

# Short sequences and cross-bifix analysis

Dragana Bajic , Cedomir Stefanovic

Department of Communications and Signal Processing

University of Novi Sad

21000 Novi Sad, Serbia and Montenegro

LMCDRA@eunet.yu, cex@uns.ns.ac.yu

**Abstract:**—An analytical approach to a search process for a set of  $M$  fixed sequences in random data covers the problems of a search for the sequences with errors, as well as of a search for the distributed sequences. It is of greater practical interest than the formerly analysed case of search for a single sequence. This paper derives statistical parameters (probability distribution function, expected value, variance) of this process, introducing a new term - “cross-bifix”. The application examples are outlined.

**Keywords:**—expected duration of a search; distributed sequences; bifix; cross-bifix

## I. INTRODUCTION

Fast and reliable frame synchronization acquisition is a crucial requirement to establish a connection between a transmitter and a receiver. A common method to obtain this is to insert periodically a synchronization sequence (frame-alignment word, FAW) in data sequence, thus indicating the boundaries between the frames of data. Then a correlation technique (or a window-sliding search) is used to find the position of FAW in the received signal.

Analytical approaches to synchronization acquisition process and methods for construction of sequences with the best aperiodic autocorrelation properties (alternatively, with minimal total simulation probability) have been the subject of numerous analyses over the past decades, e.g. [1-9]. Only few of them investigated the relationship between frame length and synchronization sequence length&structure [10, 11], showing that, if frame length exceeds “turning point”, other sequence structures might outperform the optimised ones considering the acquisition time. This is a consequence of the known result [12]: duration of search for a fixed sequence in random data is the shortest for “bifix-free” patterns, so they are the most likely to be simulated, thus prolonging the acquisition. This introduces the necessity to analyse statistically a process interesting both to mathematicians and engineers – search-for-a-fixed-sequence process. Purely mathematical achievements (at binary level) are recently summarized in [13], but older and more suitable engineering approach [12] introduced already mentioned term “bifix” – a subsequence that is both a prefix and a suffix of an observed synchronization sequence. Based upon the bifix analysis, probability density functions (p.d.f) of search process in random data [14] and in frame (where overlap region has to be

taken into account) [15] were derived. These formulae were necessary mathematical prerequisite for analytical expressions that verify simulation-study results from [11, 12].

However, analytical derivation [14, 15] of simulation curves from [11, 12] had no further application, since the exact matching of the synchronization sequence at the receiver end has been long abandoned. Current techniques perform search for sequences within the specified distance of the inserted one. At symbol level, this is equivalent to a search for a set of sequences; therefore, new analytical tools for multiple search process have to be developed. These analytical tools and their application to distributed sequences are the topic of this paper.

The next section gives a brief explanation of the derived formulae for statistical parameters of multiple search process (abstract in [16]). The comparison of the best known distributed sequences [17] and contiguous synchronization sequences of the same redundancy are performed within the third section and followed by the concluding remarks.

## II. SEARCH FOR $M$ SEQUENCES IN RANDOM DATA

The goal of the multiple search processes is to find any sequence out of  $M$  known sequences while sliding along data stream consisting of random but not equiprobable symbols. Each sequence consists of  $N$   $L$ -ary symbols. For example, search for binary ( $L=2$ ) distributed sequence  $0x0x1$  means search for  $M=4$  sequences –  $00001, 00011, 01001$  and  $01011$ .

The search starts (the first test,  $k=1$ ) when the first  $N$  received data symbols are compared to each one of  $M$  sequences. If the test fails (the received  $N$  data symbols match to none of  $M$  sequences), test position shifts or “slides” and symbols 2 to  $N+1$  are compared to  $M$  sequences, etc. Search stops if the  $k^{\text{th}}$  test succeeds, meaning that received data stream of length  $k+N-1$  satisfies “extended matching condition” (e.m.c): last  $N$  symbols of data stream equal to one of  $M$  sequences while none of these  $M$  sequences was found at any of the previous test positions. Probability of this event (i.e. probability that number of tests is exactly  $k$  or, rather, probability that one of the  $M$  sequences is found at the  $k^{\text{th}}$  test - p.d.f. of the process – is evaluated using a recursive formula:

$$\Pr\{k\} = \sum_{i=1}^M \Pr^{(i)}\{k\}$$

$$\Pr^{(i)}\{k\} = \sum_{j=1}^M \sum_{m=1}^{\min(k-1, N)} (h_{ji}^{(N+1-m)} \cdot r_i^{(m-1)} - h_{ji}^{(N-m)} \cdot r_i^{(m)}) \cdot \Pr^{(j)}\{k-m\},$$

$$\Pr^{(i)}\{1\} = r_i^{(N)}, \quad i = 1, \dots, M. \quad (1)$$

New sequence descriptors are introduced. The first one is “**cross-bifix**”. A cross-bifix is a subsequence of length  $n \leq N$  that is a suffix of  $i^{\text{th}}$  sequence and a prefix of  $j^{\text{th}}$  sequence,  $i, j = 1, \dots, M$ . The corresponding cross-bifix indicator  $h_{ij}^{(n)}$  equals to 1 if cross-bifix of length  $n$  exists, e.g. binary sequences  $P_i = \mathbf{0001}$  and  $P_j = \mathbf{0011}$  have a 3-bit cross-bifix  $h_{ij}^{(3)} = 1$ , while obviously  $h_{ji}^{(3)} = 0$ . If  $i=j$ ,  $h_{ii}^{(n)}$  denotes classical bifix indicator  $h_n$  introduced in [12]. The default values for cross-bifix indicators are:

$$h_{ij}^{(0)} = 1, \quad h_{ij}^{(N)} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, \quad i, j = 1, \dots, M \quad (2)$$

The second descriptor is a set of “**tail probability vectors**”. A tail of length  $n$  is a suffix – last  $n$  symbols of a sequence. Its probability is obtained by multiplication of the last  $n$  symbols of a sequence. The default values are  $r_i^{(0)} = 1$ ,  $i = 1, \dots, M$ ; a tail of length  $N$  is a sequence itself, so  $r_i^{(N)}$  is a probability of  $i^{\text{th}}$  sequence. If a stream of data along which a search is performed is equiprobable with  $p = 1/L$ , probability of each tail of length  $n$  is  $r_i^{(n)} = p^n$ . The formal proof of (1) can be found in [1].

