Entropy/Length Profiles, Bounds on the Minimal Covering of Bipartite Graphs, and Trellis Complexity of Nonlinear Codes

Ilan Reuven, *Student Member, IEEE*, and Yair Be'ery, *Member, IEEE*

***Abstract*—In this paper, the trellis representation of nonlinear codes is studied from a new perspective. We introduce the new concept of entropy/length profile (ELP). This profile can be considered as an extension of the dimension/length profile (DLP) to nonlinear codes. This elaboration of the DLP, the entropy/length profiles, appears to be suitable to the analysis of nonlinear codes. Additionally and independently, we use well-known information-theoretic measures to derive novel bounds on the minimal covering of a bipartite graph by complete subgraphs. We use these bounds in conjunction with the ELP notion to derive both lower and upper bounds on the state complexity and branch complexity profiles of (nonlinear) block codes represented by any trellis diagram. We lay down no restrictions on the trellis structure, and we do not confine the scope of our results to proper or one-to-one trellises only. The basic lower bound on the state complexity profile implies that the state complexity at any given level cannot be smaller than the mutual information between the past and the future portions of the code at this level under a uniform distribution of the codewords. We also devise a different probabilistic model to prove that the minimum achievable state complexity over all possible trellises is not larger than the maximum value of the above mutual information over all possible probability distributions of the codewords. This approach is pursued further to derive similar bounds on the branch complexity profile. To the best of our knowledge, the proposed upper bounds are the only upper bounds that address nonlinear codes. The novel lower bounds are tighter than the existing bounds. The new quantities and bounds reduce to well-known results when applied to linear codes.**

***Index Terms*—Bipartite graphs, complete graphs, entropy, entropy/length profiles, mutual information, nonlinear codes, trellis complexity.**

I. Introduction

A TRELLIS diagram can be viewed as an efficient representation of a code for a maximum-likelihood soft-decision decoding. Trellis description of block codes was originally introduced in 1974 by Bahl *et al*. [1] for maximum *a posteriori* decoding. In 1978, Wolf [28] used a trellis construction based on the parity-check matrix to apply the Viterbi algorithm for soft-decision decoding of block codes. In the same year, Massey [19] introduced an alternative trellis construction. The formal definition of the trellis construction for linear block codes was made by Forney [4]. All the above references confine their scope to linear codes. Muder [21] provided a rigorous graph-theoretic definition of trellises and showed that Forney's trellis is the minimal trellis representation of any linear block code. In his paper Muder also addressed nonlinear codes and pointed out that the minimal proper trellis for a nonlinear code need not be the minimal trellis. The fact that the original trellis introduced by Bahl *et al*. is isomorphic to the minimal trellis presented later by Forney and Muder was announced by Kot and Leung [13], and later by Zyablov and Sidorenko [29]. Another significant contribution to the trellis analysis of block codes is due to Kschischang and Sorokine [15]. In their paper, the authors examined the trellis structure of general block codes and concluded that minimal trellises for nonlinear codes are computationally intractable. The solutions of this minimization may result in improper or unobservable trellises. Most of the study in [15] is dedicated, however, to the construction of the minimal trellis for linear codes by forming a product of elementary trellises. These trellises correspond to the one-dimensional subcodes generated by the so-called "atomic codewords." The trellis-oriented generator matrix that was introduced by Forney [4] is composed of these codewords. In a recent survey on trellis theory for linear block codes McEliece [20] showed that among all trellises that represent a linear code, the BCJR trellis [1] has the fewest edges and the minimum vertex count at each level, simultaneously. Recently, Vardy and Kschischang [26] showed that this minimal trellis also minimizes the *expansion index* of the trellis (the total number of "bifurcations"). Moreover, they extended these results to the general class of *rectangular* [14] (*separable* [22], [23]) codes. All these complexity measures are minimized in the unique biproper trellis of these codes ([14], [26]). These results were independently obtained by Sidorenko [22].

Every block code C has a unique minimal proper trellis. If C is a linear code then this trellis is also the minimal trellis. This minimal trellis is characterized by the fact that any other trellis for C has at least as many states and as many edges as the minimal trellis in every level of the diagram. The problem of finding a trellis representation for a nonlinear block code that minimizes measures of trellis complexity (state complexity and branch complexity), even under a given coordinate permutation, appears to be computationally infeasible. The analysis of the complexity measures of a trellis representation

Manuscript received November 12, 1996; revised June 4, 1997. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, June 1997.

The authors are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat Aviv 69978, Tel-Aviv, Israel.

Publisher Item Identifier S 0018-9448(98)00656-7.

of a nonlinear code raises several difficulties that are not encountered in the analysis of trellis diagrams for linear codes.

1) The unique minimal proper trellis which exists for all block codes (under a given coordinate permutation) need not be a minimal trellis of a nonlinear code. Furthermore, the solution of this minimization problem may not be unique.
2) The minimal trellis for nonlinear codes may include several edges emanating from the same vertex with the same label (improper trellis). Improper trellises for nonlinear codes may have fewer vertices than that of proper diagrams ([15], [21]). Moreover, a minimal trellis of a nonlinear code may not be observable ("one-to-one"), i.e., the inclusion of distinct paths that represent the same codeword may reduce complexity measures (e.g., [15]).
3) The minimization of the number of vertices (states) at one level, under a given coordinate ordering, may impose an increase of the vertex count at an other level [15].

Given a coordinate ordering, the state and the branch complexity profiles of a linear block code are simple functions of the ordered *dimension/length profile* (DLP) [5], and the appropriate diagrams are easily computed. The DLP lower bound on the state complexity profile of a code under any permutation of its coordinates is determined by the (unordered) DLP. The minimal trellis of a nonlinear code may not be unique, and there is no equivalent formula to determine the state and the branch complexity profiles of its trellis diagram. Thus the problem of deriving lower and upper bounds on the trellis complexity of nonlinear codes has a considerable practical importance. Lafourcade and Vardy [16] developed new lower bounds on the state and branch complexity profiles. These bounds are obtained by dividing the index axis of the code into several sections. The authors generalize the bounds to nonlinear codes by the introduction of the *cardinality/length profile*. This quantity does not utilize the asymmetry of nonlinear codes and thus their appropriate bounds for nonlinear codes do not appear to be tight. The authors thus use combinatorial properties of the specific codes under discussion to tighten the lower bounds. The latter technique is stated to biproper trellises only.

In this paper we study the trellis complexity (state complexity and branch complexity profiles) of nonlinear codes. We employ information-theoretic measures to extend the notion of DLP to nonlinear codes by the introduction of the *entropy/length profile* (ELP). The entropy/length profiles appear to be advantageous descriptive characteristics of nonlinear codes and the suitable tools to derive tight bounds on the trellis complexity of any trellis representation of these codes. The novel bounds, which also apply to improper and non-one-to-one trellis representations, result in a significant improvement upon the existing bounds. The ordered ELP bounds the trellis complexity of nonlinear codes under a given permutation, and the trellis complexity of the code under any coordinate ordering is bounded by the unordered ELP. The prescribed bounds reduce to well-known results when applied to linear codes. These results can be reinterpreted from our new viewpoint. The new tools can also be used to derive new bounds on trellis complexity of linear codes (e.g., Corollary 4 and Theorem 10 herein) though we believe that their main contribution is to the investigation of nonlinear codes.

