Impact of Variable Length Codes on the Interleaving Gain of Turbo Systems: The Concept of Bounded Spectrum
Xavier Jaspar, Graduate Student Member, IEEE, and Luc Vandendorpe, Fellow, IEEE

Abstract—When a telecommunication system is constrained in terms of delay and complexity, it is usually wise to allow some cross-layer cooperation between the source and channel layers. In this context, the success of joint source-channel turbo techniques has been attested several times in the literature, in particular to transmit variable length code (VLC) streams which are very sensitive to error propagation. Capitalizing on previously developed performance upper bounds, this paper investigates whether the VLC can contribute to the interleaving gain of concatenated codes just as a convolutional code with non-catastrophic encoder would. To this end, the important concept of bounded VLC spectrum is introduced and is proved to be a sufficient condition for the VLC to contribute indeed to the interleaving gain. As a by-product, this concept is also proved to be closely related to non-catastrophic VLCs and, under certain assumptions, to the well known concept of statistically synchronizable VLCs.

Index Terms—Variable length code, linear code, interleaving gain, bounded VLC spectrum, statistically synchronizable VLC.

I. INTRODUCTION

The success of joint source-channel turbo (de)coding to transmit variable length code (VLC) data streams has been illustrated for the first time in [1], based on the serial concatenation of a VLC with a convolutional code (CC). This concatenation has then been improved in several directions, notably the VLC decoder [2], [3], the VLC itself [4]–[6] and the inner error correcting code (ECC) [6]–[9] to cite a few. In particular, some light has been shed recently in [10], [11] on the theoretical error correcting properties of VLCs in turbo/concatenated systems by developing and proving several performance bounds as well as an approximate expression of the global code (VLC+ECC) distance spectrum.

Capitalizing on these bounds, this paper focuses on the interaction between the VLC and the ECC. More precisely, it is devoted to the analysis of the interleaving gains [12], [13] offered by the global code, based on result from [14]. In particular, it is investigated whether the VLC, as outer code of the serial concatenation, can contribute to the interleaving gain, just as a convolutional code with non-catastrophic encoder would. To this end, the important concept of bounded VLC spectrum is introduced and is proved to be a sufficient condition for the VLC to contribute indeed to the interleaving gain. As a by-product, this concept is also proved to be closely related to non-catastrophic VLCs and, under certain assumptions, to the well known concept of statistically synchronizable VLCs [15], which gives us a simple way to test whether a given VLC has a bounded spectrum, without actually computing the spectrum. Simulation results will confirm the significance of bounded VLC spectra for the interleaving gains. Interestingly, they will also reveal that the maximum a posteriori (MAP) decoder, unlike the ML decoder, is sometimes still able to offer the interleaving gains when the VLC spectrum is not bounded, just as if it was bounded.

Most results of this work concern the properties of the VLC alone and are thus general and independent of the ECC. But Theorem 21, that is the contribution of the VLC to the interleaving gain, is however focused on the serial concatenation of a VLC with a CC. This result is straightforwardly extensible to ECCs made up of multiple parallel/serially concatenated CCs (parallel/serial turbo codes). But the extension to other ECCs requires further research.

The remainder of this paper targets experts in VLC coding and turbo techniques, and assumes a good knowledge in this field [1]–[9]. The paper is structured as follows. Section II introduces a few background results and notations, the considered communication system and several assumptions. Section III constitutes the main part and is devoted to the presentation of the theoretical results (with the proofs in Appendix). At last, simulations illustrating the theoretical results are reported in Section IV and possible extensions of this work are discussed in Section V.

II. BACKGROUND AND TRANSMISSION SYSTEM

In the following, random variable are written with capital letters and realizations with small letters. $P(z)$ is the abbreviation of the probability $P(Z = z)$. The sub-sequence $(Z_m, Z_{m+1}, \ldots, Z_n)$ is written $Z_{m:n}$, or $Z$, when $m,n$ can be omitted. $\mathcal{A}$ is the decision taken on $Z$ at the receiver. $\mathcal{A}$ is the alphabet of the source symbols and $\mathcal{Y}$ the set of VLC codewords. Given a symbol sequence $s \in \mathcal{A}^+$, $u = \text{vlc}(s) \in \mathcal{Y}^+$ is the bit sequence produced by the VLC, with the reverse operation $s = \text{vlc}^{-1}(u)$. $l(u)$ is the bit length of $u$, and $l(v) = l_\text{gcd}(v)$ is the symbol length of $s$. By extension, $l(s) = l(\text{vlc}(s))$ and $l_\text{gcd}(s) = l(s)$. $d_H(\ldots)$ and $d_S(\ldots)$ denote the Hamming and Levenshtein [16] symbol distances, $l_\text{gcd}$ is the greatest common divisor of the VLC lengths, $\text{gcd}\{l(s) : s \in \mathcal{A}\}$, and $l'_\text{gcd} = \text{gcd}\{l(s) : s \in \mathcal{A}\}$.
\(\mathcal{A}, P(s) > 0\). Besides, \(l_{\text{max}} = \max\{l(s) : s \in \mathcal{A}\}, l_{\text{min}} = \max\{l(s) : s \in \mathcal{A}, P(s) > 0\}, l_{\text{min}} = \min\{l(s) : s \in \mathcal{A}, P(s) > 0\}\). \(|\mathcal{A}|\) is the number of elements in the set \(\mathcal{A}\). At last, \(\mathbb{N}\) is the set of non-negative integers \(\{0, 1, 2, 3, \ldots\}\).

A. Prefix VLCs and Source of Symbols

Let \(\mathcal{B}\) be the binary alphabet \([0, 1]\). A finite sequence \(w = b_1, b_2, \ldots, b_m\) of code letters \(b_i \in \mathcal{B}\) is called a word over \(\mathcal{B}\), of length \(l(w) = m\). Let \(R\) be the empty word; we have \(l(R) = 0\) and \(RW = w = wR\) for any word \(w\). Let \(\mathcal{B}^+ = \{0, 1\}\), the set of all non-empty words. Let prefix \((w) = \{p \in \mathcal{B}^+ : \exists v \in \mathcal{B}^+, w = pv\}\) and suffix \((w) = \{s \in \mathcal{B}^+ : \exists v \in \mathcal{B}^+, w = vs\}\). Each element of prefix \((w)\) is a prefix of \(w\) and each element of suffix \((w)\) a suffix of \(w\). A code is a set \(\mathcal{V} \subseteq \mathcal{B}^+\). The elements of \(\mathcal{V}\) are called codewords. If no codeword is the prefix of another codeword, then the set \(\mathcal{V}\) is called a prefix variable length code (VLC). If, besides, no codeword is the suffix of another codeword, then the VLC is said reversible. In the following, only prefix VLCs are considered and we omit to specify “prefix” for convenience. For any code \(\mathcal{U} \subseteq \mathcal{B}^+\), let prefix \((\mathcal{U}) = \{u \in \mathcal{U} : \exists v \in \mathcal{U} \text{ prefix } (w)\}\), the set of all prefixes of \(\mathcal{U}\). Let us then consider a source of discrete, independent, (memoryless) and identically distributed symbols over the alphabet \(\mathcal{A}\) with \(|\mathcal{A}| = |\mathcal{V}| \geq 2\), and characterized by the probability distribution \(P(S)\). Let vlc\((\cdot)\) be a bijective mapping that associates to each symbol \(s \in \mathcal{A}\) a unique codeword \(w = \text{vlc}(s) \in \mathcal{V}\). The inverse mapping is denoted \(\text{vlc}^{-1}(\cdot)\). In the following, this mapping is always assumed implicitly; we will use concepts associated with codewords equivalently with symbols and vice versa, e.g., \(l(s) = l(\text{vlc}(s)), P(w) = P(\text{vlc}^{-1}(w))\). Let \(\mathcal{A}^+ \triangleq \bigcup_{i \geq 1} \mathcal{A}^i\) and \(\mathcal{V}^+ \triangleq \bigcup_{i \geq 1} \mathcal{V}^i\).

At last, we introduce a relation of equivalence between sub-sequences of source symbols and of VLC bits, \(s_{m:n} \equiv u_{i:j}\), which means \(s_{m:n} \equiv u_{i:j}\) refer to the same realization of the source/VLC. It implies notably \(u_{i:j} = \text{vlc}(s_{m:n}), l(s_{i:m-1}) = i - 1\) and \(l(s_{1:n}) = j\). In the following, this equivalence is generally assumed implicitly, unless otherwise stated.

B. Considered Transmission System

Consider in Fig. 1 the VLC stream is framed into finite-length VLC blocks (frames or packets) before the interleaving by \(\Pi\) of length \(N\). In this paper, let us focus on the following framing rule. Assume the VLC stream has been processed up to a certain symbol. Assume the symbol and bit indices have been shifted such that the next symbol and bit to be processed are \(S_1\) and \(U_1\). Then given a fixed \(N\), let \(F_0\) be the framing rule that forms the next VLC block with \(S_1:N\) (\(N\) is variable) subject to

\[N_b = l(S_1:N_b) \leq N, \quad N_b + l(S_{N_b+1:N}) > N,\]

where \(S_{N_b+1:N}\) becomes the first symbol of the next VLC block. After the framing process, zeros are appended to \(U_1:N_b\) to get \(U_1:N\) if \(N_b < N\). The bits \(U_1:N\) are interleaved by \(\Pi\) into \(U_1'\). The linear ECC generates the coded bits \(C\) and sends them across the channel, which is binary, symmetric, memoryless and time invariant.

