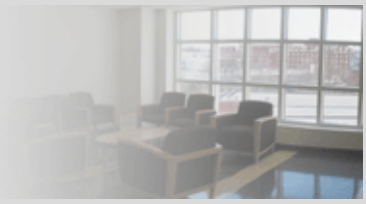
Center for Scientific Computation
And Mathematical Modeling

University of Maryland, College Park



Spread Spectrum from Two Perspectives

Shlomo Engelberg

June 2003

CSCAMM Report 03-09

http://www.cscamm.umd.edu/publications/

CSCAMM is part of the
College of Computer, Mathematical &
Physical Sciences (CMPS)

SPREAD SPECTRUM FROM TWO PERSPECTIVES

SHLOMO ENGELBERG*

Abstract. Spread spectrum techniques are often employed when transmitting information. They are widely used in wireless and cellular telephony. Spread spectrum techniques allow one to partition bandwidth, to hide transmissions, and to protect one's transmissions from being jammed.

In this article, we develop two facets of spread spectrum. We show how to understand the properties of direct sequence spread spectrum. For this purpose we make use of probabilistic arguments. We also show how to design the pseudo-random sequences that spread spectrum transmitters and receivers generally use. This causes us to consider the properties of recurrence relations (and polynomials) over the integers modulo two.

Key words. random telegraph signals, pseudo-noise sequences, spread spectrum techniques, Wiener-Kinchine theorem

1. Introduction.

1.1. General Introduction. Spread spectrum transmission works by taking a signal that has most of its energy at low frequencies and smearing the signal's spectral content in such a way that its energy is spread over a broad range of frequencies[8]. The smearing is done in such a way that to people who know how the smearing is done, it is relatively simple to "unsmear" the message, while to others it is well nigh impossible to recover the message. Often, because the power of the signal at any given frequency can be less than the power of the noise at that frequency, an uninitiated observer may even be unaware that a signal has been transmitted. This feature can be used to hide transmissions or to hide data (such as a digital watermark) inside other data[2].

The first patent on spread spectrum was taken out by the actress Hedy Lamarr and the avant-garde composer George Antheil for a device that would allow a torpedo to be controlled by radio. The patent for "a secret communication system" was awarded on August 11, 1942[9]. The main advantage of their system was that transmissions made using their system could not be detected by listening at *the* frequency of transmission. Because in spread spectrum the energy is not localized at a particular frequency there is no specific frequency at which one can eavesdrop.

The main idea behind direct sequence spread spectrum (the type of spread spectrum that we consider) is to take a signal of interest, $X(t)$, and to multiply it by a second signal $R(t)$ that satisfies $R^2(t) = 1$. That is, one multiplies $X(t)$ by a signal that is always equal to ± 1. Clearly the more frequently the sign of $R(t)$ changes the more high frequency content $R(t)$ will have. Also, after multiplying $X(t)$ by $R(t)$ the product will also have a lot of energy in its high-frequency region; the product will also have many sign changes. Demodulation—unspreading—is accomplished by multiplying the received signal $X(t)R(t)$ by $R(t)$ once again. The product is equal to $X(t)$—the original signal.

In this paper, we develop the theory of spread spectrum communications in two ways. We consider both a probabilistic approach and a deterministic approach. The probabilistic approach has the advantage of allowing us to talk about spread spectrum in a particularly nice way. The deterministic approach has the advantage of allowing us to describe how spread spectrum is actually implemented (with some of the details

*Shlomo Engelberg is with the Electronics Department, Jerusalem College of Technology, P.O.B. 16031, Jerusalem, Israel

left out). It also turns out that to understand the properties of *practical* spread spectrum systems, one must make use of some truly beautiful mathematics.

1.2. Background. In this section we recall a few results that we need in the following section. (See [3] for more information on these topics.) Ergodic processes can be thought of as random functions of time. Let $X(t)$ be an ergodic process. For each value of t the function $X(t)$ is a random variable. That is, $X(0.12)$ is a random variable, $X(0.3)$ is a different random variable, and so on for each different value of t. A *stationary* ergodic process is a process for which absolute time is unimportant; statistics of the process cannot depend directly on t: the time at which a measurement was made. If one has a stationary ergodic process, $X(t)$, then its autocorrelation, $R_{XX}(\tau)$, is defined as

$$R_{XX}(\tau) \equiv E(X(t)X(t+\tau)).$$

Note that the autocorrelation does not depend on t; it only depends on τ (which is the *difference* of the times at which the measurements were made). According to the *Wiener-Kinchine theorem*, the process's power spectral density (PSD), $S_{XX}(f)$—which gives the density of the power in the process at the frequency f—is the Fourier transform of the autocorrelation. We define the Fourier transform of a function $y(\tau)$ to be

$$\mathcal{F}(y(\tau))(f) = \int_{-\infty}^{\infty} e^{-i2\pi f\tau} y(\tau)\, d\tau.$$