The bounds on the state and the branch complexity profiles are preceded by some general new results regarding the problem of minimal covering of a bipartite graph by complete subgraphs. In [15], Kschischang and Sorokine have observed that the problem of minimizing the vertex count of a trellis representation of a code at a given index is equivalent to the aforementioned graphical problem. This interpretation of the problem involves the representation of the past/future relation of a code via a Cartesian array. In this paper we elaborate on this observation and show that minimizing the branch count between any two indices can also be treated as the same problem of covering by product-form subsets. This analysis poses a Cartesian product representation of a different relation induced by the code. This observation enables the use of unified basic proofs to establish bounds on both state and branch complexity profiles.

In addition to the practical use of the new measures, our study has also a conceptual merit. It shows the close relation between trellis representation and mutual information. The trellis complexity of a linear code is completely defined in terms of the dimensions of projections of the code and its subcodes. Recently, McEliece [20] presented an information-theoretic interpretation of the vertex count of the BCJR trellis. In this paper we elaborate this relation and we bound the trellis complexity of nonlinear codes by some information-theoretic measures. The derived relations by contrast to the corresponding relation for linear codes require relatively complex proofs based on definitions of appropriate probabilistic models. The complexity of a trellis representation for nonlinear codes cannot be expressed in closed-form terms. These trellises are subject to some anomalies peculiar to nonlinear codes. In this paper we prove, however, that the complexity of any possible trellis representation cannot be smaller than some well-defined information-theoretic measures, regardless of the trellis structure.

The paper is organized as follows. Preliminaries and basic notations are presented in Section II. The next two sections develop two distinct frameworks which are combined in the subsequent sections to derive bounds on the state and branch complexity profiles of nonlinear block codes. Section III introduces the different entropy/length profiles and develops their relationships and their basic properties. It is shown that the prescribed profiles are natural extension of the DLP to nonlinear codes. The ELP reduces to the DLP when applied to linear codes. Section IV is devoted to the problem of covering a bipartite graph by complete subgraphs. We present a lower and an upper bound on the minimum number of these constituent complete subgraphs. These bounds are phrased as the mutual information between the two disjoint subsets of vertices of the bipartite graph. In Section V, we use the general results of Section IV in conjunction with the ELP notion that was developed in Section III to present both upper and lower bounds on the state complexity profile of (nonlinear) codes described by any trellis diagram. The state complexity of a code under a given permutation is lower-bounded by the ordered ELP, and the lower bound for any coordinate ordering is a simple function of the unordered profiles. The basic theorem for a given coordinate permutation determines that the state complexity of any trellis representation of the code at any given level cannot be smaller than the mutual information between the *past* and the *future* portions of the code at this same level, under a uniform distribution of the codewords. These new bounds exhibit an appreciable improvement on the CLP bounds for nonlinear codes [16]. We also present an additional bound on the state complexity, i.e., on the maximum value of the state complexity profile. This bound is obtained by partitioning of the index axis for the code. Afterwards we develop an upper bound on the state complexity profile. We present a graph-theoretic bound that actually furnishes a constructive procedure to derive a (one-to-one) trellis diagram that meets the bound at any given index. This complexity profile may not be achieved at all the levels of the diagram simultaneously. Thereafter, we prove that this bound can be expressed as a maximum value of the mutual information between the past and the future portions of the code with respect to the probability distribution of the codewords.

Notably, as an example, we apply the bounds to the Nordstrom–Robinson code. Our upper bound disproves the conjecture of Lafourcade and Vardy [16] that the known state profile of the construction by Forney [4] and due to Vardy [24] is optimal componentwise. We devise a new construction of the code. The trellis complexity of the new construction is smaller than that of the known construction of the code at some indices. The new construction also infers that the new lower bounds are rather tight.

Section VI extends the results of Section V to the branch complexity profile, when this profile refers to the logarithm of the total number of paths through the trellis between two levels i and j, where j need not necessarily be equal to $i+1$. The lower and upper bounds on the branch complexity profile are also expressed in terms of ELP and mutual information. The lower bounds on the trellis complexity can also be considered as an additional proof to the optimality of the minimal biproper trellis for group codes ([5], [6], [14]). Moreover, these bounds indicate that this trellis also minimizes the total number of trellis branches between any two indices. Finally, concluding remarks are given in Section VII.

II. Preliminaries

In this section, we briefly overview relevant terms of trellis theory and introduce the notations that we use throughout the paper. A length-n *sequence space* W is defined by an *index set* I, $I \triangleq \{1, 2, \cdots, n\}$ and by a set of *symbol alphabets* $\{A_i: i \in I\}$. W is the Cartesian product $\prod_{i \in I} A_i$. That is, W is the set of all sequences $\boldsymbol{a} = (a_i, i \in I)$ with $a_i \in A_i$. A *block code* is a subset of W. When the constituent symbol alphabets are groups, W is called a *group sequence space*. We define the componentwise group operation in W

$$\boldsymbol{a} * \boldsymbol{b} \triangleq (a_i * b_i, i \in I)$$

where $a_i * b_i$ denotes the product of $a_i, b_i \in A_i$ under the binary group operation in A_i. Evidently, W is a direct product group under the above operation. A *group code* is a subgroup of a group sequence space W.

Let $C(n, M)$ be a block code that consists of M codewords of length n. Most of our results apply to the above general definition of block codes. For expedience we confine some results to block codes whose symbol alphabets at all the positions have the same size (e.g., a code over a field GF(q)). For these codes, when the minimum distance of the code d is of interest, the code will be denoted by the triple (n, M, d).

A trellis $T = (V, A, E)$ for a block code is an edge-labeled directed graph where V, A, and E are defined in the next few lines. V is the set of *vertices* (*states*) in the graph. Every vertex is assigned a level in the range $\{0, 1, \cdots, n\}$. The set of vertices at level i is denoted by $V_i(C)$, so V is the union of disjoint subsets, $V = \cup_{i=0}^{n} V_i(C)$. A is a finite alphabet set, and E is a set of ordered triples (v_j, α, v_k) where $v_j \in V_{i-1}(C)$, $v_k \in V_i(C)$, and $\alpha \in A_i$. These triples, called *edges*, connect vertices at level $i-1$ to those at level i. Any path from $V_0(C)$ to $V_n(C)$ defines an n-tuple $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ over $\prod_{i \in I} A_i$. In the sequel we will use the term *branches* to describe partial paths through the trellis that connect states at different levels. These states need not necessarily be at adjacent levels. The state complexity of the trellis diagram at level $i, s_i(C)$, is the logarithm of the vertex count at this level, i.e., $s_i(C) \triangleq \log |V_i(C)|$. The sequence $\boldsymbol{s}(C) = \{s_i(C), 0 \le i \le n\}$ is the *state complexity profile* of C.