By exhaustive summation, the other statistical parameters (expected value of the number of tests - duration of search) and second moments can be evaluated as well:

$$E\{k\} = T = 1 - N + [\Pr\{1\}]^{-1} \cdot \sum_{i=1}^M S_i \cdot C_i; \quad (3)$$

$$E\{k^2\} = 1 - 2 \cdot N \cdot T - N^2 + [\Pr\{1\}]^{-1} \cdot \sum_{i=1}^M (2 \cdot C_i \cdot T_i + W_i \cdot S_i) \quad (4)$$

where

$$C_{ij} = \sum_{m=1}^N r_j^{(m-1)} \cdot h_{ij}^{(N-m+1)}; \quad (5)$$

$$C_i = \sum_{j=1}^M C_{ij} = \sum_{j=1}^M \sum_{m=1}^N r_j^{(m-1)} \cdot h_{ij}^{(N-m+1)};$$

$$W_{ij} = \sum_{m=1}^N (2m-1) \cdot r_j^{(m-1)} \cdot h_{ij}^{(N-m+1)}; \quad (6)$$

$$W_i = \sum_{j=1}^M W_{ij} = \sum_{j=1}^M \sum_{m=1}^N (2m-1) \cdot r_j^{(m-1)} \cdot h_{ij}^{(N-m+1)};$$

The quantities  $S_j$  and  $T_j$ ,  $j=1, \dots, M$  can be obtained as solutions of the following sets of linear equations:

$$\mathbf{A} \cdot \mathbf{S}^T = [\mathbf{0} \quad \mathbf{0} \quad \dots \quad \mathbf{0} \quad \mathbf{1}]^T \quad (7)$$

$$\mathbf{A} \cdot \mathbf{T}^T = \mathbf{B}^T \quad (8)$$

Matrix  $\mathbf{A}$  is an  $M \times M$  matrix while  $\mathbf{B}$  is a vector:

$$\mathbf{A} = [A_{ij}], \quad A_{ij} = \begin{cases} C_{ji}/r_i^{(N)} - C_{j(i+1)}/r_{i+1}^{(N)}, & i < M \\ 1, & i = M \end{cases}, \quad j = 1, \dots, M \quad (9)$$

$$\mathbf{B} = [B_i], \quad B_i = \begin{cases} \sum_{j=1}^M 0.5 \cdot (W_{j(i+1)}/r_{i+1}^{(N)} - W_{ji}/r_i^{(N)}) \cdot S_j, & i < M \\ T, & i = M \end{cases} \quad (10)$$

If  $M=1$  and symbols are equiprobable,  $E\{k\}$  and  $E\{k^2\}$  become the same as in [12, 14], i.e. the case of a single sequence search.

In further sections, since the non-equiprobable result has just recently been developed, all the illustrative examples and evaluations would be expressed for streams of equiprobable data.

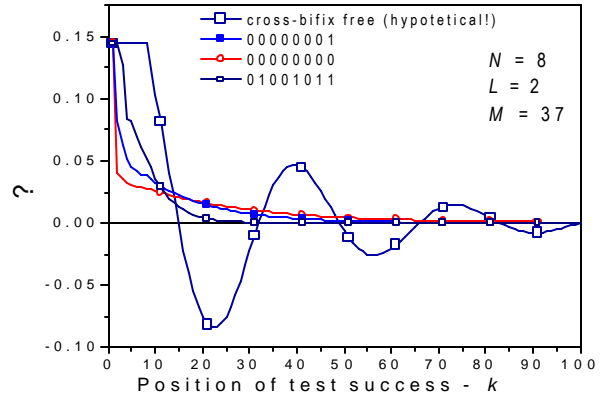


Fig. 1 An illustrative example of p.d.f. and impossible p.d.f.

As an illustrative example, p.d.f. for three sets of  $M=37$  sequences is drawn. Each set is derived from an 8-bit sequence, with all the sequences that are at Hamming distance 1 and 2. But, if all the cross-bifix matrices (except the default ones) are filled with zeros (cross-bifix free, impossible for  $N=8$ ,  $M=27$ ,  $L=2$  case), an impossible p.d.f. arises. This occurrence might be interesting for future purely theoretical and abstract research.

### III. DISTRIBUTED SEQUENCES AND CROSS-BIFIX-FREE SEQUENCES

A distributed sequence [17] of length  $N$  is specified by  $C = c_1 c_2 \dots c_{N-1}$ . Out of  $N$  symbols,  $l$  ones are fixed, taking value 0 or 1, and the remaining  $N-l$  are unconstrained data, identified by the character  $x$ . The data symbols will assume the values 0 and 1 with equal probability. By definition, the first and last positions of the sequence  $C$  are fixed symbols, e.g. for sequence  $C = 0x \dots x0$ ,  $l=3$  and  $N=7$ .

For contiguous sequences it is known that the best ones are bifix-free; analogously, a term “cross-bifix-free” sequences can be defined, for a set of sequences with all the cross-bifix indicators equalling 0, except the default ones. This sets values of (9) and (10) is:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 1 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & -1 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}, \quad \mathbf{B} = [0 \ 0 \ \dots \ 0 \ T] \quad (11)$$

from which it follows immediately:

$$S_j = 1/M; \quad T_j = T/M. \quad (12)$$

Expressions (3) and (4) are then considerably simplified, as well as expression for variance  $S^2 = E\{k^2\} - E^2\{k\}$ :

$$T = 1 - N + L^N / M \quad (13)$$

$$S^2 = (1 - 2 \cdot N) \cdot L^N / M + L^{2N} / M^2. \quad (14)$$

Putting  $M=1$ , expressions for statistical parameters of an ordinary bifix-free sequence is obtained [12, 14].

The only detailed contribution dealing with distributed sequences [17] proposes methods for construction of optimal sequences and gives a list of best one found. It is not surprising that such sequences are cross-bifix free (the

authors of [17] call them bifix-free, obviously in a cross-bifix sense). From Eq. (13) it is obvious that distributed sequences outperform contiguous ones since the best sequences should have minimal expected value of search duration [12]. The bigger  $M$ , the shorter  $E\{k\}$ .

#### IV. SEARCH IN FRAME

For the search process in the frame, the set of sequences consists of all the sequences that are within the specified distance from the inserted synchronizing sequence (FAW). If a single error per FAW is allowed (in binary case the distance is Hamming and the number of errors  $e=1$ ), set consists of  $M=N+1$  sequences. If  $e=2$ ,  $M=1+N+N(N+1)/2$ , e.t.c.

However, a frame itself contains a periodically inserted FAW, so the analyses applicable to the search in random data have to be altered. With the new analytical tool it is possible to establish the relationship between the FAW length, frame length  $F$  and number of errors per FAW allowed  $e$ , introducing the survival probability (the non-simulation probability) as a comparison criterion. The starting search position (offset regarding the real sync. position, Fig. 2), is denoted with  $S$ .