With this definition, the inverse Fourier transform is

$$\mathcal{F}^{-1}(Y(f))(\tau) = \int_{-\infty}^{\infty} e^{+i2\pi f\tau} Y(f)\, df.$$

We find that the power spectral density that corresponds to $X(t)$ is

$$S_{XX}(f) = \mathcal{F}(R_{XX}(\tau))(f).$$

As the autocorrelation is the inverse Fourier transform of the power spectral density, we find that

$$R_{XX}(0) = \int_{-\infty}^{\infty} e^{+i2\pi f 0} S_{XX}(f)\, df = \int_{-\infty}^{\infty} S_{XX}(f)\, df. \tag{1.1}$$

The autocorrelation $R_{XX}(\tau) = e^{-\mu|\tau|}$ will play a large role in the sequel. Recall that

$$\mathcal{F}(e^{-\mu|\tau|})(f) = \frac{2\mu}{(2\pi f)^2 + \mu^2}. \tag{1.2}$$

The Fourier transform of the product of two functions is the convolution of the transforms of the two functions; that is,

$$\mathcal{F}(g(t)h(t))(f) = \mathcal{F}(g(t)) * \mathcal{F}(h(t))(f),$$

where $*$ is the convolution operator defined by the integral

$$G * H(f) \equiv \int_{-\infty}^{\infty} G(f-\eta)H(\eta)\, d\eta.$$

2. The Probabilistic Approach.

2.1. The Random Telegraph Signal. In most modern communication systems, information is digitized before being transmitted. The transmitter is normally "asked" to transmit a stream of symbols. Here we assume that our data are of this form. Let us suppose that we have a signal, $X(t)$, that—like a telegraph signal—always assumes one of two values—say $\pm a$. Further suppose that one expects the signal to switch between the two possible values μ times per second in a totally random way. Also assume that the probability of a change occurring in any given time interval is independent of what happened in any other (disjoint) interval.

Let Y be the number of sign changes in an interval of time τ. Let N be the number of ("virtual") subintervals into which we have chosen to break the interval of length τ. (Later we let $N \to \infty$—which gives us our final answer.) We find that the probability of a single sign change in a subinterval of length τ/N is $\mu\tau/N + e(N)$ where $e(N) = o(1/N)$.

The probability of $M \leq N$ sign changes occuring in the interval of length τ is the probability of one sign change occuring in M of the N subintervals. (As $N \to \infty$ the chances of two or more sign changes occuring in the same subinterval is negligible.) Thus the probability of M sign changes follows the binomial distribution[3, p. 17-18]. We find that

$$P(Y = M) = \left(\begin{array}{c} N \\ M \end{array} \right) (\mu\tau/N + e(N))^M (1 - \mu\tau/N - e(N))^{N-M}.$$

Let us consider the autocorrelation of $X(t)$. Clearly

$$X(t)X(t+\tau) = \pm a^2.$$

It is obvious that the sign will be a plus when there have been an even number of sign changes between t and $t+\tau$ and it will be a minus if there have been an odd number of sign changes between the two times. We find that

$$\begin{aligned} E(X(t)X(t+\tau)) &= a^2 P(Y \in \text{even}) - a^2 P(Y \in \text{odd}) \\ &= a^2 \left(P(Y \in \text{even}) - P(Y \in \text{odd}) \right). \end{aligned}$$

Now let us calculate $P(Y \in \text{even}) - P(Y \in \text{odd})$. We find that

$$\begin{aligned} P(Y \in \text{even}) - P(Y \in \text{odd}) &= \sum_{\substack{n=0 \\ n \text{ even}}}^{N} P(Y = n) - \sum_{\substack{n=1 \\ n \text{ odd}}}^{N} P(Y = n) \\ &= \sum_{n=0}^{N} (-1)^N P(Y = n) \\ &= \sum_{n=0}^{N} \left(\begin{array}{c} N \\ n \end{array} \right) \{(-\mu|\tau|/N - e(N))^n \\ &\qquad \times (1 - \mu|\tau|/N - e(N))^{N-n}\} \\ &\overset{\text{binomial theorem}}{=} (1 - 2\mu|\tau|/N - 2e(N))^N. \end{aligned}$$