Let $P_J(\boldsymbol{c})$ denote the projection of a codeword $\boldsymbol{c} \in C$ onto $J \subseteq I$. That is, if $J = (i_1, i_2, \cdots, i_{|J|})$ and $(c_1, c_2, \cdots, c_n) = \boldsymbol{c} \in C$ then $P_J(\boldsymbol{c}) = (c_{i1}, c_{i2}, \cdots, c_{i|J|})$. The complementary set of J in I will be denoted by $I - J$. The set of the projection of all codewords onto J is denoted by $P_J(C)$. Let $B_{i,j}(C)$ denote the set of branches (paths) between states at indices i and $j > i$. Each branch is described by the triple $(v_i, P_{[i+1,j]}(\boldsymbol{c}), v_j)$ where $v_i \in V_i(C), v_j \in V_j(C)$, and $\boldsymbol{c} \in C$. Similarly, we define $b_{i,j}(C) \triangleq \log |B_{i,j}(C)|$.

For a given coordinate ordering we denote

$$s_{\max}(C) \triangleq \max_i \{s_i(C)\}$$

$$b_{\max}^k(C) \triangleq \max_{i,j} \{b_{i,j}(C): j - i = k\}.$$

The minimum $s_{\max}(C)$ and $b_{\max}^k(C)$ over all coordinate orderings is called the *state complexity* $s(C)$ and the *length-k branch complexity* $b^k(C)$ of C, respectively.

In this paper we prove some theorems regarding *any trellis diagram* of the code. This pertains to any edge-labeled graph of the foregoing structure. The diagram should have the following properties.

1) The graph has a single initial state and a single final state: $s_0 = s_n = 0$.
2) There is at least one path from the initial vertex to every vertex, and at least one path from each vertex to the final state.
3) The set of n-tuples corresponding to all the paths is identical to the set of the codewords of C.

Clearly, these requirements do not preclude improper representations of the code and not even non-one-to-one trellises.

In this paper we use basic measures of information theory. Detailed background on these measures: entropy and mutual information, can be found in any textbook on information theory (e.g., [7]). A probability space, i.e., a sample space and its probability measure will be called an *ensemble*. For an ensemble X with the sample space $\Omega = \{x_1, x_2, \cdots, x_K\}$, the probability that the outcome x will be a particular element $x_k \in \Omega$ will be denoted by $p_X(x_k)$. The *entropy* of the ensemble X is defined by

$$H(X) \triangleq -\sum_{x \in \Omega} p_X(x) \cdot \log[p_X(x)].$$

The base of the logarithm determines the units used to measure information. Our bounds do not require a particular base, albeit the choice of q-ary digits, where q is the alphabet size over which the code is defined, seems to be the natural numerical scale to present the results that apply to block codes over a fixed alphabet size. All the results here onwards use this convention. That is, the logarithms in the results for block codes over GF(q) are assumed to be to base-q, and the logarithms in other general results are taken to the same arbitrary base. Let XY be a joint ensemble with *joint entropy* $H(XY)$, the *conditional entropy* of X given Y is denoted by $H(X|Y)$. The *mutual information* between the pair of ensembles X and Y is defined by

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Further on we use this relation in conjunction with the following basic relation:

$$H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (2.1)$$

to represent the results in several equivalent forms and to show the relations between various measures. We also use the notation $I_u(X;Y)$ to designate the mutual information between X and Y under a uniform distribution of the elements of the sample space of the joint ensemble XY.

III. Entropy/Length Profiles

Let $i^- \triangleq [1, 2, \cdots, i]$ denote the special subset of I which consists of the first i consecutive indices and let $i^+ \triangleq [i+1, i+2, \cdots, n]$ be the complementary subset of i^- in I, where $0 \le i \le n$, with the convention that 0^- and n^+ are empty subsets. The set of the truncated codewords to the first i coordinates (the *past*) is denoted by $P_{i^-}(C)$, and $P_{i^+}(C)$ is the truncation to the remaining $n-i$ coordinates (the *future*).

A. ELP and Ordered ELP

Regard the (n, M) code C as an ensemble whose sample space is the set of codewords and assign each codeword a uniform probability of $1/M$. Let X_{i^-} denote a random i-tuple variable that takes on the values of the set $P_{i^-}(C)$ with a probability function that is imposed by the uniform probability of each entire codeword. Similarly, X_{i^+} will denote a random $(n-i)$-tuple variable whose sample space is the set $P_{i^+}(C)$, and X_J will denote a random $|J|$-tuple variable that takes on the values of the set $P_J(C)$ with probabilities that are induced by the uniform distribution of the codewords. In particular, $X_{[j]}$ will denote a random variable that gets the values of the projection of C onto a single index j under the same probabilistic model. Throughout this paper, when we make C into a probability space, a uniform distribution of the codewords is always assumed, unless otherwise asserted.

Definition 1: The *ordered entropy/length profile* (*ordered ELP*) of C (n, M) over GF(q) will be defined as the sequence $\boldsymbol{g}'(C) = \{\boldsymbol{g}'_i(C), 0 \le i \le n\}$ where $g'_i(C) \triangleq H(X_{i^-})$. Using the fact that $H(X_{i^-}X_{i^+}) = H(C) = \log_q M$, in conjunction with the basic relation (2.1) we have

$$g'_i(C) \triangleq H(X_{i^-}) = \log_q M - H(X_{i^+}|X_{i^-}). \quad (3.1)$$

Definition 2: The *ordered conditional entropy/length profile* (*ordered conditional ELP*) of an (n, M) code C over GF(q) will be defined as the sequence $g(C) = \{g_i(C), 0 \le i \le n\}$, where

$$g_i(C) \triangleq H(X_{i^-}|X_{i^+}) = \log_q M - H(X_{i^+}). \quad (3.2)$$

Using the fact that $H(X|Y) \le H(X)$, it is apparent that $\boldsymbol{g}(C) \le \boldsymbol{g}'(C)$, where the relational operators will denote a componentwise appropriate relation when used with vector-valued (sequence) quantities.

We will clarify these definitions by a simple example that will be used later in other contexts to demonstrate the relation between various measures.