At the receiver, the frame-MAP (maximum a posteriori) detection is considered, \(\hat{s} = \arg\max_{s} \{P(y|s)\} = \arg\max_{s} \{P(y,s)\}\) where \(y\) is the received signal, \(s\) the transmitted sequence of symbols and \(\hat{s}\) the decision taken on \(s\) at the receiver. It is approximated (for complexity reasons) by a frame-MAP joint source-channel iterative/turbo decoder.

C. Assumptions

As mentioned in the introduction, the analysis in this paper is based on theoretical results from [10], [11]. These results are based on the frame-ML (maximum likelihood) detection \(\hat{s} = \arg\max_{s} \{P(y|s)\}\) and on several assumptions.

Assumption 1 (source/VLC): Consider the source/VLC from Section II-A with at least two different symbols of non-zero probability.

Assumption 2 (global system): In addition to Assumption 1, we consider in Fig. 1 finite-length frames \(U_1:N_b = \text{vlc}(S_1:N_b)\), a uniform interleaver [13] of length \(N\), a linear ECC and a frame-ML decoder. The decoder knows (and uses) the values of \(N_b\) and \(N\) but not the value of \(N_b\). If \(N_b < N\), the decoder knows besides that the bits \(U_{N_b+1:N}\) are zeros.

Assumption 3: In addition to Assumption 2, all symbol values have non-zero probabilities.

D. Distance spectrum and bounds

Background results from [10], [11] on the distance spectrum of VLCs are now recalled in a slightly rephrased form.

Let \(\rho(U_1:1, \tilde{U}_1:1)\) be the event that \((U_{i-1}, U_{i-1}:1 \in \mathcal{V}^+ + \{R(i)\})\). This event characterizes the synchronization right before the bit \(U_i\), between the transmitted stream \(S_1:1: \equiv U_1:1\) and the decoded stream \(\tilde{S}_1:1: \equiv \tilde{U}_1:1\). Indeed, when this event occurs, both streams have a codeword starting with \(U_i\). Therefore, if there is no bit error after \(U_i\), i.e., \(\tilde{U}_1:1 = U_1:1\), then all subsequent codewords are decoded correctly (up to a possible time shift, which does not affect the Levenshtein symbol distance). This is the idea underlying the concept of error event: An error event \(e\) is generated by some errors and, in this paper, we consider that \(e\) starts with the beginning of its first erroneous codeword, say with \(U_i\) for some \(i\) with \(\rho(U_i:1, \tilde{U}_1:1)\), and ends as soon as we have again \(\rho_+ + 1(U_i:1, \tilde{U}_1:1)\) with the smallest integer \(j > i\) (thus ends with \(U_j\)). In other words, \(i\) and \(j + 1\) are the only elements of the set \(\{l : i \leq l \leq j + 1, \rho(u, \tilde{u})\}\).

Definitions 4: Under Assumption 2, \(s_{m:n} \equiv u_{i:j}\) and \(\tilde{s}_{m':n'} \equiv u_{i:j}\) form an error event \(e\) if \(u_{i:j} \neq u_{i:j}\) and \(\{i, j + 1\} = \{l : i \leq l \leq j + 1, \rho(u, \tilde{u})\}\) and we say \(e\) starts with \(u_i\) or \(s_{m:n}\). By extension of previous notations, let \(E_{l:1} = E_{l:1}(u_{i:j})\) be the symbol length, \(l(e) = l(\tilde{u}_{i:j})\) the bit length, etc. The VLC stream spectrum is under Assumption 2

\[A_{k, h, l, i, j, d}^\text{vlc} = \sum_{s_{i:j} \in \mathcal{A}_{\mathcal{V}, h, l, i, j}} P(s_{i:j}) A_{k, h, l, i, j}^\text{vlc},\]
where the conditional spectrum $A_{s,L,h,s}^{vlc}$ is

$$A_{s,L,h,s}^{vlc} = \left\{ \tilde{s}, \tilde{s} \in \mathcal{A}^+ : d_{s,L,h,s} = h, d_{s,L,h,s} = \tilde{s} \right\},$$

(3)

where $d_H(.)$ and $d_{s,L,h,}(.)$ denote the Hamming and Levenshtein [16] distance. The VLC spectrum is bounded if $\exists c < 1, \forall h, \exists t_h, v_{h} \geq t_h, A_{h,h,s}^{vlc} \leq c^h$ or $A_{h,h,s}^{vlc} = O(c^h)$. At last, the free distance is $d_f^{vlc} = \min\{h \in \mathbb{N} : A_{h,h,s}^{vlc} \neq 0, h > 0\}$. □

Given these definitions, we can upper bound the performance of the system described in Section II-B as follows.

**Theorem 5 (System performance):** The frame-ML performance of the system in Fig. 1 is upper bounded under Assumption 2 by

$$\text{FER} \leq \sum_{h} P_h \sum_{s_L,w} A_{s,L,h,w}^{ecc} N_w^h,$$

(4)

$$\text{SER}_{L} \leq \sum_{h} P_h \sum_{s_L,w} N_{v}^{\text{min}} S_{L} A_{s,L,h,w}^{ecc} N_w^h,$$

(5)

$$\text{BER} \leq \sum_{h} P_h \sum_{s_L,w} N_{v}^{\text{min}} w A_{s,L,h,w}^{ecc} N_w^h,$$

(6)

where $A_{ecc}^{ecc}, h,w$ is the distance spectrum of the ECC, $A_{s,L,h,w}^{ecc}$ the average distance spectrum of the VLC block, $P_h$ the pairwise error probability, $N_{v}^{\text{min}}$ and $N_{v}^{\text{min}}$ are respectively the minimum admissible values of $N_v$ and $N_h$ having a non-zero probability for the given framing rule $F_h$.

The exact computation of the VLC block spectrum $A_{s,L,h,w}^{ecc}$ is particularly difficult. Fortunately, it can be replaced by the so-called upper spectrum $A_{s,L,h,w}^{ecc,upp}$ and still maintain strict upper bounds in (4)-(6). For the framing rule $F_h$, this upper spectrum is upper bounded by the following expression

$$A_{s,L,h}^{ecc,upp} \leq \frac{P_{ad}}{\gcd} \sum_n \binom{N}{n} \sum_{i=0}^N T_{s,L,h,i,n}^{vlc}$$

(7)

where $P_{ad}$ is the greatest common divisor $\gcd\{l(s) : s \in \mathcal{A}^+, P(s) > 0\}$, $t_{\text{max}} = \max\{l(s) : s \in \mathcal{A}^+, P(s) > 0\}$, and $T_{s,L,h,i,n}^{vlc}$ is the coefficient of the polynomial

$$T_{s,L,h,i,n}^{vlc}(S_{L}, H, L_s, L_h, \Omega) = \sum_{s,L,h,l,i,n} T_{s,L,h,l,i,n}^{vlc} S_{L}^{s} H^{l} L_s^{l} L_h^{l} \Omega^{n}$$

(8)

$$= \sum_{n \geq 1} \left(A_{s,L,h,l,i,n}^{vlc}(S_{L}, H, L_s, L_h, \Omega) \right)^n,$$

$$A_{s,L,h,l,i,n}^{vlc}(S_{L}, H, L_s, L_h, \Omega) \}

(9)

III. STATISTICAL SYNCHRONIZATION, BOUNDED SPECTRUM AND INTERLEAVING GAINS

The concepts of statistically synchronizable VLCs and synchronizing sequences are thoroughly studied in [15]. In Section III-A, we introduce a particular flavor of these concepts and then show that statistical synchronization and bounded spectrum (see Def. 4) are equivalent. This equivalence has the nice consequence in Section III-C that Algorithm 18, which tests the statistical synchronization, tests also whether the VLC spectrum is bounded. A bounded VLC spectrum is then shown by Th. 21 in Section III-E to be a key sufficient condition for the VLC to contribute to the interleaving gains, condition which — it is worth recalling — can be tested by Algorithm 18.

A. Bounded Spectrum and Statistically Synchronizable VLC

Compared to [15], the scope is here restricted to prefix VLCs and to binary errors. In particular, neither bit insertion nor bit deletion are considered. These restrictions make the concepts hereafter slightly simpler. In addition, we make hereafter the concept of statistical synchronization depending on the considered decoder.

The following concept is a particular case of a bounded spectrum, is closely related to CCs with non-catastrophic encoder, see Section III-D, and is used in many lemmas in Appendix C to prove Th. 21.

**Definition 6:** A VLC spectrum is strongly bounded if $\exists c < 1, \exists t, \forall h, \exists y, A_{s,L,h}^{vlc} \leq c^h$.

**Property 7:** It is strongly bounded if and only if $\exists c < 1, \exists t, \forall h, \exists y, A_{s,L,h}^{vlc} \leq c^h$.

Though statistical synchronization, bounded spectrum and non-catastrophic behavior are more properly properties of the triplet (source, VLC, mapping), we will refer to them in the following with some abuse as properties of the VLC to make the presentation lighter.

Let $U_{1:1} = \text{vlc}(S_{1:1})$ be the transmitted stream without channel code and $\hat{U}_{1:1} = \text{vlc}(\hat{S}_{1:1})$ the decoded stream.