We see that as we consider ever smaller ("virtual") intervals—as $N \to \infty$—the last term tends to $e^{-2\mu|\tau|}$, and

$$R_{XX}(\tau) = E(X(t)X(t+\tau)) = a^2 e^{-2\mu|\tau|}.$$

2.2. Spread Spectrum—the Probabilistic Way. Suppose that we take an input signal, $X(t)$, and multiply it by a signal of the type of the preceding section—a "random telegraph signal," $R(t)$, with $a = 1$. Assume that the two signals are independent. Then the autocorrelation of the resultant signal, $Y(t) = X(t)R(t)$, is

$$\begin{aligned} R_{YY}(\tau) &= E(Y(t)Y(t+\tau)) \\ &= E(X(t)R(t)X(t+\tau)R(t+\tau)) \\ &= E(X(t)X(t+\tau)R(t)R(t+\tau)) \\ &\overset{\text{independence}}{=} E(X(t)X(t+\tau))E(R(t)R(t+\tau)) \\ &= R_{XX}(\tau)R_{RR}(\tau). \end{aligned}$$

As the PSD of $Y(t)$ is the Fourier transform of the autocorrelation, we find that

$$S_{YY}(f) = S_{XX} * S_{RR}(f).$$

Let us assume that $S_{XX}(f)$ is band-limited—that $S_{XX}(f) = 0$ for $|f| > B$. From (1.2) we see that

$$S_{RR}(f) = \frac{2(2\mu)}{(2\pi f)^2 + (2\mu)^2} < \frac{1}{\mu}.$$

For large μ, $S_{RR}(f)$ is quite small, and it is also nearly flat out to its bandwidth—$\mu/(2\pi)$.

Let us consider the convolution of the input and the random telegraph signal. We find that

$$\begin{aligned} S_{YY}(f) &= \int_{-\infty}^{\infty} S_{XX}(f_1)S_{RR}(f - f_1)\, df_1 \\ &\overset{\text{positivity of } S_{XX}(f_1)}{\leq} \frac{1}{\mu}\int_{-\infty}^{\infty} S_{XX}(f_1)\, df_1 \\ &\overset{(1.1)}{=} \frac{1}{\mu}R_{XX}(0) \\ &= \frac{1}{\mu}E(X^2(t)) \\ &\overset{X^2(t)=1}{=} \frac{1}{\mu}. \end{aligned}$$

As $\mu \to \infty$, we find that the energy at any given frequency tends to zero. Thus, a signal modulated in this way will not tend to interfere much with other signals that are transmitted in the same bandwidth, as long as μ is sufficiently large.

Also note that if one imagines that the energy in the random telegraph signal is zero above $f = \mu/\pi$, then from the definition of the convolution we see that the bandwidth of the modulated signal is just the bandwidth of the signal plus the bandwidth of the random telegraph signal. Assuming that the bandwidth of the random telegraph signal is much larger than the bandwidth of the signal, we find that the bandwidth of the modulated signal is (very approximately) the bandwidth of the random telegraph signal.

How does one demodulate such a signal? One multiplies the modulated signal by $R(t)$ again. As $R^2(t) = 1$, we find that

$$Y(t)R(t) = X(t)R(t)R(t) = X(t).$$

This method of demodulation has an interesting "anti-jamming" property. Suppose there is a narrow band signal in the range of frequencies that the modulated signal occupies. To demodulate the signal, we modulate the received signal—which spreads the spectrum of the interfering signal. Assuming that after multiplication by the random telegraph signal one has a low-pass filter, one finds that only the fraction of the energy from the demodulated signal that is in the baseband—that is in the low frequency range—affects the output. Thus, a narrow band signal will not interfere with our signal very much unless it is very powerful indeed. Let us now consider an example.

2.3. A Spread Spectrum Signal with Narrow Band Noise. In communication systems, one often has a carrier signal and an information signal. One generally speaks of the information bearing signal *modulating* the carrier signal. In our case we consider a carrier signal, $R(t)$, that is a random telegraph signal with $\mu = 100$ and an information bearing signal, $X(t)$, that is a random telegraph signal with $\mu = 1$. The information bearing signal is the modulating signal, and the carrier is the signal being modulated. Let our noise, $N(t)$, be a random telegraph signal with $\mu = 1$. In all cases let $a = 1$. Thus the PSD of both the signal and the noise is

$$S_{XX}(f) = S_{NN}(f) = \frac{2}{(2\pi f)^2 + 1^2}.$$

The spectrum of the signal being modulated is

$$S_{RR}(f) = \frac{200}{(2\pi f)^2 + 100^2}.$$

After modulation the transmitted signal, $Y(t) = X(t)R(t)$, will have autocorrelation

$$R_{YY}(\tau) = R_{XX}(\tau)R_{RR}(\tau) = e^{-101|\tau|}.$$

Thus the PSD of the transmitted signal is

$$S_{YY}(f) = \frac{202}{(2\pi f)^2 + 101^2}.$$

Note that this signal is spread over a wide range of frequencies.

