CDMA Mobile Communication Networks

Elvino S. Sousa
Department of Electrical and Computer Engineering
University of Toronto

1 Introduction

We are undergoing a renaissance in the field of wireless communications. The reasons for the current developments are due to recent advances in the area of VLSI which have made it possible to implement sophisticated signal processing techniques required to achieve large spectral efficiencies in mobile communications. One of the techniques that has seen great interest has been the *spread spectrum (SS)* technique. In particular one of its derivatives, *code division multiple access* (CDMA), has received ubiquitous interest as a technique for the provision of services for cellular and personal communications networks.

Spread spectrum is a modulation technique that was initially devised in the late thirties and especially during WWII as a technique for communications in the presence of jamming [1][2]. The research in spread spectrum was mostly of a classified nature until the early seventies when interest in its use for commercial applications arose [3]. A spread spectrum modulation technique may be defined as a modulation technique that uses a bandwidth which is much larger than the transmitted data rate. With such a bandwidth expansion one gains interference rejection capability. The effect is not unlike that of wideband FM, where bandwidth expansion results in benefits in interference rejection. However, in the case of SS the bandwidth expansion is much larger than is typical for FM, and this expansion is done in a certain way that requires knowledge of a coded signal at the receiver - the so-called *spreading code.*

With the bandwidth expansion a number of benefits arise. The main benefit is that of *interference rejection.* In the case of hostile interferers it offers the only means for communications with the resulting *antijamming capability.* Other benefits include *robustness against multipath fading* - an effect similar to frequency diversity which is required for channels with a small coherence bandwidth relative to the transmitted data rate. The spreading of the spectrum results in the transmitted signal having a low power spectral density over a wide band, hence a resulting benefit is that which is known as *low probability of intercept.* Another byproduct of the spreading coupled with the requirement for the knowledge of the spreading code at the receiver is a modest inherent degree of *secrecy protection.* The interference rejection capability results in the capability for simultaneous transmissions in the channel, or the capability to achieve *multiple access* in a networking context. The resulting multiplexing technique, spread spectrum multiple access (SSMA), or code division multiple access (CDMA), has advantages over traditional multiple access techniques such as time division multiple access (TDMA), or frequency division multiple access (FDMA), in that it does not require coordination of the users in time or in frequency.

Another benefit of spread spectrum signaling which has only recently been exposed is an apparent advantage as a technique for use in a cellular context; it is an efficient technique to achieve spatial reuse. On a one-cell system the SS technique is inferior to FDMA, or TDMA; however, in a multi-cell system SS allows operation with a frequency reuse cluster size of one.

There are many different types of spread spectrum techniques. Mathematically they are all based on similar principles. The process of modulation in a digital communication system is the process of mapping a symbol alphabet to a signaling constellation in some signal space. With spread spectrum modulation a subspace (of small dimension) of a signal space (of large dimension) is used to transmit each symbol. The actual subspace changes every one or a few symbols. The sequence of subspaces changes according to a code. The selection of subspaces defines a specific type of spread spectrum such as *direct sequence* (DS), *frequency hopping* (FH), *time hopping* (TH), various *hybrids* of these such as frequency hopping/direct sequence hybrid (FH/DS), etc. In all of these forms of spread spectrum we have the notion of processing gain which is directly related to the bandwidth expansion factor.

In this paper we discuss the basic characteristics of direct sequence spread spectrum and its use in a networking context. The context of a cellular or personal communications network architecture raises many issues involving well developed concepts in point to point DS systems. These issues involve synchronous versus asynchronous transmission, performance in fading, hand-offs, degree of sectorization, area coverage, etc. In section 2 we discuss the basic properties of DS systems including spreading code sequences, types of DS modulators, synchronization, symbol error probability calculation, multiple access capability, and strategies for error control. In section 3 we discuss the characteristics of generic CDMA cellular networks. Section 4 deals with the effect of realistic propagation environments and the issues of Rake receivers, resolvability of the channel, power control, and system fading margin. Section 5 summarizes some of the characteristics of the CDMA cellular network standard IS-95 and the European CODIT project, and section 7 summarizes the paper.

2 Direct Sequence Spread Spectrum

The block diagram of a basic DS system is illustrated in Fig. 1. The input $d(t)$ is a sequence of rectangular data pulses. The data signal is multiplied by a signal $c(t)$ consisting of a sequence of similar pulses (known as the spreading code), running at a significantly higher rate, to obtain a baseband spread spectrum signal. This signal is then modulated with a carrier (frequency ω_c) to yield a bandpass DS signal. The transmitted signal is affected by the interference $I(t)$ resulting in a received signal with spectra as shown in Fig. 2. The performance of the scheme depends on the power spectral density of the interference signal. For a white noise interference signal there is no advantage to using spread spectrum. At the receiver the signal is demodulated by down converting to baseband and multiplying by the spreading code $c(t)$. The multiplication by $c(t)$ results in the sprectra of the data signal being despread to its original spectra. The interference, on the contrary is spread by this operation. The resulting signal is integrated, or lowpass filtered, and sampled. Fig. 3 shows the power spectral densities at the input and output of the lowpass filter. This figure clearly illustrates the noise reduction which is achieved by the spread spectrum scheme. The amount of reduction is proportional to the bandwidth expansion factor, i.e. the number of code chips per data bit, which is appropriately called the *processing gain* (PG) of the system.

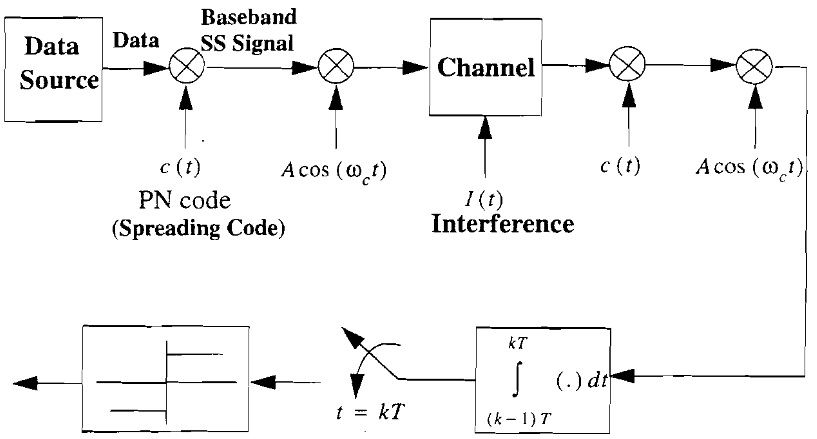


Figure 1 - Basic direct sequence spread spectrum system with no chip pulse filtering.