**Definition 8:** Given a source and a decoder, a prefix VLC is statistically synchronizable, or more properly, statistically $\rho$-synchronizable, if

$$\lim_{s \rightarrow \infty} P(\tau_i(U_{1:i:1}, \hat{U}_{1:i:1}) \leq s \mid U_{1:i} = u_{1:i:1}, \hat{U}_{1:i} = \hat{u}_{1:i:1}) = 1$$

(9)

for any integer $i$ and any $u_{i+1} \in \text{prefix}(\mathcal{Y}^+)$ of strictly positive probability, independently of any channel errors, where $\tau_i(U_{1:i:1}, \hat{U}_{1:i:1})$ is the smallest integer $t \geq i$ with $p_t(U(1:i:1), \hat{U}_{1:i:1}) = 1$.

Loosely speaking, a VLC is thus statistically synchronizable if, independently of any channel bit errors among the bits $U_{1:i-1}$, the decoder resynchronizes with high probability with the correct sequence provided that a sufficient number of correct bits ($\hat{U}_{1:i-1} = U_{1:i-1}$) are received. A VLC is (deterministically) synchronizable if $\tau_i(U_{1:i:1}, \hat{U}_{1:i:1})$ is bounded, i.e., if there exists a finite $t$ such that

$$P(\tau_i(U_{1:i:1}, \hat{U}_{1:i:1}) \leq t \mid U_{1:i} = u_{1:i:1}, \hat{U}_{1:i} = \hat{u}_{1:i:1}) = 1$$

(10)

Under Assumption 2, the key characteristic in common between statistical synchronization and bounded spectrum is the synchronizing sequence, as stated hereafter in Th. 12.

**Definition 9:** A sequence $s' \equiv u'$ is synchronizing if, for all $\hat{u} \in \mathcal{B}^+$ subject to $l(\hat{u}) \equiv 0 \pmod{\gcd}$, we have either $\hat{u} u' \in \mathcal{Y}^+$ (admissibility) or $\hat{u} u' \notin \text{prefix}(\mathcal{Y}^+)$ (inadmissibility), where $l_{gcd}$ is the greatest common divisor $\gcd\{l(s) : s \in \mathcal{A}^\pm\}$.

In other words, $\hat{u} u'$ is either an admissible sequence of VLC codewords or inadmissible. Because of this, all decoders
that consider only admissible sequences, e.g., the frame-ML decoder, resynchronize as soon as a synchronizing sequence is received error-free. Note synchronization markers, often used in compression standards, play at least the same role as the synchronizing sequence. Therefore, many results hereafter follow generally also when synchronization markers are used.

Roughly speaking, when the VLC has no synchronizing sequence, there exists intuitively at least one sequence from which we can never resynchronize, as stated contrapositively hereafter in Lemma 11. We refer to such a particular sequence as an anti-synchronizing sequence.

Definition 10: A sequence \( u^a \in \mathcal{B}^+ \) is anti-synchronizing w.r.t. sequences of non-zero probability if its length is a multiple of \( l_{\text{gcd}} \) and if \( u^a u \in \text{prefix}(\mathcal{V}^+) \setminus \mathcal{V}^+ \) for all \( u \in \mathcal{V}^+ \) with \( P(u) > 0 \).

By definition, \( u u^a v \) is also anti-synchronizing for any \( u, v \in \mathcal{V}^+ \).

Lemma 11: Given a VLC, the following propositions are equivalent under Assumption 1:

a) the VLC has at least one synchronizing sequence of non-zero probability;
b) there is no sequence \( u^a \in \mathcal{B}^+ \) that is anti-synchronizing w.r.t. sequences of non-zero probability;
c) there is no prefix \( p \in \text{prefix}(\mathcal{V}) \) that is anti-synchronizing w.r.t. sequences of non-zero probability.

See proof in Appendix A.

Theorem 12: Given a VLC, the following propositions are equivalent under Assumption 2:

a) the VLC has at least one synchronizing sequence of non-zero probability;
b) the VLC is statistically synchronizable;
c) the VLC spectrum is bounded;
d) the VLC spectrum is strongly bounded.

See proof in Appendix B.

Corollary 13: Under Assumption 3, a VLC has at least one synchronizing sequence if and only if one of the equivalences of Th. 12 is satisfied.

B. Some Examples

The following proposition is straightforward and provides a sufficient but restrictive condition to guarantee statistical synchronization.

Proposition 14: Under Assumption 2, if a VLC has a codeword of length \( l_{\text{gcd}} \) and of non-zero probability, then the sequence \( u^a \) made up of this codeword repeated \( l_{\text{max}} / l_{\text{gcd}} \) times is synchronizing, where \( l_{\text{max}} = \max\{l(s) : s \in \mathcal{S}^+\} \).

With fixed length codes (FLCs), all codewords have length \( l_{\text{gcd}} \) and thus all FLCs are statistically synchronizable. Besides, we have always \( \rho(u, U^+) \) for any \( s \equiv 0 \pmod{l_{\text{gcd}}} \), which implies the following well-known result.

Proposition 15: All FLCs are synchronizable.

By contrast, here is now a straightforward result about non-synchronizable VLCs.

Proposition 16: Under Assumption 2, no complete reversible VLC is statistically synchronizable, FLCs excepted.

A VLC is complete if it satisfies Kraft’s inequality with strict equality; when a VLC is complete, we have in particular \( \text{prefix}(\mathcal{V}^+) = \mathcal{B}^+ \), i.e., all binary sequences are admissible.

Example 17: The reversible VLC given by the codewords 000, 0010, 0100, 1010, 1011, 110 and 111, is complete. This VLC is thus not statistically synchronizable and its spectrum is not bounded under Assumption 2. Furthermore, any prefix of this VLC is anti-synchronizing.

In Section III-D, we will see that VLCs with unbounded spectrum are not ensured to have a non-catastrophic behavior. This together with Th. 12 and Proposition 16 complements the results of [17] where it is shown that reversible VLCs with bounded spectrum lead to interesting performance gains in terms of SER. This difference between bounded and unbounded spectra motivates the algorithm in the next section.

C. Bounded Spectrum Test Algorithm

This contrast, performance-wise, between bounded and unbounded spectra clearly motivates the search for necessary and sufficient conditions for bounded spectrum. So far, no such conditions have been found in the literature. Instead, we can rely on the efficient algorithm of [15, Section V] to test whether the VLC spectrum is bounded, without actually computing the spectrum. Indeed, the equivalences in Th. 12 have the nice consequence that this algorithm, which tests whether a synchronizing sequence exists, tests also whether the VLC spectrum is bounded. This algorithm is now recalled in a slightly modified form that does not consider bit deletion and that considers only synchronizing sequences of non-zero probability. A C++ implementation is provided in [18].

Let \( \mathcal{S}_0 = \{ p \in \{R\} \cup \text{prefix}(\mathcal{V}) : l(p) \equiv 0 \pmod{l_{\text{gcd}}} \} \) and, for any \( x \in \mathcal{S}_0 \), let \( \mathcal{D}(x) \) be the union of the sets \( \{d \in \mathcal{S}_0 : \exists w \in \mathcal{V} \cup \{R\}, P(w) > 0, \exists v \in \mathcal{V}^+ \setminus \{R\}, xw = vd\} \cup \{\chi : \exists w \in \mathcal{V}, P(w) > 0, xw \notin \text{prefix}(\mathcal{V}^+)\} \).

The first set contains states \( d \in \mathcal{S}_0 \) in the Balakirsky trellis [19] that are attainable from the state \( x \) by encoding a codeword \( w \) of non-zero probability. The second set contains \( \chi \) if there exists some \( w \) of non-zero probability that leads from \( x \) to an non-existing state (represented by \( \chi \)).

Algorithm 18: (existence of a synchronizing sequence of non-zero probability)

Step 1: Let \( F = \{R, \chi\} \).

Step 2: Let \( F' = \{x \in \mathcal{S}_0 \setminus F : \mathcal{D}(x) \cap F \neq \emptyset\} \).

Step 3: Appends \( F' \) to \( F \), i.e., let \( F = F \cup F' \).

Step 4: If \( F' = \emptyset \), go to Step 5. Otherwise, go to Step 2.

Step 5: A synchronizing sequence of non-zero probability exists if and only if \( F = \mathcal{S}_0 \cup \{\chi\} \).

Step 2 selects all states from which we can attain a state in \( F \) (including \( \chi \)) with a codeword of non-zero probability. By induction, \( \mathcal{S}_0 \setminus F \) contains at the end of the algorithm the set of anti-synchronizing prefixes w.r.t. sequences of non-zero probability. This explains the decision taken at Step 5 by Lemma 11.
D. Bounded Spectrum and Non-Catastrophic VLC

Definition 19: A VLC is SER\textsubscript{L},catastrophic (resp., catastrophie in average) if a finite number of bit errors can generate (in average) an infinite number of Levenshtein symbols errors.

Theorem 20: Straightforwardly\textsuperscript{3}, no VLC with bounded spectrum is SER\textsubscript{L},catastrophic in average.