Example 3.1: Consider the binary nonlinear code

$$C_1 = \{0000, 0110, 1100, 1111\}.$$

When each codeword is assigned a probability of 0.25, the possible values of the first coordinate are 0 or 1 with probability 0.5 of each event. The first two indices can be one of the following combinations: $P_{2^-}(C_1) = \{00, 01, 11\}$ with the probabilities $\{0.25, 0.25, 0.5\}$, respectively, since each of the first two prefixes appears in one codeword, whereas two codewords start with 11. Thus

$$g'_2(C_1) = -2 \cdot 0.25 \cdot \log_2 0.25 - 0.5 \cdot \log_2 0.5 = 1.5.$$

The evaluation of the complete ordered ELP of C_1 yields

i:	0	1	2	3	4
$g'_i(C_1)$ [bits]:	0	1	1.5	2	2

The first coordinate of a codeword in C_1 is completely determined by the other three, i.e., any possible combination of the last three coordinates appears in only one codeword. When the last two coordinates of the code are known, there is an uncertainty about the first two coordinates only when the last coordinates are 00 since two codewords end with this value. The complete ordered conditional ELP of C_1 is

i:	0	1	2	3	4
$g_i(C_1)$ [bits]:	0	0	0.5	$0.75 \cdot \log 3 = 1.19$	2

Definition 3: The *entropy/length profile* (*ELP*) of a code $C\ (n, M)$ over GF(q) is the sequence

$$\boldsymbol{h}'(C) = \{h'_i(C), 0 \le i \le n\}$$

where

$$h'_i(C) \triangleq \min_J\{H(X_J)\colon |J| = i\}, \qquad 0 \le i \le n. \quad (3.3)$$

Evidently, $h'_i(C)$ is the minimal entropy of a punctured code $P_J(C)$ whose length $|J|$ equals i. Obviously, $\boldsymbol{g}'(C) \ge \boldsymbol{h}'(C)$ since $|i^-| = i$.

Definition 4: The *conditional entropy/length profile* (*conditional ELP*) of a code $C\ (n, M)$ over GF(q) is the sequence

$$\boldsymbol{h}(C) = \{h_i(C), 0 \le i \le n\}$$

where

$$h_i(C) \triangleq \max_J\{H(X_J|X_{I-J})\colon |J| = i\}, \qquad 0 \le i \le n \quad (3.4)$$

is the maximal conditional entropy of any i coordinates of the code C. Clearly, $\boldsymbol{g}(C) \le \boldsymbol{h}(C)$.

Example 3.1 (continued): The code C_1 of the previous example has the following ELP and conditional ELP (see the bottom of this page).

The minimum value of $H(X_{[j]})$, i.e., $h'_1(C)$, is achieved when evaluated for the second or fourth coordinate.

Example 3.2: To illustrate the difference between the ELP and the *cardinality/length profile* (CLP) [16] we compare both profiles when applied to the code of Example 3.1 and to $C_2 = \{0000, 0010, 1100, 1111\}$. For $i = 2$, for instance, the maximum number of codewords that have two common bits is two in both codes, e.g., the first two codewords of C_1 have the same components in their first and fourth positions. The other two codewords do not have common bits at these positions. In C_2, however, the first two codewords have common bits in the first two indices (00), the other two codewords share the same bits (11) in these indices. The CLP at $i = 2$ is $\log_2 M(2; C_1) = \log_2 M(2; C_2) = 1$, where $M(l; C)$ is the maximum cardinality of any subcode of C whose codewords have the same components at $n - l$ indices. The conditional ELP measure at $i = 2$ distinguishes between the two cases: $h_2(C_1) = 0.5$ and $h_2(C_2) = 1$. The conditional ELP therefore implies the obvious observation that a different amount of uncertainty in the two codes is resolved by the knowledge of some of the components of a codeword.

B. Basic Properties of the Entropy/Length Profiles

Lemma 1: For linear codes, the ELP reduces to the inverse DLP, the conditional ELP reduces to the DLP, the ordered ELP reduces to the inverse ordered DLP, and the ordered conditional ELP reduces to the ordered DLP.

Proof: It follows immediately from the definitions of the dimension/length profiles in [5] and those of the entropy/length profiles herein. □

Likewise, when the ordered ELP profiles are applied to a group code, they reduce to the log-cardinality of the corresponding subcodes and projections of the code as defined in [6].

From (3.1) and (3.2) we have

$$H(X_{i^+}) + g_i(C) = g'_i(C) + H(X_{i^+}|X_{i^-}) = \log_q M. \quad (3.5)$$

The following lemma establishes a similar relation between the unordered profiles.

Lemma 2: Let C be an (n, M) code over GF(q). For any i, $0 \le i \le n$

$$h_i(C) + h'_{n-i}(C) = \log_q M. \quad (3.6)$$

Proof:

$$\begin{aligned} h_i(C) &\triangleq \max_J\{H(X_J|X_{I-J})\colon |J| = i\} \\ &= \max_J\{\log_q M - H(X_{I-J})\colon |J| = i\} \\ &= \log_q M - \min_J\{H(X_{I-J})\colon |J| = i\} \\ &= \log_q M - h'_{n-i}(C). \end{aligned}$$ □

When C is a linear code with dimension k, then this lemma reflects the relation between dimensions of the projection $P_{I-J}(C)$ and the subcode C_J, where C_J defines the set of all codewords whose components are all zero outside of J. In this case, $k - \dim(C_J)$ information units out of total k can be reconstructed from the knowledge of the components of the codeword in the indices within $I - J$. The rest of the information is included in C_J. The prescribed measure $H(X_J|X_{I-J})$ extends the notion of subcodes to nonlinear codes. This measure also quantifies the amount of information one lacks to retrieve the source message from the components of the codeword at the indices $I - J$.

In particular, for both linear and nonlinear codes, when $H(X_J|X_{I-J}) = 0$, the components of the code at the coordinates J are uniquely determined by the components of the codeword in the remaining positions. These $I - J$ coordinates are an *information set* for C.

Lemma 3 (Monotonicity): If C is an (n, M) code over GF(q), then for any i, $0 \le i \le n - 1$

$$\begin{aligned} &\text{a)}\quad 0 \le g'_{i+1}(C) - g'_i(C) \le 1 \\ &\text{b)}\quad 0 \le g_{i+1}(C) - g_i(C) \le 1 \\ &\text{c)}\quad 0 \le h'_{i+1}(C) - h'_i(C) \le 1 \\ &\text{d)}\quad 0 \le h_{i+1}(C) - h_i(C) \le 1. \end{aligned} \quad (3.7)$$

i:	0	1	2	3	4
$h'_i(C_1)$ [bits]:	0	$2 - 0.75 \cdot \log 3 = 0.81$	1.5	2	2
$h_i(C_1)$ [bits]:	0	0	0.5	$0.75 \cdot \log 3 = 1.19$	2

Proof:

a) From (2.1), we have

$$g'_{i+1}(C) = g'_i(C) + H(X_{[i+1]}|X_{i^-}).$$

Clearly, $0 \le H(X_{[i+1]}|X_{i^-}) \le 1$ and the lemma follows.

b)

$$\begin{aligned} g_{i+1}(C) &= H(X_{(i+1)^-}|X_{(i+1)^+}) \ge H(X_{i^-}|X_{(i+1)^+}) \\ &\ge H(X_{i^-}|X_{i^+}) = g_i(C). \end{aligned}$$

Likewise

$$\begin{aligned} g_{i+1}(C) &= \log_q M - H(X_{(i+1)^+}) \\ &\le \log_q M - [H(X_{i^+}) - 1] = g_i(C) + 1. \end{aligned}$$

These two inequalities establish b).

c) The left-hand side inequality is fairly obvious. It remains to establish the second inequality. Suppose $J = \{j_1, j_2, \cdots, j_i\}$ is the set of the i indices that minimizes $\{H(X_J): |J| = i\}$. We create J', a subset of I that incorporates J and an additional arbitrary index l that is not included in J. Clearly,

$$h'_{i+1}(C) \le H(X_{J'}) \le H(X_J) + 1 = h'_i(C) + 1.$$

d) This inequality follows immediately from c) and Lemma 2. □

The above property of the different profiles also holds for any probability distribution of the codewords.