An analytical expression for p.d.f. of the search process in a frame is obtained as a derivation of Eq. (1)[19]:

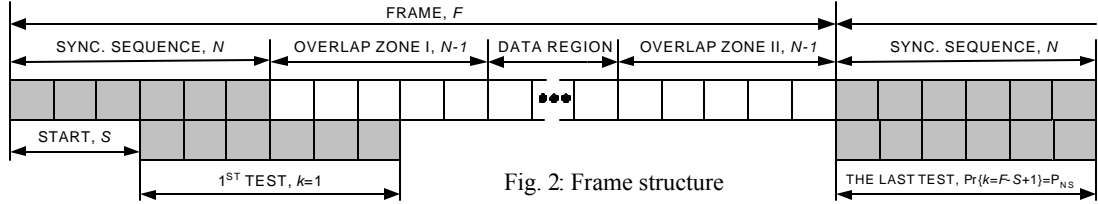


Fig. 2: Frame structure

1. Start in the first overlap region,  $1 \leq S < N$ :

$$1 \leq S < N:$$

$$\Pr^{(i)}\{k\} = \begin{cases} h_{li}^{(N-S+1-k)} \cdot r_i^{S+k-1}, & (h_{li}^{(N-S+1-d)} + h_{lm}^{(N-(k-d))}) \neq 2, \quad d \leq k-1, \quad l=1, \dots, M, \quad 1 \leq k \leq N-S+1 \\ 0, & h_{li}^{(N-S+1-k)} = 1, \quad \text{at least one } (h_{li}^{(N-S+1-d)} + h_{lm}^{(N-(k-d))}) = 2, \quad d \leq k-1, \quad l=1, \dots, M, \quad 1 \leq k \leq N-S+1 \\ \sum_{j=1}^M \sum_{m=1}^{\min(k-1, N)} (h_{ji}^{(N+1-m)} \cdot r_i^{(m-1)} - h_{ji}^{(N-m)} \cdot r_i^{(m)}) \cdot \Pr^{(j)}\{k-m\} & N-S+2 \leq k \leq F-S+1 \end{cases}$$

$$\Pr\{k\} = \begin{cases} \sum_{i=1}^M \Pr^{(i)}\{k\}, & 1 \leq k \leq F-N-S+1 \\ \sum_{i=1}^M h_{i1}^{(N+S-F-1+k)} \cdot \Pr^{(i)}\{k\} / r_i^{(N+S-F-1+k)}, & F-N-S+2 \leq k \leq F-S+1 \end{cases}$$

2. Start in the data region,  $N < S \leq F - N$ :

$$N < S \leq F - N:$$

$$\Pr^{(i)}\{k\} = \sum_{j=1}^M \sum_{m=1}^{\min(k-1, N)} (h_{ji}^{(N+1-m)} \cdot r_i^{(m-1)} - h_{ji}^{(N-m)} \cdot r_i^{(m)}) \cdot \Pr^{(j)}\{k-m\} \quad 1 \leq k \leq F-S+1$$

$$\Pr\{k\} = \begin{cases} \sum_{i=1}^M \Pr^{(i)}\{k\}, & 1 \leq k \leq F - N - S + 1 \\ \sum_{i=1}^M h_{i1}^{(N+S-F-1+k)} \cdot \Pr^{(i)}\{k\} / r_i^{(N+S-F-1+k)}, & F - N - S + 2 \leq k \leq F - S + 1 \end{cases}$$

3. Start in the second overlap region,,  $F - N < S \leq F$ :

$$F - N < S \leq F :$$

$$\Pr^{(i)}\{k\} = \sum_{j=1}^M \sum_{m=1}^{\min(k-1, N)} (h_{ji}^{(N+1-m)} \cdot r_i^{(m-1)} - h_{ji}^{(N-m)} \cdot r_i^{(m)}) \cdot \Pr^{(j)}\{k-m\} \quad 1 \leq k \leq F - S + 1 \quad (15)$$

$$\Pr\{k\} = \sum_{i=1}^M h_{i1}^{(N+S-F-1+k)} \cdot \Pr^{(i)}\{k\} / r_i^{(N+S-F-1+k)}, \quad 1 \leq k \leq F - S + 1$$

The most interesting result at this research level is the probability that the search would survive - the last term of the p.d.f. - i.e. the probability that none of the  $M$  sequences would be simulated before the real position of FAW ( $2^{\text{nd}}$  sync. sequence in Fig. 1). The search that starts immediately after the real sync. position,  $S=1$ , is the worst case for which “the non-simulation probability” or “survival probability”, equals to:

$$P_{NS} = \Pr\{F - S + 1\} = \Pr\{F\}. \quad (16)$$

Fig. 3 shows the p.d.f of 8bit sync. sequences, Jones' 00011011 [0] and Willard's 00100111 with  $F=40$ . If  $e=0$  and  $M=1$ , all bifix-free sequences, including the mentioned two ones, have the same p.d.f. These two sequences have the same p.d.f. if  $e=1$  and  $M=9$ , too. For  $e=2$  and  $M=37$ , p.d.fs are different although  $P_{NS}$  remain the same.

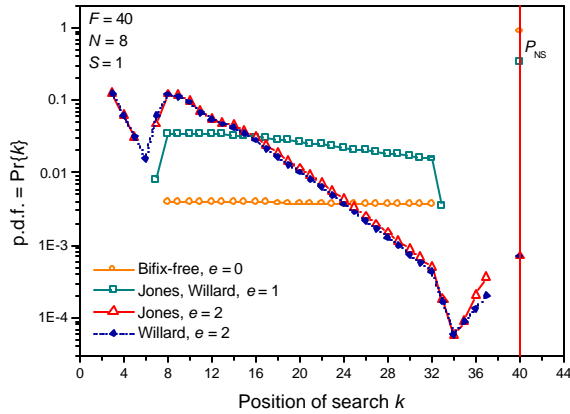


Fig. 3. P.d.f and  $P_{NS}$  of good 8-bit sequences

If the other eight-bit sequences are observed as well, it might be noticed that the survival probability is zero for  $e>0$ , except for the sequences from Fig. 3, known to be the best ones. If  $S=1$  and  $e=1$ , search would stop at  $k=1$  for the all-zeros sequence, or at one of the positions  $1 < k < 9$  for (bifix-free!) 00000001 sequence. Good sequences will survive, as well as a repetitive sequence 01010101 with considerably lower  $P_{NS}$  (Fig. 4a). If  $e=2$ , only the good sequence will survive (Fig. 4b). The same occurs if the search starts within the data region, omitting the first overlap zone – then the simulation of

not so good sequences would occur within the second overlap region.

If  $S=20$  and  $e=1$ , all-zero sequence will not survive (Fig. 5a); but for  $e=2$ , all sequences except the good ones will have  $P_{NS}=0$  (Fig. 5b).

