Design of Synchronisation Sequences for Mobile Radio Applications Using Genetic Algorithms

R. Perry, S. Sriranganathan, D. R. Bull, A. Nix

University of Bristol, UK

Abstract

This paper discusses the application of Genetic Algorithms (GAs) to the problem of searching for a set of synchronisation codes for TDMA based mobile radio systems. The problem may be formulated as a constrained optimisation problem and is transformed into an unconstrained one by including a penalty term within the fitness function. The quality of the sequences obtained, using the genetic algorithms, are compared with some previously published results. The paper discusses the sensitivity of the genetic algorithm's behaviour to the formulation of the fitness function and algorithm parameters. Using results from polyphase codes it is also shown how sets of sequences for different modulation formats can be obtained from either a BPSK or QPSK code.

I. INTRODUCTION

All current digital TDMA systems incorporate a synchronisation sequence to allow the receiver to compensate for unequal arrival times between frames. The synchronisation sequence is a relatively short sequence which is incorporated in each transmitted packet to allow variable time delays introduced during transmission to be compensated for at the receiver. The GSM system incorporates a 26 bit sequence in the centre of the transmitted packet (fig 1).

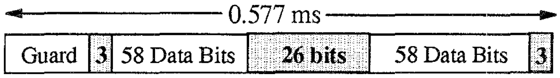

Figure 1: The GSM Transmitted Frame Structure

In many systems the existence of multipath propagation results in several replicas of the transmitted signal being received, each with a different time delay and relative attenuation. This multipath propagation is conveniently modelled as a time varying linear filter (a channel impulse response (CIR)) with symbol spaced taps, or fraction thereof, depending on the sampling rate employed at the receiver (fig 2).

The synchronisation sequence in these cases can also be used to form an estimate of this CIR. The CIR estimate is formed by cross-correlating the received signal with a complex conjugated version of the original sequence. The auto-correlation function of the synchronisation sequence should therefore have a single peak at zero time shift and be zero-valued for all other time lags of interest.

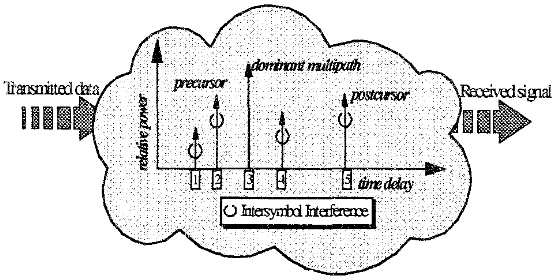

Figure 2: Model of Channel Distortion

The CIR estimate can be used for equalisation using MLSE and can also be used to initialise a channel matched filter ahead of an adaptive equaliser. In some systems (e.g. GSM) the synchronisation sequence is also used to carry out equaliser training.

To support multiple users, a set of synchronisation sequences are required. If only a single sequence was used, then for two interfering signals arriving at a receiver at a similar time, there would be no way to distinguish the contribution of each to the received signal. This problem can be avoided by choosing a *set* of sequences which are as little correlated as possible, i.e. the sequences should exhibit low cross-correlation values. Distinct training sequences are then allocated to channels using the same frequencies in cells which are close enough to interfere with one another (fig 3) [8].

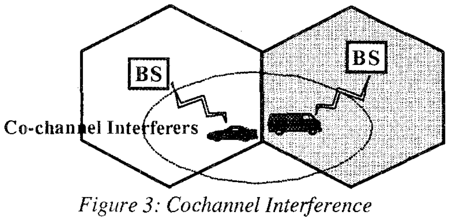

Figure 3: Cochannel Interference

In this paper, the problem of obtaining a *set* of synchronisation sequences is considered. In previous work [1,2], attention has focused on the selection of sequences on the basis of their auto-correlation

Genetic Algorithms in Engineering Systems: Innovations and Applications
12-14 September 1995, Conference Publication No. 414, © IEE, 1995

properties and for short sequence lengths an exhaustive search is satisfactory. In order to search for sets of codes however, the time required to carry out an exhaustive search becomes prohibitive. Genetic algorithms have therefore been investigated as a means of performing the search. The choice of fitness function and the effect of both sequence length and set size on the search performance was studied. Using some results for polyphase codes it is shown how MSK and π/4-DQPSK sequences can be derived from a BPSK or QPSK sequence respectively. This avoids searching for independent sequences for these modulation formats. The method of channel estimation is reviewed first.

