread-spectrum signal*, even though, when the code rate r = 1/n is very small, it follows from (11) that it utilizes *a Fourier bandwidth very much larger than 1/T Hz to carry one bit of information every T seconds*.

Example 6 (concluded): Because the trivial code consists of only two codewords, $(a_1, a_2, ..., a_n)$ and its negative, we see from (12) that the signal space has collapsed to a one-dimensional space, i.e., N = 1. The Shannon bandwidth is thus B = 1/(2T) and hence $\gamma = n$. Trivial coding of low rate r = 1/n restricts the input to the binary-antipodal modulator in such a way that its output becomes a spread-spectrum signal. In fact, comparing (6) and (12) we see that such trivial coding gives us precisely the same modulated signal as in the CDMA system of Example 2.

If one accepts our definition of a spread-spectrum signal, then the above examples allow us to draw the following two conclusions:

- From examples 3 and 5, we see that *a large ratio of Fourier bandwidth to information rate does* ***not*** *imply that a signal is spread-spectrum.*

- Example 5 teaches us the, perhaps surprising, lesson that modulating a coded sequence produced by a coding system that expands Fourier bandwidth by a large factor n generally does *not* produce a spread-spectrum signal. *The Fourier bandwidth expansion due to nontrivial coding is fundamentally different from the kind of Fourier bandwidth increase that produces a spread-spectrum system,* although example 6 shows that trivial coding can indeed produce this latter type of expansion.

It is common when considering coded CDMA systems for communication engineers to speak of doing part of the spectrum spreading with coding and part with direct-sequence multiplication–we have not infrequently so spoken ourselves. But we see now that such statements are nonsensical. These are not two parts of some single bandwidth-expansion process, but rather two very different ways of increasing Fourier bandwidth. Much of the remainder of this paper will be devoted to investigating the reasons that one might wish to do one or both of these kinds of bandwidth expansion.

4. WHY SPREAD THE SPECTRUM?

The original motivation for transmitting spread-spectrum signals was a military one, viz., to hide from an enemy the very fact that one was transmitting a signal. Today one speaks of the *low probability of interception* (LPI) of a spread-spectrum signal. The argument that spectrum spreading should provide the possibility for achieving LPI goes as follows. If the signal is confined to a small number N of dimensions within the global signal space of dimension $2WT = N\gamma$ in which all signals of bandwidth W and time-limited to $0 \le t < T$ must lie, and *if there are parameters of the signal that can be varied to create a very large number of possible choices for the N-dimensional signal space occupied by the transmitted signal*, then one can achieve LPI by selecting the values of these parameters at random. [We ignore here the role of the signal power and of the intensity of the noise in the enemy's receiver, which together essentially determine how long it takes to search a given number of dimensions simultaneously for the presence of signal.]

For the CDMA signal of example 2, there are 2^L possible choices of the binary (± 1) parameters $a_1, a_2, ..., a_L$, but changing the sign of all parameters leaves the signal in the same one-dimensional signal space. Fixing $a_1 = +1$ leaves us with 2^{L-1} choices of $a_2, a_3, ..., a_L$, each of which gives a different one-dimensional signal space. For large L, the enemy's task of finding the signal space actually used by the sender is thus akin to "looking for a needle in a haystack". A CDMA signal with a large number L of "chips" per symbol period does indeed afford low probability of interception.

For the TDMA signal of example 1, however, the only parameter that can be varied is the choice of the L consecutive symbol periods (out of a total of KL such periods) in which data symbols will be transmitted. There are only KL possible choices so that a low probability of interception can be achieved only if the product $KL = \gamma L$ is very large. [Spies in World War II frequently used this method to hide their transmission; they transmitted Morse code for only a few seconds and then went silent for very long periods.] The point is that *a large spreading factor* γ *alone is not enough to provide LPI capability, there must be many different ways to choose the signal space.* For the same spreading factor γ, a TDMA signal is much easier to intercept than a CDMA signal because there is much less freedom to choose its signal space.

