

# Trellis State Aggregation for soft decoding of Variable Length Codes

Herve Jégou  
IRISA/University of Rennes  
Email: Herve.Jegou@irisa.fr

Simon Malinowski  
IRISA/INRIA  
Email: Simon.Malinowski@irisa.fr

Christine Guillemot  
IRISA/INRIA  
Email: Christine.Guillemot@irisa.fr

Campus Universitaire de Beaulieu  
Avenue du Général Leclerc  
35042 RENNES Cedex - France

**Abstract**—This paper describes a new set of state models for soft decoding of Variable Length Codes. A single parameter  $T$  allows to trade complexity against estimation accuracy. The extrema choices for this parameter lead respectively to construct the well-known bit-level and bit/symbol trellises. For a proper choice of the parameter  $T$ , the results obtained by running a BCJR or Viterbi estimation algorithm on the proposed state models are close to those obtained with the optimum state model. The complexity is however significantly reduced. It can be further decreased by projecting the state model on two state models of reduced size, and by combining their decoding results. This combination is shown to be optimal for the Viterbi algorithm.

## I. INTRODUCTION

Variable Length Codes are widely used in modern compression codecs such as MPEG4 or JPEG2000. However, such codes are very sensitive to channel noise: a single bit error may dramatically propagate through the bitstream. In recent years, many authors have considered the problem of soft decoding or joint source/channel decoding of Variable-Length Codes (VLCs) [1][2][3][4][5][6][7][8]. Indeed, the channel model and/or measures may be exploited to correct some errors, leading to schemes that significantly outperform the instantaneous decoding of the received bitstream. Two trellises have been considered to estimate the emitted sequence: the bit-level trellis proposed in [1] and the bit/symbol trellis. Similar to convolutional codes, the estimation uses either the Viterbi algorithm [10] or the BCJR algorithm [9]. Let us recall that the Viterbi algorithm minimizes the sequence error rate (SQER). This algorithm returns the sequence with the highest probability according to the source model and the channel measures. Similarly, the BCJR algorithm optimally estimates the marginal probabilities of either the emitted bits or the internal states of the decoder. This algorithm is also known as the forward/backward algorithm in the literature. Hence, the bit/symbol trellis coupled with the BCJR algorithm [9] allows to obtain the optimum estimates minimizing either the Bit Error Rate (BER) or the Symbol Error Rate (SER) [3].

However, the number of states of the bit/symbol trellis is a quadratic function of the sequence length. The corresponding

complexity limits the range of sequence lengths to which the method applies. The complexity is actually not tractable for typical sequence lengths. In order to overcome this complexity hurdle, most of the authors (a) either use the bit-level trellis, (b) either apply suboptimal estimation methods such as sequential decoding [4] to the bit/symbol trellis. Together with the BCJR algorithm, the method (a) amounts to optimally estimating with a suboptimal Hidden Markov model. The method (b) processes an approximate estimation on a trellis which fully represents the whole transmission chain. The main advantage of the bit/symbol trellis over the bit-level trellis is its amenability to integrate a termination constraint on the number of symbols. Both trellises have been extended and adapted to take into account the Markovian (first-order) property of the source [6] [5] and have been coupled with a convolutional code in an iterative structure, e.g. in [3].

In this paper, we describe a novel set of state models and the corresponding trellises for the estimation of the Hidden Markov chain. The state model, described in Section II, is defined by both the internal state of the VLC decoder (i.e., the internal node of the VLC codetree) and the rest of the Euclidean division of the symbol clock by a fixed parameter  $T$ . Let us recall that the bit/symbol trellis internal state model is defined by both the internal node of the VLC decoder and the symbol clock instant. Therefore, the proposed approach consists in aggregating the states of the bit/symbol trellis which are distant of  $T$  symbol clocks. If  $T = 1$ , the resulting trellis is equivalent to the usual bit-level trellis proposed in [1]. If  $T$  is greater or equal than the symbol sequence length  $L(\mathbf{S})$ , the trellis is equivalent to the bit/symbol trellis. The intermediate values of this parameter ( $1 < T < L(\mathbf{S})$ ) allow to gracefully trade complexity against the estimation accuracy. The intuition behind this state aggregation is that a desynchronization error will be detected if the difference between the numbers of emitted and decoded symbols is a quantity which is not a multiple of  $T$ . The choice of the parameter  $T$  is motivated by the capability of Variable Length Codes considered to quickly resynchronize, property which has been widely studied in previous works, e.g. in [12] or [13].

