



S-38.220

Licentiate Course on Signal Processing in
Communications, FALL -97

Code Selection for CDMA Systems

Kimmo Kettunen

VTT, Information Technology
P.O.Box 1202, FIN-02044 VTT
tel. 358 9 456 5608
kimmo.kettunen@vtt.fi

Date: 13.11.1997

Contents

1	Introduction	3
2	CDMA system design aspects	3
3	General Properties of PN Sequences	3
3.1	Statistical properties of PN sequences	4
3.2	Periodic and Aperiodic Correlation Functions	4
4	Some known families of spreading sequences	6
4.1	Walsh-Hadamard codes	6
4.2	<i>m</i> -sequences	6
4.3	Sequences with low cross-correlation	8
4.3.1	Gold sequences	9
4.3.2	Kasami sequences	10
4.3.3	Bounds on Crosscorrelation of Sequences	10
5	Code Selection Criteria for CDMA Systems	11
6	Conclusions	13

1 Introduction

This report gives an overview on the code selection principles with emphasis on the CDMA system code design. Since the pseudo-noise sequence research is a vast area of study, our emphasis has been a main factor when deciding the range of issues that are covered. The report begins with some general background, mainly some CDMA system design issues and the basic properties of the PN sequences, including the various correlation functions. After that some widely used spreading sequences are presented with their basic properties, together with some well-known bounds that have a major impact on the code design. Finally, some results of the performance study done in [Kär96] is described and some conclusions are drawn from those results.

2 CDMA system design aspects

In general, the spread spectrum signals are commonly used for [Pro92]

- combatting or suppressing the detrimental effects of interference due to jamming, interference arising from other users of the channel, and self-interference due to multipath propagation;
- hiding a signal by transmitting it at low power and, thus making it difficult for an unintended listener to detect in the presence of background noise;
- achieving message privacy in the presence of other listeners.

The desired properties of PN sequences depend clearly on the target application. Since our goal is to study the code selection for CDMA systems, the main emphasis is on the control of the interference arising from other users of the channel. The detailed requirements depend on the target system type, such as whether the system is a military or civilian system, is it a cellular system, what are the code length requirements and is the mode of operation synchronous or asynchronous. These high-level requirements induce requirements in code design for the various properties of the target code, such as the auto- and cross-correlation and the pseudo-noise (balance, run distribution, power spectrum) properties. In practice a choice must be made to select some parameters to be emphasized in the design, such as the maximal vs. the mean square-sum correlation value.

3 General Properties of PN Sequences

This section gives a brief overview on the necessary background for understanding the pseudorandom sequences and their properties. The emphasis will be on the properties

relevant to CDMA code design. This includes the definitions of various cross- and auto-correlation functions as well as the key pseudo-noise properties usually required.

3.1 Statistical properties of PN sequences

In general, the SSMA chip sequences are assumed to be statistically independent, identically distributed random processes, i.e. Bernoulli sequences. Thus the sequences used for the generation of spread spectrum signals should have the same key “randomness” properties as a truly random Bernoulli sequence. These key properties, the three randomness postulates of Golomb can be characterised as follows([Vit95]):

Balance Property Relative frequencies of “0” and “1” are each 1/2.

Run Property Run lengths (of zeros or ones) are as expected in a Bernoulli sequence.

Shift Property If the random sequence is shifted by any nonzero number of elements, the resulting sequence will have an equal number of agreements and disagreements with the original sequence.

In practice, especially the short code period CDMA sequences are designed to have low time cross-correlation, so they are only quasi-orthogonal and thus the sequences in short-code CDMA systems are not statistically independent. With longer code period, the CDMA sequences behave more like Bernoulli sequences. The long-code CDMA systems are in fact sometimes called (pseudo)random (R-CDMA) systems as opposed to short-code deterministic (D-CDMA) systems.

