Notice of Violation of IEEE Publication Principles

"Viterbi Algorithm in Continuous-Phase Frequency Shift Keying"
by Hailun Tan
in the Proceedings of the International Conference on Digital Communications, 2006 (ICDT'06), August 2006

After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.

This paper is a revised version of the following paper. The lead author, Liang Miao, was not informed of the revisions.

"Application of Viterbi Algorithm in Binary Continuous Phase FSK"
by Liang Miao, Hailun Tan
in a Research Report presented at Australia National University, October 2003

Viterbi Algorithm in Continuous-Phase Frequency Shift Keying

Hailun Tan
Australian National University, Faculty of Engineering & Information Technology
u2532164@anu.edu.au

Abstract

In this paper, the theory of Viterbi Algorithm is introduced based on convolutional coding. The application of Viterbi Algorithm in the Continuous-Phase Frequency Shift Keying (CPFSK) is presented. Analysis for the performance is made and compared with the conventional coherent estimator and the complexity of the implementation of the Viterbi decoder in hardware device is analyzed. At last the relevant conclusion is given.

Index Terms – Viterbi Algorithm, Convolutional Coding, Binary Continuous Phase FSK, Markov Process, MAP, MLSE, BER

1. Introduction

1.1 Background Introduction

Viterbi algorithm (VA) can be viewed as a solution of estimation for a finite sequence from Markov process through memoryless noise channel as illustrated in Figure 1:

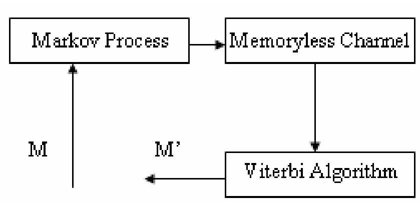

Figure 1: General Model for Viterbi Algorithm

Sequence detection with Viterbi decoding has been widely considered for the detection of signals with memory[1]. It was originally invented to decode convolutional codes[2]. So, the introduction of the Viterbi algorithm will mainly be based on the decoding process for the convolution coding.

There exist several statistical tools in VA estimation such as the Maximum A posteriori Probability (MAP)[3] and Maximum Likelihood Sequence Estimation (MLSE) [1, 4-7].

1.2 MAP and MLSE

MAP and MLSE can be both viewed as a derivation from the BAYES Estimation[8, 9]. In BAYES criterion, two notations are made:

- the priori probabilities (denoted as P (H_0) and P (H_1))
- The cost to each possible decision (denoted as C_{ij}), i, j = 0, 1, as the cost associated with the decision D_i given that the true hypothesis is H_j. Hence, the decision rule resulting from the BAYES criterion is:

$$\frac{f_{Y/H1}(Y/H_1)}{f_{Y/H_0}(Y/H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})}$$

In MAP, let the costs

Cii = 0, i = 0, 1

Cij = 1, i $\neq$ j and i, j = 0, 1

Hence, minimizing the risk is equivalent to minimizing the probability of error. Then, the decision rule is reduced to

$$P(H_1/y) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0/y)$$

In MLSE, let

$$P_0(C_{10}-C_{00}) = P_1(C_{01}-C_{11})$$

It yields:

$$P(y/H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(y/H_2)$$

1.3 Viterbi Decoder

The convolution encoder is basically a finite-state machine, and the VA decoding is done on the optimal decoder based on MLSE[4].Figure 2 shows the block diagram in which the convolutional coding and Viterbi decoding are applied.

0-7695-2650-0/06/$20.00 (c) IEEE

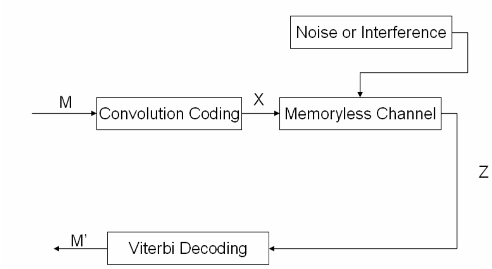

Figure 2: Viterbi Decoder for the convolutional coding

The Viterbi decoding involves the search through the trellis for the most likely sequence. i.e., to select the path with the maximum probability of P (z|x), where z is the received sequence and x is the signal which is needed to estimate. If the error probability in channel is P (1|0) =P (0|1) = p, L is the sequence length, and e is the number of error bits at the same time $P(z/x) = p^e(1-p)^{L-e}$

$$\log P(z/x) = L\log(1-p) - e\log(\frac{1-p}{p}) = -A - Be$$

, where $A = L\log\frac{1}{1-p}$, $B = \log(\frac{1-p}{p})$

They are both positive constants[1].