Since

$$g_0(C) = g'_0(C) = h_0(C) = h'_0(C) = 0$$

and

$$g_n(C) = g'_n(C) = h_n(C) = h'_n(C) = \log_q M$$

we conclude that

$$\begin{aligned} &0 \le g'_i(C) \le \min\{i, \log_q M\} \\ &0 \le g_i(C) \le \min\{i, \log_q M\} \\ &0 \le h'_i(C) \le \min\{i, \log_q M\} \\ &0 \le h_i(C) \le \min\{i, \log_q M\}. \end{aligned} \tag{3.8}$$

The dimension/length profiles of linear codes rise from 0 to the code's dimension k in k distinct unit steps. The increase of the entropy/length profiles of nonlinear codes need not be in unit steps, and hence the total number of steps may exceed $\log_q M$.

The following corollary exemplifies how the above profiles can be employed to derive results concerning nonlinear codes. We use Lemma 3 to rederive the generalization of the Singleton bound to nonlinear codes. It therefore indicates the potential benefit of the use of the above ideas in the area of nonlinear codes.

Corollary 1 (Generalization of the Singleton Bound to Nonlinear Codes [18, p. 544]): The minimum distance, d, of an (n, M) code C over GF(q) satisfies $d \le n - \log_q M + 1$.

Proof: We recall that $h_l(C) > 0$ iff $d \le l$. From Lemma 3 we have $h_{i+1}(C) - h_i(C) \le 1$ and since $h_n(C) = \log_q M$ we conclude that $h_{n-\kappa+1}(C) > 0$, where $\kappa = \lceil \log_q M \rceil$ is the smallest integer whose value is not smaller than $\log_q M$. Thus the minimum distance of the code, d, satisfies

$$d \le n - \lceil \log_q M \rceil + 1 \le n - \log_q M + 1. \qquad \square$$

The next section presents some new bounds on the minimal covering of bipartite graphs by complete subgraphs. The subsequent sections use these bounds along with the ELP concept to develop bounds on the trellis complexity of (nonlinear) block codes.

IV. Bounds on the Minimal Covering of Bipartite Graphs by Complete Subgraphs

A *bipartite* graph $G = G(A, B, C)$ is defined as a graph whose vertex set can be partitioned into two disjoint subsets A and B such that no vertex in a subset is *adjacent* to vertices in the same subset. We denote by C the *edge* set of G. Each edge $(v, v') \in C$ of a bipartite graph connects a vertex $v \in A$ with a vertex $v' \in B$. An edge (a, b) is *incident from* a and is *incident into* b. For convenience we refer to the vertices in the subset A as *initial* vertices and B is referred to as the set of *terminal* vertices. We also denote

$$C = \{(a, b): a \in A, b \in B, (a, b) \in C\} = \{c_i,\ i = 1, 2, \cdots, M\}$$

where M denotes the cardinality of the edge set. A bipartite graph is said to be *complete* if every vertex in one of the two subsets is connected to *all* the vertices of the second subset. A *subgraph* of a bipartite graph is bipartite. *A matching* in G is a set of pairwise-disjoint edges. A *maximum matching* is a matching of a maximum size. A *vertex cover* of G is a set V $(V \subseteq A \cup B)$ of vertices such that V contains at least one endpoint of every edge of G. Further background on the aforementioned fundamental concepts may be found in textbooks on graph theory (e.g., [27]).

In this section we address the problem of covering bipartite graphs by complete subgraphs. We shall be concerned with the following question: are there $l \le L$ subsets $V_1, V_2, \cdots, V_l$ of the vertex set $A \cup B$ of G such that each V_i induces a complete bipartite subgraph of G and such that each edge (a, b) of G is contained in some V_i?

In particular, the *minimal covering* of G is of major interest. The covering is said to be minimal in the sense that any other covering comprises at least the same number of constituent subsets. This problem is equivalent to that of finding a minimal covering of a two-dimensional array by product-form subsets (complete rectangles) and similar to the minimization of the total number of terms in the sum-of-products Boolean expression using a Karnaugh map. This problem is known to be NP-complete [8, [GT18]]. Thus bounds on the minimal covering are of a significant importance even in the framework of graph theory.

Motivated primarily by applications in soft-decision decoding via a trellis diagram, we henceforth present some general information-theoretic bounds on the minimal covering, taking faith in their later utility. Notably, to the best of our knowledge, information-theoretic measures have not yet been used by graph-theoreticians to constitute bounds for this problem. The problem of minimizing the vertex count of trellis representation of codes at a given index may also be formulated as a minimal covering problem [15]. Hence we use the bounds derived in this section to lower-bound and upper-bound the state complexity profile of (nonlinear) codes. In Section VI we pursue this approach further and show that the problem of minimizing the branch count between two levels of a trellis can also be presented as another application of the problem of covering a bipartite graph by complete subgraphs. Accordingly, we contrive an appropriate model that facilitate the derivation of the corresponding bounds on the branch complexity profile.

In what follows we study the problem of covering the edges of G by the minimum number, R, of complete subgraphs. We present several bounds on this number R. In the following theorems we consider a joint XY ensemble whose sample space is the edge set C. The probability distribution of XY is

$$p(x=a, y=b) = p_i, \quad \text{if } (a,b) = c_i,\ p_1 + p_2 + \cdots + p_M = 1. \tag{4.1}$$

Our bounds do not impose the requirement that every edge should be contained in exactly one subgraph. Consequently, our bounds also apply to unobservable coverings, i.e., coverings for which the precise edge in a constituent complete subgraph corresponding to each edge of the original graph cannot be recovered unambiguously. All the logarithms in this section (including those used in the information-theoretic measures) are taken to the same arbitrary base.

Theorem 1 (Lower Bound): Let $V = A \cup B$ and C be the vertex set and the edge set of a bipartite graph G, respectively. Let T be a covering of G by S complete subgraphs, then S is bounded by

$$\log S \geq I_u(X;Y) \tag{4.2}$$

where $I_u(X;Y)$ stands for the mutual information between X and Y under a uniform distribution of the elements of the joint sample space, i.e., $p_i = 1/M, i = 1, 2, \cdots, M$.

