Optimal sequences for channel estimation using discrete Fourier transform techniques

C. Tellambura, M. G. Parker, Y. J. Guo,
S. J. Shepherd and S. K. Barton

Abstract— **This paper addresses the problem of selecting the optimum training sequence for channel estimation in communication systems over time-dispersive channels. By processing in the frequency domain, a new explicit form of search criterion is found, the gain loss factor (GLF), which minimises the variance of the estimation error and is easy to compute. Theoretical upper and lower bounds on the GLF are derived. An efficient directed search strategy and optimal sequences up to length 42 are given. These sequences are optimal only for frequency domain estimation, not for time domain estimation.**

I. Introduction

For burst-transmission digital communication systems, channel estimation (CE) is required for maximum likelihood sequence estimation receivers [1] and non-iterative equalizers [2]. A typical data burst consists of several blocks of user data and a pre-determined training sequence (TS) which is used to estimate the channel impulse response (CIR). This paper addresses the problem of selecting optimal CE sequences for frequency domain processing.

CE can be done using a Wiener filter or the discrete Fourier transform (DFT). In general, to estimate L channel taps with a length N CE sequence, the Wiener filter needs to store the complex filter coefficients (which can be pre-computed given the autocorrelation function of the CE sequence) and to compute complex multiplications. Similarly, the DFT method involves sending a CE sequence whose DFT is stored at the receiver. Each bin of the N-point DFT of the received sequence is divided by the corresponding bin of the stored DFT to give an N-point vector, the inverse DFT (IDFT) of which gives the channel estimates. Table 1 compares the two methods, showing memory requirements and complexity.

Sequences with impulse-like correlation functions are suitable for CE (and other applications [3]), and the problem of finding such sequences has received a great deal of attention in the past [4]. Specifically, [5-9] consider CE given a known training sequence. Following the least-squares (LS) philosophy, [5] presents algorithms for optimal unbiased CE with aperiodic spread spectrum signals for white or non-white noise. Optimum unbiased CE given white noise is considered in [6] following a maximum-likelihood (ML) approach. For fast start-up CE, optimal training sequences of two-level, three-level, and four-level symbols (non-constant amplitude) are reported for lengths up to sixteen. ML and matched-filter CE for uplink transmission in code division multiple access systems is considered in [7], which also gives a simplified channel estimator using the DFT where the total number of unknown channel taps is equal to the length of the basic sequence. Periodic CE sequences have been studied in [8], where a construction for some poly-phase (but not binary) perfect autocorrelation sequences is given.

In [9], LS filtering (Wiener filtering) for CE is considered and optimal binary sequences up to length 22 are found by exhaustive computer search. The search criterion is:

$$F = \mathrm{tr}(P^{-1}) \quad (\mathrm{dB}) \tag{1}$$

where $\mathrm{tr}(\cdot)$ is the trace of a matrix and P is the $L \times L$ correlation matrix of the training sequence. The resulting sequences offer the best possible signal-to-estimation-error ratio (SER) at the output of the channel estimator.

This paper takes an approach similar to that of [9], but, importantly, all processing occurs in the frequency domain. This leads to an explicit expression for the search criterion, termed the gain loss factor (GLF), which is only a function of the power spectrum of the training sequence. A CE sequence is defined as *optimal* if it minimises the GLF for the frequency domain approach. Equivalently, an optimal sequence maximises the output SER. However, sequences given in this paper are *not optimal* for time domain estimation. In fact, they have marginally worse performance (Section 5). This paper provides

- a frequency domain approach (i.e., using the DFT)
- bounds on the GLF

The Work described in this paper was carried out at the University of Bradford, supported by the UK DTI/EPSRC LINK project: PC2011 "High Throughput Radio Modem" under EPSRC grant reference GR/K00318 in collaboration with Symbionics Networks Limited, and by EPSRC under grant reference GR/K48914.

C. Tellambura was with the Telecommunications Research Group, Department of Electronic and Electrical Engineering, University of Bradford, West Yorkshire, BD7 1DP, UK. He is now with the School of Computer Science and Software Engineering, Monash University, Clayton, Victoria 3168, Australia. Email: chintha@dgs.monash.edu.au.

M. G. Parker was with the Telecommunications Research Group, Department of Electronic and Electrical Engineering, University of Bradford, West Yorkshire, BD7 1DP, UK. He is now with the Code Theory Group, Inst. for Informatikk, Høyteknologisenteret i Bergen, University of Bergen, Bergen 5020, Norway. Email: matthew@ii.uib.no.

Y. J. Guo was with the Telecommunications Research Group, Department of Electronic and Electrical Engineering, University of Bradford, West Yorkshire, BD7 1DP, UK. He is now with Fujitsu Europe Telecom R&D Center Ltd. (FTRC), 2 Longwalk Rd, Stockley Park, Uxbridge, UB11 1AB. Email:Y.Guo@fujitsu.co.uk.

S. J. Shepherd is with the Telecommunications Research Centre, School of Electronics and Digital Media, University of Bradford, Bradford BD7 1DP, UK. E-mail: S.J.Shepherd@bradford.ac.uk.

S. K. Barton was with the Telecommunications Research Group, Department of Electronic and Electrical Engineering, University of Bradford, West Yorkshire, BD7 1DP, UK. He is now with the Institute of Integrated Information Systems, School of Electronic and Electrical Engineering, University of Leeds Leeds, LS2 9JT, England. E-mail: S.K.Barton@elec-eng.leeds.ac.uk

- an optimal sequence search procedure for periodic and non-periodic cases.