2.1 Spreading Code Sequences

The spreading code is generated by a shift register with feedback connections as shown in Fig. 4. The sequence generated by this circuit is periodic with a period that depends on the choice of feedback taps. For a register with n stages it is possible to choose the taps to produce a sequence of period $2^n - 1$. This is the maximum period which can be achieved. The feedback connection pattern is usually specified by a polynomial. The length of the resulting sequence depends on the properties of this polynomial. Sequences with maximum period correspond to polynomials which are called primitive. For a given m the number of primitive polynomials is given by the function $\Phi(2^m - 1)/2$, where $\Phi(x)$ is the number of integers less than x that are relatively prime to x. An important property of these sequences is that they have periodic autocorrelation functions that approximate a train of delta functions. Ref. [4] contains a comprehensive discussion of the properties of these sequences.

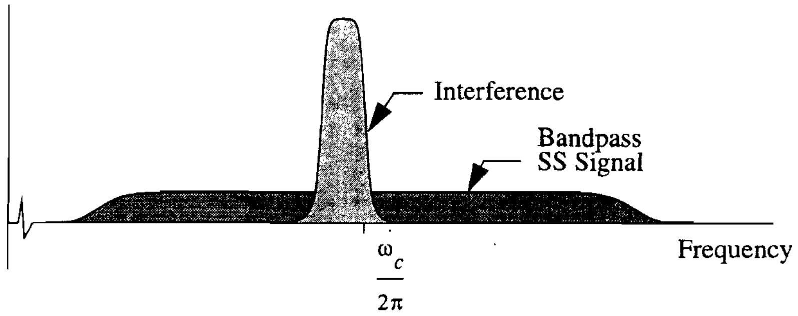


Figure 2 - Power spectral density of received signal.

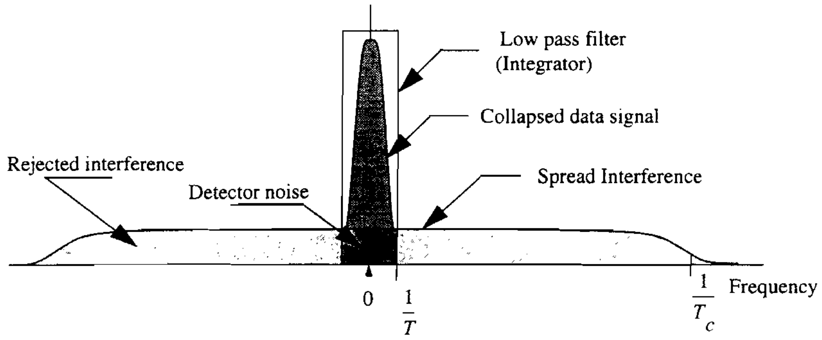
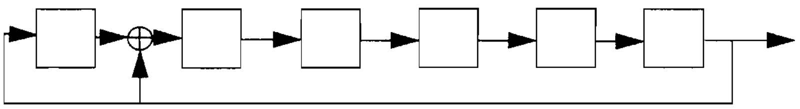


Figure 3 - Power spectral densities at integrator input.

In the design of a DS system there are a number of issues dealing with the type of PN code chosen. Do we use codes with short period or codes with long period? A short period code is one where the period is approximately equal to the data symbol period. In practice utilized code periods range in values equal to the data symbol period (e.g. 10 - 10,000 chips), to very large periods (e.g. $2^{41} - 1$ in the IS-95 standard [5]). The advantage of a short code period is that it is easier to achieve spreading code synchronization; the disadvantage is that in a networking context the mutual interference between users tends to exhibit a periodic structure. Also, in a multiple access system a unique spreading code is assigned to each user. With short codes the number of distinct user addresses is limited. In a multiple access system utilizing DS, each user is assigned a spreading code which acts like a channel. A significant amount of research has gone into the design of such families of sequences. The main criteria is to find sequences which have good autocorrelation properties such as m-sequences, but also have low cross correlations for any pair of sequences and any relative shift between the two sequences. Such families of sequences have been found and including sequences known as the Gold, Kasami, and Bent sequences [2].



$$p(x) = x^6 + x + 1$$

Figure 4 - Shift register sequence generator. Generates msequence with period 63.

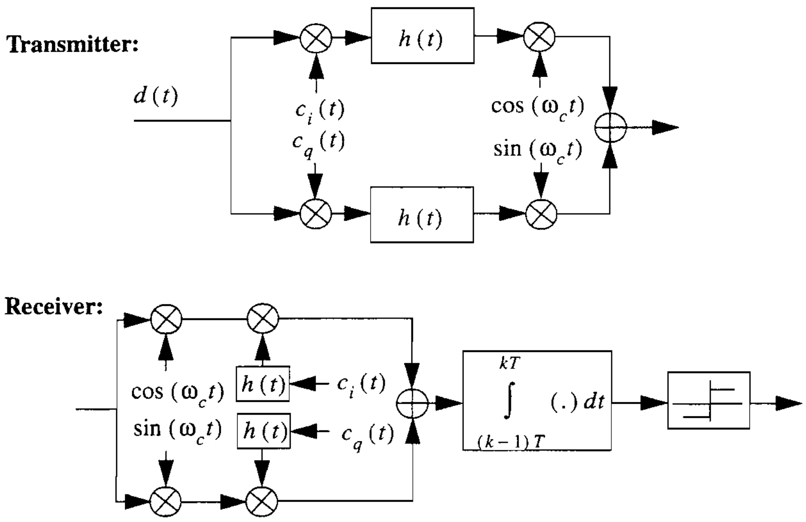


Figure 5 - QPSK bpe transmitter and receiver with chip pulse shape filtering.

2.2 Types of DS Modulators

The basic DS SS scheme that we have described above could be thought of as a BPSK modulation scheme at the chip pulse level. It is possible to generalize this scheme to obtain the analogous form of QPSK. In the case of spread spectrum there are a few possibilities. We may split the data sequence into two substreams and create a DS/BPSK signal for each stream where the carrier phases have a 90 degree phase difference. An alternative is to transmit each bit simultaneously on the two DS/BPSK systems as in a diversity scheme. The second of these alternatives is illustrated in Fig.

5. In these systems we are also filtering the chip pulses to limit the spectrum of the signal. The above system utilizes a continuous time correlator. A more modern approach is to use a chip matched filter to detect the chips individually and to use a discrete time correlator.