It means that finding the maximum likelihood path in trellis diagram is equivalent to finding the minimum hamming distance.

Figure3 (series) shows the bit sequences in the convolutional coding and decoding process. Figure 3-1 is the original signal with length of 20. Figure 3-2 is the sequence after the (2, 1, 3) convolutional coding. The sequence length is enlarged to 40 so that the distance of the signal is enlarged. Figure 3-3 is the received sequence through Additive White Gaussian Noise (AWGN) channel with Bit Error Rate (BER) of 0.1. The signals in circle are the error bits. Figure 3-4 is the sequence after the VA estimation. Obviously, the error bits are all corrected.

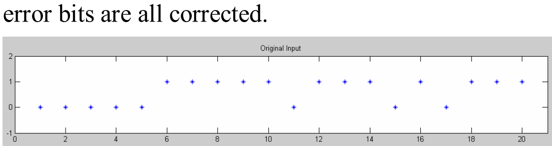

Figure 3-1: The randomly generated signal

[1] Assume the error probability rate (p) is smaller than 0.5. Otherwise, the communication does not make any sense if the error case is more likely to happen than the normal case.

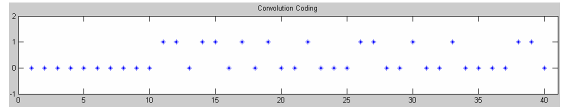

Figure 3-2: output sequence of the (2, 1 ,3) convolution coding

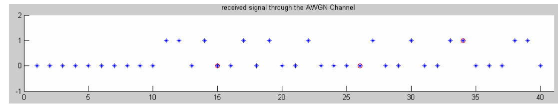

Figure 3-3: the received signal through the AWGN channel

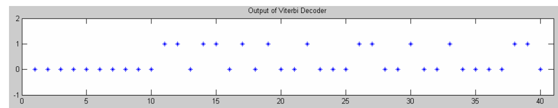

Figure 3-4: the output of the Viterbi Decoder

It is known that BER of the sequence through an AWGN with Pulse Modulation is:

$$P_e = Q(\sqrt{2R_c\, E_b/n_0}) \approx (4\pi R_c\, E_b/n_0)^{-1/2} e^{-R_c E_b/n_0}$$

If the convolution coding and VA decoding method is not applied, the BER will be:

$$P_b = Q(\sqrt{2E_b/n_0}) \approx (4\pi\, E_b/n_0)^{-1/2} e^{-E_b/n_0}$$

In [5], it has been shown that for the (2, 1, 3) convolutional coding and VA decoding, the BER will be:

$$P_b \approx \frac{M(d_{free})}{k} \cdot 2^{d_{free}} \cdot P_e^{d_{free}/2} = 2^5 P_e^{5/2}$$

Figure 4 clearly shows the performance of the coding and decoding scheme in the channels with different Signal/Noise Ratio (SNR).

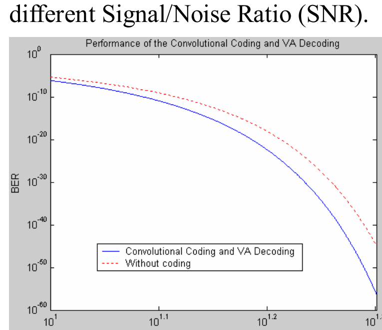

Figure 4: The performance of convolution coding and Viterbi decoding

The scheme of convolutional coding and Viterbi decoding does not show a significant improvement in BER illustrated in Figure 4. However, d_{free} will increase linearly to the constraint length and it will lead to exponential decrease of BER. So, great improvement can be expected if the constraint length is increased. However, the computational cost of the decoder will also increase exponentially with the constraint length and it will be discussed later. The trade-off should be made between the performance and the complexity.