To find optimal binary sequences of length N for estimating L channel tap coefficients, an exhaustive search may be done over 2^N possible sequences. If one follows [9], this involves computing F (inverse of an $L \times L$ matrix) for all sequences and selecting the best one, whereas in this paper, computing the GLF requires an N-point DFT. As seen later, the GLF is invariant under several operations. Since these operations partition the 2^N space into equivalence classes, the computation of the GLF for just one representative from each class is sufficient. This approach, in conjunction with an incremental DFT computation (see appendix), significantly reduces (Table 5) computational search complexity and enables a search up to $N = 42$.

The training sequence can be periodic or non-periodic and these two cases are treated in the following two sections. The channel output is a linear convolution, whereas DFT estimation requires cyclic convolution. Thus, the periodic and non-periodic cases require DFTs with cyclically-extended [10-11] and zero-padded inputs, respectively. Of these two, periodic sequences are more commonly used in practical systems. Also it is easier to search for optimal sequences among periodic sequences because they have a cyclic shift invariance property (I3, Section 3.A). This property ensures that a large fraction of codes in the search space can be eliminated from the search. Loss factors tend to be smaller for the periodic case (e.g., Tables 4 and 6).

This paper is organised as follows. Section 2 introduces the GLF, derives upper and lower bounds for it, and uses GLF invariance transformations and a bound on the GLF of a set of constant weight sequences to find optimal codes for the non-periodic case. Section 3 gives optimal codes for the periodic case. Section 4 provides two channel estimation examples. Section 5 compares the performance of time domain and frequency domain techniques. Conclusions are given in Section 6.

II. Non-periodic case

Assume the channel is represented as a finite impulse response filter with T-spaced taps, where T is the symbol period. These taps remain constant at least for the duration of the training sequence. The complex, low-pass channel impulse response is given by

$$h(t) = \sum_{k=0}^{L-1} h_k \delta(t - kT) \tag{2}$$

where $\delta(t)$ is the Dirac delta function, L is the total number of taps, and h_k is the complex tap weighting the kth delayed replica. The tap vector

$$\mathbf{h} = \{h_0, h_1, \ldots, h_{L-1}\}$$

is estimated by processing the received signal samples. Assume a sequence

$$\mathbf{b} = \{b_0, b_1, \ldots, b_{N-1}\}$$

is initially transmitted for this purpose, and $b_k \in \{1, -1\}$ (i.e., constant amplitude signals). The received signal samples are given by

$$y_k = \sum_i h_i b_{k-i} + \nu_k, \quad k = 0, 1, \ldots, M-1 \tag{3}$$

where $M = N + L - 1$ and $\{\nu_k\}$ is a white noise sequence[1] with variance σ^2, i.e., $E(\nu_k \nu_l^*) = \sigma^2 \delta(k - l)$. Since a DFT approach is used to estimate $\mathbf{h}$, $\mathbf{h}$ and $\mathbf{b}$ are augmented[2] such that $h_k = 0$ for $k = L, \ldots, M-1$ and $b_k = 0$ for $k = N, \ldots, M-1$. The M point DFT of $\{b_k\}$ is defined by the well-known relation

$$B_n = \sum_{k=0}^{M-1} b_k e^{-j2\pi nk/M}, \quad n = 0, 1, \ldots, M-1 \tag{4}$$

where $j = \sqrt{-1}$. Similarly, the DFTs of $\{h_k\}$ and $\{y_k\}$ are $\{H_n\}$ and $\{Y_n\}$, respectively. Thus,

$$Y_n = H_n B_n + \mathcal{V}_n, \quad n = 0, 1, \ldots, M-1 \tag{5}$$

where $\{\mathcal{V}_n\}$ is the DFT of the noise sequence, $\{\nu_k\}$. The channel estimate is obtained as

$$\hat{h}_k = \frac{1}{M} \sum_{n=0}^{M-1} \left(\frac{Y_n}{B_n} \right) e^{j2\pi kn/M}, \quad k = 0, 1, \ldots, M-1. \tag{6}$$

Using (5), and the fact that taking the inverse DFT of the DFT of a sequence recovers the original sequence, we have

$$\hat{h}_k = h_k + \frac{1}{M} \sum_{n=0}^{M-1} \left(\frac{\mathcal{V}_n}{B_n} \right) e^{j2\pi kn/M}, \quad k = 0, 1, \ldots, M-1. \tag{7}$$

In the absence of noise $\hat{h}_k = h_k \forall k$. Also, $\hat{h}_k$ is an unbiased estimate of h_k, i.e., $E(\hat{h}_k) = h_k \ \forall k$. A good CE sequence should minimise the variances of the error terms $\hat{h}_k - h_k$. The variance of all noise terms affecting the L useful estimates, $\{\hat{h}_0, \ldots, \hat{h}_{L-1}\}$, is given by

$$\sum_{k=0}^{L-1} E((\hat{h}_k - h_k)(\hat{h}_k - h_k)^*) = \sigma^2 \frac{L}{M} \sum_{n=0}^{M-1} \frac{1}{|B_n|^2}. \tag{8}$$

The ratio M/L can be considered as the maximum *processing gain* (PG) attainable by LS filtering, which is reduced by the GLF (inherent to $\mathbf{b}$) defined as

$$\mathcal{M}(\mathbf{b}) = \sum_{n=0}^{M-1} \frac{1}{|B_n|^2}. \tag{9}$$

Ideally, if $\mathcal{M}(\mathbf{b}) = 1$, the maximum PG is realized during the channel estimation process. For given N and L, optimum sequences are obtained when $\mathcal{M}(\mathbf{b})$ is minimised subject to the constraint:

$$\sum_{n=0}^{M-1} |B_n|^2 = NM, \tag{10}$$

which is obtained by applying Parseval's theorem to $\{b_k\}$ and $\{B_n\}$. Heuristically, a good CE sequence should have a reasonably flat spectrum. To quantify this notion, a spectral flatness measure is introduced as follows. Define the spectral max-min ratio (SMMR) of $\{B_n\}$ as

$$\mathcal{X}(\mathbf{b}) = \frac{\max\{|B_n| : 0 \le n < M\}}{\min\{|B_n| : 0 \le n < M\}}. \tag{11}$$