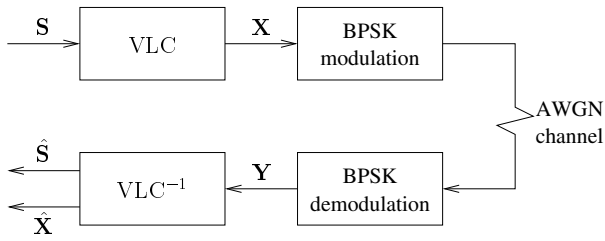


Fig. 1. Transmission setup.

The decoding complexity can be further reduced by considering separate estimations on trellises of smaller dimensions, whose parameters  $T_1$  and  $T_2$  are relatively prime. If the two sequence estimates are not equal, the decoding on a trellis with a higher value of the parameter  $T$  is then computed. It is shown in Section III that the expectation of the overall computing cost is a function of the channel noise and is decreased for most values of the signal to noise ratio.

Simulations results are provided in Section IV for a well-known source of the literature. It is shown that small values of the parameter  $T$  allow to increase significantly the accuracy of the estimation over the bit-level trellis. The performances of the approach for several error rate measures are shown to tightly approach the ones of the bit/symbol trellis for low values of  $T$  ( $T \leq 10$ ), together with a computing cost which is about  $T$  times the complexity of the bit-level trellis.

## II. STATE MODELS

The transmission setup of Fig. 1 is considered. Let  $\mathbf{S} = S_1, \dots, S_t, \dots, S_{L(\mathbf{S})}$  be a source of length  $L(\mathbf{S})$  taking its values into a finite alphabet  $\mathcal{A}$ . This source is encoded with a VLC  $\mathcal{C}$ , producing a bitstream  $\mathbf{X} = X_1, \dots, X_k, \dots, X_{L(\mathbf{X})}$  of length  $L(\mathbf{X})$ . The entropy of the source is denoted  $h$  and the mean description length (mdl) of the code  $\mathcal{C}$  is denoted  $h_{\mathcal{C}}$ . Hence the rate of the source coder is given by  $\frac{h}{h_{\mathcal{C}}}$ . The length of the shortest codeword is denoted  $l_m$  and the length of the longest codeword is denoted  $l_M$ . The bitstream  $\mathbf{X}$  is modulated using a Binary Phase Shift Keying (BPSK) modulation and is transmitted over an Additive White Gaussian Noise (AWGN) Channel, without any channel protection. The quality of the channel is characterized by its signal to noise ratio denoted  $E_b/N_0$  and expressed in decibels (dB). The decoder can then compute the probability  $\mathbb{P}(X_k = 0 \text{ or } 1 | Y_k)$  from the received measures  $\mathbf{Y} = Y_1, \dots, Y_k, \dots, Y_{L(\mathbf{X})}$ . Together with the source model, these measures are exploited to estimate the emitted bits and the corresponding symbols. These estimations are respectively denoted  $\hat{\mathbf{X}}$  and  $\hat{\mathbf{S}}$ . Note that we reserve capital letters to represent random variables and small letters to represent corresponding realizations.

The whole transmission chain is a Hidden Markov Model. We consider an estimation based on the bit clock  $k$ . Let  $N_k$  denote the random variable corresponding to the internal state of the VLC (i.e. the internal node of the VLC codetree) for

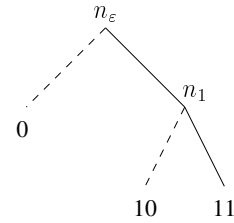
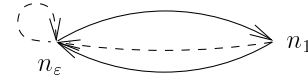


Fig. 2. Internal states  $n_\epsilon$  and  $n_1$  of the decoder.

3-a



3-b

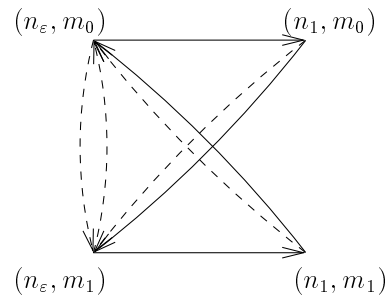


Fig. 3. Automata defining the state model for a)  $T = 1$ , b)  $T = 2$ : automaton corresponding to the bit-level trellis and the extended trellis with  $T = 2$ .