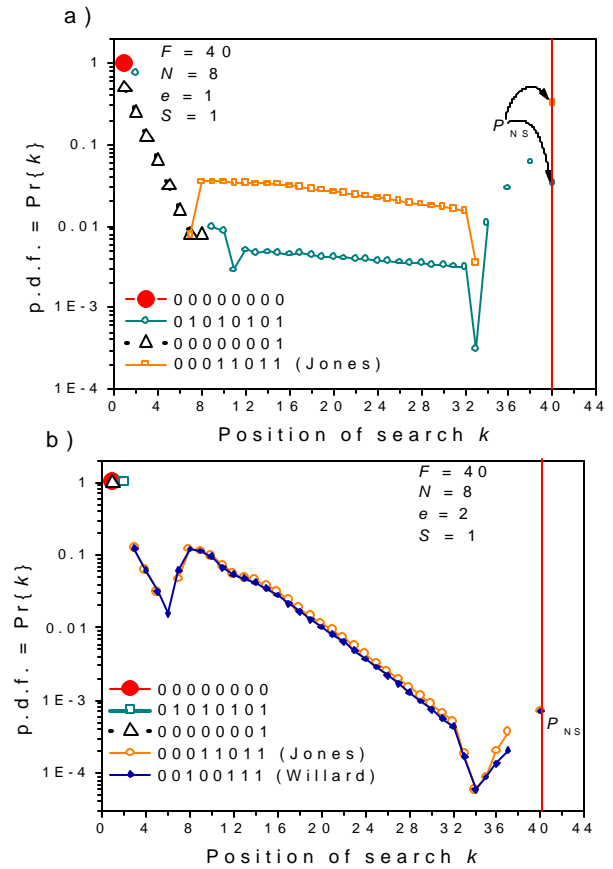


Fig. 4. P.d.f. and  $P_{NS}$  of  $N=8$  sequences,  $S=1$  a)  $e=1$ ; b)  $e=2$

From Fig. 6 the relationship between the  $F$ ,  $N$ ,  $e$  and  $P_{NS}$  can be observed. Long frames and more errors require longer FAW. However, limiting value of  $P_{NS}$  for sequences of satisfactory length is not always 1 (the limit can be easily

proven). Fig. 7 shows  $P_{NS}$  when  $e=0$  for characteristic sequences – the ones with periodic components are obviously not suitable for frame alignment. Similar results are shown, for  $e=1$ , in Fig. 8. It is interesting that in this case  $P_{NS}$  for sequences of type  $00\dots000$  and  $00\dots001$  are steady zero! At last, Fig. 10 shows p.d.f. for different bit-error rates. It should be noticed that simulation is not possible at the positions  $k = 1, 4$  and  $19$ , but channel errors make it possible.

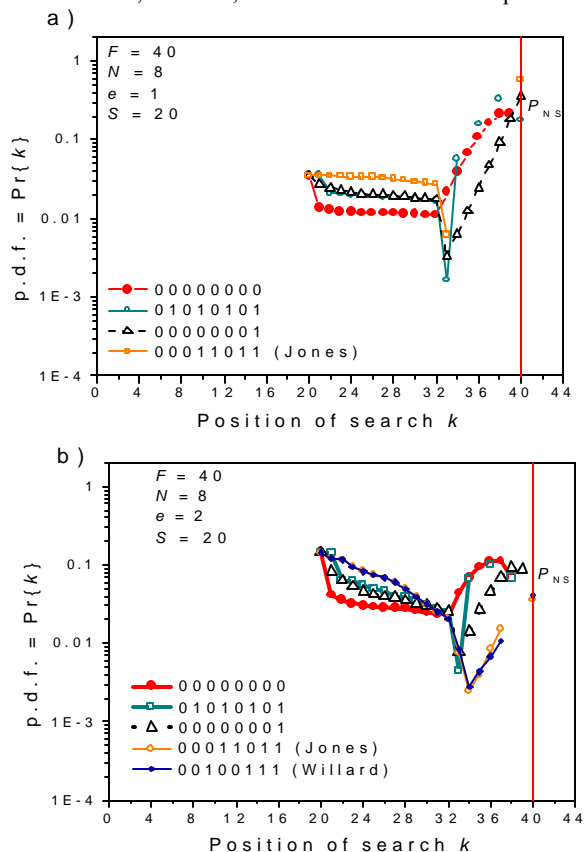


Fig. 5. P.d.f. and  $P_{NS}$  of  $N=8$  sequences,  $S=20$  a)  $e=1$ ; b)  $e=2$

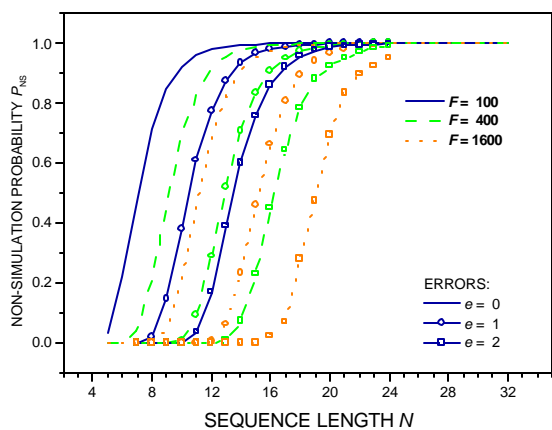


Fig. 6. Relationship between  $F, N$  and  $e$

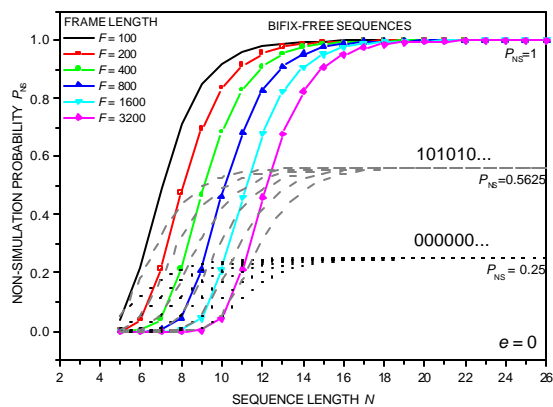


Fig. 7.  $P_{NS}$  for characteristic sequences,  $e=0$

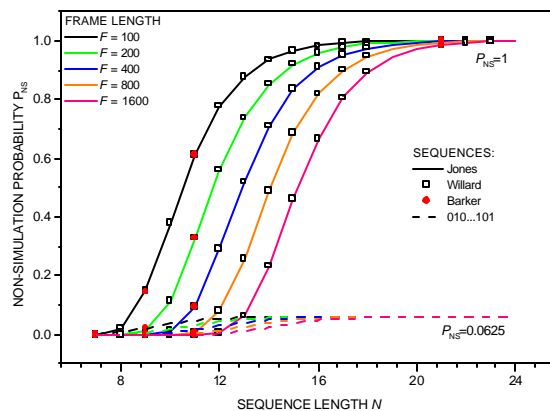


Fig. 8.  $P_{NS}$  for characteristic sequences,  $e=1$

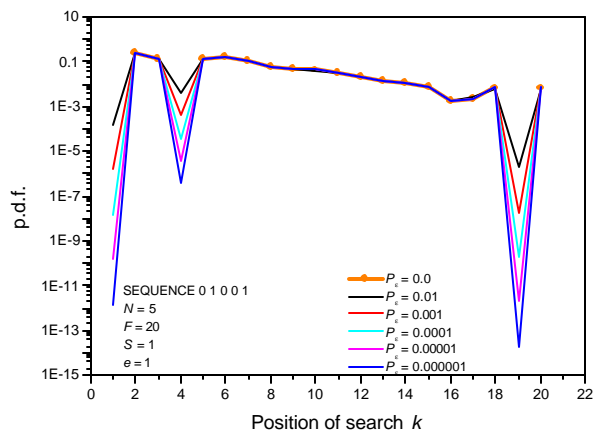


Fig. 9. P.d.f. in case of channel errors

## V. MEAN ACQUISITION TIME

This section deals with the evaluation of Häberle's curves [11], showing in brief the explanation of the model and a method for averaging the derived p.d.fs.