The rest of the paper is organized as follows. Section 2 will describe the problem I want to solve. VA will be incorporated into BCFSK in section 3. The relevant simulation and analyze on the VA in BCFSK is analyzed in section 4. The implementation issues and the respective conclusion will be presented in section 5 and 6, respectively.

2. Problem Formulation

Since the spectral characteristics of multi-level conventional orthogonal FSK occupy much wider bandwidth than those of binary FSK. Moreover, error performances of multi-level Continuous Phase Modulation (CPM) is deteriorated seriously in the fast fading environment[10], the emphasis is on the Binary Frequency Shift Keying (BFSK).

In BFSK, the frequency of a constant amplitude carrier signal is switched between two values according to the two possible message states (high (1)/low (0) tone). Depending on how the frequency variations are imported into the transmitted waveform, FSK signal will have either a discontinuous or continuous phase between bits. In this paper, I will concentrate on the Continuous-Phase FSK (CPFSK). In general, the CPFSK signal can be represented as:

$s(t) = \cos[w(u_k)t + \theta_k], kT \le t < (k+1)T$

Where w (u_k) is the frequency selected by u_k, and θ_k is some phase angle. In CPFSK, it is required the condition ($w(u_{k-1})kT + \theta_{k-1} \equiv w(u_k)kT + \theta_k \mod 2\pi$) to ensure the phase is continuous.

In order to demonstrate how it works, the simplest binary FSK is processed as follows: The input sequence u is assumed binary. w (0) and w (1) are designed as w(0) goes through an integer number of cycles in T seconds and w(1) through an odd half-integer number; i.e., $w(0)T \equiv 0$ and $w(1)T \equiv \pi \mod 2\pi$ [3].

Then a two-state process can be obtained with the state $X = \{0, \pi\}$. The transmitted signal y (t) can be written as:

$$y(t) = \cos[\cos(u_k)t + x_k] = \cos x_k \cos w(u_k)t \quad kT \le t < (k+1)T.$$

Let $s_0(t) = \cos w(0)t$ and $s_1(t) = \cos w(1)t$ as bases of the signal space, I can write

$$y_k = y_{0k}s_0(t) + y_{1k}s_1(t)$$

where

$$(y_{0k}, y_{1k}) = \begin{cases} (1,0) & if\,(u_k = 0, x_k = 0) \\ (-1,0) & if\,(u_k = 0, x_k = \pi) \\ (0,1) & if\,(u_k = 1, x_k = 0) \\ (0,-1) & if\,(u_k = 1, x_k = \pi) \end{cases}$$

If the signal is transmitted in the AWGN channel, the received signal can be denoted as:

$$Z_k = (z_{0k}, z_{1k}) = (y_{0k}, y_{1k}) + (n_{0k}, n_{1k})$$

From the equations above, I can get the trellis diagram as Figure 5 illustrates.

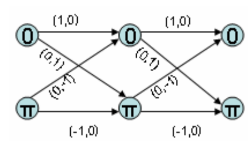

Figure 5: the trellis diagram for BCFSK

3. Viterbi algorithm in CPFSK

Based on the setting in above section, VB can be implemented in Binary CPFSK(BCPFSK). MAP will be used to select the shortest path in the trellis diagram.

It is known that the probability of bit errors (Pe) in this model is estimated as $\mathrm{Pe} = Q(\frac{\sqrt{2}}{2\sigma})$. Assume that the priori probability of the binary signal is the same (i.e.: P (1) = P (0) =0.5), it is easy to show that the probability to shift from one state to another is the same. Without losing generality, the transition between two states is defined as Figure6:

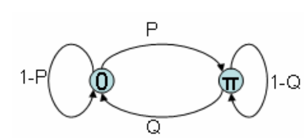

Figure 6: the state diagram of BCPFSK

Two survival sequences in the k[th] step are denoted as Γ_{k1} and Γ_{k2}. The Viterbi recursive equation can be written as the following three stages:

1. Initialize: k=0, $\Gamma_{01} = \Gamma_{02} = 0$, and assign the memory space for state sequence X (k) where X (k) is a 2×Ke array. Ke is the length of the sequence.