3.2 Periodic and Aperiodic Correlation Functions

The correlation properties of PN codes play a major part in the code design for CDMA systems, since they determine not only the level of multiple access interference, i. e. the interference arising from other users of the channel and self-interference due to multipath propagation, but also the code acquisition properties. The first one is affected by the cross-correlation properties between different codes of the family whereas the last two are affected by the auto-correlation properties, that is the correlation between time-shifted versions of the same code. There are several variants of correlation functions:

Definition 1 Let $\{u(t)\}$ and $\{v(t)\}$ be (complex valued) sequences of length (or period) N . The aperiodic cross-correlation function of the sequences $\{u(t)\}$ and $\{v(t)\}$ is defined

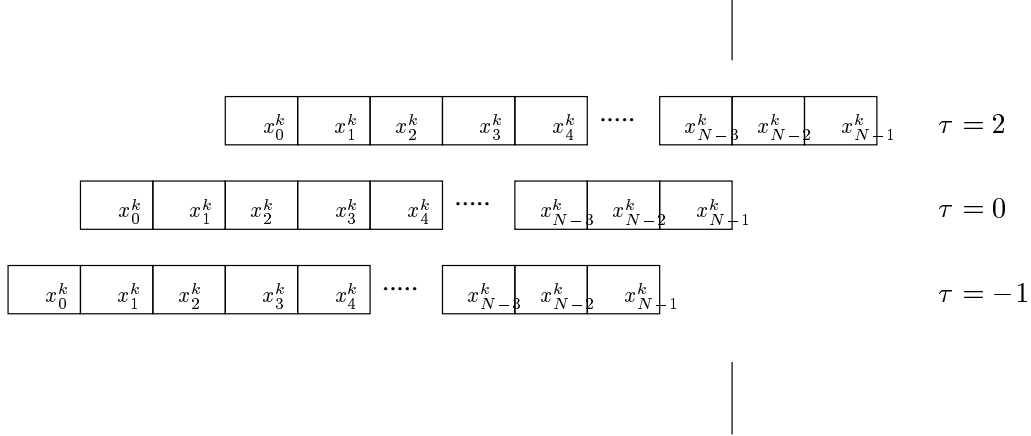


Figure 1: Aperiodic correlation

by

$$\rho_{u,v}(\tau) = \begin{cases} \sum_{t=0}^{N-1-\tau} u(t+\tau)v^*(t) & \text{if } 0 \geq \tau \geq N-1, \\ \sum_{t=-\tau}^{N-1} u(t+\tau)v^*(t) & \text{if } 1-N \geq \tau < 0. \end{cases} \quad (1)$$

The periodic (or even) cross-correlation function of the sequences $\{u(t)\}$ and $\{v(t)\}$ is defined by

$$\theta_{u,v}(\tau) = \sum_{t=0}^{N-1} u(t+\tau)v^*(t) = \rho_{u,v}(\tau) + \rho_{u,v}(\tau - N). \quad (2)$$

The odd cross-correlation function of the sequences $\{u(t)\}$ and $\{v(t)\}$ is defined by

$$\hat{\theta}_{u,v}(\tau) = \rho_{u,v}(\tau) - \rho_{u,v}(\tau - N). \quad (3)$$

When the sequences u and v are the same one speaks about *autocorrelation* functions, denoted by θ_u and $\hat{\theta}_u$. The figure 3.2 illustrates the meaning of positive and negative values of τ . The difference between even and odd cross-correlation becomes clear if one considers a spread spectrum system where a sequences is produced from a PN sequence through modulation by a data sequence. Then the odd (resp. even) cross-correlation occurs when the consequent data bits are the different (resp. the same).

The aperiodic correlations give more realistic estimates of the CDMA system performance as the periodic ones do. However, since the problem of designing sequence families with low aperiodic correlation is a difficult one, the conventional approach has been to design based on periodic correlation properties and to subsequently to analyze the aperiodic properties of the resulting design.

4 Some known families of spreading sequences

In this section we will present some of the most common PN code families, namely the Walsh-Hadamard codes, m -sequences, Gold-codes and Kasami-codes. The Walsh-Hadamard codes are orthogonal on zero code delay whereas the m -sequences, Gold-codes and Kasami-codes are non-orthogonal with varying cross-correlation properties.