It is expected that an optimal CE sequence has $\mathcal{X}(\mathbf{b}) \approx 1$, while poor CE sequences have $\mathcal{X}(\mathbf{b}) \gg 1$. Clearly, the GLF and the SMMR are closely related parameters. This is further evidenced by the bounds:

$$\frac{M}{N} \le \mathcal{M}(\mathbf{b}) \le \frac{1}{N} \left[1 + (M-1)\mathcal{X}^2(\mathbf{b}) \right]. \tag{12}$$

[1] The notation $\{x_k\}$ denotes a vector $\mathbf{x} = \{x_0, x_1, \ldots, x_{K-1}\}$ where K is clear from the context.

[2] This means the actual transmitted signal is of the form $b_0, b_1, \ldots, b_{N-1}, 0, \ldots, 0, d_0, d_1, \ldots$, where $d_k \in \{1, -1\}$ are the data bits. The number of zero symbols is at least $L - 1$.

Proof: From the classical *Cauchy-Schwarz* inequality[3], one has

$$\mathcal{M}(\mathbf{b})\sum_n |B_n|^2 \geq M^2. \tag{13}$$

Combining this with (10) gives the lower bound (LB) in (12). Also, one has

$$\mathcal{M}(\mathbf{b})\sum_n |B_n|^2 = M + \frac{1}{2}\sum_{m\neq n}\left(\lambda_{m,n} + \frac{1}{\lambda_{m,n}}\right) \tag{14}$$

where $\lambda_{m,n} = |B_m|^2/|B_n|^2$, and the right hand sum contains $M^2 - M$ terms. Thus, by definition,

$$\lambda_{m,n} + \frac{1}{\lambda_{m,n}} \leq 2\mathcal{X}^2(\mathbf{b}), \tag{15}$$

and combining this and (14) yields the upper bound (UB) in (12).

If $\mathcal{X}(\mathbf{b}) = 1$, the bounds converge and the CE sequence satisfies

$$\mathcal{M}(\mathbf{b}) = \frac{M}{N}, \tag{16}$$

which is the smallest possible value. This result is intuitively pleasing and leads to the following definition: a perfect CE sequence has unity SMMR. If a sequence has a spectral null (i.e., $B_j = 0$ for some $0 \leq j < N$), both the GLF and the SMMR are equal to infinity and the sequence is unsuitable for CE. However, the same sequence may be perfect for CE in the time domain if its autocorrelation function has $L-1$ zeros (33). This highlights a difference between time domain estimation [9] and frequency domain estimation. The former is only estimating L, whereas the latter provides M estimates (7). Of course, $M-L$ of these are simply noise terms. Therefore, it is easier to find optimal codes for the time domain approach since only L unknowns are estimated as compared to M for frequency domain estimation (Section 5).

We define the loss factor (in dB) as

$$\mu = 10\log_{10}\left[\frac{\mathcal{M}(\mathbf{b})}{M/N}\right], \tag{17}$$

which indicates the deviation from the LB. A perfect CE sequence has a loss factor of 0 dB.

A. GLF Invariance Transformations

Many GLF invariance transformations that occur for the periodic case (Section 3.A) are not realized for the non-periodic case (since the zero-padded DFT is taken). However, two sequences $\mathbf{b}$ and $\mathbf{c}$ have the same GLF provided:

I1 phase shift of π: $c_k = -b_k$

I2 time reversal : $c_k = b_{N-1-k}$.

The proof is omitted for brevity. The cyclically shifted (q positions to right) sequence given by $c_k = b_{k+q \bmod \mathrm{N}}$ will not have the same GLF.

Example: Given the above, the following $\{b_k\}$ provide the same SER performance for CE: $\{1,-1,1,1\}$, $\{-1,1,-1,-1\}$ and $\{1,1,-1,1\}$.

[3]That is, for real sequences $\{x_k\}$ and $\{y_k\}$, $(\sum x_k y_k)^2 \leq (\sum x_k^2)(\sum y_k^2)$.

B. Constant weight

Convert the $b_k \in \{-1,1\}$ into $a_k \in \{0,1\}$: $a_k = (1-b_k)/2$. If the Hamming weight of $\mathbf{a}$ is $w(\mathbf{a})$, let the sets X_l be defined as

$$X_l = \{\mathbf{a} \mid w(\mathbf{a}) = l\} \quad l = 0,1,\ldots,N. \tag{18}$$

The cardinality of X_l is $|X_l|$ and $\sum_l |X_l| = 2^N$. Below it is shown that the code search needs to be conducted only for a *few* selected X_ls. As $|X_l| \ll 2^N$ for large N, this leads to significant reduction in computation time.

Let the DFT of $\{a_k\}$ be denoted by $\{A_n\}$. As $b_k = 1-2a_k$, it follows that,

$$B_n = \frac{1-\omega^N}{1-\omega} - 2A_n \quad n = 1,2,\ldots,M-1 \tag{19}$$

where $\omega = \exp(-j2\pi n/M)$. If $w(\mathbf{a}) = W$, then $B_0 = N - 2W$. Moreover, the computation of $\{A_n\}$ is sufficient to determine the GLF.