Proof: We suppose that T is a covering of G by complete bipartite subgraphs $\{G_i, i = 1, 2, \cdots, S\}$. We define a new joint XYZ ensemble where XY is the ensemble defined by (4.1) and Z is an ensemble whose sample space is $\{G_i, i = 1, 2, \cdots, S\}$. The joint sample space of XYZ, Ψ, is the set of triples $\{(a, b, G_k): (a, b) \in C, 1 \leq k \leq S\}$. Nonzero probabilities are assigned only to triples (a, b, G_k) wherein G_k is a subgraph that comprises the edge $c = (a, b)$ of G. This sample space can equivalently be depicted as a set of pairs $\{(c, G_k), c \in C\}$. Consequently, under the above restriction of the probability function, the "outcome" takes on the values only of a reduced subset of Ψ. This subset will be denoted by $\Omega, \Omega \subseteq \Psi$. There is a one-to-one correspondence between Ω and the union of the edge sets of all the constituent subgraphs that cover G. Evidently, The cardinality of Ω is at least M, and if an edge (connected vertex pair) of G appears in several subgraphs, each such appearance is represented by a different element in Ω.

We count the total number of edges in T that correspond to each edge of G and assign probabilities to the elements of Ω according to the following rule: all the elements of the sample space that represent the same edge of G are assigned the same probability of $1/(rM)$, where r is the total number of elements in the sample space Ω that represent the specific edge of G. This rule preserves the uniform probability of the edges of G. Thus under this model, the ensemble XY still gets one of M possible values with a probability of $1/M$. The elements of Ω, however, are no longer uniformly distributed when the diagram is unobservable. Consequently, $H(Z|XY)$ is not necessarily zero. The marginal probability of Z, i.e., $\{p(G_k), 1 \leq k \leq S\}$ can be derived from the joint probability function. Thus

$$\begin{aligned} \log M &= H(XY) \leq H(XYZ) \\ &= H(Z) + H(X|Z) + H(Y|XZ). \end{aligned} \tag{4.3}$$

Clearly,

$$H(Y|XZ) \leq H(Y|X). \tag{4.4}$$

Let $\{y_j, 1 \leq j \leq |B|\}$ denote all possible terminal vertices, and let L_k, $1 \leq k \leq S$ denote the total number of initial vertices included in G_k. Using this notation

$$\begin{aligned} H(X|Z) &= \sum_{k=1}^{S} p(G_k) \cdot H(X|G_k) \\ &= \sum_{k=1}^{S} p(G_k) \cdot H(X|G_k) \cdot \sum_{j=1}^{|B|} p(y_j|G_k) \\ &= \sum_{j=1}^{|B|} \sum_{k=1}^{S} p(y_j|G_k) \cdot p(G_k) \cdot H(X|G_k) \\ &\leq \sum_{j=1}^{|B|} p(y_j) \cdot \max_k \{\log L_k \colon p(y_j|G_k) > 0\} \\ &\leq \sum_{j=1}^{|B|} p(y_j) \cdot H(X|y_j) = H(X|Y). \end{aligned} \tag{4.5}$$

The last inequality is justified by the following two observations. First, let G_k be an arbitrary subgraph in T that incorporates a specific terminal vertex y_j. Clearly, the cardinality of the set of all initial vertices in G_k is not larger than the number of all the initial vertices connected to y_j in G. Second, $H(X|y_j)$ equals the log-cardinality of the set of all initial vertices of G incident into y_j, since our probabilistic model retains the uniform distribution of the edges of the graph G.

We substitute (4.4) and (4.5) in (4.3) to obtain,

$$H(Z) \leq \log_q M - H(Y|X) - H(X|Y) = I_u(X;Y). \tag{4.6}$$

Eventually, we have the inequality $H(Z) \leq |Z| = \log S$. The use of this inequality along with (4.6) provides $S \geq I_u(X;Y)$. □

The next theorems introduce an upper bound on the minimal covering problem. That is, we assert that there exists a covering of a bipartite graph by complete subgraphs such that the total number of constituent subgraphs is not larger than a defined measure. Furthermore, we construct a diagram that achieves the bound. The first derivation is graph-theoretic. The second approach resorts to information theory. In this framework we contrive a new probabilistic model. The sample space of this model still incorporates the edge set as in Theorem 1. However, the derivation of the upper bound does not involve a uniform distribution of the edges. We define the bound as the maximum value of the mutual information between the above-mentioned ensembles X and Y, where the maximization is done over all possible probability functions over the elements of the sample space. In the following theorems we denote by $N(G)$ the maximum matching, i.e., the maximum cardinality of a subset D, $D \subseteq C$, of detached edges, such that any two distinct edges $(a_i, b_i), (a_j, b_j) \in D$ satisfy $a_i \neq a_j$ and $b_i \neq b_j$. Theorem 2 is due to a well-known result by König and Egervary (e.g., [9] and [27]).

Theorem 2: There exists a one-to-one covering of a bipartite graph G by $N(G)$ complete subgraphs $G_1, G_2, \cdots, G_{N(G)}$, i.e., $C_i \cap C_j = \emptyset, \forall i \neq j$, where C_k denotes the edge set of G_k. Moreover, every subgraph in the prescribed covering comprises either a single initial vertex (*expanding subgraph*) or a single terminal vertex (*merging subgraph*).

Proof: Consider a vertex-incident $|A| \times |B|$ matrix Q of a bipartite graph G whose vertex set $V = A \cup B$. Clearly, Q is a $(0,1)$ matrix, and a 1 in the (i,j) entry of the matrix represents an edge (a_i, b_j) of G. An expanding (complete) subgraph corresponds to elements of Q which are in the same row. Elements of Q in the same column correspond to a merging subgraph. By a theorem of König and Egervary, the minimum number of lines containing all the 1's in Q is equal to the maximal number of 1's in Q, no two of which are in the same line of the matrix (the *term rank* of the matrix), namely, $N(G)$.

Let this set of lines consist of rows R_Q and columns C_Q. In order to obtain a one-to-one covering of the edge set of G we first design expanding subgraphs using rows R_Q. Then we design the merging subgraphs using columns C_Q in which elements of rows R_Q are replaced by zeros. □

The above theorem may be rephrased in a different equivalent way: the maximum size of a matching in G equals the minimum size of the vertex cover of G. The next theorem gives an interesting information-theoretic interpretation to the cardinality of the maximum matching $N(G)$.