4.1 Walsh-Hadamard codes

The Walsh-Hadamard codes are generated as follows: Let

$$\mathbf{H}_2 = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix} \quad (4)$$

and, if $M > 2$ is a power of 2, recursively

$$\mathbf{H}_M = \begin{pmatrix} +\mathbf{H}_{M/2} & +\mathbf{H}_{M/2} \\ +\mathbf{H}_{M/2} & -\mathbf{H}_{M/2} \end{pmatrix} \quad (5)$$

Then the rows of any \mathbf{H}_M form a mutually orthogonal set of sequences. As mentioned above these sequences have zero cross-correlation when the codes are synchronous, but when asynchronous, their cross-correlation is very much dependent on the particular pair of codes used, some will have cross-correlation zero while others will have a very high correlation. In addition, these codes have a number of drawbacks [Gla96]:

- The codes do not have a single, narrow autocorrelation peak. In fact the non-peak auto-correlation is dependent on the actual code word, but it's never zero.
- The spreading is not over the whole bandwidth, but over a number of discrete frequency components. This may be seen in figure 4.1, where the amplitude spectrum of one Walsh sequence of period 64 and an m -sequence of length 63 is compared. See section 4.2 for more information about m -sequences.
- The partial cross-correlation is not always identically zero making the family unsuitable for asynchronous application.

Despite of these drawbacks the Walsh-Hadamard sequences are used in IS-95 system for channel separation. In this case a separate PN sequence is used for spreading (an m -sequence).

4.2 m -sequences

The maximum-length shift-register sequence, or m -sequence for short, is probably the most widely known PN sequence. It has a length of $n = 2^m - 1$ bits and is generated by

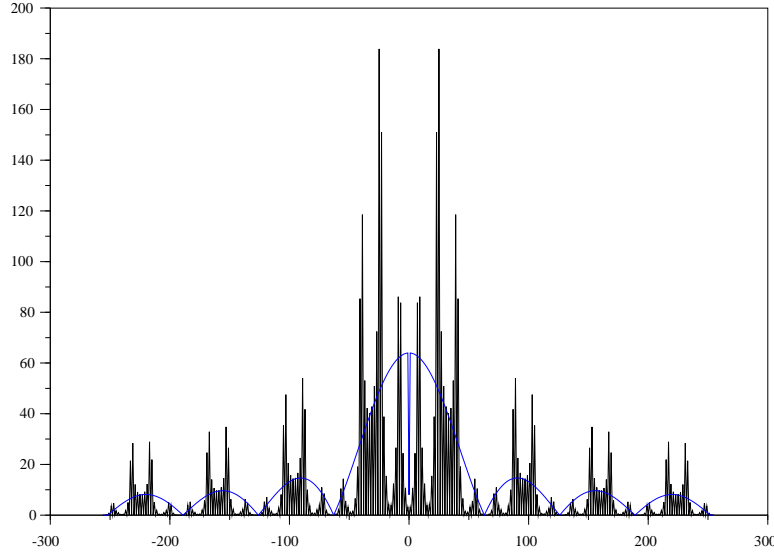


Figure 2: Amplitude spectrums of m - and Walsh sequences

an m -stage shift register with linear feedback as shown in figure 4.2. Each clock time the register shifts all contents to the left. The rightmost element of the register is updated as a linear combination of the register content, that is according the recursive formula

$$s(t + m) = \sum_{i=0}^{m-1} f_i s_{t+i} \text{ for all } t. \quad (6)$$

The coefficients f_i are called the connection variables. If we set $f_m = 1$ we may define the *characteristic polynomial* of the sequence $\{s(t)\}$ as $f(x) = \sum_{i=0}^m f_i x^i$. Not all possible connection parameter sets produce a maximal sequence. One can derive the following necessary condition for the maximality (See [Vit95]):

Theorem 1 *The sequence $\{s(t)\}$ is maximal if the characteristic polynomial of the sequence is irreducible (not factorable).*