According to Eqs. (10) and (12), an optimal CE sequence has a nearly constant amplitude spectrum, i.e.,

$$|B_n| \approx \sqrt{N} \quad n = 0,1,\ldots,M-1. \tag{20}$$

Consequently, if B_0 is far away from $\sqrt{N}$, such a sequence is unlikely to be optimal. This in turn suggests that the optimality of a sequence somewhat depends on its Hamming weight. To make this notion precise, define by $\mathcal{M}(\mathbf{b}|W)$ the GLF of a sequence with Hamming weight W ($0 \leq W \leq N$). Thus, from (10), for such a sequence satisfies

$$\sum_{n=1}^{M-1} |B_n|^2 = MN - (N-2W)^2. \tag{21}$$

Now the best case happens if $|B_n|$, $1 \leq n < M$, are all equal (which also follows from using Cauchy-Schwarz). This means that

$$\mathcal{M}(\mathbf{b}|W) \geq \frac{(M-1)^2}{MN-(N-2W)^2} + \frac{1}{(N-2W)^2}. \tag{22}$$

This bound shows the smallest GLF for a set of constant-weight sequences. For an exhaustive search of optimal sequences, only X_l for $l = \{1,2,\ldots,\lfloor N/2\rfloor\}$ need be considered at most (this follows from I1). However, this range can be further reduced by using (22), as exemplified below.

Example: Consider $N = 16$ and $L = 2$. For $W = \{1,2,\ldots,7\}$, $\mathcal{M}(\mathbf{b}|W)$ is lower bounded by the set (22) $\{3.37, 2.01, 1.50, 1.25, 1.11, 1.06, 1.21\}$. Then a computer search in X_6 yields a sequence with GLF equal to 1.21. Thus, further search is required only in X_5 and X_7 yielding minimum GLFs of 1.37 and 1.30, respectively.

C. Computer search

A rough outline of the search program is as follows:

1. In the first step, take the length N vector $\mathbf{a} = (1,1,\ldots,0)$ and $g = \infty$.
2. In the i th step, compute $\{B_n\}$ and $\mathcal{M}(\mathbf{b})$. If $\mathcal{M}(\mathbf{b}) \leq g$, then save $\mathbf{b}$ and $g = \mathcal{M}(\mathbf{b})$.
3. Update $\mathbf{a}$ keeping $w(\mathbf{a}) = W$ and repeat 2.

The above procedure is repeated for a sufficient number of Hamming weights. In the following tables, all optimal sequences are given in hexadecimal notation.

Optimal non-periodic sequences are reported in Tables 2-4. Generally, the loss factor increases with L given N. There are exceptions, however. In Table 3, the loss factor decreases from 1.07 to 0.75 dB for L from 5 to 6. This does not mean, however, that estimating 5 taps is worse than estimating 6 taps. Maximum PG varies from 6.81 to 6.19 dB. Considering loss factors, PGs of 4.94 and 4.44 dB can be realized. Thus, SER performance will be better for the $L = 5$ case.

These tables show optimal sequences for a given number of channel taps (L) and sequence lengths (N). In practice, however, it is not easy to fix N and use the optimal sequence for each L because the transmitter and receiver do not know L in advance. However, one can design conservatively for the worst-case by using $L_{\max}$ (which may be obtained by propagation measurements). Then, the optimal code for $L \geq L_{\max}$ in the above tables can be used.

D. Conjecture

Can $\mathcal{X}(\mathbf{b})$ be used as the search criterion instead of $\mathcal{M}(\mathbf{b})$, given their apparent equivalence (12)? The answer would be unequivocally yes, if one could prove that for any two sequences $\mathcal{X}(\mathbf{b_1}) \leq \mathcal{X}(\mathbf{b_2})$ iff $\mathcal{M}(\mathbf{b_1}) \leq \mathcal{M}(\mathbf{b_2})$. Currently, no proof has been found. However, the following conjecture holds for all cases observed. Let S denote the set of all length N binary sequences. Let $\mathcal{X}_{\min} = \min\{\mathcal{X}(\mathbf{b}) \mid \mathbf{b} \in S\}$ and $\mathcal{M}_{\min} = \min\{\mathcal{M}(\mathbf{b}) \mid \mathbf{b} \in S\}$. Let $A = \{\mathbf{b} \mid \mathcal{M}(\mathbf{b}) = \mathcal{M}_{\min}\}$ and $B = \{\mathbf{b} \mid \mathcal{X}(\mathbf{b}) = \mathcal{X}_{\min}\}$. Then A and B are never disjoint; i.e., $A \cap B \neq \emptyset$, where $\emptyset$ is the empty set. In other words, at least one sequence achieves both $\mathcal{X}_{\min}$ and $\mathcal{M}_{\min}$ simultaneously.

III. Periodic case

To estimate L channel taps, the TS is now a cyclic extension of the basic sequence of length N, $N \geq L$, given by

$$b_{(N-T) \bmod \mathrm{N}}, \ldots, b_{N-1}, b_0, b_1, \ldots, b_{N-1}$$

where $T \geq L-1$. A large T will facilitate receiver synchronisation but increase TS overhead. Since the TS is periodic, its convolution with the CIR is periodic and CE is possible using the DFT. The development of Eqs. (1)-(5) still applies here, but zero-padding is now replaced by cyclic extension. Further, the first T received samples are discarded and the next N are used for CE. Only N-point DFTs are needed. Again, the SMMR can be defined as

$$\mathcal{X}(\mathbf{b}) = \frac{\max\{|B_n| : 0 \leq n < N\}}{\min\{|B_n| : 0 \leq n < N\}}.$$