II. CHANNEL ESTIMATION

There are two principle methods which can be used to obtain a channel estimate. These are *(i)* using the *least sum of square errors* (LSSE) [1] and *(ii)* using *cross-correlation* [4]. It is assumed here that the channels to be estimated consist of symbol spaced multipaths. The results can however be extended to fractionally spaced multipaths by viewing the channel as a composition of interleaved symbol-spaced models.

For the correlation method, the channel estimate is formed by cross-correlating the received sampled signal with a replica of a part of the original sequence. The *circular correlation* of the synchronisation sequence is defined as [2]

$$\rho_0 = \sum_{k=0}^{k=N-1} a_k a_k^* \qquad (1a)$$

$$\rho_j = \sum_{k=0}^{N-1-j} a_k a_{k+j}^* + \sum_{k=N-j}^{N-1} a_k a_{k+j-N}^* \qquad |j| < L \quad (1b)$$

where $a_k = \exp(i\theta_k)$ is the k*th* transmitted symbol. The synchronisation sequence is therefore formed from a fixed length code, repeated at the ends for a given number of symbols. Given a sequence length N with L symbols appended to each end, the total sequence length is N+2L. We require $\rho_j = 0$ for $j = -L..L \;\; j \neq 0$ and $\rho_0 = 1$. The quality of estimate is determined by N and the number of significant paths in the CIR is assumed to be 2L or less.

In the LSSE approach, symbols are only appended to the start of the sequence. The optimisation metric is the minimisation of the sum of squared errors formed from the difference in the estimation of the received signal (output from tapped delay line filter model) with the actual received sequence i.e.

$$J(\mathbf{h}) = \sum_{k=0}^{N-1} \left| r_k - r_k^{est} \right|^2 \qquad (2)$$

where r_k and r_k^{est} are the received signal and estimate of that signal and **h** is the channel impulse response to be estimated, covering up to L symbol periods. The received signal is modelled as [1]

$$r_k = \sum_{i=0}^{L} h_i p_{k-i} + n_k \qquad (3)$$

the set $\{p_{-L}, \dots p_0, \dots\dots, p_{N-1}\}$ is the complete training sequence and n_k is noise modelled as discrete additive white Gaussian noise. Minimising (2) leads to the set of equations [1]

$$\mathbf{P}\,\mathbf{h}^{est} = \mathbf{z} \qquad (4)$$

where $\mathrm{P} = \sum_{k=L}^{N-1-L} \mathbf{p}_k \mathbf{p}_k^H \;\; ; \;\; \mathbf{z} = \sum_{\mathrm{i}=0}^{\mathrm{N}-1} \mathrm{r_i} \mathbf{p}_{\mathrm{i}}^*$

and $\mathbf{p}_k = \left[a_k \quad a_{k-1} \quad .. \quad a_{k-L} \right]$

The pre-computed inverse of the correlation matrix is stored at the receiver. If the inverse has a simple or regular structure, the processing required to obtain the channel estimate will be reduced. A *perfect* sequence is therefore defined as one which has an associated diagonal correlation matrix. For such sequences, the channel estimate can be obtained without matrix calculations. Although it is possible to find several perfect sequences for small L (L≤4), for larger L none have been found. In this paper, sequences suitable for carrying out synchronisation using the method of cross correlation are discussed. The structure of the correlation matrix P, is important when considering the training behaviour of an equaliser and was discussed in a previous paper [9].