The m -sequences have the balance, run and shift properties. The periodic (even) auto-correlation function of an m -sequence of length n is

$$\theta_u(j) = \begin{cases} n & \text{if } j = 0, \\ -1 & \text{if } 1 \leq j \leq n - 1. \end{cases} \quad (7)$$

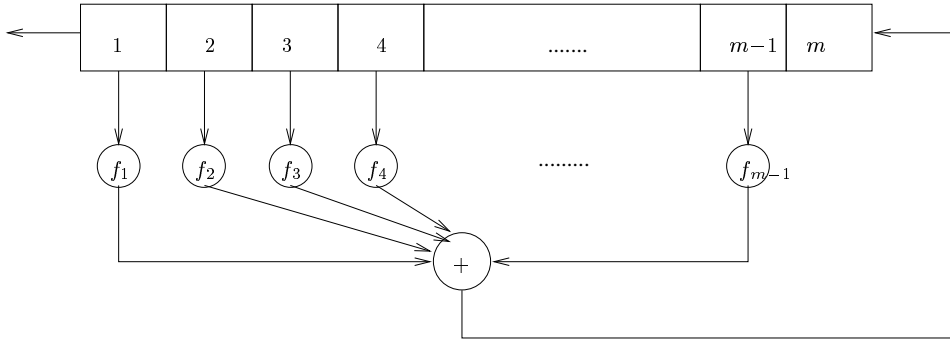


Figure 3: An m -stage shift register with linear feedback

The pseudo-noise properties as well as the autocorrelation formula can be proved directly, but they also follow naturally from some general results of the theory of finite fields, when the m -sequence is considered as a trace code. For more details see [Hel96].

The m -sequences are almost ideal when viewed in terms of their autocorrelation function. In CDMA, however, the cross-correlation properties of the sequence play a central role as seen in section 5. Unfortunately the periodic (or even) cross-correlation between any pair of m -sequences of the same period can have relatively high peaks ([Pro92]) that are undesirable in CDMA. It is of course possible to select a small subset of m -sequences that have smaller cross-correlation peak values, but the size of the set is usually too small for CDMA applications. One can use m -sequences to derive PN sequence families with better cross-correlation properties. These sequences are the subject of the next section.

4.3 Sequences with low cross-correlation

Even though the cross-correlation of m -sequences in general may have large peaks, Gold and Kasami have proved ([Gol68], [Kas66]) that certain pairs of m -sequences of length n exhibit a convenient three-valued cross-correlation function with values $\{-1, -t(m), t(m) - 2\}$, where

$$t(m) = \begin{cases} 2^{(m+1)/2} + 1 & (\text{odd } m) \\ 2^{(m+2)/2} + 1 & (\text{even } m) \end{cases} \quad (8)$$

Such sequences are called *preferred sequences*. These sequences may be utilised to produce several well known families of binary sequences with good cross-correlation properties. These families (called Quadratic Form Sequences in [Hel96]) include the Gold and Kasami sequences.