Adding the noise to our signal, we find that the received signal is $V(t) = Y(t) + N(t)$. As the noise and the signal are (by assumption) independent and zero mean, we find that

$$\begin{aligned} R_{VV}(\tau) &= E((Y(t) + N(t))(Y(t+\tau) + N(t+\tau))) \\ &= E(Y(t)Y(t+\tau)) + E(Y(t)N(t+\tau)) + E(N(t)Y(t+\tau)) \\ &\qquad + E(N(t)N(t+\tau)) \\ &= R_{YY}(\tau) + 0 + 0 + R_{NN}(\tau) \\ &= R_{YY}(\tau) + R_{NN}(\tau). \end{aligned}$$

Thus the power spectral density of the received signal is

$$S_{VV}(f) = S_{YY}(f) + S_{NN}(f) = \frac{202}{(2\pi f)^2 + 101^2} + \frac{2}{(2\pi f)^2 + 1^2}.$$

Note that at low frequencies the desired signal is swamped by the noise.

After multiplying the received signal by $R(t)$, the demodulated signal is

$$W(t) \equiv V(t)R(t) = Y(t)R(t) + N(t)R(t) = X(t) + N(t)R(t).$$

At this point it is easy to determine the power spectral density of $W(t)$. The power spectral density of $X(t)$ is already known. Furthermore, $M(t) \equiv N(t)R(t)$ is just the noise modulated by $R(t)$. As the characteristics of the noise and the signal are identical, we find that

$$S_{MM}(f) = \frac{202}{(2\pi f)^2 + 101^2}.$$

Thus, the PSD of $W(t)$ is just

$$S_{WW}(f) = S_{XX}(f) + S_{MM}(f) = \overbrace{\frac{2}{(2\pi f)^2 + 1}}^{\text{signal}} + \overbrace{\frac{202}{(2\pi f)^2 + 101^2}}^{\text{noise}}.$$

Note that here the signal power is located at low frequencies and the disturbing signal's power has been smeared out over a wide range of frequencies. If one now filters out the high frequencies, one finds that the signal power is much greater than the noise power after the filtering is done.

2.4. The Effect of Multiple Transmitters. Suppose that one has several independent spread spectrum transmitters going at the same time. To what extent will the various signals interfere with one another's reception?

If one has independent signals, $R_i(t)$, that are being modulated, then the spread spectrum receiver—which is identical to a spread spectrum transmitter—just "spreads" the unrelated signals further. Thus, after the low-pass filter at the receiver not much energy from the other signals is left.

Let us consider the case of N transmitters each transmitting a random telegraph signal, $X_i(t)$, for which $\mu = 1$ and $E(X_i(t)) = 0$ each modulating a random telegraph signal, $R_i(t)$ for which $\mu = 100$ and $E(R_i(t)) = 0$. By design all the $R_i(t)$ are independent of one another and the $X_i(t)$. Clearly

$$S_{X_iX_i}(\tau) = \frac{2}{(2\pi f)^2 + 1^2}$$
$$S_{R_iR_i}(\tau) = \frac{200}{(2\pi f)^2 + 100^2}$$

The modulated version of the i^{th} signal is just $Y_i(t) = X_i(t)R_i(t)$, and we have seen that

$$S_{Y_iY_i}(\tau) = \frac{202}{(2\pi f)^2 + 101^2}.$$

Let us now consider the signal at the i^{th} receiver. The signal as it exists in the "ether" (after all the signals are combined) is

$$V(t) = \sum_{i=1}^{N} Y_i(t).$$

The autocorrelation of this signal is

$$R_{VV}(\tau) = E(V(t)V(t+\tau)) \overset{\text{independence}}{=} \sum_{i=1}^{N} R_{Y_iY_i}(\tau) = NR_{Y_1Y_1}(\tau).$$

Consequently its PSD is

$$NS_{Y_1Y_1}(f) = \frac{N \cdot 202}{(2\pi f)^2 + 101^2}.$$

We find that our signals are quite "well mixed" and that their energy has been "smeared."

After we detect the i^{th} signal, the demodulated signal is

$$W_i(t) \equiv V(t)R_i(t) = X_i(t) + \sum_{j \neq i} X_j(t)R_j(t)R_i(t).$$

Note that the signal $R_i(t)R_j(t), i \neq j$ is a random telegraph signal that changes sign 200 times per second on average ($\mu = 200$). Thus $N_j(t) \equiv X_j(t)R_j(t)R_i(t), i \neq j$ is a random telegraph signal for which $\mu = 201$. As the $X_i(t)$ are all independent, we find that once again the autocorrelation functions can be added, and we have

$$R_{W_iW_i}(\tau) = R_{X_iX_i}(\tau) + (N-1)R_{N_1N_1}(\tau).$$

The PSD of the signal after the demodulator is

$$S_{W_iW_i}(f) = \frac{2}{(2\pi f)^2 + 1^2} + (N-1)\frac{402}{(2\pi f)^2 + 201^2}.$$

Note that the signal energy is concentrated in the low frequencies—the frequencies below about $1/(2\pi)$ Hertz. The noise in that region—where the signal is of order one—does not exceed $(N-1)/100$. Thus, if there are not too many other signals, it should not be difficult to detect the signal.