The twin brother of low probability of interception is *electromagnetic compatibility* (EMC). If it is hard to determine whether a signal is present, then it is obvious that this signal cannot be interfering substantially with other commonly present signals. [If it did interfere strongly with some commonly present signals, we could use those signals in a process to detect its presence.] The excellent EMC capability of a CDMA signal is perhaps the strongest argument that we have today for preferring it to a TDMA signal. [We leave to the reader the task of showing the EMC superiority of a CDMA signal over a frequency-division multiple-access (FDMA) signal.]

The first cousin of low probability of interception is small *inter-user interference* (IUI), which is the prerequisite for a good multiple-accessing capability. If it is difficult to detect the presence of a signal known to be in some class of signals, then shouldn't two such signals hardly interfere with one another? The answer is "yes, at least in a statistical sense!". If two such signals are selected independently at random, then the probability of substantial IUI between them will be small. But care must be taken when the total number of interfering signals is large or when users persist for a long time in using the signal determined by one random choice of parameters. Of course, the K users of a TDMA system with signals as in example 1 will experience no IUI at all when they are well synchronized. But K users (say, K on the order of L) of a CDMA system with signals as in example 2 and with the signal parameters frequently varied will experience IUI with roughly the same statistics no matter whether they are well synchronized or not. This robustness of a CDMA system with regard to inter-user interference is, of course, an important practical consideration.

5. CODING, SPREADING AND NOISE

It is time now to take a more strictly information-theoretic look at the advantages, if any, provided by spectrum spreading. To keep matters simple, we consider the single user system shown in Fig. 2. Of fundamental interest are (1) the *information rate*, R, measured in information bits (i.e., random binary digits) per second at the modulator input; (2) the *capacity,* C_W, also measured in information bits per second, of the channel created by the modulator and the band-limited additive white Gaussian noise (AWGN) waveform channel, which is the upper limit of rates R for which reliable (in the sense of arbitrarily small positive probability of error) recovery of the information bits is possible at the receiver when an appropriate coding system is used; (3) the *average power*, S, of the modulated signal; (4) the *one-sided noise power spectral density*, N_o, of the AWGN; (5) the *Fourier bandwidth*, W, of the bandlimited AWGN waveform channel (which we take without loss of essential generality as equal to the Fourier bandwidth of the modulated signal, as there is no point in transmitting anything outside this band and, if one transmits in a smaller bandwidth, then one might as well reduce the channel bandwidth accordingly); and finally (6) the *Shannon bandwidth,* B, of the modulated signal. Because W is the Fourier bandwidth of s(t), it follows from the fundamental theorem of bandwidth that $B \leq W$.

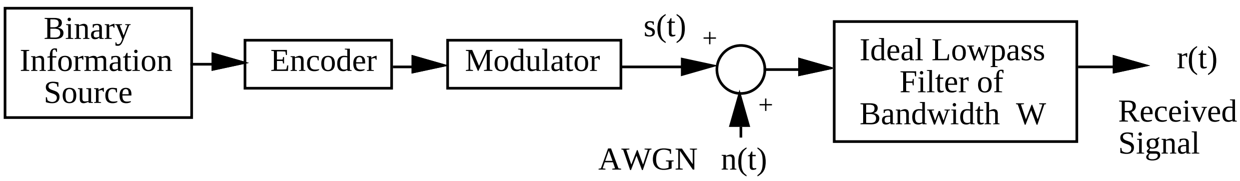


Fig. 2: The single-user communication system under study.

Shannon [1], with his penchant for getting to the heart of the matter, has given us the key relationship among these quantities, namely,

$$C_W = B \log_2(1 + \frac{S}{N_o B}) \quad \text{(bits/sec)}. \tag{14}$$

The reader may be surprised to see the Shannon bandwidth B rather than the Fourier bandwidth W in this expression for C . but he or she will find that (14) is precisely the equation that Shannon derives in [1]. It is easy to check that the right side of (14) increases monotonically with increasing B; because $B \leq W$, it follows that

$$C_W \leq W \log_2 (1 + \frac{S}{N_o W}) \quad \text{(bits/sec)} \tag{15}$$

with equality if and only if B = W, i.e., *if and only if there is no spectrum spreading!* The reason that Shannon wrote (15) with an equality sign, rather than (14), in his final expression for the capacity of the AWGN channel is that he assumed that the choice of the modulated signal was up to the sender and that thus the sender would choose a signal with B = W to obtain (maximum) capacity.