This theorem states the link between bounded spectrum and average non-catastrophic behavior. But we can state much more about bounded spectrum thanks to the equivalence with strongly bounded spectrum from Th. 12. Let us start with the analogy between a VLC with strongly bounded spectrum and a CC with non-catastrophic encoder: For such a VLC and such a CC, by Def. 6 and by [14, Th. A.1]\textsuperscript{4}, respectively, there exists \( \eta \) such that \( A_{h,l_h} \leq c^h \) and \( A_{h,l_h} = 0 \) for all \( l_h \geq \eta h \). Roughly speaking, for a given number \( h \) of bit errors, the lengths \( l_h \) of all error events of the CC and of the “most probable” error events of the VLC, are at most linearly proportional to \( h \), that is, \( l_h < \eta h \) —“most probable” means that the probability of the set of all error events of the VLC not satisfying \( l_h < \eta h \) decreases at least exponentially as \( c^h \). Since the number \( s_L \) of Levenshtein symbol errors of an error event satisfies \( s_L \leq l_h \), we can state at last a stronger result than Th. 20: The numbers of Levenshtein symbol errors of the “most probable” error events are at most linearly proportional to the numbers of bit errors. For the interested reader, Th. 26 formalizes this in a certain way in Appendix C, in order to prove Th. 21.

E. Bounded Spectrum and Guaranteed Interleaving Gains

The next theorem states essentially the interleaving gains of serial concatenations under Assumption 2 when the VLC spectrum is bounded. Since the “boundedness” of the VLC spectrum can be tested efficiently with Algorithm 18, we are given a practical tool to guarantee the interleaving gains. Note the theorem states only the (theoretical) existence and the values of the interleaving gains, not the minimum interleaving length and channel SNR required to observe these gains in practice.

Theorem 21: For the system described in Section II-B with a recursive non-catastrophic CC encoder as ECC, if \( d'_{f}^{cc} \geq 2 \) and if the VLC satisfies one of the equivalences in Th. 12, there exists under Assumption 2 some threshold independent of \( N \) such that, for any fixed SNR\textsuperscript{5} above that threshold,

\[
\begin{align*}
BER & = O\left(N^{-\left(\frac{d'_{f}^{cc}+1}{2}\right)+\epsilon}\right), \\
SER_L & = O\left(N^{-\left(\frac{d'_{f}^{cc}+1}{2}\right)+\epsilon}\right), \\
FER & = O\left(N^{-\left(\frac{d'_{f}^{cc}+1}{2}\right)+\epsilon}\right),
\end{align*}
\]

for arbitrary \( \epsilon > 0 \). See proof in Appendix C. □

\textsuperscript{3}By \( \sum_{s_L=1}^{\infty} s_L A_{s_L}^{vLC} \leq \sum_{s_L=1}^{\infty} l_h A_{s_L}^{vLC} \) and by distinguishing \( l_h < l'_h \) and \( l_h \geq l'_h \), with \( l'_h \) from Def. 4.

\textsuperscript{4}There is a typo in [14, Theorem A.1]: read \( d_{L} \geq L_{\eta} \) instead of \( d \geq L_{\eta} \).

\textsuperscript{5}If we use instead the Bhattacharyya noise parameter \( \gamma \) of the channel and the noise exponent defined as \( \alpha = -\log(\gamma) \), see details in [14], then the threshold is independent of the type of channel.

IV. Simulation Results

To illustrate the impact of the interleaving gains, let us consider the following non-optimized example: a source/VLC, framed according to \( F_h \), directly followed by a rate-1/2 regular repetition code (RC), which repeats each VLC bit twice, serially concatenated with a uniform interleaver of size \( N \) and with a recursive non-systematic punctured CC encoder with generators \( (037, 021)_8 \), where 037 is the feedback. The global code rate is fixed to \( R_c = 1/2 \), which is obtained by adapting the puncturing rate of the CC coded bits given the code rate of the chosen VLC. Finally, the VLC bits and the CC coded bits are multiplexed and sent across the channel. Though Th. 21 deals only with the serial concatenation, the developments used in the proof extend easily to other concatenations as well. For example, a double serial concatenation of a VLC with two CCs was considered in [17, Th. 6] and a hybrid concatenation of a VLC serially concatenated with a multiple parallel turbo code was considered in [17, Th. 3]. The proof of Th. 21 extends easily to these systems. It extends also easily to the particular interleaving gains obtained with reversible VLCS in [17, Th. 5] if \( d'_{f}^{vLC} \geq 3 \) and in [17, Th. 6] if \( d'_{f}^{vLC,m} \geq 3 \) — but the extension for \( d'_{f}^{vLC} = 2 \) and \( d'_{f}^{vLC,m} = 2 \) is less straightforward and requires further investigation.

Corollary 22: Since \( SER \leq FER \),

\[
SER = O\left(N^{-\left(\frac{d'_{f}^{vLC}}{2}\right)+\epsilon}\right)
\]

for arbitrary \( \epsilon > 0 \), under the assumptions of Th. 21. □

Beyond Th. 21, what happens when the VLC spectrum is not bounded? Firstly, note VLCS with unbounded spectrum are rare. In particular, it is proved in [20] that “almost all” complete VLCS have a synchronizing sequence, which under Assumption 3 is equivalent to having a bounded spectrum by Corollary 13. Nevertheless, when the VLC spectrum is not bounded, we have always at least the interleaving gain on the FER offered by the ECC since \( SER_L \leq SER \leq FER \). Besides, simulation-based results in Section IV will reveal some unexpected properties of VLCS with unbounded spectrum when the frame-MAP decoder is used.
Note the output of the RC can be viewed as a new VLC with a free distance twice as large, thus by (11),

$$\text{SER}_{\text{L}} = O\left( N^{-\frac{2d_{\text{fc}}^{\text{L}}}{2} + \epsilon} \right) = O\left( N^{-\frac{d_{\text{fc}}^{\text{L}}}{2} + \epsilon} \right).$$

(14)

This interleaving gain on the SER_{L} is illustrated in Fig. 2, with the English alphabet source (see [4] and references therein), followed by either a Huffman VLC with $d_{\text{fc}}^{\text{L}} = 1$ or a reversible VLC (RVLC) with $d_{\text{fc}}^{\text{L}} = 2$. As we can see, the SER_{L} decreases as expected as $N^{-1}$ in Fig. 2 when the Huffman VLC is used, and as $N^{-2}$ when the RVLC is used.

Let us now examine how significant it is to assume a bounded VLC spectrum—or one of the equivalences from Th. 12—in order to ensure the interleaving gains (10)–(12). Th. 21 shows that it is a sufficient condition (among others), but is it a necessary condition? The following paragraphs answer this question partially for the ML and MAP decoders.

Let us consider again the same system, but this time with the frame-ML decoder and with the MAP decoder, for two VLCs and two interleaver sizes $N = 1000$ (dashed lines) and $N = 10000$ (solid lines).

Let us elucidate what lies behind this difference: Why, to ensure (10)–(12), is the spectrum boundness not a necessary condition with the MAP decoder, while it seems to be necessary with the ML decoder? To guarantee the interleaving gain, the key requirement that emerges from the developments underlying Th. 21 and Th. 26 is essentially that the probability of long error events must be “small”. When the VLC spectrum is bounded, this requirement is trivially satisfied. When the VLC spectrum is not bounded, it is not necessarily satisfied but it might well be, though, if the frame-MAP decoder is used (instead of the frame-ML decoder) and if the symbols are not distributed as

$$P_{\text{ml}}(s) = 2^{-K l(s)}, \quad \text{with } K \sum_{s \in A} 2^{-K l(s)} = 1.$$  

(15)

To understand this, let us consider the error event formed by the pair $(S_{1:n}, \tilde{S}_{1:n})$ subject to $d_{\text{H}}(S_{1:n}, \tilde{S}_{1:n}) = h$. For the complete RVLC and the source distribution in the first column in Table I, we can show that $d_{A}(S_{1:n}, \tilde{S}_{1:n})$ increases monotonically with $n$ as $n \to \infty$ for realizations of $S_{1:n}, \tilde{S}_{1:n}$ in the typical set. In other words, with high probability and for sufficiently large $n$, $d_{A}(S_{1:n}, \tilde{S}_{1:n})$ is large enough to enable decoder is used. Such a difference is quite surprising because the difference between ML and MAP can only be explained by the residual redundancy in the VLC stream, which is very small in this case (close to 1%). Besides, notice that with the non-reversible VLC, there is almost no difference in SER_{L} between the ML and MAP decoders.

Let us consider again the same system, but this time with the frame-ML decoder and with the MAP decoder, for two VLCs and two interleaver sizes $N = 1000$ (dashed lines) and $N = 10000$ (solid lines).

Note the output of the RC can be viewed as a new VLC with a free distance twice as large, thus by (11),

$$\text{SER}_{\text{L}} = O\left( N^{-\frac{2d_{\text{fc}}^{\text{L}}}{2} + \epsilon} \right) = O\left( N^{-\frac{d_{\text{fc}}^{\text{L}}}{2} + \epsilon} \right).$$

(14)

This interleaving gain on the SER_{L} is illustrated in Fig. 2, with the English alphabet source (see [4] and references therein), followed by either a Huffman VLC with $d_{\text{fc}}^{\text{L}} = 1$ or a reversible VLC (RVLC) with $d_{\text{fc}}^{\text{L}} = 2$. As we can see, the SER_{L} decreases as expected as $N^{-1}$ in Fig. 2 when the Huffman VLC is used, and as $N^{-2}$ when the RVLC is used.

Let us now examine how significant it is to assume a bounded VLC spectrum—or one of the equivalences from Th. 12—in order to ensure the interleaving gains (10)–(12). Th. 21 shows that it is a sufficient condition (among others), but is it a necessary condition? The following paragraphs answer this question partially for the ML and MAP decoders.