The above scheme can be modified to obtain an OQPSK version by inserting a halfchip delay element on the quadrature channel. It is also possible to obtain DPSK versions of the above, although the relative performance of the DPSK system to the BPSK system tends to be worse than in the case of narrow band systems. The advantages of the QPSK versions lie in a greater degree of robustness against tone interference. In the case of a BPSK system the effect of a tone interferer is dependent on the carrier phase of the interferer with respect to the carrier phase of the signal. There are also issues of constant envelope signal design which, depending on practical implementation issues, may be an important factor.

2.3 Synchronization

The receiver of a spread spectrum system requires the generation of a local spreading code signal that is time synchronized with the spreading code signal used at the transmitter, in addition to the usual carrier synchronization required for coherent demodulation. The spreading code synchronization consists of two steps: coarse synchronization where the time of the spreading code is brought within one chip (usually called *acquisition),* and fine synchronization where a tracking circuit continuously corrects errors in the code phase *(tracking).* The acquisition step is performed during link setup and when the tracking circuit looses lock. There are various scenarios in the code acquisition process. First is the degree of uncertainty. This could range from complete uncertainty over the whole period of the spreading code to the case where there is a smaller degree of uncertainty around a nominal time that is obtained from the system.

The main performance parameter of the acquisition circuit is the average acquisition time. This time depends on the length of the uncertainty interval in terms of the number of chips, the examination time to test a particular code phase, the signal to interference ratio, and the synchronization circuit architecture (serial or parallel correlation). The synchronization system may derive the carrier or PN code from the actual information signal or from a side channel which carries this signal. In most cases it is not energy efficient to transmit a side signal containing the carrier information. However, in the case of point-to-multipoint systems, such as in the forward link of a cellular system it is advantageous to transmit a strong signal since it can be used by many mobiles - such a signal is called a *pilot signal* and can be used to obtain both PN code and carrier synchronization.

2.4 Symbol Error Probability

The probability of symbol error depends on the power of the received signal and on the form of the interference, i.e. whether the interference is white Gaussian noise, colored Gaussian noise, tone interference, multiple access interference, etc. For moderate to large processing gains (say greater than 20), and error probabilities higher than 10^{-4}, we may derive a closed form expression which is reasonably accurate. Let us consider DS/BPSK (similar results may be obtained for other forms of DS), and a bandpass interference signal $Y(t)$. The received signal is

$$r(t) = \sqrt{2P}\left(\sum_{k=-\infty}^{\infty} d_{\lfloor \frac{k}{N} \rfloor} c_k h(t - k\tau)\right)\cos(\omega_c t + \theta) + Y(t) \tag{1}$$

where c_k is the PN code sequence, d_k is the data sequence, P is the received signal power, and $Y(t)$ is a bandpass interference process. We write this process in terms of two low pass processes as

$$Y(t) = Y_I(t)\cos(\omega_c t + \theta) + Y_Q(t)\sin(\omega_c t + \theta) \tag{2}$$

If the process is $Y(t)$ the above representation is possible and the power spectral density for $Y_I(t)$ is

$$S_I(f) = \text{LowPass}\,[S_Y(f - f_c) + S_Y(f + f_c)] \tag{3}$$

where $S_Y(f)$ is the power spectral density of $Y(t)$. If we use a correlator receiver, the output of the detector is

$$\pm ANE_h + 2\int_0^T Y(t)\, c(t)\cos(\omega_c t)\, dt \quad = \text{Signal+ Noise} \tag{4}$$

where the sign depends on the detected bit. For moderate to large processing gain and symbol error probabilities greater than 10^{-4} the above integral may be modeled as a Gaussian random variable. The probability of error can then be obtained as

$$P_{error} = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{N_{0eff}}}\right) \tag{5}$$

where $N_{0eff} = \frac{1}{E_h}\int_{-\infty}^{\infty} |H(f)|^2 S_I(f)\, df$ is an equivalent white noise power spectral density for the interference, and $E_h = \int_{-\infty}^{\infty} |H(f)|^2\, df$. As an example, if $Y(t)$ is a tone close to the carrier with power P_Y then $S_I(f) = P_Y \delta(f)$ and $\frac{E_b}{N_0} = \frac{NP_S}{P_Y}$ where P_S is the power of the signal.

2.5 Multiple Access Capability

The classical technique to achieve multiplexing on a channel is to assign orthogonal signal sets to the different users. From a practical standpoint two schemes for achieving orthogonality are commonly used: orthogonality in time (TDMA), and orthogonality in frequency (FDMA). Direct sequence spread spectrum offers an alternative to these. With the capability for interference rejection we can allow a number of signals on the channel. Each user transmits a DS signal with a unique spreading code. In general the signals are not orthogonal and the number of signals supported depends on the error probability requirement, and on a one-cell system is less than the corresponding number for FDMA or TDMA. The resulting multiple access scheme is typically called spread spectrum multiple access (SSMA), or code division multiple access (CDMA). Originally the latter term was used in cases where there was an attempt to design the spreading codes to achieve low crosscorrelations, hence maximize the number of users for a given error rate, whereas the term SSMA was used in cases where the codes behaved completely random. Recently, however, the term CDMA has achieved widespread usage regardless of the type of spreading codes.

We may classify CDMA systems by the spreading sequence length as either short sequence or long sequence systems. Short sequence systems utilize sequences with period in the order of the data symbol whereas long sequence systems utilize sequences with periods that are much greater than the data symbol period. The advantage of using short sequences is that the synchronization and design of the receiver is simpler. However with short sequences the interference in a CDMA system tends to be correlated with a resulting lower performance in the case of the standard correlator receiver. Also with short sequences the address space for the different users is limited.

Another major classification of CDMA systems is whether the system is synchronous or nonsynchronous at the chip level, i.e. whether the chip level transitions for the different users occur in synchronism. For a nominal spreading bandwidth of 1 MHz the chip duration is approximately 1 μs, hence it is difficult to achieve chip synchronous operation unless the link is a point to multipoint link or the link distances are small. With chip synchronous operation it becomes possible to utilize orthogonal spreading codes so as to achieve a large multiple access capability and at the same time achieve the benefits of spread spectrum such as robustness to external interference.