As before, it is expected that the optimal CE sequence has $\mathcal{X}(\mathbf{b}) \approx 1$, while poor CE sequences have $\mathcal{X}(\mathbf{b}) \gg 1$. The modification of (12) to this case yields the bounds

$$\frac{1}{N}\left[1 + (N-1)\mathcal{X}^2(\mathbf{b})\right] \geq \mathcal{M}(\mathbf{b}) \geq 1. \tag{23}$$

A. GLF Invariance Transformations

Two sequences $\mathbf{b}$ and $\mathbf{c}$ have the same GLF provided:
I1 phase shift of π: $c_k = -b_k$
I2 time reversal : $c_k = b_{N-1-k}$
I3 cyclic shift : $c_k = b_{k+q \bmod \mathrm{N}}$.
I1 to I3 derive from the properties of the DFT. The spectrum of a cyclically shifted (q positions to right) sequence is given by $C_n = B_n e^{j2\pi nq/N}$, leading to the same GLF.

Example: Given the above, the following $\{b_k\}$ provide the SER performance for CE: $\{1,-1,1,1\}$, $\{-1,1,-1,-1\}$, $\{1,1,-1,1\}$, $\{1,1,1,-1\}$, and $\{-1,1,1,1\}$.

I1 to I3 coupled with the weight analysis (Section 2.B) and incremental DFT (see appendix) substantial reduces computation time for the code search. Table 5 compares the optimised code search with a simple exhaustive search.

B. Code search

Table 6 shows optimal sequences for $25 \leq N \leq 42$. Provided the cyclic extension is longer than the tail of the CIR, the entire CIR can be estimated (i.e., the GLF is not a function of L). I1 to I3 partition the 2^N sequence space into equivalence classes. In most cases, several equivalence classes achieve the minimum GLF. However, only one optimal sequence for a given N is reported here. $\mathcal{K}$ indicates the number of equivalence classes with $\mathcal{M}_{\min}$. Generally, μ should decrease for increasing N, but one should expect smaller μ values for N such that $\sqrt{N}$ is an integer (e.g., $N = 36$).

Partial searches have also been conducted for $43 \leq N \leq 47$ (Table 7).

C. m-sequences

m-sequences can be used for CE [9]. Since they are characterised by a two-valued autocorrelation function, for $N = 2^P - 1$, the power spectrum is [3]

$$|B_n|^2 = \begin{cases} 1 & n = 0 \bmod N \\ N+1 & \text{otherwise.} \end{cases} \tag{24}$$

Thus, the GLF for an m-sequence is

$$\mathcal{M}(\mathbf{b}) = \frac{2N}{N+1}. \tag{25}$$

For $N \gg 1$, $\mathcal{M}(\mathbf{b}) = 2$. That is, an m-sequence has 3 dB worse SER performance than a perfect CE sequence for frequency domain processing. Table 8 compares m-sequences with the optimal sequences found by code search, which show a 0.6 to 2.5 dB improvement over m-sequences. It should be noted that m-sequences are nearly perfect when used with the time domain approach [9].

D. Schroeder's formula

In [12], a formula yielding binary sequences of arbitrary length with low autocorrelation coefficients for non-zero shifts is given. Thus, a binary sequence of length N is generated according to

$$b_k = 1 - 2\left\lfloor \frac{k^2}{2N} \right\rfloor_{\bmod 2} \tag{26}$$

where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. This is obtained by discretised Newman phases [13].

Table 9 compares optimal sequences and those of (26). Except for $N = 12$, (26) generates good but non-optimal sequences. For $N = 31$, the optimal sequence saves about 1.5 dB in SER. It is strongly suspected in [14] that the Newman phases are optimal for the $N = 12$ case. The table entry for $N = 12$ appears to confirm this.

IV. Channel estimation examples

Here, two examples are provided to compare periodic optimal sequences found by computer search with m-sequences. Firstly, the average distortion-to-noise ratio given by

$$\gamma_{\text{dis}} = \frac{1}{\sigma^2} \sum_k E\left| h_k - \hat{h}_k \right|^2 \tag{27}$$

is computed as a function of input signal-to-noise ratio defined by

$$\gamma_{\text{in}} = \frac{1}{\sigma^2} \sum_k |h_k|^2. \tag{28}$$

Secondly, performance degradation is computed when a linear equalizer is implemented with channel estimates.

Fig. 1 shows $10 \log_{10}(\gamma_{\text{dis}})$ for a channel estimator in a typical data-quality telephone channel. The CIR used is that given by the discrete channel tap weights $\{f_k\}$ found in [1, Fig. 10-2-5(a)] and the channel span is 11 symbols. Fig. 1 shows the distortion for two m-sequences and optimal sequences of lengths 15 and 31 bits under varying input SNR, γ_{in}. The optimal sequences gain about 3 dB noise margin over the m-sequences in the frequency domain. As noted by an anonymous reviewer, if this m-sequence is used in the time domain, the loss factor is 0.064 dB. In this case, the optimal sequence, used in the frequency domain, is about 0.25 dB worse than the m-sequence used in the time domain. It should also be mentioned that a channel estimator based on m-sequences can be implemented easily, even free of multiplications in some cases [9].