III. POLYPHASE CODES

A polyphase code is a sequence of symbols with phases θ given by $\{\theta\} = \{pM\pi/N\}$ where p is some integer and M and N are relatively prime. In [2] the construction of complex codes with zero valued circular correlation functions for all non-zero lags was described. In the design of sequences here, the phases of the transmitted alphabet are restricted to a finite number. However, it is shown that for M-PSK modulation formats, sequences can be found with zero auto-correlation values for a limited number of lags. The phase invariance property of polyphase codes is also used to show how sequences for one modulation scheme can be transformed into a different format with identical correlation properties. E.g. The transmitted phases for 8-PSK are selected as $\theta_k = \pi k^2/4$. Substituting in (1b) gives

$$\rho_j = \sum_{k=0}^{N-1-j} \exp\left(\frac{\pi i}{4} \left(k^2 - (k+j)^2 \right) \right) + \sum_{k=N-j}^{N-1} \exp\left(\frac{\pi i}{4} \left(k^2 - (k+j-N)^2 \right) \right) \qquad (5)$$

Expanding the second term

$$\exp\left(\frac{\pi i}{4} \left(k^2 - (k+j)^2 + 2(k+j)N - N^2 \right) \right) = \exp\left(\frac{\pi i}{4} \left(k^2 - (k+j)^2 \right) \right) \exp\left(\frac{\pi i}{4} \left(2(k+j)N - N^2 \right) \right) \qquad (6)$$

Choosing $N^2/4$ even then

$$\exp\left(\frac{\pi i}{4}\left(2(k+j)N-N^2\right)\right)=1 \qquad (7)$$

and therefore

$$\rho_j=\sum_{k=0}^{N-1}\exp\left(\frac{\pi i}{4}\left(k^2-(k+j)^2\right)\right)=\exp\left(\frac{-\pi ij^2}{4}\right)\sum_{k=0}^{N-1}\exp\left(\frac{-\pi ijk}{2}\right) \qquad (8)$$

$$\rho_j=\exp\left(\frac{-\pi ij^2}{4}\right)S_{N-1}(j) \qquad (9)$$

where $$S_\alpha(j)=\sum_{k=0}^{\alpha}\left(\exp\left(\frac{-\pi ij}{2}\right)\right)^k$$

$\Rightarrow S_\alpha(j)=0 \quad j=1,2,3$ provided α is a multiple of four. The sequence obtained can also be used for π/4-DQPSK since the phases are located alternately on the two phase shifted QPSK constellations (fig 4).

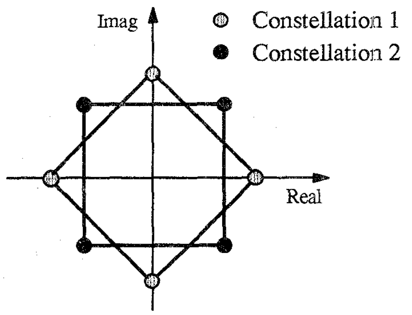

Figure 4: π /4-DQPSK Signal Constellation

It is shown next that the modified sequence in (10)

$$b_k=a_k\exp\left(\frac{i2\pi qk}{N}\right) \qquad (10)$$

will have identical correlation properties to the sequence $\{a_k\}$ with some arbitrary phase shift (phase invariance property). Substituting (10) in (1b)

$$\rho_j=\sum_{k=0}^{N-j-1}a_k a_{k+j}^*\exp\left(\frac{i2\pi q}{N}(k-k-j)\right)$$
$$+\sum_{k=N-j}^{N-1}a_k a_{k+j-N}^*\exp\left(\frac{i2\pi q}{N}(k-k-j+N)\right)$$

$$\rho_j=\exp\left(\frac{-i2\pi qj}{N}\right)\left(\sum_{k=0}^{N-j-1}a_k a_{k+j}^*+\sum_{k=N-j}^{N-1}a_k a_{k+j-N}^*\right)=0$$
$$j=1,2,3 \qquad (11)$$

This property of polyphase codes can be exploited to transform BPSK and QPSK sequences into MSK and π /4-DQPSK sequences with certain restrictions on the sequence length. The MSK sequence is ordinarily obtained by linearly increasing the phase of the transmitted signal according to the sign of the current bit. The phase at time instant n is given by

$$\phi_n=\frac{\pi}{2}\left(\sum_{i=0}^{n-1}a_i\right)+\frac{\pi}{2}a_n \qquad (12)$$

The *in-phase* and *quadrature* data streams are given by the sine and cosine of the phase ϕ_n. Consequently the two parallel streams are alternately either ±1 or 0. GMSK modulation is similar to the MSK scheme except that the data stream is filtered with a Gaussian low-pass filter. The effect of this is to limit the phase increase each bit period. As the time bandwidth product of the filter is reduced, the severity of the filtering reduces the eye opening. (fig 5).