Let us consider again the same system, but this time with the source and the two VLCs listed in the first three columns of Table I. Both VLCs are Huffman VLCs. The former has no particularity. The latter is a complete RVLC; its spectrum is thus not bounded, by Proposition 16. The SER_{L} is shown in Fig. 3, with the frame-ML decoder and the frame-MAP decoder. At first sight, these results are surprising and do not sound correct: With the ML decoder, the VLC performs better than the RVLC, but it is vice versa with the MAP decoder.

When the ML decoder is used (Fig. 3(a)), the interleaving gains are as expected: There is an interleaving gain with the VLC (bounded spectrum)—the SER_{L} decreases as $N^{-1}$—but there is no interleaving gain with the complete RVLC (unbounded spectrum). This corroborates the significance of bounded VLC spectra, i.e., that the spectrum “boundness” seems to be a necessary condition to ensure the interleaving gains (10)–(12) when the ML decoder is used, and explains why the complete RVLC performs so badly in Fig. 3(a).

When the MAP decoder is used (Fig. 3(b)), there is surprisingly an interleaving gain with both VLCs—the SER_{L} decreases as $N^{-1}$. This shows that the spectrum “boundness” is not a necessary condition when the MAP decoder is used.

This difference between the ML and MAP decoders concerning the spectrum boundness leads to a huge difference in SER_{L} with the complete RVLC: there is no interleaving gain when the ML decoder is used but there is one when the MAP decoder is used. Such a difference is quite surprising because the difference between ML and MAP can only be explained by the residual redundancy in the VLC stream, which is very small in this case (close to 1%). Besides, notice that with the non-reversible VLC, there is almost no difference in SER_{L} between the ML and MAP decoders.

Let us elucidate what lies behind this difference: Why, to ensure (10)–(12), is the spectrum boundness not a necessary condition with the MAP decoder, while it seems to be necessary with the ML decoder? To guarantee the interleaving gain, the key requirement that emerges from the developments underlying Th. 21 and Th. 26 is essentially that the probability of long error events must be “small”. When the VLC spectrum is bounded, this requirement is trivially satisfied. When the VLC spectrum is not bounded, it is not necessarily satisfied but it might well be, though, if the frame-MAP decoder is used (instead of the frame-ML decoder) and if the symbols are not distributed as

$$P_{\text{ml}}(s) = 2^{-K l(s)}, \quad \text{with } K \sum_{s \in A} 2^{-K l(s)} = 1.$$  

(15)

To understand this, let us consider the error event formed by the pair $(S_{1:n}, \tilde{S}_{1:n})$ subject to $d_{\text{H}}(S_{1:n}, \tilde{S}_{1:n}) = h$. For the complete RVLC and the source distribution in the first column in Table I, we can show that $d_{A}(S_{1:n}, \tilde{S}_{1:n})$ increases monotonically with $n$ as $n \to \infty$ for realizations of $S_{1:n}, \tilde{S}_{1:n}$ in the typical set. In other words, with high probability and for sufficiently large $n$, $d_{A}(S_{1:n}, \tilde{S}_{1:n})$ is large enough to enable
the MAP decoder to discard $S_{1:n}$ in favor of $S_{1:n}$, unlike the ML decoder. This explains intuitively why the probability of long error events is small and thus why there is an interleaving gain with this source/RVLC and the MAP decoder in Fig. 3(b). Put differently, this tends to show that this source/RVLC is statistically synchronizable when the MAP decoder is used.

One important condition, though, is that the symbols must not be distributed as in (15). Otherwise, the VLC frames are indeed equiprobable, the frame-MAP and frame-ML detections are equivalent under Assumption 2, the a priori distance $d_A(...)$ always equals zero and there is no interleaving gain with the complete RVLC, as illustrated in Fig. 4 with the source distribution given in the last column in Table I.

Actually, the closer the symbol distribution is to (15), the larger the interleaver size will have to be in order to notice the expected interleaving gain — because $d_A(S_{1:n}, S_{1:n})$ will increase slower with $n$. Similarly, the more reliable the channel is, the larger the interleaver size will have to be — because $d_A(S_{1:n}, S_{1:n})$ will have to be larger in order to “compete” with the higher channel reliability.

**V. EXTENSION TO FRAMING/MULTIPLEXING**

So far we have envisaged the framing rule $F_0$. Extensions to other framing rules are possible, sometimes straightforwardly. One extension is considered hereafter. See [10] for others.

Let us consider we would like to multiplex two sources of symbols, a high importance source $S^1$ containing sensitive and important information, and a low priority source $S^2$ containing some additional information. We take a strong VLC with a large free distance for $S^1$ and a weak VLC with small free distance for $S^2$. How the chosen VLCs for $S^1$ and $S^2$ will affect each other in terms of interleaving gains if we frame and multiplex them into one multiplexed VLC block of length $N$? More formally, consider two VLC streams, $V^1$ and $V^2$, with bounded spectra and respective free distances $d_f^1$ and $d_f^2$. Each stream $V^i = \text{vlc}(S_1 S_2 S_3 \ldots)$ is framed into a sub-block of length $N^i$ similarly to $F_0$ with $N_b^i = l(S_{1:N_b^i}) \leq N^i$, $N_b^i + l(S_{N_b^i+1}) > N^i$, (16)

where the $N^i$’s grow linearly with $N$ subject to $N^1 + N^2 = N$ and are known at the decoder. Given $U_{1:N_b}^i = \text{vlc}(S_1 S_{N_b} \ldots)$, we then append $N^1 - N_b^1$ zeros to form the sub-block $U_{1:N_b}^i$, and concatenate the sub-blocks as $U_{1:N} = U_{1:N_b}^1 U_{1:N_b}^2$.

How much does multiplexing alter the robustness of each stream? We can partially answer this question in terms of interleaving gains, by extending previous results. Let us focus on the impact of $V_2$ on $V_1$, i.e., how the interleaving gain on $V_1$ is altered when $V_1$ is multiplexed with $V_2$. Straightforwardly, if $d_f^2 \geq d_f^1$, the interleaving gain is unchanged. So let us focus on $d_f^2 < d_f^1$. Consider for example a multiplexed VLC block serially concatenated with $J \geq 1$ parallel concatenated recursive non-catastrophic CC encoders. By extension of the developed results, we can show the interleaving gains on $V_1$ are unaltered when $d_f^2 \geq 3$, or when $d_f^2 \geq 2$ and $J \geq 2$, or when $d_f^2$ is even, $d_f^2 = 1$ and $J \geq 3$. They are degraded by one unit when $d_f^2$ is odd, $d_f^2 = 1$ and $J \geq 3$. And it is self-evident that they are completely degraded when $d_f^2 = 1$ and $J = 1$. So only the cases $(d_f^2 = 1, J = 2)$ and $(d_f^2 = 2, J = 1)$ remain, which unfortunately are not as straightforward. For example, the interleaving gain is degraded by at least one unit when $d_f^2$ is odd, $d_f^2 = 1$ and $J = 2$. But further information requires additional developments (here beyond scope) that are worth considering since multiplexing is heavily used in practice.

**VI. CONCLUSIONS**

How much VLCs contribute to the so-called interleaving gains in concatenated systems has been stated and proved in Th. 21 for the serial concatenation of a VLC with a convolutional code. Extending this to the concatenations with multiple parallel and/or serial turbo codes is straightforward, but extending to other codes requires further research. To summarize, when the frame-ML decoder is used (source statistics unknown at the decoder), one notable sufficient condition in Th. 21 is that the VLC must have a bounded spectrum, a condition that can be tested with the simple Algorithm 18. Interestingly, simulation results have illustrated that this condition seems to be also necessary. By contrast, when the frame-MAP decoder is used (source statistics known at the decoder), simulation results have highlighted that this condition is not necessary — that is, Th. 21 holds also when the spectrum is not bounded — if the source statistics do not satisfy equation (15). Developing further these simulation-based conclusions is certainly worth considering for future works. Nevertheless, in a broadcast system where both MAP/ML decoders may be used, the safest solution is to consider the weakest system (ML decoder) and thus to ensure that the VLC has a bounded spectrum.

Another interesting result of this work is the concept of bounded VLC spectrum, which we proved to be closely related to some flavor of statistically synchronizable VLCs, which themselves relate to the Balakirsky trellis. It would be interesting to investigate how statistical synchronization on other trellises influence the distance spectrum and the interleaving gains. In this context, statistical synchronization on the full symbol-clock trellis [2], is certainly one promising possibility for SER-performance, e.g., by means of markers that would ensure the synchronization of the symbol clock.
(at least statistically) —synchronizing sequences ensure the statistical synchronization of only the VLC codewords.

To conclude with a practical viewpoint, we encourage code designers not to change radically their preferred design techniques but at least to check and make sure that the VLC has a bounded spectrum (thanks to Algorithm 18 and Th. 12) and to analyze the interleaving gains when they design the concatenation of a VLC with a channel code. Beyond this, maximizing the probability of the set of synchronizing sequences of the VLC is another interesting approach and a subject for future research because this lowers the constant $c$ in Def. 6 (by extension of the proof of Th. 12), which should have a positive impact on the average length of the error events of the VLC and on the minimum interleaver length above which the interleaving gains are noticeable.