For a data sequence I_n, the equalizer output can be written as [1, eq. (10-2-56)]

$$\hat{I}_n = q_0 I_n + \sum_{k \neq n} I_k q_{n-k} + \eta_n \tag{29}$$

where $\{q_n\}$ denotes the convolution of the impulse response of the equalizer and the channel, and η_n the filtered additive noise. Thus, for a given channel, $\{q_n\}$ can be computed for given taps, which are themselves computed using the channel estimates and [1, eq. (10-2-57)]. The 31-bit optimal code is given in Table 6. The bit error rate (BER) P_b is obtained by computing the probability $\text{Re}(\hat{I}_n) < 0$ given $I_n = 1$. Since the output contains a non-Gaussian term, the BER cannot be computed in terms of Gaussian tail probabilities. An infinite series developed by Beaulieu [15, eq. (30)] has been used for this purpose. Fig. 2 shows the performance degradation of a 27-tap linear equalizer. For comparison, the BER of the equalizer given perfect knowledge of the CIR is also shown. For the m-sequence and optimal sequence, the SER performance of the equalizer at $P_b = 10^{-4}$ degrades by 4.5 and 1 dB respectively, compared to the ideal. At low BERs, the equalizer performance strongly depends on the distortion introduce by CE. The optimal sequence estimator keeps this distortion down to a minimum and the resulting equalizer performs close to the ideal.

V. A comparison of time domain and frequency domain techniques

Performance differences between the frequency domain (FD) and time domain (TD) [9] techniques are discussed here. Both rely on least squares filtering, and hence should give the same level of performance for comparable cases. Suppose L channel taps are to be estimated using N channel measurements. The following comments directly apply to the periodic case, where a cyclic extension of length $L-1$ is utilised. Then the maximum achievable SER for either case is given by

$$\text{SER} = 10 \log_{10} \left(\frac{N}{L} \right) \quad \text{(dB)}. \tag{30}$$

With the FD approach, the output SER is less than this maximum by the loss factor

$$\mu = 10 \log_{10} \left(\sum_n \frac{1}{|B_n|^2} \right) \quad \text{(dB)}. \tag{31}$$

Naturally, if $\mu = 0$ dB, such a sequence would be perfect. Various sequences given in this paper come close this ideal to different degrees. For example, in Table 6, for $N = 36$, $\mu = 0.19$ dB.

For the TD approach using a length N CE sequence, the normalised output SER is [9]

$$\text{SER} = 10 \log_{10} \left(1/\text{tr}(P^{-1}) \right) \quad \text{(dB)} \tag{32}$$

where P is an $L \times L$ autocorrelation matrix. Now if the autocorrelation sequence $\phi(k)$ of the TS is an impulse function, its SER, given by (32), achieves (30). In fact, it is sufficient to have $\phi(k) = 0$ for $k = 1, \ldots, L-1$, for a sequence to be perfect for estimating L channel taps. Note that $\phi(k)$ is the cyclic autocorrelation.

Therefore, a length N symbol sequence used for estimating L channel taps is *perfect* for the time domain if

$$\phi(k) = 0 \quad \text{for} \quad k = 1, 2, \ldots, L-1. \tag{33}$$

The sequence is *perfect* for the frequency domain if

$$|B_n| = \sqrt{N} \quad \text{for} \quad n = 0, 1, \ldots, N-1 \tag{34}$$

which yields $\mu = 0$ dB. In either case, the best achievable SER is given by (30). So both approaches should result in maximum SER and as such are equivalent. It appears that TD estimation gets closer to (or achieves) (30) than FD estimation in all cases. Nevertheless, the sequences given in this paper are quite close to (30) as can be seen from the tables. For example, for $N = 40$ the periodic code achieves 0.24 dB (Table 6) within the upper bound (30).

One point to note is that for a given N and L, optimal codes may be easier to find in the TD than in the FD, because the former involves minimising $L-1$ auto-correlation values (33) whereas the latter involves converging all N spectrum amplitudes to a constant (34). Note that if

$$\phi(k) = 0 \quad \text{for} \quad k = 1, \ldots, N-1 \tag{35}$$

then such a sequence satisfies (34), being perfect for FD estimation of L channel taps. Thus, (34) implies (33), but not vice versa.

That is, a sequence optimised for the TD is not necessarily optimal for the FD. For instance, consider the $N = 16$ sequence constructed in Fig. 6 [9] for estimating 5 taps. For TD estimation, this sequence achieves the ideal performance (30) of an SER of 5 dB (16/5). If the same sequence is used for FD estimation, the output SER is found to be (from (30) and (31)) 1.35 dB. Similarly, m-sequences are nearly-perfect for TD estimation, but incur a 3 dB performance penalty when used for FD estimation. There is no contradiction here, but a sequence that satisfies (33) does not necessarily satisfy (34).

What happens when a sequence optimised for the FD is used in TD estimation? Consider the sequence given in Table 6 for $N = 36$. For the FD approach, this sequence incurs a 0.19 dB gain loss when estimating L channel taps, where $1 \leq L \leq 36$. Note that this loss is fixed, not depending on L. The cyclic autocorrelation function of this sequence can be computed as

$$\phi(k) = \begin{cases} 36 & \text{if } k = 0, \\ -4 & \text{if } k = 10, 27, \\ 4 & \text{if } k = 17, 21, \\ 0 & \text{for all other } k. \end{cases} \tag{36}$$

Thus this code is *perfect* for TD estimation of L taps where $1 \leq L \leq 10$, even though it was found using the FD approach. Its implementation is trivial in the time domain and could be a very useful sequence for many applications. For TD estimation, the loss factor (i.e. the difference between (30) and (32)) of this sequence is plotted in Fig. 3. The loss factor for the FD is an upper bound for the loss factor for the TD. Based on this and other numerical experiments, when a given sequence is used for either TD or FD estimation, one has

$$\mathrm{SER}_{\mathrm{FD}} \leq \mathrm{SER}_{\mathrm{TD}} \leq \frac{N}{L}. \tag{37}$$

Unfortunately, we have not been able provide a formal proof for this. Assuming this is true, one can claim that our optimal codes for FD can also be used for TD estimation, with a performance gap less than μ dB given in our tables. In other words, in decibels,

$$\mathrm{SER}_{\mathrm{TD}} - \mathrm{SER}_{\mathrm{FD}} \leq \mu. \tag{38}$$

It appears that the FD estimation is inherently inferior to the TD estimation. However, the performance gap is less than μ dB. Therefore, the performance of the sequences given in this paper is quite close to the ideal performance (30). Also, the FD approach may be suited to applications where FFT processing is used anyway, such as orthogonal frequency division multiplexing systems.