Theorem 3: Make the edge set C of a bipartite graph G into a joint XY ensemble that takes one of M possible values. *Assign each edge a probability* p_m, $1 \leq m \leq M$, with $p_1 + p_2 + \cdots + p_M = 1$. Then

$$\max_{p_1, p_2, \cdots, p_M} \left\{ I(X;Y) \colon \sum_{m=1}^{M} p_m = 1 \right\} = \log N(G). \tag{4.7}$$

Proof: We use the construction of Theorem 2 to facilitate the probability allocation. We assign the edge set of the subgraph G_k of the above construction probabilities p_l^k where $1 \leq l \leq M_k$ and M_k is the total number of edges in G_k. Clearly,

$$\sum_{k=1}^{N(G)} M_k = M$$

and

$$\sum_{k=1}^{N(G)} \sum_{l=1}^{M_k} p_l^k = 1.$$

We define a joint ensemble XYZ whose sample space Ω is the set $\{(c, G_k), c \in C, c \in G_k\}$. Clearly, the cardinality of this sample space is M.

$$\begin{aligned} I(X;Y) &= H(XY) - H(X|Y) - H(Y|X) \\ &\leq H(XY) - H(X|YZ) - H(Y|XZ). \end{aligned}$$

Using the fact that all the subgraphs are either expanding or merging, we have the key observation

$$H(X|YZ) + H(Y|XZ) = -\sum_{k=1}^{N(G)} \sum_{l=1}^{M_k} p_l^k \log \frac{p_l^k}{\sum_{l=1}^{M_k} p_l^k}$$

for any choice of probabilities for the elements of the Ω. Thus

$$\begin{aligned} I(X;Y) &\leq -\sum_{k=1}^{N(G)} \sum_{l=1}^{M_k} p_l^k \log p_l^k + \sum_{k=1}^{N(G)} \sum_{l=1}^{M_k} p_l^k \log \frac{p_l^k}{\sum_{l=1}^{M_k} p_l^k} \\ &= -\sum_{k=1}^{N(G)} \left\{ \sum_{l=1}^{M_k} p_l^k \right\} \log \left\{ \sum_{l=1}^{M_k} p_l^k \right\} \leq \log N(G). \end{aligned}$$

Consequently,

$$\max_{p_1, p_2, \cdots, p_M} \left\{ I(X;Y) \colon \sum_{m=1}^{M} p_m = 1 \right\} \leq \log N(G). \tag{4.8}$$

Conversely, the maximum value of $I(X;Y)$ is not smaller than $\log N(G)$. This value is attained when the edges of the maximum matching D are assigned a uniform probability of $1/N(G)$ and all the other edges are assigned zero probability. That is,

$$\max_{p_1, p_2, \cdots, p_M} \left\{ I(X;Y) \colon \sum_{m=1}^{M} p_m = 1 \right\} \geq \log N(G). \tag{4.9}$$

The right-hand side of inequality (4.9) is exactly the mutual information between the initial vertices and the terminal vertices of the edges of D under the preceding assignment of probabilities. Combining (4.8) and (4.9), we deduce (4.7). □

The above maximization problem does not have, in general, a unique solution, and therefore the foregoing choice of probability distribution is just one of the possible solutions. Eventually, we conclude that the minimal number R of complete subgraphs that cover a bipartite graph satisfies the following inequality:

$$I_u(X;Y) \leq \log R \leq \max_{p_1, p_2, \cdots, p_M} \left\{ I(X;Y) \colon \sum_{m=1}^{M} p_m = 1 \right\}.$$

V. Bounds on the State Complexity Profile of (Nonlinear) Block Codes

In the following sections we digress to investigate trellis complexity of (nonlinear) block codes. It is noteworthy to remark that though the edge count seems to be a more appropriate figure of merit to assess the complexity of the Viterbi algorithm [20], the state complexity and the branch complexity profiles still reflect the complexity of the soft-decision decoding. The total complexity can be determined by the above profiles. As we have already mentioned, nonlinear codes need not admit a minimal trellis, and some efficient representations of such codes may not be proper or one-to-one. The requirement to proper trellises is more restrictive than the confinement to one-to-one trellises since a proper trellis is necessarily a one-to-one representation of the code. The CLP-based bounds due to Lafourcade and Vardy [16] are stated for one-to-one trellises. Their improved bounds for specific codes, bounds that use combinatorial properties of the specific code, are confined to biproper trellises. We consider these requirements somewhat restrictive due to the fact that improper trellises and unobservable diagrams may have a reduced trellis complexity (e.g., [15], [21]). All the bounds established herein apply to any trellis representation of a code. The diagram need not be proper or one-to-one.

Any (n, M) block code C poses a relation between the set of the past and the future projections of its codewords at each level. Since $C \subseteq P_{i^-}(C) \times P_{i^+}(C), \forall i$, C can be considered as a code of length two over the alphabet $P_{i^-}(C) \times P_{i^+}(C)$ [15]. Therefore, the minimization of the vertex count at a given level i is equivalent to finding a minimal product-form covering of C at index i. Any trellis representation for C imposes a decomposition of $P_{i^-}(C)$ and $P_{i^+}(C)$ as follows. Let P_{ij} denote the label set of the trellis branches incident from the origin and terminated at the jth state at level i. Similarly, we denote by F_{ij} the labels of branches concocting the jth state at level i to the terminal state at level n. Clearly,

$$P_{i^-}(C) = \bigcup_{j=1}^{S} P_{ij} \qquad P_{i^+}(C) = \bigcup_{j=1}^{S} F_{ij} \qquad C = \bigcup_{j=1}^{S} P_{ij} \times F_{ij}$$

where S denotes the vertex count at level i. In general, the subsets $\{P_{ij}, j = 1, 2, \cdots, S\}$ may not be disjoint. The same is with the future subsets. If there exists a covering of the relation C at index i by disjoint subsets, i.e.,

$$P_{ip} \cap P_{iq} = \emptyset, F_{ip} \cap F_{iq} = \emptyset, \forall 1 \le i \le n-1, 1 \le \{p, q\} \le S$$

the code is said to be rectangular [14] (separable, [23]) under the given symbol order.

The following theorems are deduced from the general bounds presented in the preceding section. In these theorems, when we evaluate entropy quantities, we consider C as an ensemble whose sample space is the set of the codewords, and we assign each codeword a uniform probability of $1/M$ as explained in Section III-A. We also use the nomenclature of this section.

Theorem 4 (State Complexity Profile Under a Given Coordinate Permutation): The state complexity profile (of any trellis representation) of an (n, M) block code C is bounded by

$$s_i(C) \ge I_u(X_{i^-}; X_{i^+}), \qquad 0 \le i \le n. \tag{5.1}$$

Particularly, block codes over GF(q) satisfy

$$s_i(C) \ge I_u(X_{i^-}; X_{i^+}) = g_i'(C) - g_i(C), \qquad 0 \le i \le n. \tag{5.2}$$

Group codes meet this bound with equality.

In a vector form this bound reads $\boldsymbol{s}(C) \ge \boldsymbol{g}'(C) - \boldsymbol{g}(C)$.