Here we must in honesty point out that we have been somewhat cavalier in writing (14) without stating the precise condition for which this expression gives the true capacity. This condition is that all the coefficients s_1, s_2, ... s_N in the expansion (1) must be independently controllable by the choice of the modulator input symbols. This is indeed the case for all of the signals in the above six examples with the exception of the PPM modulated signal in example 3. We see from (8) that in fact for this signal only one of the N = M coefficients can be non-zero so they are certainly not independently controllable. For such modulation systems, the expression in (14) gives only an upper bound on capacity–which is why PPM modulation is not "energy efficient" except for high "signal-to-noise ratios."

It is important not to draw the wrong conclusion from (14) and (15). The real question is not whether spectrum spreading can increase capacity (it never can!), but whether spectrum spreading, which may be desirable for other reasons such as those considered in the previous section, necessarily entails a substantial loss of capacity for the used Fourier bandwidth W. This time the answer is more subtle and more interesting. As B increases without limit, the right side of (14) tends to 1.44 S/N_o. Thus,

$$C_W \leq C_\infty = 1.44 \frac{S}{N_o} \quad \text{(bits/sec)} \tag{16}$$

with near equality when the Shannon bandwidth B (and hence also the Fourier bandwidth W) is very large. For instance, when

$$B \geq 4 \frac{S}{N_o}, \tag{17}$$

the capacity given by (14) is at least 90% of C_∞. As long as the Shannon bandwidth is large enough to satisfy (17), then no matter how large a spreading factor is used, the capacity will be at least 90% of the maximum capacity achievable with the given Fourier bandwidth W. Spectrum spreading cannot increase capacity, but it need not reduce it significantly.

The Shannon bandwidth B is always proportional to what we will call the *symbol rate*, R_m, of the modulator, which we define as the number of symbols per second that are output from the modulator. But the information bit rate, R, *cannot exceed the capacity,* C, of the channel if the system is to operate reliably, and C, in turn must satisfy (16). How then with this fixed upper bound on R can we always achieve the necessary large Shannon bandwidth B or, equivalently, necessary large R_m? The answer is to use a code with sufficiently low *code rate* r, measured in modulation symbols per information bit, because $R = R_m \cdot r$ and hence

$$R_m = \frac{R}{r} \quad \text{(symbols/sec).} \tag{18}$$

Because the Shannon bandwidth is proportional to R_m, we see that *coding (except in trivial cases such as that in Example 6) increases both the Shannon bandwidth and the Fourier bandwidth by a factor proportional to the reciprocal of the code rate* r. This is the true nature of the "bandwidth expansion" due to coding. *Spectrum spreading*, on the other hand, *increases the Fourier bandwidth but not the Shannon bandwidth*, which is the true nature of its "bandwidth expansion".

6. EFFECTS OF FILTERING ON BANDWIDTH

Filtering has interestingly different effects on Shannon bandwidth and Fourier bandwidth. Consider first linear filtering. If the input signal is given as in (1) by the linear combination of the orthonormal functions $\phi_i(t)$, $1 \leq i \leq N$, then the output signal of a linear filter, whether time-invariant or time-varying, must be the same linear combination of the functions $\lambda_i(t)$, $1 \leq i \leq N$, where $\lambda_i(t)$ is the response of this filter to $\phi_i(t)$. The signals $\lambda_i(t)$, $1 \leq i \leq N$, will in general be neither orthogonal nor normalized to unit energy, but they nonetheless span the signal space needed to represent the filter output signal so that this space has dimension at most N. We conclude that *linear filtering can never increase Shannon bandwidth*.