VI. Conclusions

Channel estimation using a known training sequence is required in various communication systems. It has been shown that, for CE with the DFT, optimal sequences must have the smallest GLF. By exploiting invariance properties of the GLF and the bounds on the GLF for constant-weight sequences, optimal sequences up to length 42 have been found. While the sequences are optimal for frequency domain CE, they are marginally worse than optimal codes for time domain CE [9]. Interestingly, for any sequence it appears that the SER for the FD case forms a lower bound on the SER for the TD case.

VII. Acknowledgement

The authors wish to thank the reviewers for their detailed comments and the associate editor Dr. Mitra for her corrections/comments, which led to several improvements of the manuscript.

Appendix

Generation of constant-weight sequences

Let $\mathbf{a} = \{a_0, a_1, \ldots, a_{N-1}\}$, $a_k \in \{0, 1\}$. Let the Hamming weight of $\mathbf{a}$ be W and $K = N - W$ be the number of 0's. Below it is shown how distinct $\mathbf{a}$ can be generated while keeping $w(\mathbf{a}) = W$ and excluding the $N - 1$ cyclic shifts of $\mathbf{a}$. This will yield nearly an N-fold reduction in computations (the periodic case) compared to simply generating all $\mathbf{a}$ with $w(\mathbf{a}) = W$.

Let the set of indices $0 \leq j_1 \leq j_2, \ldots, j_W \leq N-1$ denote the locations of 1's, i.e., $a_{j_k} = 1$ for $k = 1, \ldots, W$. Without loss of generality, consider only sequences with $j_1 = 0$ (this follows from I1 in Section 3.A). Define $n_k = j_{k+1} - j_k - 1$ for $k = 1, \ldots, W-1$ and $n_W = N - 1 - j_W - j_1$. Clearly, for given N and W, any $\mathbf{a}$ can be uniquely described by the set $\{n_k\}$.

Example: A sequence $\{1, 0, 0, 1, 1, 1, 0, 1, 0, 0\}$ is described by $n_1 = 2$, $n_2 = 0$, $n_3 = 0$, $n_4 = 1$, and $n_5 = 2$.

Thus, all constant-weight sequences of weight W can be generated by all combinations of $0 \leq n_k \leq K$ for $1 \leq k \leq W$ satisfying

$$\sum_{k=1}^{W} n_k = K. \tag{A.1}$$

However, to avoid the generation of cyclically-shifted versions of each sequence, the solutions of (A.1) are accepted only if they satisfy

$$n_1 \geq \{n_k \mid k = 2, 3, \ldots, W\} \tag{A.2}$$

and

$$n_2 \geq n_W. \tag{A.3}$$

Using the invariance operations (cyclic shift and time reversal, Section 3.A), the search space can be partitioned into equivalence classes and just one sequence from a class needs to be considered. Consider any sequence described by $n_1', n_2', \ldots, n_W'$ satisfying (A.1). Now if $n_1' \geq \{n_k' \mid k = 2, 3, \ldots, W\}$, then it will be accepted for computation of the GLF. Otherwise, there exists a $k1$ such that $n_{k_1}' \geq \{n_k' \mid k = 1, 2, 3, \ldots, W\}$, for some $1 < k1 \leq W$. Thus, a circular shift of this sequence will satisfy (A.2).

Example: A sequence $\{1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0\}$ will not satisfy (A.2), but its shifted version $\{1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1\}$ will.

Similarly, (A.3) will eliminate some of the time-reversal (I3) equivalent sequences.

In the software, the sequences are generated using nested loops governed by $n_1, n_2, \ldots, n_W$, with n_1 controlling the outer-most loop. The n-th DFT bin can be expressed as

$$B_n = f_1(n_1) + f_2(n_1, n_2) + \cdots + f_W(n_1, n_2, \ldots, n_W).$$

At the beginning of the loop controlled by n_{i+1} $(1 \leq i \leq W-1)$, the partial sum

$$\sum_{k=1}^{i} f_k(n_1, \ldots, n_k)$$

can be computed and reused for all possible combinations of $n_{i+1}, \ldots, n_W$. This allows for a time-efficient incremental DFT computation.

References

[1] J. G. Proakis, *Digital Communications.* New York: McGraw Hill, 3rd ed., 1995.

[2] P. Butler and A. Cantoni, "Noniterative automatic equalization," *IEEE Trans. Commun.*, vol. COM-23, pp. 621–633, June 1975.

[3] M. R. Schroeder, *Number theory in science and communication.* Berlin: Springer, 2nd ed., 1984.

[4] L. Bömer and M. Antweiler, "Perfect N-phase sequences and arrays," *IEEE J. Select. Areas. Commun.*, vol. 10, pp. 782–789, May 1992.

[5] T. Felhauer, "Digital signal processing for optimum wideband channel estimation in the presence of noise," *IEE Proceedings-F*, vol. 140, no. 3, 1993.

[6] A. P. Clark, Z. C. Zhu and J. K. Joshi, "Fast start-up channel estimation," *IEE Proceedings-F*, vol. 131, pp. 375–381, July 1984.