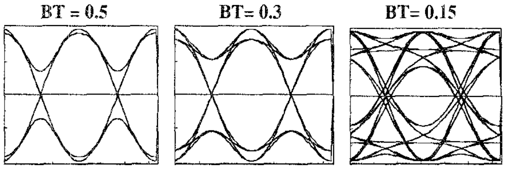

Figure 5: Eye Diagrams for Coherent GMSK

The MSK sequence is derived from a BPSK sequence according to the following mapping

$$b_k=\exp\left(\frac{\pi k}{2}\right)a_k \qquad (13)$$

The new sequence is detected (as OQPSK [4]) and with appropriate toggling of the signs of alternate pairs of data bits, the resulting sequence is a transformed BPSK sequence which when MSK modulated has the desired correlation properties (provided the sequence length is a multiple of 4).

A π/4-DQPSK sequence is obtained through the mapping

$$b_k=\exp\left(\frac{\pi k}{4}\right)a_k \qquad (14)$$

applied to a QPSK synchronisation sequence with the restriction that the sequence length is a multiple of 8. By linearly increasing the phase, the symbols will alternately lie on one of the two phase-shifted constellations, as in fig 4. The sequence is then detected as if originally transmitted as π/4-DQPSK and the detected sequence then gives the correct starting code.

The ability to derive sequence codes from one modulation format is useful because (i) time consuming searches can be avoided and (ii) sequence searching using GAs may be adversely affected by certain modulation formats.

IV. SEQUENCE SEARCH USING GENETIC ALGORITHMS

Due to the time required for an exhaustive search, genetic algorithms were identified as a means of performing the search for sets of codes [7]. For the problem considered here, the sequences are naturally represented by a set of decimal numbers which are each transformed to a binary string and then joined to form the chromosome structure. To obtain the sequence quality, the strings are mapped onto the desired modulation format before evaluation of the fitness function. The sets of codes are denoted as (N,L,M) where N is the fundamental length of the sequence, L is the number of symbols appended to each end and M refers to the number of sequences used to form the set. Ideal sets of codes are defined as sequences with zero auto-correlation values at all non-zero lags less than or equal to L. Given that no code may be used more than once, the total number of solution sets is given by

$$2^N \times 2^N - 1 \times 2^N - 2 \times .....\times 2^N - M + 1$$

E.g. for a code set (16,5,4) there are 1.84×10^{19} possible combinations of codes. For the design of synchronisation sequences, two properties which must be optimised have been identified, i.e. the auto- and cross-correlation function values. In order to obtain an accurate channel estimate a sequence with a zero-valued auto-correlation function for non-zero lags is desired. Since for certain systems this is a requirement, e.g. GSM, then the search procedure is complicated by the need to incorporate this as a *constraint*. The problem is then one of constrained optimisation, i.e. minimising the correlation between sequences while ensuring that their auto-correlation properties are ideal for a fixed time window. In general, there are no further restrictions on any member of the population space, except for codes which are simply cyclic shifts of one another. In practice (without guarding against such an occurrence) solution sets only rarely contained cyclically shifted codes.

It should be noted that the auto-correlation cost term is dependent only on individual genes whereas the cross-correlation cost term for each gene is highly dependent on the values of the other genes. Since GAs generally require low epistasis in order to function well, it was thought that this could potentially inhibit the GAs effectiveness [6].

The simplest approach to deal with a constraint is to employ a penalty function within the evaluation routine [10]. This has the advantage of simplicity, but is also potentially problematic since the penalty may prevent the acceptance of good building blocks contained within illegal individuals which are assigned relatively poor fitness values. Given that the majority of individuals contained in the search space are illegal, there is a real danger that the first feasible solution, encountered in the search, will ultimately lead to convergence without any effective optimisation of the cross correlation properties.