Depending on how one defines Fourier bandwidth, e.g., if one uses the "3 dB bandwidth," linear filtering can increase Fourier bandwidth by emphasizing parts of the spectrum where the input signal is weak but substantially non-zero, but this is inessential. More essentially, *linear filtering generally reduces the Fourier bandwidth* of a signal by suppressing parts of the spectrum where the signal is strong. If such a signal had Shannon bandwidth B essentially equal to its Fourier bandwidth W, then it follows from the fundamental theorem of bandwidth that its Shannon bandwidth would also be reduced by such linear filtering. Suppose, however, that s(t) has a flat spectrum and is a well-spread signal (i.e., $\gamma >> 1$) with good LPI capability so that the fraction $1/\gamma$ of the $2WT\gamma$ available dimensions that are actually used for the signal space appear to be quite randomly chosen. For the linear filtering of this signal performed by its transmission through a channel with multipath propagation, it is hard to imagine that the filtered versions $\lambda_i(t)$, $1 \leq i \leq N$, of the original basis functions $\phi_i(t)$, $1 \leq i \leq N$, of the signal space would not still be linearly dependent. With high probability, one expects that the signal space still has dimension N, i.e., that the Shannon bandwidth of s(t) is unchanged, even though the Fourier bandwidth may have been substantially decreased. *Linear filter generally does not decrease the Shannon bandwidth of a well-spread signal with a good LPI capability*, which is why such signals tend to be robust against the deleterious effects of multipath propagation.

Nonlinear filtering is quite a different matter. We saw in Example 5 that (nontrivial) coding of a binary information sequence increases, by a factor equal to the reciprocal of the code rate, both the Shannon and the Fourier bandwidths of the signal at the output of a binary antipodal modulator fed by that sequence compared to that for an uncoded sequence with the same information rate (i.e., as in Example 4 with $n = 1$). But linear (over the finite field GF(2)) coding of a binary (0 or 1) sequence is a discrete-time nonlinear (over the field of real numbers) filtering of that sequence considered as a binary (±1) sequence, which implies that the modulator output for the coded input sequence is a nonlinearly filtered version of the modulator output for the uncoded sequence. This illustrates that *nonlinear filtering of a signal generally increases both its Shannon and its Fourier bandwidth*. Aside from the nonlinear filtering resulting from coding, it appears difficult to say much more in general about the effects of the kinds of nonlinear filtering that one might encounter in typical spread-spectrum systems.

7. TWO PARADOXES

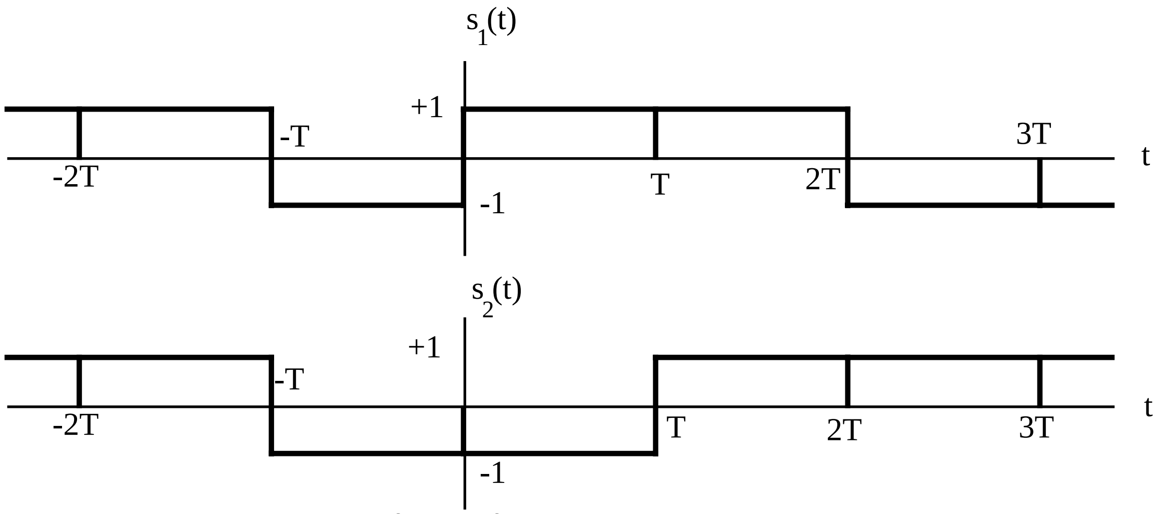
We now consider two paradoxes that arise in CDMA systems and show how the concepts previously developed, particularly the notion of Shannon bandwidth, can be used to resolve them.