the bit clock  $k$ . The full state model is defined by the set of tuples  $(n_k, t_k)$ , where  $n_k$  is the realization of  $N_k$  and where  $t_k$  denotes the symbol clock instant corresponding to the bit clock instant  $k$ . Note that the set of values  $(t_k)_{1 \leq k \leq L(\mathbf{X})}$  defines the boundaries of the codewords in the bitstream. The internal states of the automaton associated to a given VLC are defined by the internal nodes of the codetree. It also corresponds to the set of codeword prefixes, including the void prefix  $\epsilon$  which corresponds to the root node in the codetree. For the code  $\mathcal{C}_0 = \{0, 10, 11\}$ , the set of prefixes is  $\{\epsilon, 1\}$ . The set of states of the decoder is denoted  $\{n_\epsilon, n_1\}$ , as depicted in Fig. 2. The symbol clock  $t_k$  is within the interval of integers  $[k/l_M, k/l_m]$ . Hence the total number of states of the full state model is quadratic as the sequence length (equivalently the bitstream length) increases. The resulting computing cost is not tractable, even for moderate values of the sequence length  $L(\mathbf{S})$ . In the bit-level trellis [1], this state model is reduced to the random variable  $N_k$  only. The transitions of this model are triggered by the source symbol probabilities and are represented with an automaton, as depicted in Fig. 3-a for the code  $\mathcal{C}_0$ .

Let us now consider an integer  $T$  such that  $1 \leq T \leq L(\mathbf{S})$ . We consider the state model defined by the set of tuples  $(n_k, m_k)$ , where the value  $m_k = t_k \bmod T$  is the rest of the Euclidean division of  $t_k$  by  $T$ . The corresponding

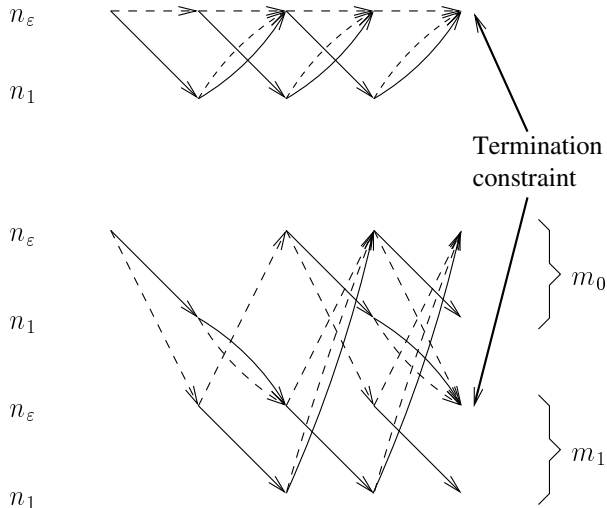


Fig. 4. Trellises for  $T = 1, 2$ : the bit-level trellis and the extended trellis with  $T = 2$ . The termination constraint is also depicted (here,  $L(\mathbf{S})$  is assumed to be odd).

random variable is denoted  $M_k$ . Note that  $T = 1$  amounts to considering the bit-level trellis of [1]. This state model is represented as an automaton, as depicted in Fig. 3-b for the code  $C_0$ . The transitions which trigger a symbol, i.e. those which terminate in the state  $n_\varepsilon$ , modify the modulo  $M_k$  as  $M_k = M_{k-1} + 1 \bmod T$ . Hence, the transition probabilities of this automaton are given by

$$\mathbb{P}(N_k = n_k, M_k = m_k | N_{k-1} = n_{k-1}, M_{k-1} = m_{k-1}) = \begin{cases} \mathbb{P}(N_k = n_k | N_{k-1} = n_{k-1}) & \text{if } n_k \neq n_\varepsilon \text{ and} \\ & m_k = m_{k-1} \\ \mathbb{P}(N_k = n_k | N_{k-1} = n_{k-1}) & \text{if } n_k = n_\varepsilon \text{ and} \\ & m_k = m_{k-1} + 1 \bmod T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where the probabilities  $\mathbb{P}(N_k = n_k | N_{k-1} = n_{k-1})$  are deduced from the source statistics. Note that the transition probabilities  $\mathbb{P}(N_k | N_{k-1})$  are the ones used in the bit-level trellis. These quantities are equal to 0 if the internal node of the decoder  $N_k$  can not be reached from the node  $N_{k-1}$ . The decoding trellises corresponding to the proposed model are depicted in Fig. 4.

The proposed state model allows to preserve an information on the symbol clock. In order to exploit this information, the decoder has to know the modulo value  $m_{L(\mathbf{x})} = L(\mathbf{s}) \bmod T$  of the number of emitted symbols. This information corresponds to a termination constraint, as depicted in Fig. 4. If this value is not given by the syntax elements of the source coding system, the transmission cost of this value is greater than or equal to  $\log_2(T)$  bits. Note that the knowledge of this value has a lower cost than the one of transmitting the exact number of emitted symbols in the bit/symbol trellis. In the following, the perfect knowledge of the quantity  $m_{L(\mathbf{x})}$  is

assumed to be known by the decoder.