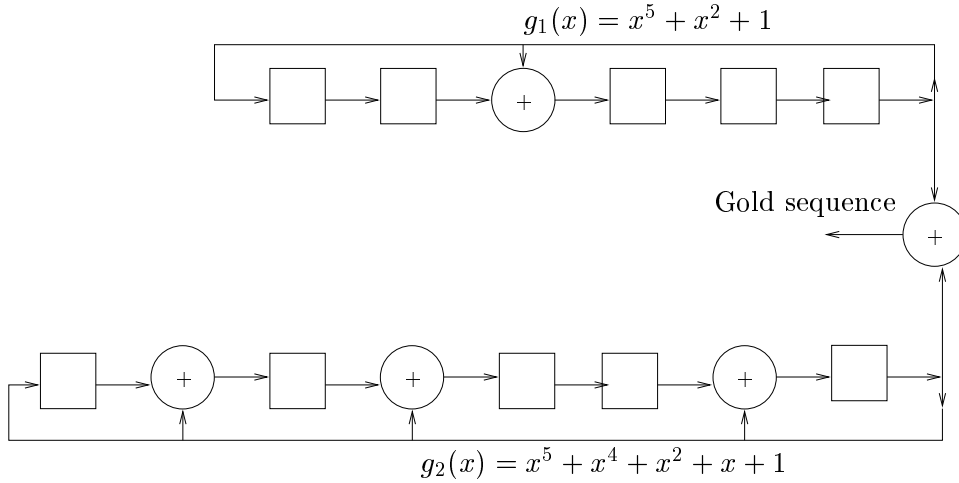


Figure 4: Generation of Gold sequences of length 31

4.3.1 Gold sequences

Given a preferred pair of sequences of period $n = 2^m - 1$, say $s = \{s(t)\}$ and $r = \{r(t)\}$, we can construct a new family of sequences by taking the modulo-2 sum of s with the n cyclicly shifted versions of r or vice versa. Thus, we obtain n new periodic sequences with period n . If we include also the original sequences s and r we obtain a family of $n + 2$ sequences. The resulting sequences are called *Gold sequences*. Figure 4.3.1 shows the shift registers for generating a preferred pair of sequences corresponding the polynomials

$$\begin{aligned} g_1(x) &= x^5 + x^2 + 1 \\ g_2(x) &= x^5 + x^4 + x^2 + x + 1. \end{aligned}$$

In this case, there are 33 different sequences, corresponding to the 33 relative phases of the two m -sequences.

The cross-correlation function for any pair of sequences in the family of Gold sequences generated with a given preferred pair was proven by Gold to be three-valued with possible values $\{-1, -t(m), t(m) - 2\}$, where $t(m)$ was defined in Eq. 8. The Gold codes from “different” Gold code groups have however bad correlation properties, even when synchronised.

With the obvious exception of s and r , the set of Gold sequences does not contain maximum-length shift-register codes of length n . Thus the auto-correlation functions are not two-valued. Similar to the cross-correlation function, the off-peak auto-correlation function is three-valued and takes the values from the set $\{-1, -t(m), t(m) - 2\}$. Hence the off-peak values of the auto-correlation function are upper-bounded by $t(m)$.

Table 4.3.1 (from [Pro92]) illustrates the differences between m - and Gold sequences. For various values of m , the table shows the values of the peak-cross-correlation between

Table 1: Cross-correlation of m - and Gold sequences

m	$n = 2^m - 1$	Number of m -sequences	θ_c	$\theta_c/\theta_u(0)$	$t(m)$	$t(m)/\theta_u(0)$
3	7	2	5	0.71	5	0.71
4	15	2	9	0.60	9	0.60
5	31	6	11	0.35	9	0.29
6	63	6	23	0.36	17	0.27
7	127	18	41	0.32	17	0.13
8	255	16	95	0.37	33	0.13
9	511	48	113	0.22	33	0.06
10	1023	60	383	0.37	65	0.06
11	2047	176	287	0.14	65	0.03
12	4095	144	1407	0.34	129	0.03

the m -sequences (θ_c) and the values of the upper bound for Gold sequences ($t(m)$). Also included are the ratios $\theta_c/\theta_u(0)$ and $t(m)/\theta_u(0)$ that illustrate how for Gold sequences the peak cross-correlation value is a large percentage of the peak auto-correlation value.

4.3.2 Kasami sequences

A procedure similar used in the generation of Gold sequences can be used to generate a smaller set of $M = 2^{m/2}$ binary sequences of period $n = 2^m - 1$, where m is even. In this procedure, we begin with an m -sequence s and we form a binary sequence r by taking every $2^{m/2} + 1$ bit of s . In other words, the sequence r is formed by decimating s by $2^{m/2} + 1$. It can be verified that the resulting r is periodic with period $2^{m/2} - 1$.