The first paradox concerns a two-user CDMA system (with no multipath propagation) where both users have as their spreading waveform a simple rectangular pulse of duration equal to the symbol period T, which they modulate with binary (±1) data. Typical transmitted signals for users 1 and 2 are shown in Fig. 3a.

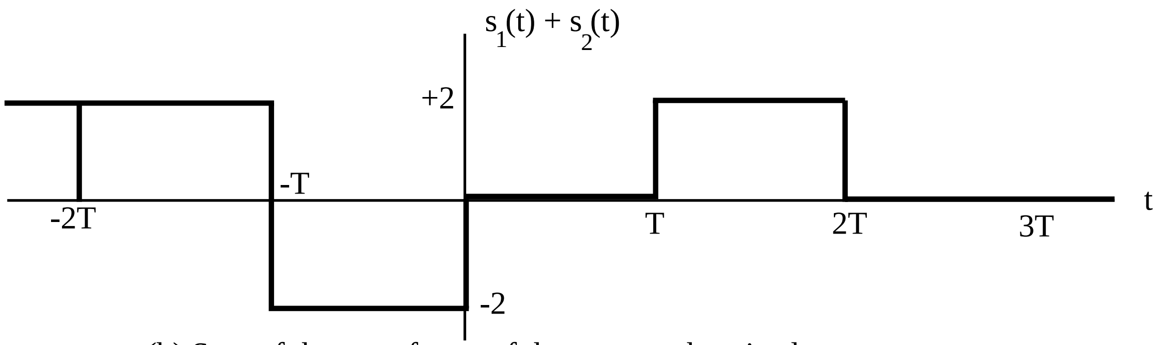
Suppose first that the two users are perfectly synchronized so that the received signal is the sum of the two signals (shown in Fig. 3b) plus Gaussian noise that we suppose to be negligibly small. This continuous-time channel is precisely equivalent to the discrete-time noiseless binary adder channel in which each user sends +1 or -1 and the output is the real sum of the inputs. The symmetric capacity of this channel (i. e., the upper limit of the total rate at which arbitrarily reliable operation is possible with all users sending information at the same rate) is well-known to be 3/2 bits/symbol [12, p. 392-393]. The users, by proper coding of their information, can simultaneously communicate as reliably as desired to the receiver, each at an information rate as close to 3/4 bits/symbol as desired, but can do this at no higher common rate.

Suppose next that the two users are out of synchronization by one-half a symbol period so that the sum of their two signals is as shown in Fig. 3c. We note that, at any time instant where the two signals reinforce one another, the receiver can correctly detect the sent (±1) symbols of both users. Moreover, at the next time instant thereafter when the receive signal changes value, the receiver detects the location of one symbol edge for one user and hence knows the locations of all symbol edges for that user; the next subsequent change of the received signal not located at a symbol edge for that user locates all the symbol edges of the other user. The receiver can now detect perfectly the entire sent (±1) data sequences of both users, since knowing the symbol-edge locations tells the receiver which user's signal has changed at each subsequent instant of change in the received waveform. Because with probability 1 the two users' symbols will reinforce at some point, it follows that the two users can now simultaneously communicate reliably to the receiver, each at the information rate of 1 bit per symbol. *The lack of synchronization has caused the symmetric capacity to increaes from 3/2 bits/symbol to 2 bits/symbol!*