On the decoder side, two soft decoding algorithms are considered, respectively the BCJR [9] and the Viterbi [10] algorithm. The BCJR algorithm is applied with the appropriate boundary initializations. Using the same notations as in [9], it amounts to assigning  $\alpha_0(n, m) = 1$  if  $n = 0$  and  $m = 0$ , otherwise  $\alpha_0(n, m) = 0$ . The quantity  $\beta_{L(\mathbf{x})}$  is set so that the ending internal state of the decoder is equal to the state  $n_\varepsilon$  and so that the termination constraint  $m_{L(\mathbf{x})}$  is enforced. To obtain an estimate  $\hat{\mathbf{X}}$  of the emitted bitstream, a hard decision is processed on the output of the BCJR algorithm. In the Viterbi algorithm, the ending state is also forced to be the state  $(n_\varepsilon, m_{L(\mathbf{x})})$ . For both algorithms, the source model probabilities  $p_t$  are given by Eqn. 1. Note that the BCJR and the Viterbi algorithms respectively lead to minimize the BER and the sequence error rate (SQER). They will be used accordingly in the simulations.

### III. COMBINED TRELLIS DECODING

In this section, we propose an approach allowing to further reduce the complexity of the decoder. This approach is motivated by the following equivalence, verified if  $T_1$  and  $T_2$  are relatively prime:

$$L(\mathbf{S}) \bmod (T_1 T_2) = m \Leftrightarrow \begin{cases} L(\mathbf{S}) \bmod T_1 = m \bmod T_1 \\ L(\mathbf{S}) \bmod T_2 = m \bmod T_2. \end{cases} \quad (2)$$

Note that, if  $T_1$  and  $T_2$  are not relatively prime, the converse is not verified. This property is exploited by the algorithm described in this section. The corresponding approach will be referred to as combined trellis decoding.

Now, let us recall that if  $T = 1$ , the resulting trellis is equivalent to the bit-level trellis. If  $T$  is greater or equal than  $L(\mathbf{S}) - \frac{L(\mathbf{X})}{l_M} + 1$  (*a fortiori* to  $L(\mathbf{S})$ ), the trellis is equivalent to the bit/symbol trellis. The intermediate values of  $T$  amount to considering trellises whose complexity is lower than the one of the bit/symbol trellis. Since the number of states is almost proportional to the quantity  $T$ , the computing cost  $D_T$  corresponding to the trellis of parameter  $T$  can be approximated as

$$D_T \approx T \times D_{\text{bal}}, \quad (3)$$

where  $D_{\text{bal}}$  denotes the computing cost of the bit-level trellis.

The rationale behind combined trellis decoding is to use two trellises of parameters  $T_1$  and  $T_2$  instead of the trellis of parameter  $T = T_1 \times T_2$ . We will also assume that the great common divisor of  $T_1$  and  $T_2$  is 1. This condition is not strictly required, but taking some parameters  $T_1$  and  $T_2$  which are not relatively prime induces a suboptimality in terms of computing cost.

According to Eqn. 2, if a symbol sequence satisfies the termination constraint for both trellises  $T_1$  and  $T_2$ , the termination constraint is also verified for the trellis of parameter  $T_1 \times T_2$ . Hence, the set of symbol sequences which are valid

for trellis  $T$  is the same as the set of symbol sequences which are valid for the two trellises  $T_1$  and  $T_2$ .

The decoding of a sequence proceeds as follows:

- 1) the Viterbi algorithm is applied to both trellises  $T_1$  and  $T_2$ . Note that these two decoding algorithms may be run in parallel. They respectively provide the estimated sequences  $\hat{\mathbf{S}}_1$  and  $\hat{\mathbf{S}}_2$ .
- 2) If  $\hat{\mathbf{S}}_1 = \hat{\mathbf{S}}_2$ , the decoded sequence is assumed to be correct. This sequence satisfies the termination constraint of the trellis of parameter  $T_1 \times T_2$ . The decoded sequence is used as the estimate of the emitted sequence.
- 3) Else, the Viterbi algorithm is applied to the trellis of parameter  $T_3$ , with  $T_3$  chosen so that

$$T_1 + T_2 < T_3 \leq T_1 \times T_2. \quad (4)$$

For example, we may choose

$$T_3 = T_1 \times T_2, \quad (5)$$

but this choice is not strictly required. The corresponding estimate is denoted  $\hat{S}_{T_3}$ .