APPENDIX

A. Proof of Lemma 11

Proof of Lemma 11: The implications (a)$\Rightarrow$(b)$\Rightarrow$(c) follow straightforwardly from Def. 9 and Def. 10. The implication (c)$\Rightarrow$(a) follows almost completely from the “necessity” part of the proof of [15, Th. 1] by rewriting the VLC codewords with a new code alphabet $X = \mathcal{B}^{k}$.\hfill $\blacksquare$

B. Proof of Th. 12

Let us prove successively the equivalence (a)$\Leftrightarrow$(b) and the implications (a)$\Rightarrow$(d), (d)$\Rightarrow$(c) and (c)$\Rightarrow$(a).

Proof of Th. 12, equivalence (a)$\Leftrightarrow$(b): It follows almost completely as a particular case of [15, Corollary 1] by rewriting again the VLC codewords with a new code alphabet $X = \mathcal{B}^{k}$.\hfill $\blacksquare$

Proof of Th. 12, implication (a)$\Rightarrow$(d): If there exists a synchronizing sequence $s_{l_{1}}$ of non-zero probability, then the VLC stream spectrum $A_{h,l_{1}}^{\text{vlc}}$ can be upper bounded as follows. Let us restrict the summation to sequences $s_{l_{1}}$ of non-zero probability (the others do not contribute to the spectrum). Then, $l(s_{l_{1}}) < l_{\text{max}}$, thus $A_{h,l_{1}}^{\text{vlc}} \leq l_{\text{max}}/l_{h}$ and $A_{h,l_{1}}^{\text{vlc}} \leq l_{\text{max}}/l_{h} P(s_{l_{1}})$. Note $P$ is the probability of sequences $s_{l_{1}}$ such that $A_{h,l_{1}}^{\text{vlc}} > 0$. Let us find a set that includes these sequences so that we can upper bound $P$ by the probability of this set. To do this, let split the sub-sequence $s_{l_{1}}$ into $|l_{s} - 1|/k$ parts of symbol length $k$, plus a possible residue of less than $k$ symbols. Notice $A_{h,l_{1}}^{\text{vlc}} = 0$ when at least $h + 1$ parts of $s_{l_{1}}$ contain $s_{l_{1}}$, because we have then $h + 1$ copies of the synchronizing sequence, thus at least one of them is not affected by any of the $h$ bit errors, which resynchronizes the decoder and thus prevents any error event from going further. Consequently, the set of $s_{l_{1}}$ such that at most $h$ parts of $s_{l_{1}}$ contain $s_{l_{1}}$ includes the set of sequences $s_{l_{1}}$ subject to $A_{h,l_{1}}^{\text{vlc}} > 0$. Therefore, $P \leq c^{\ell_{1}}(l_{s} - 1)/k - h$ with $c^{\ell_{1}} = 1 - P(s_{l_{1}}) < 1$. Then, to summarize, $A_{h,l_{1}}^{\text{vlc}} = l_{\text{max}}/l_{h}$ and $c^{\ell_{1}}(l_{s} - 1)/k - h$ for some $\eta$, where we have used $n^{\eta} \leq n^{k}/k!$. At last, note $h^{2}/h! \leq n^{k}$, let $K = e^{c^{\ell_{1}}(l_{s} - 1)/k - h}$ and let $c^{\ell_{1}} = e^{n^{k}/k}$. Then, $A_{h,l_{1}}^{\text{vlc}} = n^{k} e^{n^{k}/k} (K \eta e^{n^{k}/k})^{h}$. It follows that there exists $c < 1$ and $\eta$ such that for all $l_{s} \geq l_{h}$, $A_{h,l_{1}}^{\text{vlc}} \leq c^{\ell_{1}}$, which concludes the proof by Property 7.

Proof of Th. 12, implication (d)$\Rightarrow$(c): This is straightforward, by definition.

For convenience, we say in the following that a pair of sequences $(u, \bar{u})$ starts an error event if there exists a pair $(u', \bar{u}')$, $l(u') = l(\bar{u}') > 0$, such that $(u, u', \bar{u}, \bar{u}')$ forms an error event. Besides, let $l_{h}^{0}$ be the threshold above which (i) any integer $l$ multiple of $l_{h}^{0}$ can be written as a summation of lengths of codewords of non-zero probability; and (ii) any integer $l$ multiple of $l_{h}^{0}$ can be written as a summation of lengths of codewords. Such a threshold always exists. Let $l_{s}^{0} = \lfloor(l_{h}^{0} + 2l_{\text{max}})/l_{\text{min}}\rfloor$.

Lemma 23: Under Assumption 1, if a VLC has no synchronizing sequence of non-zero probability, then there exists a pair $(u_{1}, \bar{u}_{1})$ that starts an error event such that $u_{1} \in \mathcal{V}^{+}$ with $P(u_{1}) > 0$, $u_{1}$ is a synchronizing w.r.t. sequences of non-zero probability, $l(u_{1}) = l(\bar{u}_{1}) \leq l_{h}^{0} + 2l_{\text{max}}$ and $l_{s}^{0}(u_{1}) = l_{h}^{0}$.

In words, the decoder can reach an anti-synchronizing sequence with a bounded number of bit errors.

Proof: By Lemma 11, there exists a prefix $p \in$ prefix($\mathcal{V}^{+}$) that is anti-synchronizing w.r.t. sequences of non-zero probability. Note $l(p) < l_{\text{max}}$ and $l(p) \equiv 0 \pmod{l_{h}}$. Let $u \in \mathcal{V}^{+}$ be the shortest (in bit length) sequence of codewords with $l(u) \equiv l_{h}$ and $l(u) \equiv 0 \pmod{l_{h}}$. Such a sequence always exists by definition of $l_{h}^{0}$. Besides, still by definition of $l_{h}^{0}$, there exists a sequence $u \in \mathcal{V}^{+}$ of length $l(u) = l(\bar{u}) > 0$.

At last, let $i$ be the largest integer with $p_{i}(u, \bar{u})$ and let $j = l(\bar{u})$. Note $l(u) \equiv l_{h}^{0} + 2l_{\text{max}}$. Therefore the pair formed by $u_{1} = u_{i+j}(p)$ and $\bar{u}_{1}$ forms an error event.

Lemma 24: Under Assumption 1, given $u_{1} \in \mathcal{V}^{+}$ with $l(u_{1}) \geq l_{h}^{0} + 2l_{\text{max}}$ and given a pair $(u, \bar{u})$ that starts an error event with $u \in \mathcal{V}^{+}$, there exist an integer $l_{s} \leq l_{h}^{0} + 2l_{\text{max}}$ and a sequence $u_{i+j}(p)$ such that $(u, u_{i+j}(p), \bar{u}, \bar{u}_{i+j}(p))$ forms an error event.

Proof: Let $s \in$ suffix($\mathcal{V}^{+}$) be a suffix such that $u_{s} \in \mathcal{V}^{+}$. Such a suffix always exists. Note $l(s) < l_{\text{max}}$ and $l(s) \equiv 0 \pmod{l_{h}}$ since $l(u) \equiv l(\bar{u}) \equiv 0 \pmod{l_{h}}$. Let $i$ be the smallest integer with $i \geq l_{h}^{0} + 2l_{\text{max}}$ and $u_{i+j}(p) \in \mathcal{V}^{+}$. By definition of $l_{h}^{0}$, there exists an error event $u_{i+j}(p)$ of length $i - l(s)$. At last, let $\bar{u}_{i} = u_{i} - u_{i+j}(p)$ and let $l_{s}$ be the smallest integer with $p_{i}(u_{1}, u_{1+j}(p)) \equiv 0 \pmod{l_{s}}$. Note $l_{s} < l_{h}^{0} + 2l_{\text{max}}$ and $(u_{i+j}(p), u_{i+j}(p), u_{i}, \bar{u}_{i})$ forms an error event.

Proof of Th. 12, implication (c)$\Rightarrow$(a): Let us prove the contrapositive of (c)$\Rightarrow$(a), that is, the spectrum is not bounded if the VLC has no synchronizing sequence of non-zero probability. Let $l \geq l_{h}^{0}$. We now focus on the quantity $A_{l} \equiv \sum_{k=1}^{l_{h}^{0} + l_{\text{max}}} A_{h, l_{1}}^{\text{vlc}}$, which can be rewritten as $A_{l} = \sum_{s_{l_{1}} \in \mathcal{A}} P(s_{l_{1}}) A_{l_{1}}^{\text{vlc}}$, with $A_{l_{1}}^{\text{vlc}} \equiv \sum_{s_{l_{1}} \in \mathcal{A}} \sum_{l_{1} \equiv l_{h}} A_{h, l_{1}}^{\text{vlc}}$. Notice that $A_{l}$ is bounded, i.e., $A_{l} = O(c^{\ell_{1}})$ with $c < 1$, if and only if the VLC

For example, for $l_{h}^{0} = 2$, the VLC codeword u_{1}s_{8} = 00011011 becomes $u_{1}s_{1} = abcd$ with $a = 00$, $b = 01$, $c = 10$ and $d = 11$. In terms of desynchronization on the Balalisky trellis, insertions/deletions of code letters have then the same impact as binary errors.
spectrum is bounded. We can therefore prove that the VLC spectrum is not bounded by proving that \( A^h \) is lower bounded by a constant independent of \( l \), which we will do for \( l \geq 2l^h \).