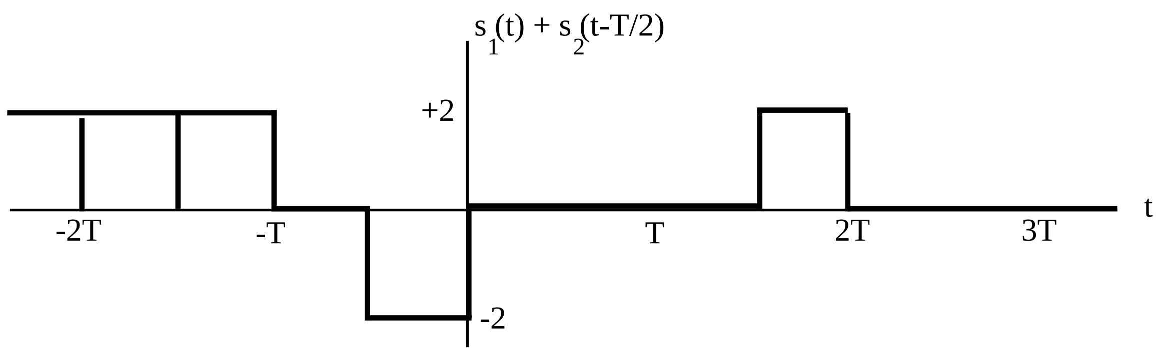
Any communication engineer who has struggled to obtain synchronization in some digital communication system will have trouble accepting the fact that lack of synchronization can actually increase channel capacity--but it is true here! Even if we had taken the Gaussian noise to be small but not negligible, we would have been led to the same conclusion–the capacity with synchronization would be nearly 3/4 bits/symbol, but the capacity without synchronization would be close to 1 bit/symbol; including noise does not resolve the paradox. The resolution of the paradox comes from consideration of Fig. 3(c) where we see that, to represent the sum of the two unsynchronized signals, we need two orthonormal functions (each a rectangular pulse of width T/2 seconds) per symbol period compared to only one such function (a rectangular pulse of width T seconds) for the sum of the two synchronized signals. *The Shannon bandwidth of the sum signal in the unsynchronized case is twice that in the synchronized case!* This increase in Shannon bandwidth occurs because rectangular pulses do a rather poor job of filling out their Fourier bandwidth. To test this reasoning, we replaced the rectangular functions as illustrated in Fig. 3 with sinc functions (as in Fig. 1) of Fourier bandwidth W = 1/(2T), i.e., equal to the Shannon bandwidth B = 1/(2T) of each user's signal. In this case, the sum signal has this same Fourier bandwidth W regardless of the synchronization between the two users and hence, by the fundamental theorem of bandwidth, the sum signal must also have the same Shannon bandwidth 1/(2T). We then calculated capacity for various time shifts between the users' signals by a Monte Carlo method and found that *capacity was virtually independent of this time shift.* The communication engineer may still be somewhat surprised that the lack of synchronization has not decreased capacity in this case, but he or she should be gratified to hear that at least it does not increase capacity.



(a) Waveforms of the two users.



(b) Sum of the waveforms of the two synchronized users.



(c) Sum of the waveforms of the two users when out-of-synch by half of a symbol period.

Fig. 3: The paradox where lack of synchronization increases capacity.