The fitness function was initially chosen as

$$fitness = \sum_{l=0}^{M} \sum_{m=l+1}^{M} \left(\sum_{n=-L}^{L} \left| R_{lm}(n) \right|^2 \right)^{1/2} + W \sum_{l=0}^{M} \left(\sum_{n=0}^{L} \left| R_{ll}(n) \right|^2 \right)^{1/2} \qquad (15)$$

where $R_{lm}(n)$ is the circular correlation between the *lth* and *mth* sequences and W is a weighting factor used to impose a penalty for non-ideal auto-correlation properties of the sequences (*cost function 1*). The value of W was varied in order to determine the effect of the penalty function on the performance of the GAs. It should be noted that there are more cross-correlation terms than auto-correlation terms and therefore without a weighting factor, the cross-correlation cost term will dominate the cost function.

Sequences using BPSK, QPSK and MSK (GMSK) modulation formats were obtained using the GAs with ideal autocorrelation properties. For the GMSK modulation format, an MSK approximation to the transmitted GMSK sequence is used at the receiver [3,4]. This reduces the complexity of the channel impulse response estimation i.e. the stored reference sequence is alternately ±1 or ±j. However, the autocorrelation properties using this approximation are non-ideal due to the intersymbol interference introduced by the GMSK modulation. The best results obtained for BPSK codes satisfying the autocorrelation results are shown in Table 1.

Code (N,L,M)	Sequence Nos. (Decimal)
(16,5,4)	36680, 12976, 3729, 47172
(16,5,6)	5099, 10463, 47687, 52559, 6488, 3530
(16,5,8)	41251, 4833, 57622, 18242, 35060, 62600, 62005, 5036
(12,5,4)	2593, 3463, 209, 751
(12,5,6)	1587, 2126, 2237, 740, 3672, 1672
(12,5,8)	2615, 1217, 101, 751, 2270, 740, 1969, 353

Table 1: Effect of set size on BPSK codes

The quality of the sequences is independent of M, i.e. each subset of four codes from the set of 6 or 8 codes was found to have a similar cost function to the set of 4. A crossover rate of 0.6 was used, the mutation rate was varied from 0.001 to 0.005 and the population size was varied between 50 and 70.

The effect of the weighting factor W on the convergence behaviour of the best and average cost values of the population when searching for the BPSK code set (16,5,4) is shown in figs 6a and 6b. The stepped convergence curve in 6a is due to the large weighting factor applied to the autocorrelation term, i.e. each improvement in this term produced a large drop in the total cost. This effect is less pronounced for smaller W as shown in 6b for W=1.0; note however, that the autocorrelation term was non-zero for the best code set obtained in this case.

It was found that the convergence of the GAs was sensitive to the choice of W, although little degradation in the final minimum cost was observed for W≥10, i.e. the use of a harsh penalty function did not significantly degrade performance, but some form of penalty was necessary to avoid convergence to a non-feasible solution.

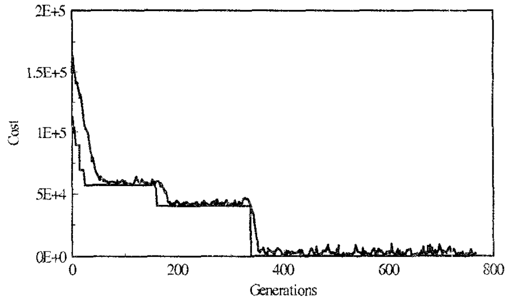

Figure 6a: Convergence Behaviour of the Best and Average Values of the Cost function W=10000

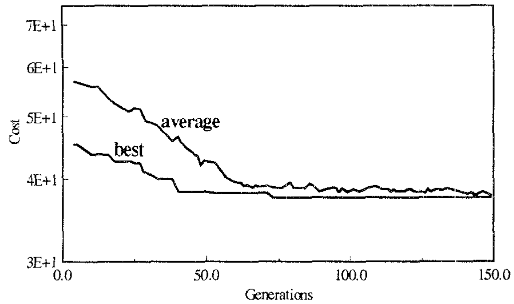

Figure 6b: Convergence Behaviour of the Best and Average Values of the Cost function W=1.0

The best convergence curves for the different values of W are shown in fig.7. The stepped convergence here is emphasised by the choice of a logarithmic scale.