[7] B. Steiner and P. Jung, "Optimum and suboptimum channel estimation for the uplink of CDMA mobile radio systems with joint detection," *Euro. Trans. Telecom.*, vol. 5, pp. 39–50, Jan.-Feb. 1994.

[8] A. Milewski, "Periodic sequences with optimal properties for channel estimation and fast start-up equalization," *IBM J. Res. Develop.*, vol. 27, pp. 426–431, 1983.

[9] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proceedings-F*, vol. 138, pp. 371–378, Aug. 1991.

[10] J. M. Cioffi and J. A. C. Bingham, "A data-driven multitone echo canceller," *IEEE Trans. Commun.*, vol. 42, pp. 2853–2869, Oct. 1994.

[11] S. K. Barton I. R. Johnson, S. J. Shepherd, and P. W. J. Van Eetveld, "Simulation and analysis of the distortion generated by the bulk-FFT demultiplexer," *Signal Processing*, vol. 54, no. 3, pp. 285–294, 1996.

[12] M. R. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," *IEEE Trans. Inform. Theory.*, vol. IT-13, pp. 85–89, 1970.

[13] D. J. Newman, "An L1 extremal problem for polynomials," *Proc. Amer. Math. Soc.*, vol. 16, pp. 1287–1290, Dec. 1965.

[14] D. R. Gimlin and C. R. Patisaul, "On minimizing the peak-to-average power ratio for the sum of N sinusoids," *IEEE Trans. Commun.*, vol. 41, pp. 631–635, Apr. 1993.

[15] N. C. Beaulieu, "The evaluation of error probabilities for intersymbol and cochannel interference," *IEEE Trans. Commun.*, vol. 39, pp. 1740–1749, Dec. 1991.

C. Tellambura received his BSc degree with honors from the University of Moratuwa, Sri-Lanka, in 1986, MSc in electronics from the King's College, UK, in 1988, and PhD in Electrical Engineering from the University of Victoria, Canada, in 1993. He was a post-doctoral research fellow with the University of Victoria (1993-94) and the University of Bradford (1995-96). His research interests include coding, communication theory, modulation, equalization and wireless communications.

Dr Y. Jay Guo (SM'96) received his Ph. D degree on antennas and scattering from Xian Jiaotong University, P. R. China, in 1987. From 1989 to 1994, he was working as Research Fellow at the University of Bradford, United Kingdom. During this period, he developed a number of high efficiency and low cost flat lens and reflector antennas. In 1994, Dr Guo was promoted to Senior Research Fellow and later Lecturer to manage various research projects and supervise postdoctoral research assistants/fellows and Ph. D students, and to lecture on Telecommunications, when he developed his interests in mobile communications, particularly in CDMA cellular systems and radio local area networks (RLAN). In recognition of his research achievement, he was awarded an official Ph. D by the University of Bradford in 1997. Since July 1997, Dr Guo has been with Fujitsu Europe Telecom R & D Centre as Principle Engineer, working on base station techniques for the third generation mobile communications systems.

Dr Guo has published over eighty research papers in academic journals and at international conferences, and he is a senior member of IEEE.

Dr Simon Shepherd was educated at Edgbaston College, Kenilworth School, the Royal Naval Air Engineering School Lee-on-the-Solent, the Britannia Royal Naval College Dartmouth and the Royal Naval Engineering College Manadon. He holds a First Class Honours degree in Engineering and a Doctorate in Cryptography. He served as a Commissioned Officer in Her Majesty's Navy as a weapons and communications specialist where he held a number of appointments ranging from sea-going front line duty in aircraft carriers to shore duties in Naval Intelligence. After 12 years service, he left the RN and joined British Aerospace plc where he was Senior Systems Engineer in the Special Projects Division responsible for computer modelling and development of future systems. He was appointed to his current position as Lecturer in Cryptography and Computer Security in the University of Bradford in 1991.

Dr Shepherd is a Member of the London Mathematical Society, a Senior Member of the Institution of Electronic & Electrical Engineers, a Fellow of the Institute of Mathematics and a Fellow of the Arcadian Society. He is Technical Director of Cerise Innovation Technology plc, Technical Director of Cedar Maritime Designs Ltd and Chief Executive of Silicon Alchemy Ltd. His research interests include Cryptography, Signals & Communications Intelligence, Algebraic & Computational Number Theory, and Coding and Information Theory.

Stephen K. Barton(SM'92) received the B.Sc. (Eng.) in 1970 from University College London, and the M.Sc. in Telecommunications Systems in 1974 from the University of Essex. From 1970 to 1976 he was employed by Marconi Research Laboratories, principally on high speed fast acquisition modems for satellite TDMA. From 1976 to 1980 he was with Her Majesty's Government Communications Centre, working on GaAs FET oscillators and conformal antennas. From 1980 to 1985 he was with the Rutherford Appleton Laboratory, where he originated the Communications Engineering Research Satellite project. From 1985 to 1989 he was with Signal Processors Ltd. working on Adaptive TDMA modems and digital receivers generally. From 1989 to 1998 he was with the University of Bradford, and since 1998 he is with the University of Leeds. His research interests include multi-carrier modulation/demodulation, Wireless LANs, CDMA, channel equalisation and near/far resistant detection algorithms. From 1993 to 1996 he was chairman of Working Group 3: Broadband Systems, of the European Commission Co-operation in Science and Technology programme: COST 231: Evolution of Land Mobile Radio (including personal) Communications. Since 1997 he is chairman of Working Group 1: Radio System Aspects, of COST 259: Wireless Flexible Personalised Communications.

Prof. Barton was admitted a Fellow of the Institute of Electrical Engineers (UK) in 1993.