We see that spread spectrum techniques can be used to send more than one signal in a given frequency range without the signals interfering with each other very much. This way of using the same band of frequencies for several channels is used in some cellular telephones networks, and the implementation of these ideas is referred to as CDMA—Code Division Multiple Access[6].

3. Spread Spectrum—The Deterministic Approach.

3.1. Introduction. In section 2 we considered an input signal modulating a second—much wider bandwidth—random signal. In general, using true random signals in a communication system is very difficult. Since both the transmitter and the receiver must use the random signal, the signal must be recorded and shipped to both the transmitter and the receiver. Additionally both the transmitter and the receiver must store the random signal.

One of the forerunners of modern spread spectrum *did* use random noise. During World War II a system known as SIGSALY was used to allow secure communications[1]. (The system was used to allow F. D. R. and Winston Churchill to communicate over a secure channel.) In order that the transmitter and receiver would both have access to the random noise, phonograph records of the noise were couriered to the transmitter and the receiver stations.

In modern systems, rather than using true random noise, a periodic deterministic signal that is noise-like is used. Suppose that one has a signal $X(t)$ that is derived from a sequence of bits. That is, let

$$X(t) = x(n), \quad nT \leq t < (n+1)T, \quad n \geq 0$$

where $x(n) = \pm 1$ for each $n \geq 0$. We consider multiplying this signal by another signal $R(t)$ that may change values (also between ± 1) every T/N seconds in a way that is "deterministically random." To demodulate this signal all that someone who knows $R(t)$ needs to do is to multiply the received signal by $R(t)$. Here the trick is designing $R(t)$. We consider this problem by considering the problem of how to design a sequence y_n that takes the values $+1$ and -1 in a random-seeming way.

It turns out that a very simple way to generate such a sequence is to design a feedback system in which all of the signals take one of two possible values. We consider the symbols $\{0, 1\}$, $\{F, T\}$, and $\{+1, -1\}$. We will see how to analyze and design feedback systems with a very interesting property: their output is, in a sense that we will make precise shortly, very similar to noise. Such generators are widely used as a source of pseudo-noise (PN) sequences[4].

Before we start designing PN sequence generators, we must define arithmetic on our sets. We generally think of all three of our sets as representing the same two "objects." We consider the 0 of the first set, the F of the second set and the $+1$ of the third set to represent one "object" and the 1, T, and -1 of the sets to represent a second "object." Addition on the set $\{0, 1\}$ is defined by the rules

$$\begin{aligned} 0 + 0 &= 0 \\ 1 + 0 &= 1 \\ 0 + 1 &= 1 \\ 1 + 1 &= 0. \end{aligned}$$

If one replaces the 0's by F's and the 1's by T's, then this addition is the logical exclusive *or*, the XOR, operation. If one replaces the 0 by $+1$ and the 1 by -1, then one finds that this "addition" is ordinary multiplication.

Multiplication on $\{0, 1\}$ is defined by

$$\begin{aligned} 0 \cdot 0 &= 0 \\ 1 \cdot 0 &= 0 \\ 0 \cdot 1 &= 0 \\ 1 \cdot 1 &= 1. \end{aligned}$$

We see that for the set $\{F, T\}$ multiplication is the familiar logical *and* operation. For the set $\{+1, -1\}$, the operation does not correspond to any familiar operation.

Our addition and multiplication are both associative and commutative, and multiplication has higher precedence than addition. The set $\{0, 1\}$ can be identified with the integers modulo two. With this identification made, our multiplication and addition are precisely multiplication and addition modulo two. For most of this section it will be convenient to think of the two symbols as $\{0, 1\}$. However, when needed, we can always convert them to the elements of the set $\{+1, -1\}$ or $\{F, T\}$.

3.2. Finite State Machines. Consider a device that changes its state periodically and whose next state is a function of its current state. A simple example is a

device with two states, 0 and 1, that changes from state 0 to state 1 and from state 1 to state 0. Thus, if one knows that the device starts in state 0, then one knows that the state of the device progresses as follows

$$\{0, 1, 0, 1, ...\}.$$

Note that the state of this device varies in a periodic way.

In general if one has a "machine" whose next state is a function of its previous state and for which there are only a finite number of states, then the progression of states—for any initial state—is periodic.

The periodicity is a consequence of the fact that the device has only a finite number of states and that the next state depends only on the current one. Suppose that the device has N states. Then after the device has changed states $N + 1$ times, one knows that at least one state occured twice. Because the next state is a function only of the previous state, the states must repeat themselves.