2. For step k:

$if\,(y_{0k}, y_{1k}) = (1,0)$

$\Gamma_{k1} = \min\{[\Gamma_{k-1,1} - \ln(1-P) - \ln(1-Pe)], [\Gamma_{k-1,2} - \ln(Q) - \ln(Pe)]\}$

$\Gamma_{k2} = \min\{[\Gamma_{k-1,1} - \ln(P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(1-Q) - \ln(Pe)]\}$

$if\,(y_{0k}, y_{1k}) = (0,1)$

$\Gamma_{k1} = \min\{[\Gamma_{k-1,1} - \ln(1-P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(Q) - \ln(Pe)]\}$

$\Gamma_{k2} = \min\{[\Gamma_{k-1,1} - \ln(P) - \ln(1-Pe)], [\Gamma_{k-1,2} - \ln(1-Q) - \ln(Pe)]\}$

$if\,(y_{0k}, y_{1k}) = (0,\text{-}1)$

$\Gamma_{k1} = \min\{[\Gamma_{k-1,1} - \ln(1-P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(Q) - \ln(1-Pe)]\}$

$\Gamma_{k2} = \min\{[\Gamma_{k-1,1} - \ln(P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(1-Q) - \ln(Pe)]\}$

$if\,(y_{0k}, y_{1k}) = (\text{-}1,0)$

$\Gamma_{k1} = \min\{[\Gamma_{k-1,1} - \ln(1-P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(Q) - \ln(Pe)]\}$

$\Gamma_{k2} = \min\{[\Gamma_{k-1,1} - \ln(P) - \ln(Pe)], [\Gamma_{k-1,2} - \ln(1-Q) - \ln(1-Pe)]\}$

3. Sequence Γ_{k1}, Γ_{k2} and state sequence X (k) are stored, and stage 2 is executed for step k+1. With the finite state sequences X (k), the shortest complete path will be got and the corresponding signal sequence can be estimated.

4. Simulation Evaluation

4.1 Performance Analysis

It is known that the probability of any error event starting at time k may be upper-bounded or lower-bounded[2, 3]. In the model of CPFSK, the bound can be calculated as follows: two modulated signal are assumed to be:

$$s_0(t) = \sqrt{2E_b}\cos(w(0)t + \theta_1)$$

$$s_1(t) = \sqrt{2E_b}\cos(w(1)t + \theta_2)$$

From step k, the error events can happen at step k, k+1, k+2, k+3… These can be seen in Figure 7. (e.g.: the correct path should be 0-0-0-0, but due to the error events in the middle, the path could become $0-\pi-\pi-0$)

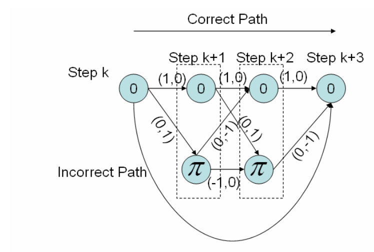

Figure 7: The analysis of the error bound

The probability can be obtained from the Euclidean Distance between the incorrect sequence and the correct sequence. For the sequence of step k+1, the Euclidean Distance is $2\sqrt{E_b}$, the probability for this particular error event is $Q(\frac{\sqrt{E_b}}{\sigma})$. It is the lower bound of Pe[2]. The probability is accumulated for all the error events in subsequent steps. The range of Pe can be obtained as

$$Q(\frac{\sqrt{E_b}}{\sigma}) < Pe < Q(\frac{\sqrt{E_b}}{\sigma}) + Q(\frac{\sqrt{3E_b}}{\sigma}) + \cdots Q(\frac{\sqrt{(2k+1)E_b}}{\sigma})$$

The right part of this formula is bounded because $Q(x)$ decreases rapidly with x. So, the Pe can be estimated as $Q(\frac{\sqrt{E_b}}{\sigma})$. Since the one-side white noise power spectral density (denoted as n_0) $n_0 = \frac{1}{2}\sigma^2$, so the Pe can be simplified as:

$$Pe = Q(\frac{\sqrt{2E_b}}{2\sqrt{n_0}})$$

However, the Coherent Detection of Binary FSK in [11] is

$$Pe = Q(\frac{\sqrt{E_b}}{\sqrt{n_0}}),$$

So Pe (probability of bit error) is reduced from $Q(\frac{\sqrt{E_b}}{\sqrt{n_0}})$ to $Q(\frac{\sqrt{E_b}}{\sqrt{2n_0}})$ when the VA is used in BCPFSK.