In the expression of \( A^h \), let us restrict the summation to all \( s_{1:t} \) of non-zero probability that start with \( s^1 = \text{vlc}^{-1}(\hat{u}^1) \), where \( \hat{u}^1 \) is given by Lemma 23 — note \( P(s^1) > 0 \). Let \( l_1 = l_2(s^1) \). If we assume temporarily that \( A_{1|s_1, t} \geq 1 \) for such sequences \( s_{1:t} \), then \( A^h \) is trivially lower bounded by \( A^h \geq \sum_{s_{1:t} \in \text{vlc}}: s_{1:t} = s_1 \) \( P(s_1) \) \( A_{1|s_1, t} \geq P(s_1) \), which is a constant independent of \( l \).

So the proof boils down eventually to showing that \( A_{1|s_1, t} \geq 1 \) for sequences \( s_{1:t} \) of non-zero probability that start with \( s^1 \). This follows from previous lemmas. Let \( l_2 = l - l_2(s^1) \); note \( l_2 \geq l_1 \) since \( l \leq 2l^h \). Given any such \( s_1 \), there exists by Lemma 23 a pair \((s^1_1, \hat{u}^1)\) that starts an error event (with some abuse of notation). By appending \( s_{1:t-1} \), the pair \((s^1_1, \hat{u}^1)\) starts also an error event. At last, given \( s_1 \), there exists by Lemma 24 some \( \hat{u}^1 \) — note \( l_3(s_1, \hat{u}^1) = l^h \) and \( l(s_1, \hat{u}^1) > l^h + 2l_{\text{max}} \) — such that the pair \((s^1_1, \hat{u}^1)\) forms an error event \( e \) for some \( l_3 \) subject to \( l_3 = l - l_{\text{max}} + 1 \leq l \). Note \( d_H(e) \leq d_H(\hat{u}^1) + d_H(\hat{u}^1) \) and thus \( d_H(e) \leq 2l^h + 4l_{\text{max}} \) by Lemma 23 and Lemma 24. Consequently, \( A_{1|s_1, t} \geq 1 \), which concludes the proof.

C. Proof of Th. 21

Lemma 25: Under Assumption 2, there exist \( \epsilon \) such that, for all \( h \), either \( A_{\text{vlc}}^l = 0 \) or \( A_{\text{vlc}}^h > \epsilon^h \). See proof in [10].

Theorem 26: Under Assumption 2, for bounded VLC spectra, there exists \( \mu \) such that

\[
\sum_{s_{1:t} \geq 1} s_{1:t} A^\text{vlc}_{s_{1:t}, h} \leq \mu h A_{\text{vlc}}
\]

for all \( h \). Note the analogy with [14, Th. A.1].

It follows from Th. 12 and Th. 26 that all bounded spectra satisfy (17) under Assumption 2.

Proof: When \( A_{\text{vlc}}^l = 0 \), (17) holds trivially. Let us thus prove (17) for \( A_{\text{vlc}}^l > 0 \). Let \( c \) and \( \eta \) be values from Def. 6 of a strongly bounded spectrum. Then, for \( \mu' \geq \eta \),

\[
\sum_{s_{1:t} \geq 1} s_{1:t} A^\text{vlc}_{s_{1:t}, h} \leq \sum_{s_{1:t} \geq 1} \sum_{l_1} l_1 A^\text{vlc}_{l_1, h} \leq \mu' h A_{\text{vlc}} + \sum_{l = \mu' h} \sum_{s_{1:t}} s_{1:t} A^\text{vlc}_{s_{1:t}, h} \leq \mu' h A_{\text{vlc}} + \epsilon^h.
\]

\[
\sum_{s_{1:t} \geq 1} s_{1:t} A^\text{vlc}_{s_{1:t}, h} \leq (\mu - 1) h A_{\text{vlc}} + \epsilon^h,
\]

where \( \epsilon^h < \epsilon^h < A_{\text{vlc}}^h \leq h A_{\text{vlc}}^h \), which concludes the proof.

Lemma 27: If the VLC spectrum is strongly bounded, then under Assumption 2 the spectrum of the VLC block \( A^h_{s_{1:t}, h} \) satisfies

\[
\sum_{s_{1:t}} s_{1:t} A^h_{s_{1:t}, h} \leq \mu h \theta^h \left( \frac{N}{[h/d^h]} \right)
\]

where \( \mu, \theta^h \) are constants independent of \( h \) and \( N \).

Proof: From the expression (8) of \( T_{s_{1:t}, h, l, k, n} \)

\[
\sum_{s_{1:t}, l, k, n \geq 1} \sum_{m=1}^n \sum_{n'=1}^n A^\text{vlc}_{s_{1:t}, l, k, s_{1:t}, h, l, k, n} = \sum_{s_{1:t}, l, k, s_{1:t}, h, l, k, n \geq 1} \sum_{m=1}^n \sum_{n'=1}^n A^\text{vlc}_{s_{1:t}, l, k, s_{1:t}, h, l, k, n} \left( \frac{[h/d^h]}{[h/d^h]} \right)
\]

With a few manipulations and by using Th. 26,

\[
\sum_{s_{1:t}, l, k, s_{1:t}, h, l, k, n \geq 1} \sum_{m=1}^n \sum_{n'=1}^n A^\text{vlc}_{s_{1:t}, l, k, s_{1:t}, h, l, k, n} \leq \mu h \sum_{l_1, h_1, h_2, \ldots, h_n \geq 1} \prod_{l=1}^n A^\text{vlc}_{h_1, h_2, \ldots, h_n}\]

Given \( c \) and \( \eta \) from Def. 6, let \( K'' = \sum_{l} \epsilon^l \), \( K' = K'' + 2\eta^h \) and \( K = 2K' \). Then

\[
A^h_{\text{vlc}} = \sum_{l=1}^{\eta h} A^h_{l, h} \leq 2h^h + \sum_{l_h = \eta h}^{\eta h} \epsilon^h \leq 2^{\eta h} + K' \leq 2h^h + K'' + K = K'' + 2K'
\]

Eventually, the lemma follows with \( \theta^h = 4K''/\text{gcd}; \) this can be proved by checking the case \( \epsilon_{\text{vlc}}^h = 2h^h \). Therefore, since the summation in (22) involves \( h_{\text{vlc}}' = \frac{1}{h_{\text{vlc}}''} \leq 2^{-1} \text{ terms} \), we have

\[
\sum_{s_{1:t}, l, k, h, l, k, n \geq 1} \sum_{m=1}^n \sum_{n'=1}^n A^\text{vlc}_{s_{1:t}, l, k, s_{1:t}, h, l, k, n} \leq \mu h K'' \text{gcd} \sum_{n=1}^n \left( \frac{[h/d^h]}{[h/d^h]} \right)
\]

Lemma 28: If the VLC spectrum is strongly bounded, then under Assumption 2 the quantities \( A^h_{l, h} \) and \( A^\text{vlc}_{s_{1:t}, h} \) are all upper bounded by \( \theta^h \left( \frac{[h/d^h]}{[h/d^h]} \right) \), where \( \theta \) is a constant independent of \( h \) and \( N \).

Proof: This follows from Lemma 27 with \( \theta = \max \{3\mu, \theta^h, 3\theta^h\} \).

Lemma 29: The interleaving gains (10)–(12) in Th. 21 hold for \( d^\text{vlc} \geq 3 \).

Proof: This lemma is an extension of [14, Th. 8.4]; the proof hereafter relies heavily on the results from [14] and uses the same notations. Let \( n \) denote the length of the global (concatenated) code in Fig. 1, \( n = N/r_{cc} \) where \( r_{cc} \) is the rate of the CC. Let \( \overline{T}_{n,k} \) be the error rate we would like to upper bound, either BER, SERL, or BER. By Th. 5. \( n_{\text{min}} \) \( T_{n,k} \leq k(n) \sum_{n=1}^{N/2} B_{h}^n P_{h} \) and \( B_{h}^n \) = \( \sum_{l} B_{h}^l \), where \( A_{h}^l \) is the spectrum of the block-equivalent CC, \( B_{h}^l \) = \( d_A^l \) and \( k(n) \) = \( 1/(N - n_{\text{max}}) \), the BER, \( B_{h}^l \) = \( \sum_{s_{1:t}} s_{1:t} A_{h}^l \) for the SERL, and \( B_{h}^l \) = \( A_{h}^l \) for the BER. The values of \( k(n) \) are upper bounds on \( (N_{\text{min}}-1)^{-1} \) and \( (N_{\text{min}}) \) from Th. 5. Let \( B_{h}^n \) denote \( \sum_{s_{1:t}} s_{1:t} A_{h}^l \), let \( L_{2} \)
be the trellis length of the CC and note that $A_{d,h}^{[2]} \neq 0$ implies $d \leq \mu h$ by [14, Th. A.1] for some $\mu$. Then, by applying Lemma 28 on $B_{d}^{[1]}$ and [14, Th. A.3] on $A_{d,h}^{[2]}$, 

$$
\sum_{d=1}^{H} \sum_{h=1}^{D} \frac{B_{d,h}^{[2]}}{(d)^{\mu h}} \leq \sum_{d=1}^{H} \sum_{h=1}^{D} \frac{\theta^{d}(d/2)^{\mu h}}{d} \sum_{j=0}^{d/2} \left( \frac{\eta h}{d-j} \right)^{d-j},
$$

(24)

for some $\mu$, $\theta$ and $\eta$ independent of $n$ and $h$. This expression corresponds exactly to [14, eq. (7.3)]. The rest of the developments in [14, Section VII] then holds straightforwardly.