Clearly, the expectation of the computing cost of the proposed decoding scheme  $D_{\text{mtd}}$  is given by

$$D_{\text{mtd}} = T_1 D_{\text{bal}} + T_2 D_{\text{bal}} + \rho T_3 D_{\text{bal}} \quad (6)$$

where  $\rho = \mathbb{P}(\hat{\mathbf{S}}_1 \neq \hat{\mathbf{S}}_2)$ . This method is worthwhile in terms of computing cost if  $D_{\text{mtd}} < T_3 \times D_{\text{bal}}$ , in other terms if

$$\rho < \rho^* = 1 - \frac{T_1 + T_2}{T_3}. \quad (7)$$

Note that the choice of the proposed algorithm is a function of the probability that the two estimators return the same sequence estimate. Therefore, it is a function of the channel noise: higher is this noise, lower is the probability that the estimators agree.

*Theorem:* Let  $T_1$  and  $T_2$  be two relatively prime integers and let  $T_3 = T_1 \times T_2$ . Then for the Viterbi algorithm, the combined decoding algorithm of parameters  $T_1$  and  $T_2$  is equivalent to the full trellis model of parameter  $T_3$ .

*Proof:* Note that proving the theorem amounts to proving

$$\hat{\mathbf{S}}_1 = \hat{\mathbf{S}}_2 \Rightarrow \hat{\mathbf{S}}_3 = \hat{\mathbf{S}}_1 = \hat{\mathbf{S}}_2. \quad (8)$$

Let us also underline that the probability of a sequence, computed by the Viterbi algorithm on a trellis of parameter  $T$  does not depend on  $T$ . Let us assume that, if two sequences have the same probability, then a subsidiary rule is applied to select one sequence among the two. For instance, the lexicographical order can be chosen as a comparison rule. Such a rule ensures the Viterbi algorithm behavior to be deterministic. Let

$$\mathcal{S}_T \triangleq \{s' / L(s') \bmod T = L(s) \bmod T\} \quad (9)$$

	Prob.	Huffman	RVLC <sub>1</sub>
	0.33	11	00
	0.30	10	11
	0.18	00	010
	0.10	011	101
	0.09	010	0110
Entropy	2.139	-	-
Avg. length	-	2.19	2.46
Free distance	-	1	2

TABLE I  
VARIABLE LENGTH CODES USED IN SIMULATIONS.

be the set of sequences satisfying the termination constraint for the trellis of parameter  $T$ . From Eqn. 2, we deduce that if  $T_3 = T_1 \times T_2$  with  $T_1$  and  $T_2$  relatively prime, then

$$\mathcal{S}_{T_3} = \mathcal{S}_{T_1} \cap \mathcal{S}_{T_2}, \quad (10)$$

hence,

$$\mathcal{S}_{T_3} \subseteq \mathcal{S}_{T_1}. \quad (11)$$

Moreover, since we have assumed that  $\hat{\mathbf{S}}_1 = \hat{\mathbf{S}}_2$ , we have

$$\hat{\mathbf{S}}_1 \in \mathcal{S}_{T_3}. \quad (12)$$

The estimate  $\hat{\mathbf{S}}_1$  provided by the Viterbi algorithm applied on the trellis of parameter  $T_1$  is then such that

$$\hat{\mathbf{S}}_1 = \arg \max_{s' \in \mathcal{S}_{T_1}} \mathbb{P}(s' | \mathbf{X}) \quad (13)$$

$$= \arg \max_{s' \in \mathcal{S}_{T_3}} \mathbb{P}(s' | \mathbf{X}) \quad (14)$$

$$= \hat{\mathbf{S}}_3, \quad (15)$$

where the subsidiary rule may be used in the selection of the maximum. This concludes the proof. ■

Therefore, as we will verify in simulations (see Section IV), the combined trellis decoding approach can advantageously substitute the equivalent single trellis decoding for a large set of signal to noise ratio  $E_b/N_0$ .

#### IV. SIMULATION RESULTS

To validate the proposed approach, the source and the codes of [14] have been considered. This source and the corresponding VLC can be considered as the reference choice for joint-source channel decoding of VLC-encoded memoryless sources. The source is a 5-symbols memoryless source defined by its stationary probabilities:

$$\mathbb{P}(S_t = a_i)_{a_i \in \mathcal{A}} = \{0.33, 0.30, 0.18, 0.10, 0.09\}.$$