Now, by taking $n = 2^m - 1$ bits of the sequences s and r , we form a new set of sequences by adding, modulo-2, the bits from s and all $2^{m/2} - 2$ cyclic shifts of the bits from r . By including s in the set, we obtain a set of $2^{m/2}$ binary sequences of length $n = 2^m - 1$. These are called the (small set of) Kasami sequences. The autocorrelation and cross-correlation functions of these sequences take on values from the set $\{-1, -(2^{m/2} + 1), 2^{m/2} - 1\}$. Hence, the maximum cross-correlation value for any pair of sequences from the set is

$$\theta_c = 2^{m/2} + 1 \tag{9}$$

4.3.3 Bounds on Crosscorrelation of Sequences

There are several bounds on the cross-correlation of sequences known. The most commonly used are the Welch bound and the Sidelnikov bound. The Welch bound applies to

complex-valued sets with no further constraints on their sequence elements while the Sidelnikov bound only applies to complex roots-of-unity sequences with a constant amplitude. These bounds are useful when estimating the cross-correlation behaviour of different code families.

A lower bound for the peak cross-correlation value for any pair of binary sequences of period n developed by Welch [Hel96] for θ_{\max} is

$$\theta_{\max} \geq n \sqrt{\frac{M-1}{Mn-1}}, \quad (10)$$

where M is the size of the family and $\theta_{\max} = \max\{\theta_a, \theta_c\}$ (θ_a is the maximum absolute sidelobe value of the even auto-correlation function). For large values of n and M , this bound is well approximated with \sqrt{n} . Sets achieving this value are said to be asymptotically optimal with respect to Welch bound. The (small set of) Kasami sequences for instance have this property.

The Sidelnikov bound states that for any set with $m \geq n$ ([Kär96]):

$$\theta_{\max} > \sqrt{2n-2}. \quad (11)$$

Sidelnikov bound is typically tighter than Welch bound by a factor of $\sqrt{2}$ for a large set size. The main impact of these bounds is that they dictate the limits within which all code designs must lie. Thus it is not possible to independently design the correlation value and the set size, but it is necessary to allow the increase of the maximum absolute correlation value in order to increase the set size for given code length.

There also exists a bound which connects the auto-correlation and cross-correlation properties of sequences thus preventing the independent design of the desired auto-correlation and cross-correlation values. This bound is ([Kär96]):

$$\left\{ \frac{\theta_c^2}{n} \right\} + \frac{n-1}{n(M-1)} \left\{ \frac{\theta_a^2}{n} \right\} \geq 1. \quad (12)$$

5 Code Selection Criteria for CDMA Systems

In CDMA systems, the traditional method in spreading code selection for the reduction of multiple-access interference is to use the maximum absolute value of the periodic (even) cross-correlation function. The average signal-to-noise power ratio (SNR) performance of an asynchronous DS/SSMA system investigated in [Kär96] was somewhat surprisingly found to be about the same with various kinds of linear spreading code families of equal period and set size, although the corresponding maximum periodic cross-correlation parameters for those families differ considerably. Since all cross-correlation values, not just peak values, affect the system performance, it is proposed in [Kär96] that a natural design parameter for system performance measure is the mean-square cross-correlation value.

Table 2: Correlation parameters for CDMA families (adapted from [Kär96])

Family	p	θ_c	$r_{i,j}$	$\sum_{k=0}^{p-1} \theta_{i,j}^2(k)$	$\sum_{k=0}^{p-1} \hat{\theta}_{i,j}^2(k)$
Gold	31	9	1674	895	1031
m -sequence	31	9	1742	991	927
Gold	63	17	7398	3807	4479
Kasami	63	9	6982	3519	3439
m -sequence	63	23	8006	4031	4079
Gold	1023	65	2093414	1043583	1081759
Kasami	1023	33	2094766	1010815	1105951
m -sequence	1023	95	2059262	1047551	1002431