Numerous contributions dealing with CDMA has already explained the Markov-state diagram of the acquisition process. The model shown in Fig. 10 is similar, except for the two major differences: the first one is a different nature of the observed process, so the probabilities are basically different; the second difference is avoidance of z-transform for the expected value evaluation. Although there are numerous methods for solving the so-called "the first passage time problem" (fundamental matrix for absorbing processes, fundamental matrix for ergodic processes, Mason's formula, z-transform [21-25]), an extremely simple and computationally the most efficient method published long ago<sup>1</sup> [26] is preferred. This method does not enable the evaluation of the further moments or density functions, but a) our aim is to revisit the Häberle's curves and b) it is extremely rare to find an application where these statistical properties are evaluated if the model is complicated. The model itself is drawn for the worst case – *resynchronization* – when the search that starts immediately after the real sync. sequence position. Very bold line with probability 1 shows the renewal [23] property of the acquisition process, thus enabling a simple evaluation of acquisition time.

Each test is denoted by circular state, with corresponding dwelling time equalling to single bit duration. If a marker sequence is simulated at an incorrect position during the search, the same position in the next frame is tested (square states, with dwelling time depending upon the position in frame). This test is considered as statistically independent, as it does not overlap the previous tests. Finally, at the correct position, verification must be made, denoted by the rectangular state.

The parameters of the model are:

- $q_{Si}$  - the sync. sequence simulation probability at the  $i^{\text{th}}$  frame position, for the statistically independent tests,  $i = 1, \dots, F-1$ ;
- $t_i$  - dwelling time at the  $i^{\text{th}}$  position in frame, if a simulation has occurred at the position  $i$ ,  $i = 1, \dots, F-1$ ;
- $q_i$  - average sync. pattern simulation probability at the  $i^{\text{th}}$  position for the sliding window procedure (statistically dependent tests; if the search process is performed shifting not symbol by symbol, but frame by frame, these tests would be statistically independent as well, equalling to  $q_i$ );
- $q_{SF}$ ,  $q_F$  - probability that sync. sequence is found at its correct position.

The evaluation of these parameters is shown within the following subsections.

The simplest evaluation is the probability of the statistically independent tests. These probabilities depend upon the position within the frame and, in an error-free case, equal to

$$q_i = \begin{cases} \sum_{j=1}^e h_{1j}^{(n)} \cdot 0.5^{N-n}, & 1 \leq i < N \\ \sum_{j=0}^e \binom{N}{e} \cdot 0.5^N, & N \leq i \leq F-N, \\ \sum_{j=1}^e h_{j1}^{(n)} \cdot 0.5^{N-n}, & F-N < i \leq F \end{cases} \quad (16)$$

It is supposed that data bits are scrambled and therefore equiprobable.

Mean dwelling time at a specific position in frame if a simulation has occurred depends upon the monitoring and verification algorithms. Fig. 11 presents a state diagram of a general case, where indices "I" and "O" denote verification and monitoring, respectively, and the probabilities are statistically independent. In this case, monitoring and verification use the same marker, so indices in Eqs. (18), (19) and (21) are replaced with "S". Fig. 12 presents a general case of verification. Choices of parameters "J" and "K" reduce the general cases either to more common "reset" or to "random-walk" algorithms.

The corresponding values are published elsewhere [27] and just repeated here:

$$t_i = \frac{U_{Oi} \cdot V_{li}}{V_{Oi}} + (U_{li} + V_{li}) \quad (17)$$

$$V_{Oi} = \begin{cases} p_{Si}^K \cdot t_{Si}^{N-K} \cdot \frac{1-t_{Si}}{1-t_{Si}^{N-K+1}}, & t_{Si} \neq 1 \\ 2^{-K} \cdot (N-K+1)^{-1}, & t_{Si} = 1 \end{cases}$$

$$V_{li} = \begin{cases} q_{Si}^{J-1} \cdot t_{Si}^{M-J} \cdot \frac{1-t_{Si}}{1-t_{Si}^{M-J+1}}, & t_{Si} \neq 1 \\ 2^{-J+1} \cdot (M-J+1)^{-1}, & t_{Si} = 1 \end{cases}$$

$$U_{Oi} = \begin{cases} q_{Si}^{-1} \cdot \left( 1 + p_{Si}^K \cdot \frac{t_{Si}}{1-t_{Si}} - (N-K+1) \cdot V_{Oi} \right), & t_{Si} \neq 1 \\ 2^{-1} \cdot (1 + 2^{-K} \cdot (N-K+1)), & t_{Si} = 1 \end{cases}$$

$$U_{li} = \begin{cases} p_{Si}^{-1} \cdot \left( 1 + q_{Si}^{J-1} \cdot \frac{t_{Si}}{1-t_{Si}} - (M-J+1) \cdot V_{li} \right), & t_{Si} \neq 1 \\ 2^{-1} \cdot (1 + 2^{-J+1} \cdot (M-J+1)), & t_{Si} = 1 \end{cases}$$

<sup>1</sup> Unnoticed, due to the printing errata

$$p_{S_i} = 1 - q_{S_i}; \quad t_{S_i} = \frac{p_{S_i}}{q_{S_i}}; \quad (18)$$

The cases  $t_{S_i} = 1$  are of no practical importance.