This enables to draw a few conclusions. Firstly, the ensemble threshold $c_{0}$ (defined in [14]) is finite for $d_{f}^{\text{vlc}} \geq 2$, by [14, Lemma 8.5]. Secondly, it follows straightforwardly from [14, Corollary 5.2] and [14, Lemma 8.6] that

$$
\sum_{h=1}^{\infty} B_{h}^{[2]} P_{h} = O\left(N^{-\left[d_{f}^{\text{vlc}}-1\right]^{-1}} + \epsilon\right)
$$

(25)

for $d_{f}^{\text{vlc}} \geq 3$. To conclude, note $k^{(n)} = 1$ for the BER and $k^{(n)} = O(N^{-1})$ for the FER, $S_{\text{LE}}$.

**Proof of Th. 21**: The interleaving gain on the FER follows from Lemma 29 for $d_{f}^{\text{vlc}} \geq 3$ and from the inequality FER $\leq 1$ for $d_{f}^{\text{vlc}} = 2$. Nevertheless, let us first develop an upper bound on the FER; this bound will then be used to tackle the BER and the $S_{\text{LE}}$. Recall the notations and the intermediate results from the proof of Lemma 29. Besides, let $H$ be the number of bit errors among the coded bits and let $D_{n}$ be, as in [14], a fixed sequence of integers satisfying $D_{n}/n^t \to 0$ and $\log(n)/D_{n} \to 0$ as $n \to +\infty$, for all $t > 0$. For example, $D_{n} = \log^{2}(n)$. By definition, the FER is equal to $P(H \geq 1) = P(1 \leq H \leq D_{n}) + P(H > D_{n})$. Note $P(H > D_{n}) \leq \sum_{h > D_{n}} B_{h}^{[2]} P_{h}$. Since the ensemble threshold $c_{0}$ is finite for $d_{f}^{\text{vlc}} \geq 2$, it follows from the proof of [14, Theorem 5.1] that there exist constants $n_{0}, K, \delta > 0$ such that $\sum_{h > D_{n}} B_{h}^{[2]} P_{h} \leq K e^{-\delta D_{n}}$ for $n > n_{0}$, and thus

$$
\text{FER}^{(n)} \leq P(1 \leq H \leq D_{n}) + K e^{-\delta D_{n}}
$$

(26)

where \( \text{FER}^{(n)} \) recalls the dependency on $n = N/r_{\text{cc}}$.

Let $W$ be the number of bit errors among the VLC bits.

Note $A_{w,h}^{[2]} \neq 0$ implies $w \leq \mu h$ by [14, Th. A.1] for some constant $\mu$. Therefore, $H \leq D_{n}$ implies $W \leq \mu D_{n}$ and

$$
\text{BER}^{(n)} \leq (N_{\text{min}}^{-1} \mu D_{n} P(1 \leq H \leq D_{n}) + K e^{-\delta D_{n}} \\
\leq k^{(n)} \mu D_{n} \text{FER}^{(n)} + K e^{-\delta D_{n}}
$$

(27)

$\delta$ is the integer parameter of the SER-L.

where $N_{\text{min}}^{-1}$ is the minimum number of bit errors, the bit errors for which the SER-L is defined.

Lastly, we observe that $A_{d,h}^{[2]} \neq 0$ implies $d \leq \mu h$ by [14, Th. A.3] for some constant $\mu$. Then, by applying Lemma 28 on $B_{d}^{[1]}$ and [14, Th. A.3] on $A_{d,h}^{[2]}$, 

$$
\sum_{d=1}^{H} \frac{B_{d}^{[2]}}{(d)^{\mu h}} \leq \sum_{d=1}^{H} \frac{\theta^{d}(d/2)^{\mu h}}{d} \sum_{j=0}^{d/2} \left( \frac{\eta h}{d-j} \right)^{d-j},
$$

(24)

for some $\mu$, $\theta$ and $\eta$ independent of $n$ and $h$. This expression corresponds exactly to [14, eq. (7.3)]. The rest of the developments in [14, Section VII] then holds straightforwardly.

This enables to draw a few conclusions. Firstly, the ensemble threshold $c_{0}$ (defined in [14]) is finite for $d_{f}^{\text{vlc}} \geq 2$, by [14, Lemma 8.5]. Secondly, it follows straightforwardly from [14, Corollary 5.2] and [14, Lemma 8.6] that

$$
\sum_{h=1}^{\infty} B_{h}^{[2]} P_{h} = O\left(N^{-\left[d_{f}^{\text{vlc}}-1\right]^{-1}} + \epsilon\right)
$$

(25)

for $d_{f}^{\text{vlc}} \geq 3$. To conclude, note $k^{(n)} = 1$ for the BER and $k^{(n)} = O(N^{-1})$ for the FER, $S_{\text{LE}}$. To conclude, note that all realizations in the VLC block realizations that contain $s_{\text{LE}}^{*}$, i.e., it may contain a sub-sequence of $s_{\text{LE}}^{*}$ but it cannot contain an occurrence of $s_{\text{LE}}^{*}$ entirely. Let then $\mathcal{P}^{(n)}$ be the set of VLC block realizations that contain an $s^{n}$-free sequence of at least $D_{n}$ symbols, and let $\mathcal{P}^{(n)}$ be the complement of $\mathcal{P}^{(n)}$. Let us first investigate the set $\mathcal{P}^{(n)}$. Note that, given $W \leq \mu D_{n}$, the bit errors affect at most $W$ occurrences of $s_{\text{LE}}^{*}$ and we need at most $W$ occurrences of $s_{\text{LE}}^{*}$ (loose upper bound) to resynchronize. Thus in $\mathcal{P}^{(n)}$, by construction, occurrences of $s_{\text{LE}}^{*}$ are separated from each other by at least $D_{n}$ symbols (loose upper bound). Therefore, the total symbol length of the error events is upper bounded by $WD_{n} + Wk \leq (1 + D_{n})\mu D_{n}$ and thus $S_{\text{LE}}^{*} \leq \min\left((1 + D_{n})\mu D_{n}, s_{\text{LE}}^{*} \leq s_{\text{LE}}^{*} \right)$ since $s_{\text{LE}}^{*}(a,b) \leq \min\left(\{s_{a}(a), s_{b}(b)\} \right)$ when $(a) = (b)$. Consequently, for $n = N/r_{\text{cc}}$, large enough, no sequence in $\mathcal{P}^{(n)}$ can lead to $S_{\text{LE}}^{*} > s_{\text{LE}}^{*}$ given at most $W \leq \mu D_{n}$ bit errors, and thus

$$
P(S_{\text{LE}}^{*} > s_{\text{LE}}^{*} | W \leq \mu D_{n}) \leq P(\mathcal{P}^{(n)}).
$$

(30)

To conclude, note that all realizations in $\mathcal{P}^{(n)}$ contain an $s^n$-free sequence of at least $D_{n}$ symbols and that the probability to have such an $s^n$-free sequence is trivially upper bounded by $N(1 - P(s_{\text{LE}}^{*}))^{D_{n}}$ for $N$ sufficiently large. Therefore, $P(\mathcal{P}^{(n)}) \leq N(1 - P(s_{\text{LE}}^{*}))^{D_{n}} = O(e^{-D_{n}}) = O(e^{-\delta D_{n}})$ for some constant $c < 1$ and $\delta' = -\log(c) > 0$. This, together with (30), concludes the proof. \hfill \blacksquare
ACKNOWLEDGMENT
The reviewers have made interesting comments to improve the level of this paper. Thank you very much!

REFERENCES


Xavier Jaspar was born in Ottignies-Louvain-la-Neuve, Belgium, in 1980. He worked on joint source-channel coding in 2002 as a visiting scientist at the Electrical and Computer Engineering Department at McGill University, Montreal, Canada, and from 2003 to 2008 at the Communications and Remote Sensing Laboratory at the Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium. He received the Engineering degree in Applied Mathematics (summa cum laude) in 2003 and the Ph.D. degree in 2008, from UCL. Since 2008, he has been working at Laboratoire, a technical competence and research center of GDF SUEZ in energy processes and energy use.

Luc Vandendorpe was born in Mouscron, Belgium, in 1962. He received the Electrical Engineering degree (summa cum laude) and the Ph.D. degree from the Université catholique de Louvain (UCL) Louvain-la-Neuve, Belgium in 1985 and 1991 respectively. Since 1985, he is with the Communications and Remote Sensing Laboratory of UCL. In 1992, he was a Visiting Scientist and Research Fellow at the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands. Presently he is full Professor and head of the EE department of UCL. He is mainly interested in digital communication systems: equalization, joint detection/synchronization for CDMA, OFDM (multicarrier), MIMO and turbo-based communications systems (UMTS, xDSL, WLAN, etc.) and joint source/channel (de)coding. In 1990, he was co-recipient of the Biennial Alcatel-Bell Award from the Belgian NSF. In 2000, he was co-recipient (with J. Louveaux and F. Deryck) of the Biennial Siemens Award from the Belgian NSF. He is or has been TPC member for IEEE VTC Fall 1999, IEEE Globecom 2003 Communications Theory Symposium, the 2003 Turbo Symposium, IEEE VTC Fall 2003, IEEE SPAWC 2005 and IEEE SPAWC 2006. He was co-technical chair (with P. Duhamel) for IEEE ICASSP 2006. He is an elected member of the Sensor Array and Multichannel Signal Processing committee of the Signal Processing Society and associate editor of the IEEE Trans. on Signal Processing. L. Vandendorpe is a Fellow of the IEEE.