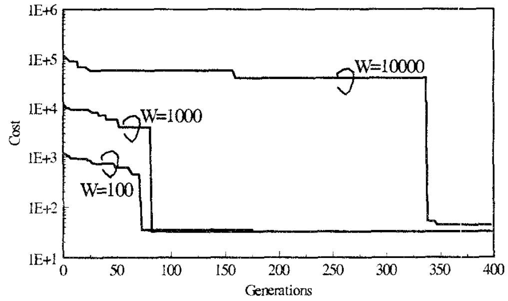

Figure 7: Effect of Variation of Penalty Function on convergence to BPSK (16,5,4) code set

For sequence searching it was found that the success of the GA was highly dependent on the length of the sequence and the number of sequences sought. For this reason sequences of lengths 14 and 10 could not be obtained.

The GAs were also used to obtain sets of codes, neglecting their cross-correlation properties. In this case ideal sets of codes could be easily obtained for sequences of length 12 and 16. However, for sequence lengths of 10 and 14 no ideal sets have, so far, been obtained. It was observed that for the cases of non-convergence the initial population had a good initial best individual, which consequently led to premature convergence. For the cases that did converge, the initial *best* value was inferior.

In [11] it was observed that penalties which are functions of the distance from feasibility are better performers than those which are merely functions of the number of violated constraints. In (15) the penalty is a function of the distance of the individual from feasibility. However, by viewing each non-zero value for each lag of the autocorrelation function as a broken constraint, a second cost function, formed by summing the number of broken constraints (weighted by some penalty factor) was used in the evaluation routine (*cost function 2*). It is noted that the number of broken constraints as defined here, could also be interpreted as a looser measure of distance from feasibility. For the sequences that did converge, this was found to give more rapid convergence with similar minimum cost. The best convergence curves obtained are shown in fig 8.

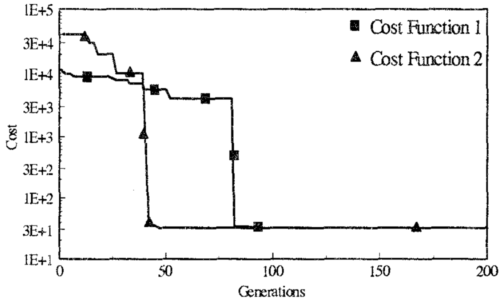

Figure 8: Effect of Cost Function on Convergence to BPSK set (16,5,4)

Sets of codes for QPSK sequences (satisfying the autocorrelation constraints) have also been obtained using cost function 1 as for the BPSK codes (Table 2). The results are shown for codes of length 16 bits (or 8 symbols). The binary codes are mapped to 8 symbol sequences to which 3 symbols are repeated at both ends.

Sequence Set	Decimal Codes
(8,3,4)	40737, 56028, 11631, 33664
(8,3,6)	36367, 60270, 24806, 9184, 24800, 25122
(8,3,8)	44779, 12191, 15923, 23582, 7965, 6879, 8740, 41952
(8,3,10)	57573, 7644, 30578, 36751, 4375, 14398, 31676, 9377, 52746, 11112

Table 2: QPSK Codes

A set of eight MSK codes (suitable for the GSM system [4]) were also obtained (N=16 L=5). The population size was set to 45, a crossover rate of 0.9 and mutation rate of 0.003 were used, 60000 trials were required for convergence. The cross-correlation properties for three of the sequences are shown in fig 9. The codes obtained were (5356, 63304, 63913, 55869, 2231, 11793, 22687, 1622).

Using the cost function in (15) the sequences obtained here were found to have superior cross-correlation properties to those obtained in [4]. The average values for the cross correlation terms from [4] and from the

GAs are compared in fig 10 together with the best code set obtained from a random search over 10^6 iterations.