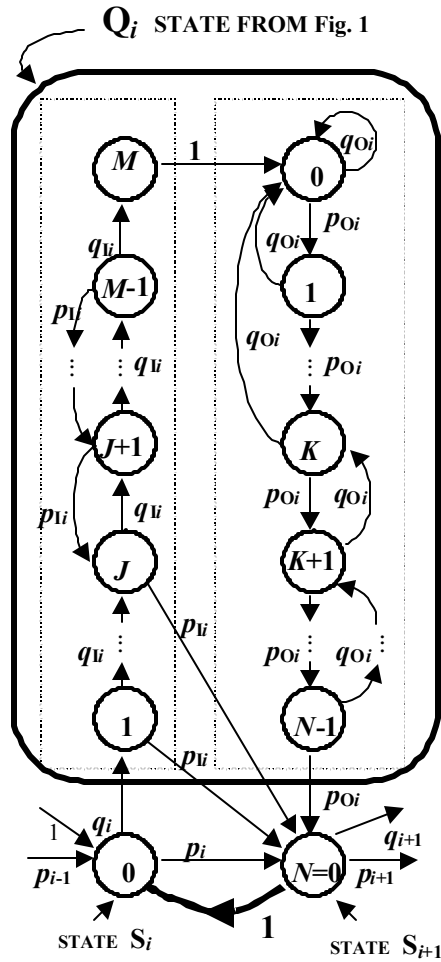
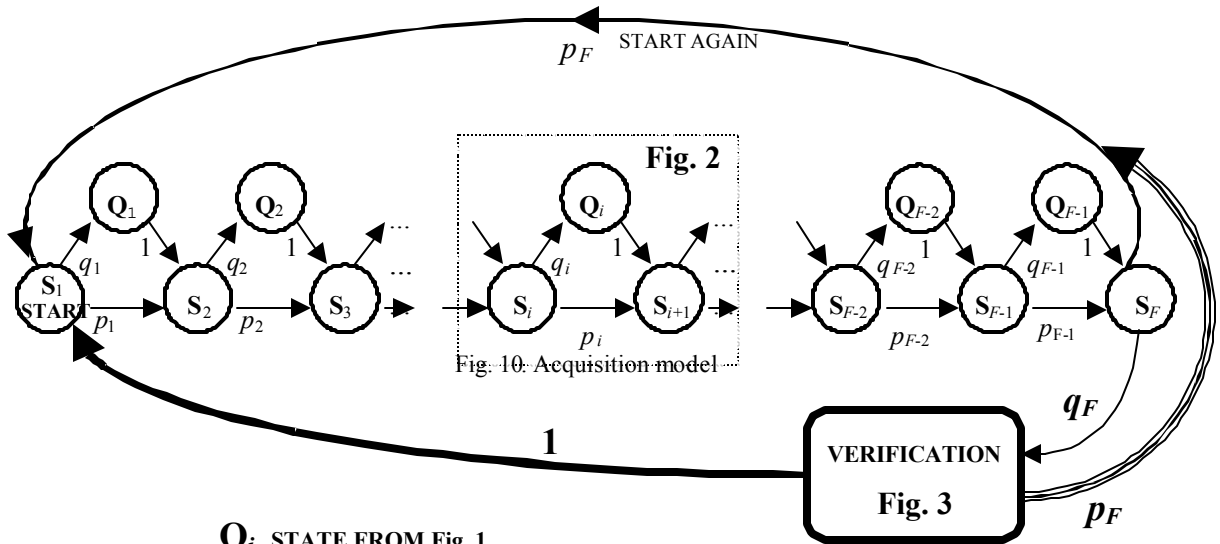


Fig. 11. Verification and monitoring

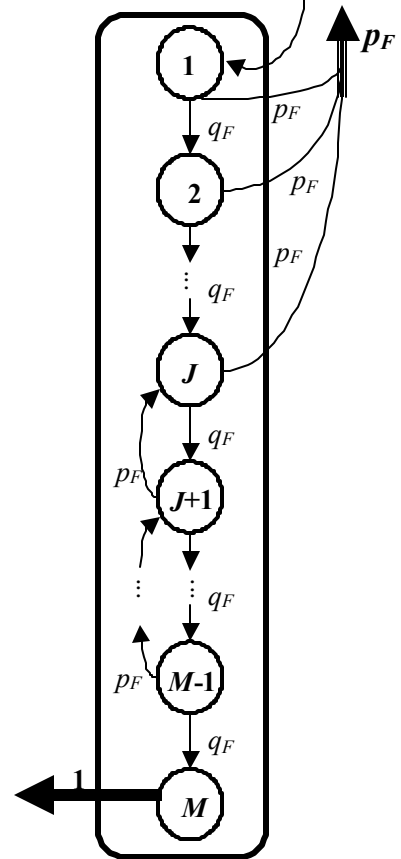


Fig. 12. Verification - general model

In a usual example of reset counters where  $J=M-1$  and  $K=N-1$ , expected value of resynchronisation time (acquisition with  $S=1$ ) equals to:

$$E\{t_{ACQ}\} = \frac{\frac{F-1}{F} + \sum_{i=1}^{F-1} q_i \cdot t_i + \frac{1-q_{SF}^M}{1-q_{SF}} \cdot \left( q_{SF} + \frac{1-q_{SF}}{F} \right)}{q_{SF}^M} \quad (19)$$

with the Eq. (17) reduced simply to

$$t_i = \frac{1-q_{Si}^{M-1}}{1-q_{Si}} + q_{Si}^{N-2} \cdot \left( \frac{1}{(1-q_{Si})^M} - 1 \right) \quad (20)$$

The only missing quantity is average simulation probability,  $q_i$

Within the previous part, the probability that a sync. marker would be found (simulated) for the first time at a specific position  $i$  in frame, given that the search has started at the position  $S$ . We shall denote this conditional probability  $\Pr\{k/S\} = \Pr\{i-S+1/S\}$ .

Since  $k-1$  tests have been made prior the successful one, this probability can be regarded as product of unsuccessful probabilities, with the probability of success at the frame position  $i$  (test no.  $k$ ):

$$\frac{\Pr\{k/S\}}{p_R(1-q_R((k-1)/S)) \cdot q_R(k/S)} = (1-q_R(1/S)) \cdot (1-q_R(2/S)) \cdot \dots \quad (21)$$

These probabilities can be evaluated using a recursion:

$$q_R(k/S) = q_R(i-S+1/S) = \begin{cases} \Pr\{1/S\}, & k=1 \\ \frac{\Pr\{1/S\}}{\prod_{m=1}^{k-1} (1-q_R(m/S))}, & 2 \leq k < F-S; \quad 1 \leq S \leq F \end{cases} \quad (22)$$

The variable  $i$  denotes the position in frame; variable  $k$  is number of test.

It is interesting to note that in [10] a formula (22) was stated and the conditional probabilities  $q_R(k/1)$  obtained using a simulation study. Our  $\Pr\{k/S\}$  formula was unknown at that time.

What we need is an average probability of simulation at a specific position within the frame. When a search has started, an eventual simulation leads to jumping to the next frame, so the search at the next frame has a fresh start. The averaging might be explained as follows:

The average probability of simulation at  $i$  equals to the probability the simulation has occurred, if the process has reached  $i^{\text{th}}$  position:

$$q_i = \frac{\Pr\{\text{simulate at } i/i \text{ is reached}\}}{\Pr\{i \text{ is reached}\}} = \frac{\Pr\{\text{simulate at } i \text{ and } i \text{ is reached}\}}{\Pr\{i \text{ is reached}\}} \quad (23)$$

Then it follows:

$$\Pr\{i \text{ is reached}\} = \sum_{m=1}^i \Pr\{\text{start}=m\} \cdot \Pr\{\text{come to } i\} = \sum_{m=1}^i p(m) \cdot \prod_{j=1}^{i-m} (1-q_R(j/S)) \quad (24)$$

$p(m) = \Pr\{\text{start}=m\}$  can be found using the recursion:

$$\begin{aligned} p(m) &= \Pr\{\text{start}=m\} = \Pr\{\text{start}=1\} \cdot \Pr\{\text{simulate at } m\} + \\ &\Pr\{\text{start}=2\} \cdot \Pr\{\text{simulate at } m\} + \dots \\ &\dots + \Pr\{\text{already at } m\} \cdot \Pr\{\text{simulate at } m\} = \\ &= \sum_{j=1}^{m-1} p(j) \cdot \Pr\{m-j+1/j\} + p(m) \cdot q_R(1/m), \end{aligned} \quad (25)$$