Its entropy equals 2.139. The statistics of this source are given in Table I together with the two VLCs used to validate the approach. The first VLC is a simple Huffman code and is defined as

$$\mathcal{C}_{11} = \{11, 10, 00, 011, 010\}.$$

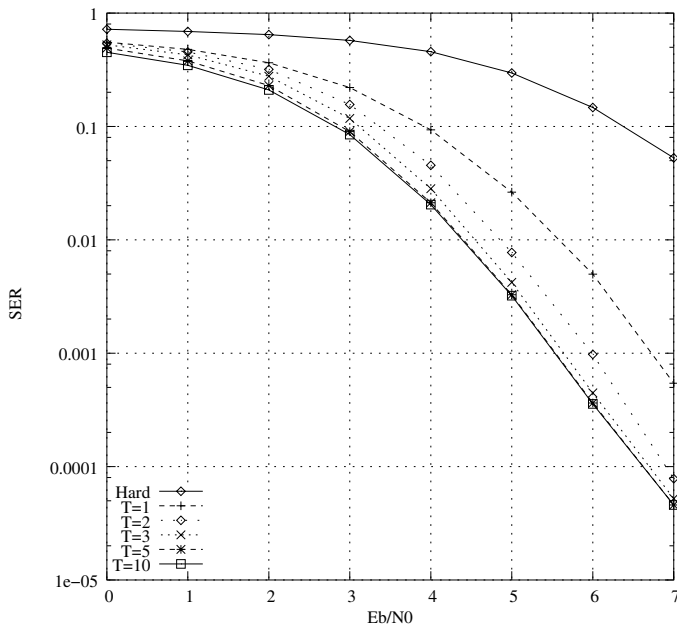


Fig. 5.  $SER_H$  results for hard decoding and soft decoding of code  $C_{12}$  with different values of the aggregation parameter  $T$ .

The second VLC is a Reversible VLC with a free distance of 2 and is defined as

$$C_{12} = \{00, 11, 010, 101, 0110\}.$$

The mdl of these codes are respectively equal to 2.19 and to 2.46 bits per symbol. The notations  $C_{11}$  and  $C_{12}$  are the ones of the original paper [14]. The realizations of the source are packetized into blocks of  $L(\mathbf{S}) = 100$  symbols. For each parameter set (VLC  $C_{11}$  or  $C_{12}$ ,  $E_b/N_0$  and  $T$ ), the results are averaged over  $10^5$  channel realizations. The error-resilience of our transmission scheme is measured by the BER, the SQER and two distinct SERs. The first SER measure is defined as the normalized Hamming distance between the emitted sequence and the received one. If the received sequence is smaller than the emitted one, missing symbols are assumed to be in error. The second SER measure is given by the Levenshtein distance [15] normalized by the number of symbols. These SERs are respectively denoted  $SER_H$  and  $SER_L$  in the following.

The Viterbi algorithm has been used for all the error-resilience measures but the BER. For this measure, we used the BCJR algorithm in order to obtain the marginal bit error probabilities. The maximum of these posterior marginals has been used to estimate the emitted sequence  $\mathbf{X}$ .

Fig. 5 presents the  $SER_H$  as a function of  $E_b/N_0$  for different values of the parameter  $T$  for the code  $C_{12}$ . The performances of hard decision decoding are also depicted. For this latter, no concealment is applied, i.e. the backward decoding ability of Reversible VLC has not been exploited. Note that this choice only affects the performances of the Hard decoding algorithm (and not the soft decoding results). Tables II and III complement the results presented in Fig. 5 by giving both the

		Code $C_{11}$				
$E_b/N_0$		3	4	5	6	7
Hard		0.022865	0.012507	0.005956	0.0023855	0.0007641
T=1		0.022748	0.012424	0.005918	0.0023750	0.0007606
T=2		0.022558	0.012074	0.005524	0.0021017	0.0006483
T=3		0.022221	0.011757	0.005349	0.0020542	0.0006399
T=5		0.021870	0.011588	0.005303	0.0020477	0.0006390
T=10		0.021812	0.011577	0.005303	0.0020477	0.0006390
		Code $C_{12}$				
$E_b/N_0$		3	4	5	6	7
Hard		0.022930	0.012548	0.005971	0.0023964	0.0007760
T=1		0.011446	0.004060	0.001048	0.0001878	0.0000219
T=2		0.011002	0.003423	0.000720	0.0001101	0.0000120
T=3		0.010529	0.003101	0.000643	0.0000989	0.0000115
T=5		0.009968	0.002930	0.000620	0.0000972	0.0000114
T=10		0.009813	0.002908	0.000618	0.0000970	0.0000114

TABLE II  
BER FOR HARD DECODING AND SOFT DECODING (MAXIMUM OF POSTERIOR MARGINALS) WITH DIFFERENT VALUES OF THE AGGREGATION PARAMETER  $T$ .

