CHANNEL ESTIMATION USING DISCRETE FOURIER TRANSFORM TECHNIQUES*

C. Tellambura
School of Computer Science and Software Engineering
Monash University, Australia
Matthew G. Parker
Code Theory Group, Inst. for Informatikk
University of Bergen
Norway
S. K. Barton
Institute of Integrated Information Systems
School of Electronic and Electrical Engineering
University of Leeds Leeds, LS2 9JT, England

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. Subject areas: Modulation and Synchronisation and/or Signal Processing for Communications

1 Introduction

For burst-transmission digital communication systems, channel estimation (CE) is required for maximum likelihood sequence estimation receivers [1] and noniterative 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 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.

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]. For instance, [5-8] 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 nonwhite 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. Milewski [7] provides a construction for some poly-phase (but not binary) perfect autocorrelation sequences.

In [8], LS filtering for CE is considered and optimal binary sequences up to length 22 are found by exhaustive computer search. The search criterion is:

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

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 [8], but, importantly, all processing occurs in the frequency domain. This leads to an explicit expression for the search

*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/K 00318 in collaboration with Symbionics Networks limited.

criterion, termed the gain loss factor (GLF), which is only a function of the power spectrum of the training sequence. A CE sequence is *optimal* if it minimises the GLF. 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 4). This paper provides a frequency domain approach (i.e., using the DFT), bounds on the GLF and an optimal periodic sequence search.

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

2 Channel estimation

We assume that the channel is represented as a finite impulse response filter with T-spaced taps, where T is the symbol period. The channel remain constant at least for the duration of the training sequence. The complex, low-pass channel impulse response (CIR) 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. We envision a data transmission of isochronous (time division multiple access) or asynchronous packets. Each packet contains information data and overhead symbols for several purposes(e.g., channel estimation and synchronisation). In our case we are only interested in the channel estimation problem. The packet structure is therefore

$$b_t, \dots, b_{N-1}, b_0, b_1, \dots, b_{N-1}, d_1, d_2, \dots, d_q$$

where $t \leq N - L + 1$. Here $b_k \in \{1, -1\}$ are for channel estimation and d_ks are information data. As we can see, TS is now a cyclic extension of the basic sequence of length N, $N \geq L$. A small t will facilitate receiver synchronization but increase TS overhead. Since the TS is periodic, its convolution with the CIR is periodic and CE is possible using the DFT.

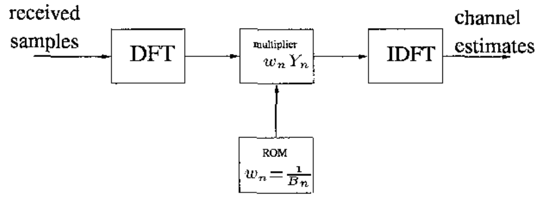

Figure 1: Channel estimation using optimal sequences.

Fig. 1 shows the channel estimator. We assume the received signal samples contain a white gaussian noise with variance σ^2. Both DFT and IDFT are a set of orthogonal transforms, so will not change the correlation characteristics of the noise. We can therefore show that [9] the variance of all noise terms affecting the L useful estimates, $\{\hat{h}_0, \dots, \hat{h}_{L-1}\}$, is given by

$$\sum_{k=0}^{L-1} E\left|\hat{h}_k - h_k\right|^2 = \sigma^2 \frac{L}{M} \sum_{n=0}^{N-1} \frac{1}{|B_n|^2}. \tag{3}$$

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

$$\mathcal{M}(\mathbf{b}) = \sum_{n=0}^{N-1} \frac{1}{|B_n|^2}. \tag{4}$$

Ideally, if $\mathcal{M}(\mathbf{b}) = 1$, the maximum PG is realized during the channel estimation process. 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

$$\mathcal{X}(\mathbf{b}) = \frac{\max\{|B_n| : 0 \leq n < N\}}{\min\{|B_n| : 0 \leq n < N\}}. \tag{5}$$

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, GLF and SMMR are closely related parameters. This is further evidenced by the bounds [9]:

$$1 \leq \mathcal{M}(\mathbf{b}) \leq \frac{1}{N}\left[1 + (N-1)\mathcal{X}^2(\mathbf{b})\right]. \tag{6}$$

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

$$\mathcal{M}(\mathbf{b}) = 1, \tag{7}$$

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 GLF and SMMR are equal to infinity and the sequence is unsuitable for CE. Define the loss factor (in dB) as

$$\mu = 10 \log_{10}\left[\mathcal{M}(\mathbf{b})\right]. \tag{8}$$

A perfect CE sequence has a loss factor of 0 dB.

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 : $c_k = b_{k+q \bmod N}$.

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.

B. Constant weight

Convert the $b_k \in \{-1,1\}$ into $a_k \in \{0,1\}$: $a_k = (1-b_k)/2$. 2^N $\{b_k\}$s can be translated to $\{a_k\}$s. If the Hamming weight of $\mathbf{a}$ is $w(\mathbf{a})$, let the sets

$$X_l = \{\mathbf{a} \mid w(\mathbf{a}) = l\} \quad l = 0,1,\ldots,N. \tag{9}$$

The size 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,

$$B_n = \frac{1-\omega^N}{1-\omega} - 2A_n \quad n = 1,2,\ldots,M-1 \tag{10}$$

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 GLF.