Let us consider the various system options in greater detail. In an orthogonal CDMA system we may represent the set of spreading codes for the various users by a matrix where each row is the spreading code of a given user. Let N be the number of chips per data symbol, then the maximum number of spreading codes is N and the set of codes forms a Hadamard matrix. For a given bandwidth the multiple access capability of this scheme is the same as that of TDMA or FDMA. If $N > 2$ then Hadamard matrices exist only if N is a multiple of 4. In general there are many Hadamard matrices of a given dimension, for N equal to a power of 2 a popular construction is obtained recursively as follows: $\boldsymbol{H}_0 = \mathbf{1}$, and

$$\boldsymbol{H}_n = \begin{bmatrix} \boldsymbol{H}_n & \boldsymbol{H}_n \\ \boldsymbol{H}_n & -\boldsymbol{H}_n \end{bmatrix} \quad (6)$$

At this point the simplest approach is to assign a row of the above matrix to each user and have the code repeat for each transmitted symbol. The drawback here is that the rows of the above matrix don't have good randomness properties, e.g. the first row consists of all 1's. Also since the codes are periodic with period equal to the data symbol period the interference at the detector will exhibit periodic behaviour. If we take any ± 1 valued N component vector v_0 and multiply all rows of the above matrix we still have a Hadamard matrix. If we perform the same multiplication with a v_0 which changes every data symbol then we have an orthogonal CDMA scheme with arbitrarily long period spreading codes. The resulting set of orthogonal spreading codes may be obtained by masking the rows of the above Hadamard matrix with successive N-chip portions of a long spreading code, also referred to as a masking sequence. In this scheme the matrix rows correspond to separate channels in FDMA or separate time slots in TDMA.

Next we consider a nonsynchronous CDMA scheme with short spreading sequences. In this case it is feasible to design sequence sets which minimize the crosscorrelations between the various codes over all possible relative time shifts between the sequences and at the same time yield sequences with good randomness properties (i.e. delta function type of autocorrelation function). Such sequence sets have been found - examples are the Gold, and Kasami sequences. Fig. 6 shows a circuit which generates the set of Gold sequences of period 15. The shift registers can be initialized in 17 distinct ways to yield 17 sequences in the set. In general for shift registers with n stages the number of sequences in the set is $2^n + 1$.

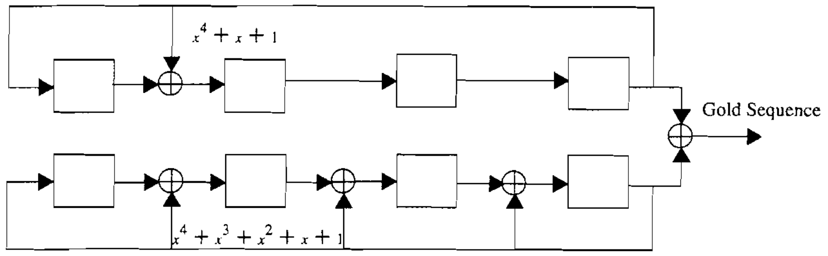


Figure 6 - Gold sequence generator.

Even though the Gold sequences are frequently used there are two weaknesses in the design criteria: i) the design criteria consider periodic crosscorrelations between the sequences, where in practice, due to the modulation by the data symbol the interference caused by one user is a result of the crosscorrelation of one code in the set with another code which has been partially inverted. Another weakness is that the optimization considers the peak value of the correlations over all code shifts as opposed to some average value. Nonetheless, since no tractable better criterion is available, these sequences are chosen in many applications. The exact probability of error as a function of the number of active users is possible to determine in principle but difficult to compute in practice.

Now we consider the case of nonsynchronous operation with very long sequences. In this case it is very difficult to design a set which minimizes any measure of crosscorrelations. In this case most sequence sets are equivalent and can usually be modelled as Bernoulli processes. For a moderate to large processing gain and if the number of users is not too small the error probability may be computed using (5) [6] [7] where

$$S_I(f) = N_0 + \frac{1}{2T_c} |H(f)|^2 \sum_{i=1}^{K-1} A_i^2 \tag{7}$$

For BPSK modulation the error probability is then

$$P_{error} = \frac{1}{2} erfc\left(\sqrt{\mathrm{SNR}_0}\right) \tag{8}$$

where

$$\mathrm{SNR}_0 = \frac{E_b}{N_0 + \frac{\Psi}{N}\sum_{k=1}^{K-1} E_{b,k}} \qquad (9$$

K is the number of users, $E_{b,k}$ is the energy per bit for the k^{th} user, E_b is the energy pe bit for the user of interest, N is the processing gain, N_0 is the one-sided power spectra density of the background noise, and

$$\Psi = \frac{1}{E_h^2 T_c}\int_{-\infty}^{\infty} |H(f)|^4 df \qquad (10$$

is a factor which depends only on the chip pulse shape: E_h, T_c, and $H(f)$ are the energy duration, and Fourier transform of the chip pulse. As an example, for a rectangular chip pulse (time limited) $\Psi=2/3$, and for a sinc chip pulse $\Psi = 1$. For different types o modulation, including coded modulation schemes, the bit error probability is given b a similar expression to (8) except that the function $\frac{1}{2}erfc(\sqrt{.})$ is replaced by some othe function $f(.)$. Eq. (8) illustrates the necessity for power control in CDMA systems, t support the maximum number of users subject to a maximum probability of error th energy per bit of the various signals at the receiver (or received power) should b constant over all users. Also if the power of the signals is sufficiently large then we ca neglect the background noise N_0.

For a given application the important parameter is the multiple access capability. Thi depends on the error probability requirement which depends on the type of service. A given error probability requirement translates into a nominal SNR requirement SNR_0 The multiple access capability is then

$$K = \frac{N}{\Psi \cdot \mathrm{SNR}_0} + 1 \qquad (11$$

As an example, if SNR_0 = 6 dB, and we use a sinc chip pulse then $K = \frac{N}{4} + 1$. Thus fo a processing gain of N (bandwidth expansion factor) we can accommodate approxi mately 25% of N users. For orthogonal schemes the number of users would be equa where

$$\mathrm{SNR}_0 = \frac{E_b}{N_0 + \frac{\Psi}{N}\sum_{k=1}^{K-1} E_{b,k}} \tag{9}$$

K is the number of users, $E_{b,k}$ is the energy per bit for the k^{th} user, E_b is the energy per bit for the user of interest, N is the processing gain, N_0 is the one-sided power spectral density of the background noise, and

$$\Psi = \frac{1}{E_h^2 T_c}\int_{-\infty}^{\infty} |H(f)|^4 df \tag{10}$$

is a factor which depends only on the chip pulse shape: E_h, T_c, and $H(f)$ are the energy, duration, and Fourier transform of the chip pulse. As an example, for a rectangular chip pulse (time limited) $\Psi=2/3$, and for a sinc chip pulse $\Psi = 1$. For different types of modulation, including coded modulation schemes, the bit error probability is given by a similar expression to (8) except that the function $\frac{1}{2}erfc(\sqrt{.})$ is replaced by some other function $f(.)$. Eq. (8) illustrates the necessity for power control in CDMA systems, to support the maximum number of users subject to a maximum probability of error the energy per bit of the various signals at the receiver (or received power) should be constant over all users. Also if the power of the signals is sufficiently large then we can neglect the background noise N_0.