		Code $C_{11}$				
$E_b/N_0$		3	4	5	6	7
Hard		0.080669	0.045259	0.021918	0.0088432	0.0028500
T=1		0.079589	0.044600	0.021570	0.0087394	0.0028137
T=2		0.079300	0.042541	0.018933	0.0068562	0.0020429
T=3		0.076878	0.040229	0.017664	0.0065245	0.0019844
T=5		0.074049	0.038910	0.017336	0.0064846	0.0019787
T=10		0.073586	0.038831	0.017333	0.0064846	0.0019787
T=100		0.073586	0.038831	0.017333	0.0064846	0.0019787
		Code $C_{12}$				
$E_b/N_0$		3	4	5	6	7
Hard		0.231274	0.154074	0.085248	0.0377131	0.0127488
T=1		0.044184	0.015951	0.004146	0.0007423	0.0000862
T=2		0.043053	0.013330	0.002723	0.0004012	0.0000427
T=3		0.040578	0.011558	0.002274	0.0003265	0.0000391
T=5		0.037005	0.010413	0.002105	0.0003104	0.0000383
T=10		0.035684	0.010212	0.002087	0.0003091	0.0000383
T=100		0.035672	0.010212	0.002087	0.0003091	0.0000383

TABLE III  
 $SER_L$  RESULTS FOR HARD DECODING AND SOFT DECODING (VITERBI) WITH DIFFERENT VALUES OF THE AGGREGATION PARAMETER  $T$ .

BER and the  $SER_L$  for the same parameter sets as in Fig. 5 and for both codes  $C_{11}$  and  $C_{12}$ .

The results obtained evidence the great advantage of increasing slightly the complexity (by the small factor  $T$ ). Fig. 5 shows that the gain obtained in terms of  $E_b/N_0$  for  $SER_H = 10^{-3}$  is about 1.2 dB if  $T = 5$  against the bit-level trellis ( $T = 1$ ). The results obtained for  $T = 5$  and  $T = 10$  are almost identical and are close to those obtained with the bit/symbol trellis. These results evidence the fact that low values of  $T$  ( $T \leq 5$ ) are sufficient to grasp the residual redundancy of the termination constraint. Table II confirms the interest of the approach for the BER measure. For values of  $E_b/N_0$  greater than 5 dB, the number of erroneous bits is lower than half the one obtained with the bit-level trellis.

Code $C_{11}$					
$E_b/N_0$	3	4	5	6	7
Hard	0.993520	0.936080	0.728570	0.406460	0.154080
T=1	0.993140	0.934890	0.727800	0.405200	0.153300
T=2	0.990450	0.917310	0.682300	0.358910	0.130400
T=3	0.989530	0.911980	0.674360	0.355210	0.129560
T=5	0.989100	0.910720	0.673450	0.355040	0.129520
T=10	0.989070	0.910700	0.673450	0.355040	0.129520

Code $C_{12}$					
$E_b/N_0$	3	4	5	6	7
Hard	0.996190	0.954080	0.768680	0.445690	0.173560
T=1	0.742780	0.379170	0.117140	0.022370	0.002650
T=2	0.668240	0.291030	0.074880	0.012750	0.001440
T=3	0.639750	0.267090	0.068340	0.011590	0.001380
T=5	0.624730	0.259120	0.066650	0.011420	0.001370
T=10	0.622450	0.258260	0.066540	0.011410	0.001370

TABLE IV

SQER FOR HARD DECODING AND SOFT DECODING (VITERBI) WITH DIFFERENT VALUES OF THE AGGREGATION PARAMETER  $T$ .

In table III and table IV, significant gains in terms of the Levenshtein distance and the SQER are also evidenced.

Fig. 6 illustrates the complexity reduction brought by the combined trellis decoding algorithm to reduce the complexity of the soft decoding for equal decoding performance. It is shown that a lower computing cost is obtained with this approach as long as  $E_b/N_0$  is greater than 0.65 dB, which is the case of most practical channels.

## V. CONCLUSION AND PERSPECTIVES

In this paper, the two following main contributions have been proposed. First, a new set of states models has been proposed for soft decoding of VLCs. As the model associated to the optimal bit/symbol trellis and unlike the one associated to the bit-level trellis, this model is able to exploit a termination constraint on the decoder side. The complexity is however significantly reduced in comparison with the optimal decoding scheme. Therefore this model may advantageously replace the bit-level trellis used in state-of-the-art iterative structures with VLCs such as described in [7]. The other main contribution, the combined trellis decoding approach, further reduces the overall complexity required to obtain the the same performances as the optimal decoding.