The second paradox that we will consider is that of a fully synchronized CDMA system with additive white Gaussian noise (AWGN) for which each user has a (±1) signature sequence that determines his symbol waveform in the manner described for one user in Example 2. We suppose that L is large so that each user sends a well-spread signal with $\gamma = L$. It follows then from our discussion in Section 5 that if only one user were active then he could not be using the channel at the maximum capacity consistent with his Fourier bandwidth. But, it is shown in [13] that when the signature sequences of the K active users meet the Welch bound with equality, then the symmetric capacity of this CDMA system is exactly the same as for a hypothetical single user with average power equal to the total power of the K users and who exploits the Fourier bandwidth to its fullest. *The individually bandwidth-inefficient users combine to create a system with maximum bandwidth efficiency!* Again this seems counterintuitive at first. However, again we see that the sum signal will have the same Fourier bandwidth (because of the sinc functions used a chip waveforms) as the signal of an individual user but will have a much greater Shannon bandwidth. In fact, as soon as the number K of users is on the order of L or greater, the Shannon bandwidth will generally be the same as the Fourier bandwidth–*the sum of the well-spread signals of the users is not a spread-spectrum signal at all*, which is why it can potentially make optimum use of the Fourier bandwidth. A closer study of the condition for meeting the Welch bound with equality for such a CDMA system shows in fact that this is just the condition for the sum signal to have Shannon bandwidth exactly equal to the Fourier bandwidth and for the sum signal to have the same power in each dimension of the signal space, which is what is needed in the signal of a single user for the AWGN channel to obtain the maximum capacity for a given Fourier bandwidth.

8. CONCLUDING REMARKS

It should be apparent that we have in this paper barely scratched the surface of the information theory of spread-spectrum systems. At best, we have pointed out the starting direction for a long journey. In particular, much additional thought needs to be given to spread-spectrum multiple-access systems, i.e., multiple-access systems in which each of several users sends a spread-spectrum signal in the same band and the sum of these signals is received. It is hardly a guess that, as in our resolution of the paradoxes described in Section 7, a very interesting quantity will be the Shannon bandwidth of this sum signal. We are far from being able to offer a coherent information-theoretic treatment of spread-spectrum multiple-access systems, even when we restrict the channel to be the bandlimited additive white Gaussian channel for the sum signal. And we have not even begun to take into account matters of paramount practical interest such as multipath propagation of each signal in the sum, time variation of the multipath channels, unequal user signal powers, and imperfect synchronization. Nonetheless, it is our conviction that until the information theory of spread-spectrum systems is worked out in enough generality to deal with such issues, the many arguments about which type of spread-sprectrum multiple-access system is better than another (say, offers greater "bandwidth efficiency") will continue to generate more heat than light.

ACKNOWLEDGEMENT

This paper profited materially from old fashioned personal discussions with Prof. I. Bar-David of the Technion–Israel Institute of Technology during a period when he was visiting at the ETH Zürich and from our e-mail discussions with Prof. S. Shamai, also of the Technion..

REFERENCES

[1] C. E. Shannon, "A Mathematical Theory of Communications," Bell System Tech. J., vol. 27, pp. 379-423 and 623-656, July and October, 1948.
[2] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, pp. 10-21, Jan. 1949.
[3] J. L. Massey, "Coding and Modulation for Code-Division Multiple Accessing", in *Proc. Third Int. Workshop on Digital Signal Processing Techniques Applied to Space Communications*, ESTEC, Noordwijk, 1992, pp. 3.1-3.17.
[4] J. L. Massey, "Deep-Space Communications and Coding: A Marriage Made in Heaven," in

Advanced Methods for Satellite and Deep Space Communications (Ed. J. Hagenauer), Lecture Notes in Control and Information Sciences No. 182. Heidelberg and New York: Springer, 1992, pp. 1-17.
[5] R. W. Emerson, "Self Reliance, in *Essays*, 1841.
[6] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," AIEE Transactions, p. 617, April 1928.
[7] D. Gabor, "Theory of Communication," J. IEE, vol. 93, part 3, no. 26, p. 429, 1946.
[8] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
[9] D. Slepian, "On Bandwidth," *Proc. IEEE*, vol. 64, pp. 292-300, 1976.
[10] D. L. Kreider, R. G. Kuller, D. R. Ostberg and F. W. Perkins, *An Introduction to Linear Analysis*. Reading, MA: Addison-Wesley, 1966.
[11] W. B. Davenport, Jr., and W. L. Root, *Random Signals and Noise*. New York: McGraw-Hill, 1958.
[12] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.
[13] M. Rupf and J. L. Massey, "Optimum Sequence Multisets for Symbol-Synchronous Code-Division Multiple-Access Channels," to appear in *IEEE Trans. Inform. Th.*, vol. IT-40, 1994.