For a given application the important parameter is the multiple access capability. This depends on the error probability requirement which depends on the type of service. A given error probability requirement translates into a nominal SNR requirement SNR_0. The multiple access capability is then

$$K = \frac{N}{\Psi \cdot \mathrm{SNR}_0} + 1 \tag{11}$$

As an example, if $\mathrm{SNR}_0 = 6$ dB, and we use a sinc chip pulse then $K = \frac{N}{4} + 1$. Thus for a processing gain of N (bandwidth expansion factor) we can accommodate approximately 25% of N users. For orthogonal schemes the number of users would be equal to N and the error probability would depend only on the background noise. In general the relationship between SNR_0 and the bit error rate depends on the modulation and error correction coding schemes.

2.6 Forward Error Correction

Error correction coding schemes are invariably used in spread spectrum modulation. A general characteristic of error correction coding schemes for narrowband systems is the resulting bandwidth expansion. With spread spectrum modulation this bandwidth expansion is absent. What is typically done is to expand the coded symbol rate at the expense of the spreading factor, while maintaining the chip rate constant. The only consideration in the use of error correcting codes is the implementation cost. This fact means that powerful codes such as orthogonal or ReedMuller (RM) codes which normally have a large bandwidth expansion factor are quite suitable for spread spectrum.

For a given system bandwidth let us compare an uncoded system with a coded system. We assume that the bandwidth expansion factor in both systems is N. For the uncoded system we use a spreader with a factor of N. Thus each input block of n information symbols produces nN chips on the channel. Now let us assume that we use an $(2^n, n)$ orthogonal code then for the coded system the same block of information symbols will produce a block of 2^n symbols at the output of the error correction encoder. If we use a spreader with a factor of $nN/2^n$ then we have a block of nN chips transmitted on the channel (same as for the uncoded system). Thus both systems have the same bandwidth. The uncoded system achieves the bandwidth expansion by the SS spreading, whereas the coded system achieves the bandwidth expansion by both spreading and error coding. It will be seen that even though the coded system has a lower spreading factor the error correction code more than compensates for the lower spreading factor and the coded system has a lower error rate for a given value of SNR_0.

The code words in the orthogonal code may be mapped to rows of a Hadamard matrix such as that defined by (6). The optimum receiver for the coded system uses a matched filter to detect a real vector of 2^n components. This vector is correlated with each row of the Hadamard matrix and the row which yields the maximum correlation is the decoded word. The correlation with all rows of the Hadamard matrix may be performed efficiently using a Fast Hadamard Transform algorithm. For uncoded BPSK the bit error rate is given by (5). For the orthogonal coded system the bit error rate is given by [8].

$$P_e = \frac{2^{n-1}}{2^n - 1}\left(1 - \int_{-\infty}^{\infty} \frac{e^{-x^2}}{\sqrt{\pi}}\left(\frac{1}{2}erfc\left(-x - \frac{\sqrt{nE_b}}{\sqrt{N_{oeff}}}\right)\right)^{2^n - 1} dx\right) \qquad \textbf{(12)}$$

The two schemes are compared in Fig. 7 for the case of a coded system with $n = 6$. For an error rate of 10^{-3} the coded scheme performs better than the uncoded scheme by 3 dB.

As the spreading factor increases we can use codes with lower rate and achieve greater coding gains. In fact with CDMA it becomes possible to approach the Shannon capacity for the channel. As the spreading ratio increases lower rate codes with higher performance can be used. The result is that for a given bit error rate the number of users normalized by the system bandwidth increases to a limit determined by the Shannon capacity.

Similar coding gains may be obtained by using other popular codes such as convolutional codes. A very effective technique is the use of a combination of two codes known as a concatenated code. For example an orthogonal and a convolutional code are used in the reverse channel of the IS-95 CDMA standard [5].

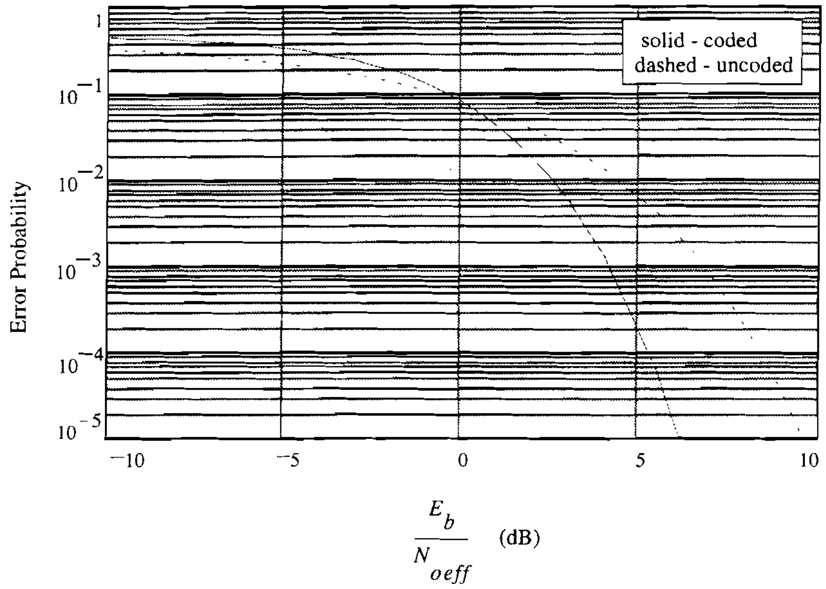


Figure 7 - Bit error rate for uncoded system and coded system using an (64,6) orthogonal code.

Another very effective combination is an RM or orthogonal code and a ReedSolomon code.