It should be remembered that  $\Pr\{1/m\} = q_R(1/m)$ , so

$$p(m) = \frac{\sum_{j=1}^{m-1} p(j) \cdot \Pr\{m-j+1/j\}}{1 - \Pr\{1/m\}}, \quad (26)$$

$$p(1) = 1$$

$$\begin{aligned} \Pr\{\text{simulate at } i \text{ and } i \text{ is reached}\} &= \\ &= \sum_{m=1}^i \Pr\{\text{start}=m\} \cdot \Pr\{\text{simulate at } i\} = \\ &= \sum_{m=1}^i p(m) \cdot \Pr\{i-m+1/m\} = \\ &= p(i) \cdot (1 - \Pr\{1/i\}) + p(i) \cdot \Pr\{1/i\} = p(i) \end{aligned} \quad (27)$$

Finally, the required mean simulation probability is:

$$q_i = \begin{cases} \Pr\{1/1\}, & i=1 \\ \frac{\sum_{m=1}^i p(m) \cdot \Pr\{i-m+1/m\}}{\sum_{m=1}^i p(m) \cdot \prod_{j=1}^{i-m} (1-q_R(j/S))}, & i>1 \end{cases} \quad (28)$$

In order to show the difference, the probabilities are plotted in Fig. 13:

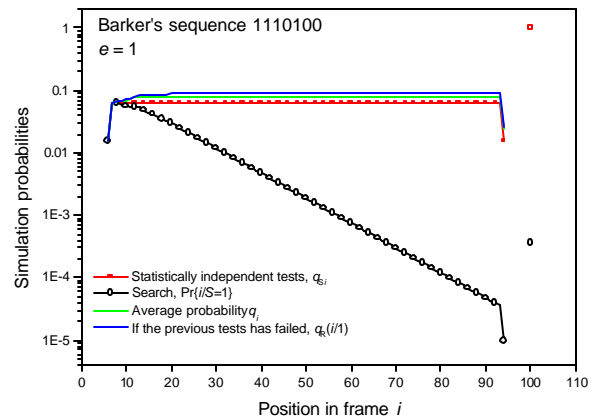


Fig. 13. Different types of simulation probabilities



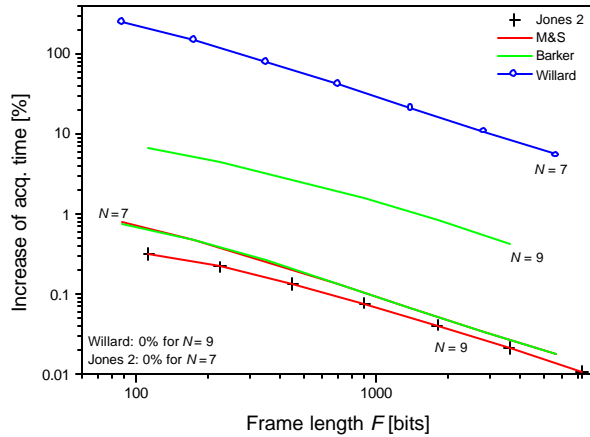


Fig. 14. Increase of Acquisition time

Fig. 14 shows that famous properties of Barker sequence does not mean that it would have better  $E\{t_{ACQ}\}$  then any other sequence: Jones' sequence outperforms it. Figs. 15. and 16 presents the final result – the curves if the errors are allowed. It is interesting to see that minimal marker length ( $N = 1$  if  $e = 0$ , and  $N = 3$  if  $e = 1$ ) lie along the same line. It is obvious (and predictable) that  $E\{t_{ACQ}\}$  is considerably longer if a distance from a correct marker is allowed in error-free condition, but achieves minimum if transmission is erroneous.

The result presented in Fig. 17 is interesting: 17a shows the acquisition time related to Jones 0 sequences ( $N=9, e=2$ ), while 17b shows the cross-bifix spectrum (number of non-zero elements in non-default cross-bifix matrices). The best (smallest) acquisition time have sequences with the least non-zero elements, a criterion that could be used for sequence optimization.

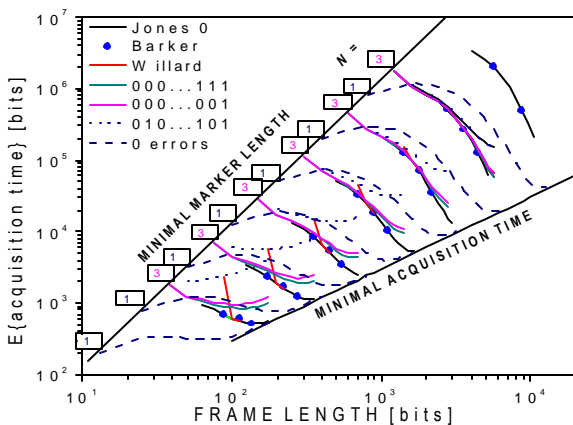


Fig. 15. Haerle curves in error-free condition

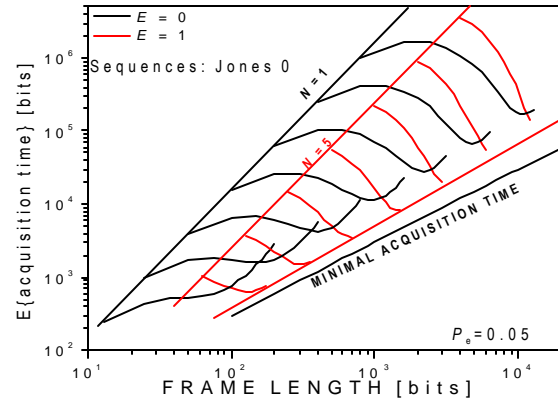


Fig. 16. Haerle curves with transmission errors

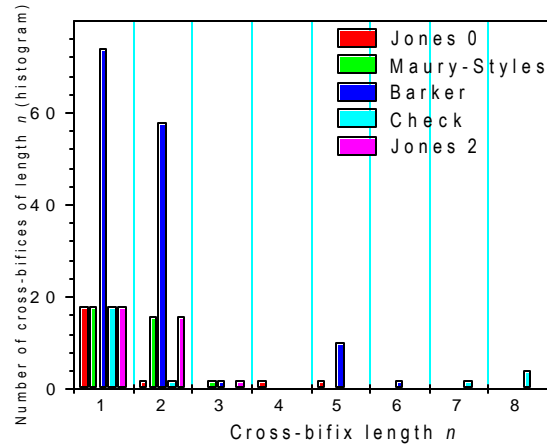
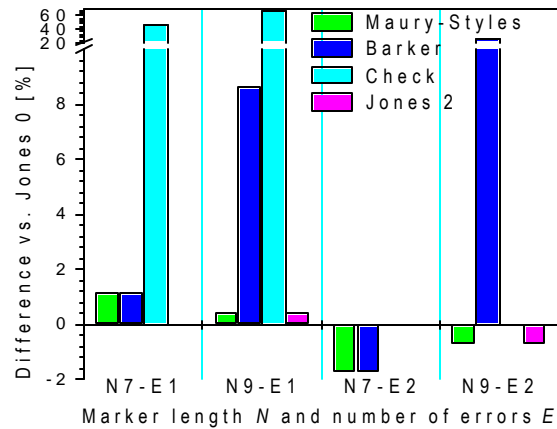


Fig. 17.  $N=9, e=2$  sequences: a) Acquisition time related to Jones 0 sequence; b) Corresponding cross-bifix spectrum

## VI. CONCLUSION

A method for sequence analysis is shown and illustrated by derivation of the mean acquisition time for frame re-alignment. This mathematical tool might have other applications, and further research is encouraged by the post-presentation discussion [27, 28, 29] and comments of the anonymous reviewer [19], with a suggestion to turn attention to QPSK sequences.