The proposed state models could be easily extended to first-order Markovian source models by adding the modulo information to the state model proposed in [5]. The estimation of  $\hat{S}$  could also be processed using a list-Viterbi decoding of the sequence. The estimates  $\mathbb{P}(S_t = m)$  could be subsequently derived from the output of this algorithm. We expect such an approach to improve the results in terms of SER.

## REFERENCES

[1] V. B. Balakirsky, "Joint source-channel coding with variable length codes," in *Proc. Intl. Conf. Inform. Theory, ISIT*, 1997, p.419.  
[2] A. Murad and T. Fuja, "Joint source-channel decoding of variable length encoded sources," in *Proc. Inform. Theory Workshop, ITW*, June 1998, pp. 94–95.

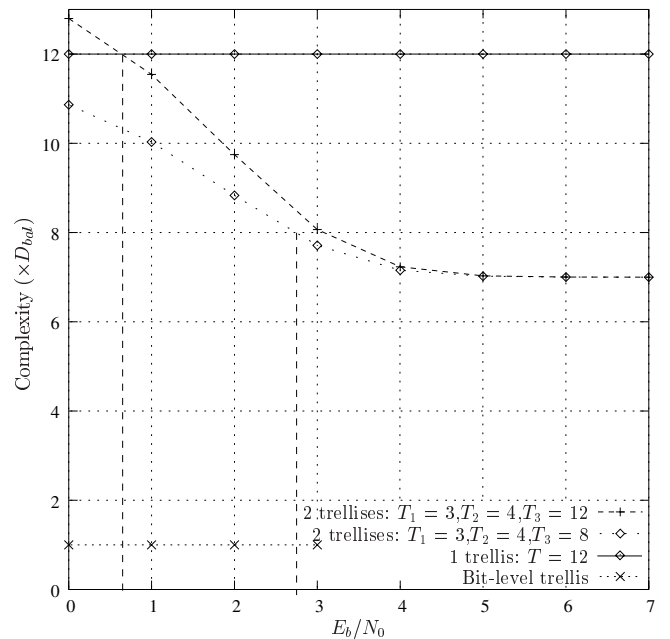


Fig. 6. Computing cost of the combined trellis decoding approach versus  $E_b/N_0$  against the computing cost of a single trellis decoding approach for parameters  $(T_1 = 3, T_2 = 4, T_3 = 12)$  and  $(T_1 = 3, T_2 = 4, T_3 = 8)$ . The corresponding cut-off values of  $E_b/N_0$  are also depicted and are respectively obtained for  $\rho^* = 0.418$  and  $\rho^* = 0.125$ .

[3] R. Bauer and J. Hagenauer, "Symbol by symbol map decoding of variable length codes," in *Proc. 3rd ITG Conf. Source and Channel Coding, Munich, Germany, Jan. 17.-19. 2000*, 2000. [Online]. Available: citeseer.nj.nec.com/bauer00symbolbysymbol.html  
[4] M. Bystrom, S. Kaiser, and A. Kopansky, "Soft source decoding with applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 10, pp. 1108–1120, Oct. 2001.  
[5] C. Weidmann, "Reduced-complexity soft-in-soft-out decoding of variable length codes," in *Proc. Intl. Conf. Inform. Theory, ISIT*, July 2003, yokohama, Japan.  
[6] R. Thobaben and J. Kliewer, "Robust decoding of variable length encoded markov sources using a three-dimensional trellis," *IEEE Trans. Commun.*, pp. 787–794, July 2003.  
[7] X. Jaspard and L. Vandendorpe, "New iterative decoding of variable length codes with turbo codes," in *Proc. Intl. Conf. Commun., ICC*, June 2004, paris, France.  
[8] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, "Joint source-channel turbo decoding of entropy coded sources," *IEEE J. Select. Areas Commun.*, vol. 19, no. 9, pp. 1680–1696, Sept. 2001.  
[9] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, pp. 284–287, Mar. 1974.  
[10] A. Viterbi, "Error bounds for convolution codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, no. 13, pp. 260–269, 1967.  
[11] E. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, no. 1, pp. 269–271, 1959.  
[12] J. Maxted and J. Robinson, "Error recovery for variables length codes," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 6, pp. 794–801, Nov. 1985.  
[13] G. Zhou and Z. Zhang, "Synchronization recovery of variable length codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 219–227, Jan. 2002.  
[14] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 158–162, Feb. 1995.  
[15] Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," *Soviet Physics Doklady*, vol. 10, pp. 707–710, 1966.