3 CDMA Cellular Networks

A cellular network consists of a three level architecture: a *switch* which performs the overall control of the network, *base stations* which are located in the various radio cells and communicate with the mobiles by radio and with the switch by a set of point to point links using hardwired or microwave trunks, and the *mobiles*. Full duplex communication is achieved using frequency division duplex (FDD). The two links in the duplex are fundamentally different in that the base to mobile link is a point-to-multipoint link (known as the *forward link*) and the mobile to base link is a multipoint-to-point link (the *reverse link*). The implication here is that it is easy to implement synchronous CDMA in the forward link but not in the reverse link.

With narrow band modulation schemes a fundamental concept in frequency reuse is the concept *of frequency re-use cluster*. Cells are grouped into clusters, of for example 7 cells, and within each cluster the whole set of available radio channels is used - each channel is used precisely once within the cluster. For spectral efficiency the cluster size should be made as small as possible. However, as the cluster size is decreased cochannel interference, with the associated loss in the quality of transmission, increases.

The use of CDMA in a cellular network was initially proposed in [9]. Spread spectrum, with its interference rejection capability, allows for operation of a cellular system with a cluster size of I. The whole frequency band is used in each cell. The number of users per cell depends on the level of multiuser interference. Multiuser interference arises from in-cell interferers and out-of-cell interferers. The nature of these types of interference on the forward and reverse channels is different. In general, to minimize interference we may attempt to orthogonalize the signals so as to reduce multiuser interference. For the set of signals transmitted from one base station this is possible in practice but the multipath propagation tends to destroy the orthogonality to some extent. For the signals in the reverse channel (and assuming no multipath propagation) it is possible to orthogonalize the signals in principle but difficult to achieve in practice. If we consider signals from different cells then it is impossible even in principle and assuming single path propagation to orthogonalize the whole set of signals and achieve high spectral efficiency.

To minimize the interference level and to maintain a minimum performance level, power control should be used. Each terminal adjusts its power so that the received power at a base station is constant over all mobiles. In cellular systems typically the bandwidths assigned to the forward and reverse channels are equal. In principle the capacity of the system is limited by the capacity of the reverse channel due to its multipoint-to-point nature, although this may not necessarily be so in a given implementation; hence in this paper we consider the capacity of the reverse channel.

Assuming a one-cell system, the capacity is given by (11) - let this capacity be denoted by K_1. For a multiple cell system there will be inter-cell interference in addition to the in-cell interference, hence for a given error rate we must reduce the number of users per cell to a value $K_m \leq K_1$. The exact value of K_m depends on the positions of the users within the cell: if all users are near their respective base stations then $K_m \approx K_1$. The other extreme is where most users are near their respective cell boundaries where the inter-cell interference is the largest. In general the total intercell interference will depend on the number of users per cell, their positions within their cells, and the propagation power loss law, e.g. r^{-4} over flat ground. If we model the terminal positions as random quantities then the inter-cell interference is a random variable. The variance of this random variable decreases as the number of users increases. If the number of users per cell is a few tens then we may approximate the intercell interference by its mean. The inter-cell interference is then $\alpha K_m P$ where α is a parameter which depends on the propagation power loss law ($\alpha < 1$), K_m is the number of users per cell and P is the common received power of a user at its base station. For an ideal propagation environment with an r^{-4} power loss law $\alpha \approx 0.4$. The total interference at a base station is

Total Interference = In-Cell Interference + Inter-Cell Interference

The in-cell interference is $(K_m - 1)\,P$, hence the total interference is approximately $((1 + \alpha)\,K_m - 1)\,P$ and the capacity per cell may be obtained by rederiving (11) from (9) by replacing the in-cell interference term in the denominator of (9) with the total interference, which yields

$$K_m = \left(\frac{N}{\Psi \cdot \mathrm{SNR}_0} + 1\right) F \tag{13}$$

where $F = 1/(1 + \alpha)$. The parameter F is a frequency re-use factor. It represents the reduction in cell capacity which is required to reduce inter-cell interference to acceptable levels. For narrow band systems F corresponds to the reciprocal of the cluster size.

From eqs. (8) and (9) and (13) the capacity of the CDMA system is limited by interference. Various enhancements to the network can be made which lead directly to a decrease in interference and a corresponding increase in capacity. We now discuss some of these.

3.1 Sectorized Antennas

Let us assume that directional antennas are used where there are G directional antennas per cell which cover the cell and have non-overlapping beams. If the sources of the interferers are uniformly distributed in space then the amount of interference at each antenna at a base station is reduced by the factor G^{-1}. Hence from (9) assuming that N_0 is negligible the SNR_0 is increased by a factor of G. Hence we can increase the number of users by the factor G and still maintain the required SNR constraint. In practice the antenna radiation patterns are not ideal, there is some sector to sector interference and the interference reduction is somewhat less than the above.

As the mobile terminal moves from sector to sector a handoff is performed in the sense that the transmission/reception of the signal switches to a different antenna sector. If the antenna beamwidth is very small then we have a large capacity per cell but handoffs occur very frequently as a terminal travels across the cell, and this may lead to a degradation in performance depending on the implementation of the handoff. Hence there is a practical upper limit to the number of sectors per cell.

Another technique to reduce interference is to use adaptive antennas. These antennas typically contain a number of elements whose inputs/outputs are combined with the amplitudes and phases adjusted dynamically in such a way as to null out the interferers and maximize the signal to interference ratio. These techniques have the potential to offer very large increases in capacity but require a large degree of signal processing capability at the receiver.

3.2 Non Continuous Traffic Sources

Traffic sources may be classified into three basic types: circuit switched with continuous transmission at a constant rate; circuit switched with bursty transmission or nonconstant rate transmission, and packet switched. For the second type of traffic the throughput of the system utilizing multiple access techniques such as FDMA or TDMA can be increased by using statistical multiplexing techniques where the number of logical channels serving a group of users is less than the number of users. Spread spectrum offers a natural method to implement this type of statistical multiplexing. If the power of the transmitted signal is reduced when the bit rate of the signal decreases then the interference in the network is reduced and the capacity of the network (in terms of the number of traffic sources supported) is increased. In the case of voice signals the bit rate concentrates on two main values: full rate, corresponding to a talk spurt, and zero, corresponding to a pause. For a typical conversation approximately 33% of the time is occupied with talk spurts and the remaining time is occupied with pauses. We refer to this value as the *voice activity factor.*

To decrease interference we may turn off the signal during the speech pauses. The difficulty here is to resynchronize when the next talk spurt arrives. One technique is to introduce a slot structure to the system and initiate transmissions only at the beginning of a slot.