According to Eq. and (6), an optimal CE sequence has a nearly constant amplitude spectrum, i.e.,

$$|B_n| \approx \sqrt{N} \quad n = 0,1,\ldots,N-1. \tag{11}$$

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 $(0 \le W \le N)$. Thus, using Parseval's theorem, for such a sequence

$$\sum_{n=1}^{N-1} |B_n|^2 = N^2 - (N-2W)^2. \tag{12}$$

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

$$\mathcal{M}(\mathbf{b}|W) \ge \frac{(N-1)^2}{N^2 - (N-2W)^2} + \frac{1}{(N-2W)^2}. \tag{13}$$

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 (13), 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 (13) $\{3.37, 2.01, 1.50, 1.25, 1.11, 1.06, 1.21\}$. 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 $\mathcal{M}(\mathbf{b}) \le g$, 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.

I1 to I3 coupled with the weight analysis (Section 2.B) and incremental DFT allow substantial improvement in computation time for the code search. More details can be found in [9]. Table 1 shows optimal sequences for $25 \le N \le 42$. Provided the cyclic extension is longer than the tail of the CIR, the entire CIR can be estimated (i.e., 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 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., $N = 36$).

Table 1: Optimal periodic CE sequences

Length (N)	Code	$\mathcal{K}$	μ (dB)	
25	E2CC21	3	0.49	1.75
26	16C8701	8	0.51	1.98
27	392B841	4	0.40	1.64
28	D724301	8	0.53	1.72
29	E9A1881	4	0.50	1.72
30	131129F1	15	0.51	2.08
31	3B446161	16	0.31	1.46
32	5230F641	21	0.39	1.73
33	849B88E1	17	0.28	1.65
34	1D18F4241	22	0.43	1.81
35	22917E461	16	0.31	1.65
36	5908973C1	33	0.19	1.38
37	19C4848BA1	26	0.32	1.66
38	1D70852361	28	0.37	1.86
39	68892D7381	4	0.22	1.50
40	4D73607281	24	0.24	1.62
41	198AD1D3401	33	0.29	1.62
42	08C65A2F881	20	0.32	1.75

3 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

$$\gamma_{\text{dis}} = \frac{1}{\sigma^2}\sum_k E\left|h_k - \hat{h}_k\right|^2 \quad (14)$$

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

$$\gamma_{\text{in}} = \frac{1}{\sigma^2}\sum_k |h_k|^2. \quad (15)$$

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

Fig. 2 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. 2 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 [8].

4 Time domain vs. frequency domain techniques

Performance differences between the frequency domain (FD) and time domain (TD) [8] 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

$$\text{SER} = 10\log_{10}\left(\frac{N}{L}\right) \quad \text{(dB)}. \quad (16)$$

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

$$\mu = 10\log_{10}\left(\sum_n \frac{1}{|B_n|^2}\right) \quad \text{(dB)}. \quad (17)$$

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

For the TD approach, the normalised output SER is given by [8]

$$\text{SER} = 10\log_{10}\left(1/\text{tr}(P^{-1})\right) \quad \text{(dB)} \quad (18)$$

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 (18) achieves (16). 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 time domain if

$$\phi(k) = 0 \quad \text{for} \quad k = 1, 2, \ldots, L-1. \quad (19)$$

This sequence is *perfect* for frequency domain if

$$|B_n| = \sqrt{N} \quad \text{for} \quad n = 0, 1, \ldots, N-1 \quad (20)$$

which yields $\mu = 0$ dB. In either case, the best achievable SER is given by (16). So both approaches should result in maximum SER and as such are equivalent. It appears that TD estimation gets closer to or achieves (16) in all cases than FD estimation. Nevertheless, the sequences given in this paper are quite close to (16) as can be seen from the tables. For example, for $N = 40$ the periodic code achieves 0.24 dB (Table 1) within the upper bound (16).

One point to note is that for given N and L, optimal codes may be easier to be found for TD than for FD, because the former involves minimising $L-1$ autocorrelation values (19) whereas the latter involves converging all N spectrum amplitudes to a constant (20). Note that if

$$\phi(k) = 0 \quad \text{for} \quad k = 1, \ldots, N-1 \quad (21)$$

then such a sequence satisfies (20), being perfect for FD estimating L channel taps. Thus, (20) implies (19), but not vice versa.

That is, a sequence optimised for TD is not necessarily optimal for FD. For instance, consider $N = 16$ sequence constructed in Fig. 6 [8] for estimating 5 taps. For TD estimation, this sequence achieves the ideal performance (1) 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 (16) and (18)) 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 (19) does not necessarily satisfy (20).

5 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 [8]. Interestingly, for any sequence it appears that the SER for the FD case forms a lower bound on the SER for the TD case. Also, the FD approach may be suited to applications where FFT processing is used anyway, such as orthogonal frequency division multiplexing systems.

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] 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.

[8] 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.

[9] C. Tellambura et al., "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, 1996. submitted.

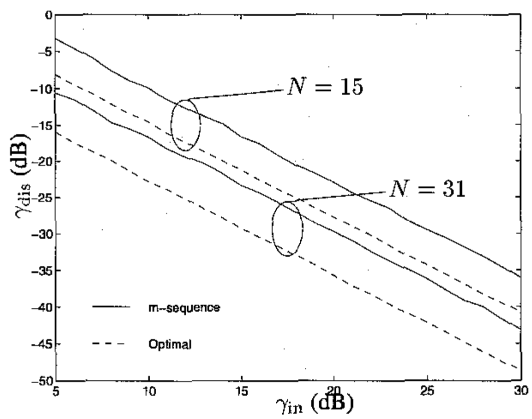

Figure 2: The output distortion for channel estimation using m sequences and optimal sequences. BPSK modulation is used.