In the previous example the device had two states, and we found that when the initial state was zero the sequence of states was periodic with period two. There are many other devices of the type described here—devices with a finite number of states, and collectively they are referred to as *finite state machines*.

3.3. Modulo Two Recurrence Relations. Consider an N^{th} order recurrence relation of the form

$$y_k = a_1 \cdot y_{k-1} + \cdots + a_N \cdot y_{k-N}, n \geq N$$

subject to the initial conditions

$$y_{N-1} = b_{N-1}, ..., y_0 = b_0,$$

where all of the operations are performed in modulo two arithmetic on elements of $\{0, 1\}$. Note that this recurrence relation can be considered a finite state machine where the state of the machine at time $k - 1$ is the set of values

$$\{y_{k-N}, ..., y_{k-1}\}.$$

We see that the state of the machine at time k is

$$\{y_{k-N+1}, ..., y_{k-1}, y_k = a_1 \cdot y_{k-1} + \cdots + a_N \cdot y_{k-N}\},$$

which is a function of the machine's state at time $k - 1$.

Note that if the initial state of the machine is $\{0, ..., 0\}$, then the machine's output will remain zero forever. As the total number of states is 2^N, the maximal period is not more than 2^N. Because zero is not part of a periodic solution, it is more correct to say that the maximal period does not exceed $2^N - 1$. In fact, there are sets of coefficients for which solutions take $2^N - 1$ steps before returning to a previous state. These turn out to be the solutions of interest to us.

3.4. A Simple Example. Consider the recurrence relation

$$y_k = y_{k-1} + y_{k-2}.$$

If we start with $y_1 = y_0 = 0$, then we find that $y_k \equiv 0, k \geq 0$. If we start with $y_1 = 1, y_0 = 0$, then we find that

$$y_2 = 1 + 0 = 1$$

$$y_3 = 1 + 1 = 0$$
$$y_4 = 0 + 1 = 1$$
$$y_5 = 1 + 0 = 1.$$

Note that in the calculation of y_5 we use 1 and 0 just as we did when we calculated y_2. We find that the recurrence relation returned to its initial state, and the sequence of states must have period 3. The progression of the set of states for this finite state machine is

$$\{\{0,1\},\{1,1\},\{1,0\},\{0,1\},...\}$$

Note that the sequence $\{y_n\}$ is

$$\{0,1,1,0,1,1,0,1,1,...\}.$$

Both of these sequence are *maximal length sequences*: they have periods that are as large as possible. We have already seen that this finite state machine has four states; hence the longest possible period is three. The only way to avoid this sequence is by starting off with zero initial conditions.

3.5. Maximal Sequences. Before we try to figure out how to choose coefficients in such a way that the resulting recurrence relation has a maximal length solution, let us consider some of the properties that such solutions have.

In any finite state machine over the set $\{0,1\}$, each possible state of the machine has a complementary state—a state for which zeros are ones and ones are zeros. Thus we can break up the set of 2^N states into two sets of 2^{N-1} states such that the two sets are complements of one another. In our previous example the states $\{0,0\}$ and $\{0,1\}$ are the complements of the states $\{1,1\}$ and $\{1,0\}$ respectively. It is clear that when one considers any given position within a state, and one asks "in how many states does the position have a one and in how many states does it have a zero?" that because of complementarity this position within the state has a one 2^{N-1} times, and it has a zero and 2^{N-1} times.

As y_n is one of the numbers that makes up the state of the system, and as we know that the system passes through all states save the all-zero state, we find that y_n must be equal to 1 a total of 2^{N-1} times in each period and it must be equal to 0 a total of $2^{N-1} - 1$ times in each period.

As we pointed out above, one will have a maximal length solution (if one exists) whenever the initial conditions are not identically zero. This is because the solution cycles through all states save the all-zero state in the course of a maximal length solution. Also, because the equation is linear the sum of any two sequences that solve the equation must itself solve the equation.

Combining these two results, we find that all solutions are (cyclic) shifts of the maximal length solution or the all-zero solution. Also the sum of any two solutions must itself be the maximal length solution at yet another shift or the all-zero solution.