For a given number of users in a cell the number of users which are transmitting a talk spurt at a given time is a Binomial distributed random variable with parameter V, the speech activity factor. For a large number of users the variance of this becomes small relative to the mean. In such a case the interference in the network is reduced to a fraction (slightly greater than V) of the interference in the case of continuous stream transmissions V, depending on the number of interferers and link outage probability.

CDMA can also accommodate constant traffic streams with different rates in a natural way. For a given error probability (performance level) the important parameter is $\mathrm{SNR}_0 = E_b/N_{0eff}$, hence if the data rate is lowered then we can achieve a given value of E_b with a lower transmitted power. This property makes it easy to integrate services with different data rates into one transmission scheme with a fixed chip rate and system bandwidth. The transmitted power of a given user is proportional to its transmitted data rate.

3.3 CDMA Capacity Equation

With the above two methods for interference reduction the sum in the denominator of (9) is reduced by the factor V/G. This value must be accounted for in rederiving (11). Also, (11) was derived assuming uncoded BPSK modulation where N is the number of chips per data bit, the processing gain. In the case of the use of error correcting codes the relevant parameter is still the processing gain, however it is now more generally defined as the ratio of the system bandwidth to the data rate, W/R. If we use a high degree of filtering the chip pulses become sinc pulses and the parameter ψ approaches unity. Accounting for all of these factors, the capacity equation becomes

$$K_m = \left(\frac{W}{R \cdot \mathrm{SNR}_0} + 1\right)\frac{FG}{V} \tag{14}$$

As an example we compute the capacity for a system such as the IS-95 system. Let $W = 1.25$ MHz, $R = 9600$ bits/s, $G = 3$, $V = 0.4$, $F = 0.71$, and $\mathrm{SNR}_0 = 7$ dB. The capacity is then $K_m = 144$ users/cell. As a comparison with the analog cellular system AMPS in the same bandwidth we can have 41.6 30KHz channels. If the frequency reuse cluster size is 7 then the number of users per cell is 42/7 = 6. Thus for the above parameters the capacity advantage of the IS-95 system over AMPS is 130/6 = 24. References [10][11][12][13] contain a more detailed analysis of the capacity.

There has been some controversy as to the capacity of CDMA systems relative to narrowband cellular systems. Part of the reason for this is the validity of the various approximations made in calculating the parameters in (14). The calculation of these parameters depends on the positions of the terminals and on the accuracy of various averaging operations. However the most important parameter and the one that causes the greatest uncertainty is the operating value of SNR, SNR_0. For each 3 dB change in the value of this parameter the capacity calculation changes by a factor of 2. For a given level of performance (error rate), the required value of SNR_0 depends on the propagation environment and the effectiveness of the power control scheme. For typical error rates (e.g. 10^{-3}) such a value may range anywhere from 3 dB to 15 dB. This corresponds to a range of 12 dB, or a capacity range by a factor of 1:16. If we assume that a more realistic requirement given the fading process is $SNR_0 = 10$ dB then the above capacity advantage over AMPS would be approximately 12.

4 Signal Propagation

The propagation environment has a large effect on the capacity of a CDMA network. The actual effect of the environment depends on the velocity of the mobile terminal and the effectiveness of the power control scheme in tracking changes in the power of the received signal. As the mobile travels the signal fades according to two component processes - a slow fading process due to shadowing and a fast fading process due to multipath propagation. To maximize the number of users, a power control scheme which attempts to maintain a constant received power must be used. The variations in signal strength due to shadow fading are frequency independent and the variations due to multipath are frequency dependent. Frequency independent variations can be compensated for based on the received signal power on the forward channel. The mobile adjusts its transmitted power based on the received signal strength. This type of compensation is sometimes called *openloop* power control. Variations due to multipath propagation can not be compensated by the above technique since the channel coherence bandwidth is smaller than the frequency separation for the forward and reverse channels; i.e., the fast fading processes in the forward and reverse channels are independent.

Considering multipath fading, the channel may be modelled with the following impulse response

$$h(t) = \sum_{i=0}^{n} a_i e^{j\theta_i} \delta(t - \tau_i) \tag{15}$$

Over small distances the a_i's remain relatively constant but the phases θ_i vary unpredictably. If the relative delays between any two paths are greater than τ then the signals in the different paths are uncorrelated (resolvable) and the received signal power is relatively constant over short distances. If the relative delays are small then the baseband signals are correlated and the RF signal varies drastically over small distances due to constructive destructive interference.

The optimum receiver for the multipath channel is the Rake receiver [14]. It consists of a set of correlators (sometimes called the Rake fingers) which demodulate the signals on the various resolvable components. The outputs of the various correlators are then combined to achieve what we may call *path diversity.* For optimum performance the required number of fingers in the Rake receiver depends on the delay spread of the channel and the bandwidth of the system - it is given approximately by DW, where D is the delay spread and W is the system bandwidth.

With a Rake receiver the effect of multipath propagation on the spectral efficiency of a CDMA network can be practically mitigated if the total received power on all paths can be maintained constant. The relevant parameters here are the system bandwidth (or chip rate) which determines the variability of the power of the received signal with distance, the channel delay spread which determines the number of received signal components, and also the variability of the received signal, the velocity of the mobile terminal, the speed of power control adjustments, and the number of fingers in the Rake receiver.

After we have done our best in system design, taking all of the above factors into account (Rake receiver, power control, etc.), we have a system where the SNR at the receiver is still not constant. The SNR may vary with distance as shown in Fig. 8. To maintain acceptable performance the SNR must be greater than some value SNR_m most of the time. On the other hand, in the capacity formula the relevant parameter is the average SNR$-SNR_0$. The difference $SNR_0 - SNR_m$ is the *fading margin.* The larger this value the lower the spectral efficiency of the system. This concept of fading margin can be generalized to systems with any type of modulation. The fading margin is a fundamental parameter in the design of any radio system. The reduction of the fading margin is extremely important to achieve a high spectral efficiency.

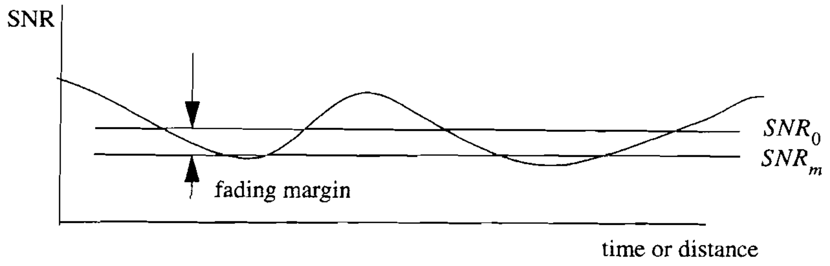


Figure 8 - Residual SNR variation due to imperfections in power control scheme and implementation of the Rake receiver.