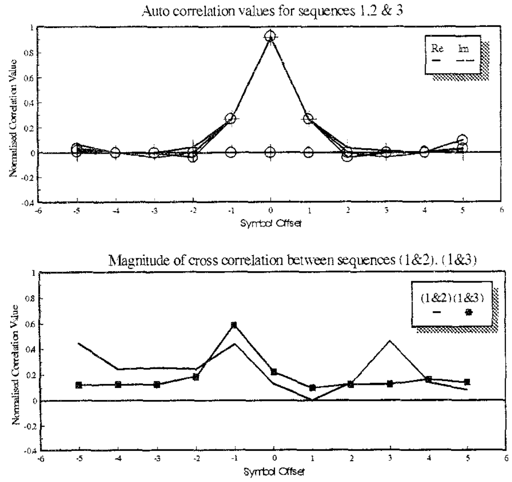

Figure 9: Cross Correlation properties for the MSK Codes

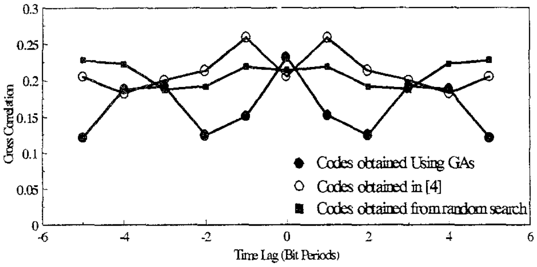

Figure 10: Average Values of the Cross Correlation terms for the sequences obtained using GAs and in [4]

The random search over the entire population space performed marginally worse in terms of the cross-correlation properties, but was unable to satisfy the constraints on the autocorrelation terms. Clearly forming a smaller search space of codes possessing ideal autocorrelation properties would allow these constraints to be satisfied.

VI. CONCLUSIONS

The design of synchronisation sequences for mobile radio systems has been described. Genetic algorithms have been used to carry out searches for optimal families of codes. In order to carry out this search, a penalty function was used to allow constraints on the autocorrelation values of the sequences to be included. It was shown that the severity of the penalty did not significantly degrade performance. The convergence of the GAs was however, found to be sensitive to the penalty term. A second penalty function based on the number of constraints broken was shown to provide superior convergence rates. For certain code lengths, sequence sets could not be found. Overcoming these limitations is the topic of on-going work. Codes for the GSM system were obtained with superior cross correlation properties to the codes in [4] and those obtained by random search.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the members of the Centre for Communications Research, University of Bristol. We also wish to thank EPSRC and Hewlett Packard Laboratories for their support of this work. In particular we would like to thank Mu Li for helpful discussions.

REFERENCES

[1] S. N. Crozier, D. D. Falconer, S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Procs. Part F vol. 138 No. 4 August 1991 pp. 371-378*

[2] D. C. Chu "Polyphase Codes with Good Periodic Correlation Properties," *IEEE Trans. Information Theory, vol. 20 pp. 531-532 July 1972*

[3] G. Benelli, A. Garzelli, F. Salvi "Simplified Viterbi Processors for the GSM Pan-European Cellular Communications System," *IEEE Trans Veh. Tech. vol. 43 No. 4 November 1994*

[4] G. D'Aria, R. Piermarini, V. Zingarelli "Fast Adaptive Equalizers for Narrow-Band TDMA Mobile Radio," *IEEE Trans Veh. Tech. vol. 40 pp 392-403 May 1991*

[5] E. Kreyszig "Advanced Engineering Mathematics," *John Wiley & Sons 1988 5th Edition pp 1035-1036*

[6] D. Beasley, D. R. Bull, R. R. Martin "An overview of genetic algorithms: Part 1, Fundamentals," *University Computing, 1993, 15, pp 58-69*

[7] Grefenstette, J. J. GENESIS: a system for using genetic search procedures. In: *Proceedings of the 1984 Conference on Intelligent Systems and Machines, pp. 161-165. (1984)*

[8] M. Mouly, M-B. Pauet "The GSM System for Mobile Communications," *Europe Media Duplication S. A.1993*

[9] R. Perry, S. , D. Bull, A. Nix "Design of Synchronisation Sequences for Mobile Radio Channel Impulse Response Estimation," *2nd International Workshop on Mobile Multimedia Communications, 11-15th April 1995.*

[10] Z. Michalewicz, C. Z. Janikow, "Handling Constraints in Genetic Algorithms," *Proc. 4th International Conference Genetic Algorithms, pp 151-157*

[11] J. T. Richardson, M. R. Palmer, G. Liepins, M. Hilliard "Some Guidelines for Genetic Algorithms with Penalty Functions," Proc. 3rd *International Conference Genetic Algorithms, pp 191-197*