The average signal-to-noise power ratio at the output from a BPSK type asynchronous DS/SSMA receiver of the j -th user can be expressed in terms of the sum of the average interference parameters (AIP) $r_{i,j}$ of K simultaneous users, and the SNR of the AWGN channel (E_b/N_0) as follows

$$\text{SNR}_j = \left\{ \frac{N_0}{2E_b} + \frac{1}{6p^3} \sum_{i=1, i \neq j}^K r_{i,j} \right\}^{-1} \quad (13)$$

where the average interference parameter is defined as $r_{i,j} = 2\mu_{i,j}(0) + \mu_{i,j}(1)$. In the case of binary sequences the expression for $\mu_{i,j}(n)$ can be defined simply as

$$\mu_{i,j}(n) = \sum_{m=1-p}^{p-1} C_{i,j}(m)C_{i,j}(m+n) \quad (14)$$

where p is the period of the sequence. On the other hand, the average interference parameter can be expressed in form ([Kär96])

$$r_{i,j} = \sum_{k=0}^{p-1} \theta_{i,j}^2(k) + \sum_{k=0}^{p-1} \hat{\theta}_{i,j}^2(k) + \mu_{i,j}(1). \quad (15)$$

From 15 and 13 it can be predicted that the average (SNR) performance of an asynchronous, BPSK type DS/SSMA system depends strongly on the square-sums of both even and odd cross-correlation values and thus on the mean-square cross-correlation values of both even and odd functions.

Various unnormalised cross-correlation parameters are presented in table 5 (from [Kär96]) for some well-known linear code families of a large range of periods. Each pair of sequences was selected randomly from the code families studied. It is clear that some other choice of sequences would give slightly different numerical values. In any case, some conclusion can be drawn from the table. The most important observation is that both the average interface parameter values and the square sums of the even CCF values are approximately equal for all well-known linear code families of equal period, even though the absolute maximum values of the even CCFs (θ_c) differ considerably. This can be seen for instance when comparing the values of average interface parameter and the square sums of the even CCF values with θ_c for codes of period 1023. The close average interface parameter values for different families of same length implies that the SNR performance of asynchronous DS/SSMA systems for those code families is approximately equal.

6 Conclusions

Traditionally in the spreading sequence analysis and design the role of the absolute maximum cross-correlation value has been in a major position. If one considers the average SNR performance presented in section 5 one can argue that in fact a mean square-sum of the cross-correlation function is a more accurate design parameter. Also the small performance differences between the different code families imply that the careful selection of the spreading code is not very essential from the system performance viewpoint.

Of course, the average SNR performance is not the only nor the best design goal. Although this kind of analysis is a standard approach in the analysis of the communication systems, the simple channel model used and the use of SNR instead of the bit error probability (BEP) impact the accuracy of the results. According to [Kär96] this inaccuracy should not affect the conclusions in a major way, and this seems to be a feasible assumption. One should also be aware that although the multiuser interference reduction is usually a major design goal, there may be other requirements, such as the maximum code acquisition time or the multipath fading performance that have impact on the code design.

References

- [Gol68] R. Gold, "Maximal Recursive Sequences with 3-valued Recursive Cross Correlation Functions," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 154-156, January 1968.
- [Kas66] T. Kasami, "Weight Distribution Formula for Some Class of Cyclic Codes," *Tech. Report No. R-285*, Univ. of Illinois, 1966.
- [Vit95] A. J. Viterbi, "CDMA: Principles of Spread Spectrum Communication," *Addison-Wesley*, 1995.

- [Pro92] J. G. Proakis, "Digital Communications (3rd Ed.)," *McGraw-Hill, Inc*, 1995.
- [Kär96] K. Kärkkäinen, "Code Families and Their Performance Measures for CDMA and Military Spread-Spectrum Systems," *Acta Univ. Oul. C 89*, 1996.
- [Hel96] T. Hellesteth, P. Vijay Kumar, "Sequences with Low Correlation," *preprint*, 1996.
- [Bha81] V. K. Bhargava, D. Haccoun, R. Matyas, P.P. Nuspl, "Digital Communications by Satellite," *John Wiley & Sons, Inc.*, 1981.
- [Gla96] J. Glas, "The Principles of Spread Spectrum Communication," <http://cas.et.tudefl.nl/glas/ssc/techn/techniques.html> Last modified on August 29,1996.