5 CDMA Standards

Currently the most developed CDMA system is that specified by the IS-95 star [5]. This is a Telecommunications Industries Association (TIA) standard for se generation cellular systems which is a wideband version and competitor of the TI standard, IS-54. The bandwidth of the system is 1.25 MHz. Orthogonal CDMA 64 channels is used in the forward link. Non-synchronous CDMA is used in the re link. Synchronization in the forward link is facilitated by a pilot tone which cont large power relative to the power of the signal of a user. Signals in the forward ch are multiplexed using orthogonal spreading codes specified by a Hadamard m The multiplexed signal is then masked by a spreading code that is also transmitt the pilot tone. The pilot tones for different base stations correspond to different of the same PN code.

As all other digital cellular systems this system makes use of powerful forward correcting codes along with interleaving. A rate 1/2 convolutional code is used i forward link. The reverse link uses a concatenated coding scheme consisting ((64,6) orthogonal code and a rate 1/3 convolutional code. Power control at the m terminal is achieved in two ways: coarse power control which is based on the p of the pilot tone and fine power control which is achieved by the transmission (800 bit/s stream on the forward link, where each bit signals a power adjustment a mobile by ±1 dB. The development of this system has resulted in the implement of a number of sophisticated techniques to enhance network performance. An these we mention the implementation of fine power control, the so called *soft han* algorithm, Rake receivers, and variable rate speech coders.

Another major CDMA project is the European project *code division testbed* (COI [15]. The goal is to design a third generation system under the guidelines of the m communications program UMTS. The goal is to integrate various services at var bit rates in a single modulation format. To achieve these goals the modulation sch allows for different chip rates, data rates, processing gains, and transmitted pc levels. A major difference from the IS-95 system is the use of nonsynchronous CD for both the forward and reverse channels. Another difference is the less strin requirement for synchronization of the various base stations.

6 Summary

We have discussed some of the basic principles of CDMA networks. We concentr on direct sequence spread spectrum and discussed important system issues such as capability for interference suppression, calculation of error probability, multiple ac capability, synchronization, use of error control codes, the use of CDMA in a cell type of network including the need for power control, and its performance in fac channels. Due to space constraints we have only discussed some of the fundamen of the subject. The literature in this subject has grown immensely in the past ten years, as can be attested by the existence of international conferences and journal special issues solely dedicated to the topic. Much of the commercial interest in CDMA in the past five years has centered on its application to cellular systems where the capacity issue is of utmost importance. As a result much of the emphasis on research centered on capacity issues. Other issues such as coexistence with other systems, flexibility in system design, hand-off implementation, and operation under blanket spectrum allocation are likely to take on greater importance in the future.

7 References

[1] R. C. Dixon, *Spread Spectrum Systems,* John Wiley and Sons, 1984.

[2] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications,* Computer Science Press, Rockville, Md., 1984.

[3] D. L. Schilling and R. L. Pickholtz and L. B. Milstein, "Spread spectrum goes commercial", *IEEE Spectrum,* August 1990, pp. 40-45.

[4] S. W. Golomb, *Shift-Register Sequences,* San Francisco: Holden Day, 1967.

[5] IS-95, Wideband spread spectrum digital cellular system dual-mode mobile station base station compatibility standard, Telecommunication Industries Association.

[6] E. S. Sousa, "Interference modeling in a direct sequence spread spectrum packet radio network, *IEEE Trans. on Commun.,* Sept., 1990, pp. 1475-1482.

[7] E. S. Sousa, "The effect of clock and carrier frequency offsets on the performance of a direct sequence spread spectrum multiple-access system", *IEEE J. on Select. Areas in Commun.,* vol. 8, No. 4, May 1990, pp. 580-587.

[8] W. C. Lindsay and M. K. Simon, *Telecommunication Systems Engineering,* Prentice Hall, New Jersey, 1973.

[9] G.R. Cooper and R.W. Nettleton, "A spread-spectrum technique for high capacity mobile communications", *IEEE Trans. on Veh. Tech.,* vol. VT7, No. 4, Nov. 1978.

[10] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver and C. E. Wheatley, "On the capacity of a cellular CDMA system", *IEEE Trans Veh. Techn.,* vol. VT40, pp. 303-312, May 1991.

[11] R. Padovani, "Reverse link performance of IS-95 based cellular systems", IEEE Personal Communications Magazine, vol. 1, no. 3,1994, pp. 28-34.

[12] A. J. Viterbi, A. M. Viterbi, and E. Zehavi, "Performance of power-controlled wide-band terrestrial digital communication", *IEEE Trans on Commun.*, vol. 41, April 1993,pp.559-569.

[13] C. Kchao, and G. L. Stuber, "Analysis of a direct-sequence spread-spectrum cellular radio system", *IEEE Trans. on Commun.*, vol. 41, 1507-1516.

[14] R. Price and P. E. Green, Jr., "A communication technique for multi-path channels", *Proc. IRE,* pp. 555-570, Mar. 1958.

[15] A. Baier, U. C. Fiebig, W. Granzow, W. Koch, P. Teder, and J. Thielecke, "Design study for a CDMA-based third-generation mobile radio system", *IEEE J. Select. Areas on Commun.,* vol. 12, no. 4, May 1994, pp. 733-743.

Elvino S. Souza was born in Graciosa, Açores, Portugal, in December 28, 1956. He re ceived the B.A.Sc. degree in engineering science, and the M.A.Sc. degree in electrica engineering from the University of Toronto in 1980 and 1982 respectively, and th Ph.D. degree in electrical engineering from the Universty of Southern California i 1985. Since 1986 he has been with the department of Electrical and Computer Engin eering at the University of Toronto where he is presently an Associate Professor. Sinc 1986 he has been a Natural Sciences and Engineering Research Council of Canad (NSERC) University Research Fellow. He has performed research in spread spectrum systems and radio networks since 1983. His current interests are in the areas of sprea spectrum systems, mobile communications, and indoor wireless communications wit emphasis on the high level performance of the network and signal processing issues af fecting such a performance. At the University of Toronto he has developed a graduat course in mobile communications and also teaches a graduate course in error contro codes. In recent years he has been invited to give lectures and short courses on sprea spectrum and wireless communications in a number of different countries. He is th technical program chairman for the Sixth IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), September, 1995, Toronto, Ca nada.