Let us define the *autocorrelation* of a sequence y_n with period M by

$$R_{yy}(k) = \sum_{j=0}^{M-1} h(y_j)h(y_{j+k})$$

where

$$h(y) = \begin{cases} +1 & y = 0 \\ -1 & y = 1 \end{cases}.$$

(The function $h(y)$ takes us from the set $\{0,1\}$ to the set $\{+1,-1\}$.) The autocorrelation equals the number of places where one period of y_j and one period of y_{j+k} have a common zero or a common one less the number of places where one sequence has a zero and the other has a one. Note that $h(x)h(y) = h(x+y)$. Thus, we find that

$$R_{yy}(k) = \sum_{j=0}^{M-1} h(y_j + y_{j+k}).$$

Because $y_j + y_{j+k}$ is also a solution of the equation, we know that if $0 < k < 2^N - 1$, then the number of ones is 2^{N-1} and the number of zeros is $2^{N-1} - 1$. Clearly, if $k = 0$ then $y_j + y_j \equiv 0 (\mathrm{mod}\, 2)$ and $R_{yy}(0) = 2^N - 1$. We see that the autocorrelation is

$$R_{yy}(k) = \sum_{j=0}^{M-1} h(y_j + y_{j+k}) = \begin{cases} 2^N - 1 & k = 0 \\ -1 & 0 < k < 2^N - 1 \end{cases} .$$

That is, two shifted sequences are almost uncorrelated. If they were totally uncorrelated for $k \neq 0$, they would be (periodic) deterministic white noise. As it is, they are as nearly uncorrelated as possible with a periodic sequence whose period is an odd number. This is why these sequences are so useful as pseudo-noise sequences.

3.6. Determining the Period. In order to analyze the sequences a given recurrence relation produces, we make use of the z-transform of the sequence. Suppose that one has a sequence of the form

$$\{a_0, a_1, ..., a_n, ...\}.$$

The z-transform associated with the sequence is

$$a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n} + \cdots .$$

This is a formal polynomial in z^{-1} and such polynomials are multiplied in the normal fashion. (Note that multiplication by z^{-1} shifts the sequence by one and inserts a zero in the first place.)

It is easy to check for periodic sequences in this format. Suppose that one has a periodic sequence with period N. Then its z-transform, which we denote by $A(z)$, must be

$$A(z) = a_0 + \cdots + a_{N-1} z^{-(N-1)} + a_0 z^{-N} + \cdots + a_{2N-1} z^{-(2N-1)} + \cdots ,$$

which satisfies the equation

$$(1 + z^{-N}) A(z) = a_0 + \cdots + a_{N-1} z^{-(N-1)}.$$

Clearly any sequence that is periodic with period N will satisfy an equation of this type, and any sequence that satisfies such an equation is periodic with period (at most) N.

Suppose that a sequence is defined by the recurrence relation

$$y_n = a_1 \cdot y_{n-1} + \cdots + a_N \cdot y_{n-N}, \quad n \geq 0,$$

with the initial conditions

$$y_{-1} = b_{N-1}, ..., y_{-N} = b_0.$$

We find that the z-transform of y_{n-k} is

$$Z(y_{n-k})(z) = z^{-k}Y(z) + z^{-k+1}b_{N-1} + \cdots + b_{N-k}$$

for $k \leq N$. If we set all the b_i to zero save b_0 which we set to one, then

$$Y(z) = a_1 z^{-1}Y(z) + \cdots + a_N(z^{-N}Y(z) + 1).$$

Making use of the fact that in our arithmetic the additive inverse is addition, we find that

$$Y(z)(1 + a_1 z^{-1} + \cdots + a_N z^{-N}) = a_N.$$

In order for the recurrence relation to be truly N^{th} order a_N must equal 1. Thus, we find that

$$Y(z)(1 + a_1 z^{-1} + \cdots + a_N z^{-N}) = 1.$$

Now suppose that

$$(1 + a_1 z^{-1} + \cdots + a_N z^{-N})R(z) = 1 + z^{-M}$$

where the degree of $R(z)$ as a polynomial in z^{-1} is $M - N > 0$. Then we find that

$$Y(z)(1 + z^{-M}) = R(z),$$

which shows that the sequence y_n is periodic with period no longer than M.

We have now turned the problem of determining the period of the solution of a modulo-two recurrence relation with particular initial conditions into the problem of determining whether or not a particular polynomial is a factor of $1 + z^{-M}$.

We are interested in finding maximal length sequences. We have shown that if such a sequence exists, then it cycles through all possible states save the all-zero state. Thus, to check for such a sequence it is sufficient to check that the sequence that begins in the state $b_0 = 1, b_i = 0, i \neq 0$ is maximal.

3.7. An Example. Consider our previous example

$$y_n = y_{n-1} + y_{n-2}.$$

Given the initial conditions $y_{-1} = 0, y_{-2} = 1$, we find that the z-transform of the sequence satisfies

$$Y(z) = z^{-1}Y(z) + z^{-2}Y(z) + 1.$$

It follows that

$$Y(z)(1 + z^{-1} + z^{-2}) = 1.$$

Multiplying both sides of the equation by $1 + z^{-1}$, we find that

$$Y(z)(1 + z^{-3}) = 1 + z^{-1}.$$

This means that y_n must have period three and the pattern that repeats must be $1, 1, 0$; that is

$$\{y_0, y_1, ...\} = \{1, 1, 0, 1, ...\}.$$

3.8. Some Conditions for Maximality. We now present two conditions for maximality without proof. The proof of the first condition is given in [5]. For the proof of the second condition see the references in [5].