At last, although “marker synchronization” always work, there are cases when careless approach lead to suboptimal solutions [30-32].

## REFERENCES

- [1] J.L. Massey: “Optimum Frame Synchronization”, *IEEE Trans. on Comm.*, pp. 115-119, Vol. COM-20, April 1972.
- [2] R. A. Scholtz, “Frame synchronization techniques,” *IEEE Trans. Commun.*, vol. COM-28, pp. 1204-1213, Aug. 1980.
- [3] P.T. Nielsen: “Some Optimum and Suboptimum Frame Synchronizer for Binary Data in Gaussian Noise”, *IEEE Trans. on Comm.*, pp. 770-772, Vol. COM-21, June 1973.
- [4] G.L. Lui and H.H. Tan: “Frame Synchronization for Gaussian Channels”, *IEEE Trans. on Comm.*, pp. 818-828, Vol. COM-35, August 1987.
- [5] T. Sekimoto and H. Kaneko: “Group Synchronization for Digital Transmission Systems”, *IRE Trans on Commun. Syst.* pp. 381-390, December 1962.
- [6] K. Brayer: “Frame synchronization for binary data transmission”, *Electronics Letters*, pp. 392-393, Vol. 7, July 1971.
- [7] O. Brugia and M. Decina: “Reframing statistic of P.C.M. multiplex transmission”, *Electronics Letters*, pp. 625-627, Vol. 7, July 1971.
- [8] D.E. Dodds, S.-M. Pan and G.W. Wacker: “Statistical distribution of PCM Framing Times”, *IEEE Trans. on Comm.*, pp. 1236-1241, Vol. COM-36, November 1988.
- [9] J. Lindner, “Binary sequences up to length 40 with best possible autocorrelation function,” *Electron. Lett.*, vol. 11, p. 507, 1975.
- [10] M.N. Al-Subbagh and E.V. Jones: “Optimum patterns for frame alignment”, *IEE Proc. part F - Commun. Radar & Signal processing*, Vol. 135 (6), pp 594-603, December 1988.
- [11] H. Häberle: “Frame synchronizing PCM systems”, *Electrical Communications*, Vol. 44, No. 4, pp. 280-287, 1969.
- [12] P.T. Nielsen: “On the Expected Duration of a Search for a Fixed Pattern in Random Data”, *IEEE Trans. on Inf. Theory.*, pp. 702-704, Vol. IT-19, September 1973.
- [13] T. McConell: “The Expected Time to Find a String in a Random Binary Sequence” [<http://barnyard.syr.edu/cover.pdf>], January 2001.
- [14] D. Bajic, D. Drajić: “Duration of search for a fixed pattern in random data: Distribution function and variance”, *Electronics Letters*, 1995, Vol. 31. No. 8, pp 631-632.
- [15] D. Bajic, D. Drajić: “Search Process: The Analysis”, *Facta Universitatis*, Series: Electronics and Energetics, 1996. Vol. 9 , No. 2, pp 191-205.
- [16] D. Bajic, J. Stojanovic and J. Lindner: “Multiple Window-sliding Search”, Proceedings of 2003 IEEE International Symposium on Information Theory (ISIT-2003), Yokohama, Japan, June 2003.
- [17] A. J. de Lind van Wijngaarden, T. J. Willink: “Frame Synchronization Using Distributed Sequences”, *IEEE Transactions on Communications*, 2000, Vol. 48, No. 12, pp 2127-2138.
- [18] C. Stefanovic and D. Bajic: “Search in non-equiprobable data”, submitted
- [19] D. Bajic: “On survival probability of alignment sequences”, Proceedings of 2004 International Symposium on Information Theory and Applications, (ISITA-2004), Parma, Italy, October 2004.
- [20] S. Mason and H. Zimmerman: **Electronic Circuits, Signals and Systems**, Wiley, New York, 1960.
- [21] D.L. Isaacson and R.W. Madsen: **Markov Chains Theory and Applications**, Wiley, 1976.
- [22] R.W. Wolfs: **Stochastic modeling and the theory of queues**, Prentice Hall International Inc, Englewood Cliffs, NJ, 1989.
- [23] R.A. Howard: **Dynamic Programming and Markov Processes**, MIT Press, Cambridge, Massachisets, 5<sup>th</sup> edition, 1969.
- [24] J.G. Kemeny and J.L. Snell: **Finite Markov Chains**, D. van Nostrand Company inc., 1960.
- [25] D. Bajic: "New simple method for solving first passage time problem", *Electronics Letters*, 1991, Vol. 27. No. 16, pp 1419-1421.
- [26] D. Bajic, J. Stojanovic: “Frame -alignment procedures for STM-1 frame”, *IEE Proceedings, Communications*, Vol. 150, No. 1, February 2003, pp. 37-44.
- [27] Dragana Bajic, Jakov Stojanovic: “Distributed Sequences and Search Process”. *IEEE International Conference on Communications ICC2004*, Paris, France, June 2004, CT08-6, ISBN: 0-7803-8534-9
- [28] N. Noels, H. Steendam and M. Moeneclaey: “Pilot-symbol assisted iterative carrier synchronization for burst transmission *IEEE International Conference on Communications ICC2004*, Paris, France, June 2004, CT08-5, ISBN: 0-7803-8534-9
- [29] Heon Huh, Tarkesh Pande and James V. Krogmeier: “Decoder-Assisted Frame Synchronization in the

- Presence of Phase/Frequency Noise" *IEEE International Conference on Communications ICC2005*
- [30] D. Bajic, V. Šenk, M. Despotovic: "Subsets of the STM-1 frame-alignment signal: a monitoring analysis", *IEE Proceedings, Communications*, Vol. 149, No. 5, October 2002, pp. 242-248.
- [31] D. Bajic, J. Stojanovic: "In-frame procedures for STM-1 frame", Proceedings of ICC2001, Vol. 9, pp 2768-2773, Helsinki Finland, June 2001.
- [32] \*D. Bajic, J. Stojanovic: "Frame-alignment monitoring for STM-1 frame", Proceedings of ICC2001, Vol. 10, pp 2957-2961, Helsinki Finland, June 2001.