In order to verify my deduction, the relevant simulation of BCFSK is processed in Matlab. Figure 8 depicts that when the scale of Bit Error Rate (BER) is

[2] The error only happens in the first step of sequence.

under 0.01dB, the detection based on Viterbi Algorithm improves SNR by 3dB, which matches the mathematical analysis above. Actually, the VA algorithm gives an ideal performance, which gives a benchmark for the sub-optimum decoding methods.

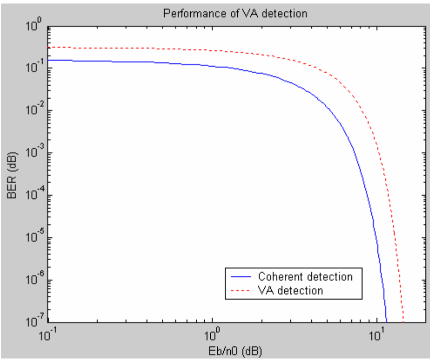

Figure 8: the Performance of VA in BCFSK

4.2 Memory Cost

It can be seen that due to the recursive feature, VA does not need much storage space in estimation process of MAP or MLSE. Most of the memory will be allocated for the storage of the state sequence X (k). The storage size of the sequence is $m \times L$ where m is the number of states and L is the truncated length of the sequence.

4.3 Computational Cost

In BCFSK, the decoder will do the job of plus-compare-select twice for each additional step. Similarly, the job of plus-compare-select will be done m times in m states trellis decoding process for each input sequence. Hence, the total cost for the sequence with length (L) and M states will be $O(ML)$. In some high-speed Digital Signal Processor such as TMS320C series and ADSP21000 series, with processing speed of thousands of MIPS (million instructions per second), it is not difficult to implement VA in perspective of hardware.

5 Implementation issues

There are several issues that need to address in the future so that VA can be applied into more Digital Signal Processing (DSP) applications.

5.1 The sequence length

From this model, the decision of the estimation won't be made until the entire signal is received. Hence, it leads to the problem that VA will not work if the state sequence is infinite. It is necessary to truncate the survivors to some manageable length (L) under such circumstance. L should be chosen large enough, so that at time K = n*L, (n is positive integer). All the survivors are in the same state sequence. Hence, the length L is related to the power of the noise in the channel obviously. On the other hand, it should not be too long otherwise it will introduce the unnecessary delay and lower the efficiency of decoding. The trade-off has to be made between accuracy and efficiency.

5.2 Last step in BCPFSK

The state sequence X (k) is estimated only when the survivor length Γ_{k+1} is calculated because the longer the sequence is, the more memory will be hogged. On the other hand, the probability of decision error decreases accordingly. At the last step (denoted as K_{last}), there are still m states left for the next step. So there are $\Gamma_{Klast+1}$, $\Gamma_{Klast+2}$ to make these sequences to converge at the same node at step K_{last}. If the decision is made only according to Γ_{klast}, the probability of the error in the last state will be much greater than that in previous states.

Some dummy sequence will be used to drive the state sequence to go on at the end of the truncated sequences. If the length of the dummy sequence is Le, then Le should be chosen large enough so that at step K_{last} + Le, all the m sequences will converge at the same node at step $K_{last} + L_e$. Under such circumstance, the decision can be made with state sequence stored in the X (k) because at step $K_{last}+L_e$, all the state sequences from step 1 to step $K_{last} + L_e$ are the same. The length Le depends on the number of the states m and the power of the noise[3]

5.3 Initial state

The algorithm is required to start with knowing the initial state X (0), but it might not be satisfied in practice. If K_e is the state that all the possible sequence in CFSK converges into, then VA can take K_e as the

[3] It means that if the number of states is large or the noise is significant, it will take a longer dummy sequence (L_e) for all m sequences to converge into one state as the final state.

initial state, which is proved to be finite in[1].