A necessary condition for maximality is that the polynomial that multiplies $Y(z)$ when all the initial conditions save for b_0 are zero, a polynomial that we shall call $Q(z)$, must not be reducible in our arithmetic. That is, it must be impossible to factor the polynomial in our arithmetic. In our previous example $Q(z) = 1 + z^{-1} + z^{-2}$. Clearly, if this polynomial factors, it must have two linear factors. The linear factors in our arithmetic are just z^{-1} and $z^{-1} + 1$. If z^{-1} were one of the two linear factors, then there could be no constant term. Hence both the linear terms must be $1 + z^{-1}$. However, squaring that term gives us $1 + z^{-2}$. This is not our polynomial. Thus, our polynomial is not reducible and our sequence may be maximal. As we have already seen, the sequence is indeed maximal.

The second condition is somewhat more interesting. It makes use of the theorem that says that in our arithmetic, every irreducible polynomial of degree $N > 1$ is a factor of the polynomial $z^{-(2^N-1)} + 1$. This implies that the maximal period that can correspond to an N^{th} degree irreducible polynomial, $Q(z)$, is $2^N - 1$—something that we know to be true. However, it says much more. The theorem says that every N^{th} order recurrence relation for which $Q(z)$ is irreducible corresponds to a sequence that has period $2^N - 1$ (among other, possibly shorter, periods). Clearly any shorter period of length M must divide $2^N - 1$—otherwise the solution cannot also be periodic with period $2^N - 1$. If we know that $2^N - 1$ is a prime number, then we know that there can be no shorter periods for no integer (other than one and the number itself) divides a prime number. (Primes of the form $2^N - 1$ are known as *Mersenne primes.*) This shows that if $2^N - 1$ is prime and if $Q(z)$ is irreducible, then the sequence generated by the recurrence relation is maximal. In our example, the polynomial is irreducible and of degree 2. As $2^2 - 1 = 3$ is a Mersenne prime, we find that the polynomial corresponds to a maximal sequence whose period is three—just as we found.

There are many tables of irreducible polynomials that can be used to generate maximal sequences. See [5, p. 62-65] and [7] for such tables.

3.9. What We Have Not Discussed. We have presented two views of the theory behind spread spectrum techniques. In the first view, we developed the properties of spread spectrum techniques using probability theory. In the second view, we developed a practical way of generating sequences that are an almost perfect realization of the spreading sequences discussed in §2.

In our deterministic approach to spread spectrum, we have not discussed two important topics. In §2.4 we saw that when one has uncorrelated modulated functions $R_i(t)$, one can transmit many signals at once without the signals interfering too much with one another too much. We have not considered how to design PN generators that produce "uncorrelated" PN sequences. This problem is addressed in [5] (in a limited fashion) and in [4].

Another practical problem—perhaps *the* practical problem—in the design of spread spectrum systems is how to cause the receiver to reach a state at which it is synchronized to the transmitter. There are many ways to deal with this problem[4], but we do not consider them here.

REFERENCES

[1] J. V. Boone and R. R. Peterson, *The start of the digital revolution: SIGSALY secure digital voice communications in World War II*, http://www.nsa.gov/wwii/papers/start_of_digital_revolution.htm, last visited December 12, 2002.

[2] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, *Secure spread spectrum watermarking for multimedia*, IEEE Trans. on Image Processing, 6 (1997), pp. 1673-1687.

[3] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, NY, 1958.

[4] J. P. F. Glas, *Non-Cellular Wireless Communication Systems*, Ph.D. Thesis, Delft University, http://cas.et.tudelft.nl/~glas/thesis/, last examined April 15, 2002.

[5] S. W. Golomb, *Shift Register Sequences*, Holden-Day Inc., San Francisco, CA, 1967.

[6] M. Hendry, http://www.bee.net/mhendry/vrml/library/cdma/cdma.htm, last examined April 15, 2002.

[7] New Wave Instruments, *Linear feedback shift registers*, http://www.newwaveinstruments.com/resources/articles/m_sequence_linear_feedback_shift_register_lfsr.htm, last examined April 15, 2002.

[8] J. G. Proakis, *Digital Communications*, Fourth ed., McGraw-Hill Book Compayn, New York, NY, 2001.

[9] T. Rothman, *Plausibility—the invention of spread spectrum technology*, http://godel.ph.utexas.edu/~tonyr/spread_spectrum.html, last examined December 10, 2002.