Figure 9 depicts the problems in 5.2 and 5.3. It shows the output state sequence of the Viterbi decoding in convolutional coding. All the sequences converge together for the first 12 steps but diverge after step 12. So, the initial state can be decided as state 1. Actually, the states in first 12 steps can be precisely decided in this case. So to the problem in 5.3, the initial state can be the state in any step between step 0 and step 12. To the problem in section5.2, the sequences diverge from step 12 to step 16. The extra more steps are required after step 16 to make these sequences converge again so that all the decision on path selection can be made precisely.

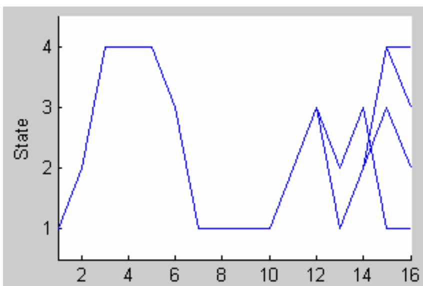

Figure 9: The output state sequence in convolution coding

6 Conclusion

VA is applied in MLSE and MAP in this paper. In perspective of MLSE, the performance is related to the constraint length of the convolutional coding. The performance will be significantly improved when the constraint length is long enough. In the perspective of MAP, VA gives an ideal BER for the traditional coherent decoder. Though Viterbi Decoder is not widely used in CPFSK, its performance gives a benchmark for the other simpler sub-optimum decoder.

VA is also analyzed in the perspective of memory requirement and computational cost. The recursive characteristic of the algorithm makes it relatively easy to implement in some high-speed programmable hardware devices. The concept of VA is to adopt extra memory to enlarge the distance of the signal sequence or the states. For MLSE, convolutional coding encodes the signal sequence with the encoding ratio of ½ and the constraint length of 3.Hence, the maximum distance of the binary signal is enlarged from 1 to 2.For MAP, the application of phase and frequency of FSK modulation can be utilized as a ½ ratio encoding method for the binary sequence to enlarge the signal distance. VA fully depends on the constraint characteristics of the signal. The statistical methods such as MLSE and MAP are used to extract the original signal with the presence of noise. So, the memory feature and the statistical methods make the signal more robust to the noise.

With these characteristics, VA has a bright future in wireless communication. For example, Direct Sequence Code Division Multiple Access (DS-CDMA) is an active research area with the objective of efficient channel sharing and utilization strategies. A fundamental problem in cellular DS-CDMA systems is interference. While some research has already shown that the channel coding/decoding algorithms provide a significant improvement in the performance of Multi-user detectors. VA will be widely used to improve the performance of the detectors.

Reference:

[1]. G. David Forney, J., *Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference.* IEEE, Trans, Inf. Theory, 1972.

[2]. Viterbi, A.J., *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm.* IEEE Trans. Information Theory, 1967. **13**(2): p. 260-269.

[3]. G. David Forney, J., *The Viterbi Algorithm.* Proceedings of IEEE, 1973. **61**(3).

[4]. Proakis, J.G., *Digital Communications*: McGraw Hill.

[5]. Zhigang Cao, Y.Q., *Communication Theory*: Tsinghua University Press.

[6]. Piret, P., *Convolutional Codes, an algebraic approach*: MIT Press.

[7]. Chari, M.J.R.a.S., *Demodulation of Cochannel FSK Signals Using Joint Maximum Likelihood Sequence Estimation.* IEEE, Trans., 1993. **2**: p. 1412 - 1416

[8]. Narkat, M., *Signal Detection and Estimation*: Artech House.

[9]. Dai, M., *Information Theory*: Tongji University Press.

[10]. Jun Horikoshi, K.K.a.T.O. *Error Performance Improvement of Bandlimited QTFSK Assisted by an Efficient Coded Modulation and Viterbi Sequence Estimation*. in *IEEE, Trans, Infomation. Theory*. 1994.

[11]. Rappaport, T.S., *Wireless Communications, Principles and Practice*: Prentice Hall.