Proof: This theorem follows immediately from Theorem 1. The ensemble X_{i^-} and X_{i^+} correspond to X and Y of Theorem 1, respectively. The ensemble Z of Theorem 1 takes on the trellis states at level i when applied to this theorem. In Theorem 1 we did not require that every edge should be represented by a single subgraph. Thus in this theorem we do not confine ourselves to one-to-one trellises. □

For linear codes, $H(X_{i^-})$ reduces to the dimension of the projection of the code onto the first i indices (inverse ordered DLP), $\tilde{k}_i(C)$, and $H(X_{i^-}|X_{i^+})$ is actually the dimension of the subcode of C that all its components are zero beyond the ith index, $k_i(C)$. The subtraction of these two terms yields the relation between the state complexity profile of linear codes and the mutual information between the past and the future portions of the code [20]

$$s_i(C) = I(X_{i^-}; X_{i^+}) = \tilde{k}_i(C) - k_i(C).$$

Notably, equality may hold in (5.1) also for nonlinear codes. From the proof of Theorem 1 we have the necessary and sufficient conditions for this equality to hold.

Corollary 2: There exists a (biproper) trellis representation for an (n, M) code C whose state complexity at level $i, s_i(C)$, satisfies $s_i(C) = I(X_{i^-}; X_{i^+})$ iff the following two conditions hold.

1) C is i-separable [23] at level i (rectangular at level i).
2) The past/future array representing the codewords of C [15] comprises disjoint rectangles with the same number of points. That is,

$$\begin{aligned} &|\{(\boldsymbol{a}, \boldsymbol{b})\colon \{(\boldsymbol{a}, \boldsymbol{b}), (\boldsymbol{a}, \boldsymbol{b_p}), (\boldsymbol{a_p}, \boldsymbol{b})\} \subset C\}| \\ &\quad = |\{(\boldsymbol{a}, \boldsymbol{b})\colon \{(\boldsymbol{a}, \boldsymbol{b}), (\boldsymbol{a}, \boldsymbol{b_q}), (\boldsymbol{a_q}, \boldsymbol{b})\} \subset C\}| \\ &\qquad \forall p, q \text{ such that } (\boldsymbol{a_p}, \boldsymbol{b_p}) \in C, (\boldsymbol{a_q}, \boldsymbol{b_q}) \in C. \end{aligned}$$

Proof: From the proof of Theorem 1 it follows that equality holds in (5.1) iff

$$H(X_{i^+}|Z) = H(X_{i^+}|X_{i^-})$$
$$H(X_{i^-}|Z) = H(X_{i^-}|X_{i^+})$$

and

$$\log |Z| = H(Z).$$

The first two equalities impose requirement 1), and the last equality is equivalent to the second requirement. □

For example, the state complexity profile of the $\mathcal{N}_{16}$ Nordstrom–Robinson code in its *standard bit-order* ([4] and [24]) meets the bound of Theorem 4 with equality. This is not true for every coordinate ordering of this code notwithstanding the fact that it is rectangular under any bit-order [23]. Clearly, the second requirement does not necessitate that all the trellis states at level i will have the same number of incident branches from the origin. The mandatory requirement is that the same amount of codewords pass through any state at level i.

The state complexity profile of a code depends on the ordering of its coordinates [4]. Recently, a few studies (e.g., [2], [11], [12], and [25]) introduced "efficient" and even optimal coordinate orderings for some linear codes. These constructions have a reduced trellis complexity. There is no general efficient algorithm for generating a minimal trellis over all possible coordinate permutations. However, the trellis complexity under any given permutation is bounded by the unordered DLP profile [5], which can be achieved for some codes (e.g., [2], [11]) componentwise. The following theorem extends this idea to nonlinear codes.

Theorem 5 (State Complexity Profile, any Coordinate Ordering): Denote by Π any coordinate permutation on the index set I. The state complexity profile of any trellis representation for an (n, M) block code C over GF(q) under any coordinate ordering, denoted by ΠC, is bounded by

$$\begin{aligned} s_i(\Pi C) &\geq h'_i(C) - h_i(C) = \log_q M - h_i(C) - h_{n-i}(C) \\ &= h'_i(C) + h'_{n-i}(C) - \log_q M. \end{aligned} \quad (5.3)$$

Proof: We recall the relations between the ordered entropy/length profiles and the unordered profiles: $\boldsymbol{h}'(C) \leq \boldsymbol{g}'(C)$ and $\boldsymbol{h}(C) \geq \boldsymbol{g}(C)$. These relations were alluded to in Section III. The use of these inequalities in the previous bound (5.2) yields $s_i(\Pi\, C) \geq h'_i(C) - h_i(C)$. The other forms of the bound are derived from Lemma 2. □

For linear codes this bound reduces to the DLP bound on the state complexity profile [5].

It is evident that the foregoing bounds are not smaller (and usually larger) than the corresponding CLP bounds [16] since $\log_q M(l; C)$ of this reference is not smaller than $h_l(C)$ for any l. The use of the ELP bound for nonlinear codes is advantageous since it utilizes the inherent asymmetry of the code, unlike the CLP-based bounds.

Example 5.1: The $\mathcal{A}_{12}$ Hadamard code (cf. [18, p. 39]) is an $(11, 12, 6)$ nonlinear code. Using Theorem 5, we obtain the lower bound on the state complexity profile under any bit-order (see the first expression at the bottom of this page). The bound is listed in terms of the cardinality of the vertex count.

The CLP bound [16] for this case is six states at indices 3–8. The ordered conditional ELP evaluated for any coordinate ordering, coincides with the unordered profile at all the indices except for $i = 6$. At this index the ordered conditional ELP is zero for some coordinate orderings. The lower bound for a given coordinate ordering (Theorem 4) is as listed except for the indices 5 and 6. The bound on the vertex count at these indices may be one of the following, depending on the specific permutation: $12, 12$; $11, 12$; or $12, 11$; respectively. This code is rectangular [23]. It is easily verified that the minimal biproper trellis under every coordinate permutation achieves the corresponding profile componentwise.

Example 5.2: The largest linear double-error-correcting code of length 11 includes 16 codewords. The corresponding nonlinear code, the $\mathcal{B}_{12}$ Hadamard code [18, p. 39], consists of 24 codewords and it has the above correction capability. The parameters of this code are hence $(11, 24, 5)$. Theorem 5 yields the following lower bound on state complexity profile of the code under any coordinate ordering (see the second expression at the bottom of this page).

The CLP bound is 12 states at the indices 4 and 7, and six states at levels 5 and 6.

Example 5.3: The Nordstrom–Robinson code $\mathcal{N}_{16}$ is a $(16, 256, 6)$ nonlinear code. The following profiles illustrate the improvement of the ELP lower bound (Theorem 5) upon the corresponding CLP-based lower bound [16]. These profiles bound the total number of vertices under any coordinate ordering (see the top of the following page).

We see that the improvement of the proposed bound for this code is by a factor of more than 2. The use of *ad hoc* combinatorial properties of the code [16] improves the ELP bound only at indices 7 and 9 to 96 states instead of 95. The evaluation of the ELP bound is, however, much more systematic and simple.

All the codes in the above three examples have been shown to be rectangular codes under any bit-order [23]. However, many other important codes do not have this property. These codes include the $(9, 20, 4)$ conference matrix code [18, p. 57], the $(10, 40, 4)$ Best code [3], and the $(10, 72, 3)$ as well as $(11, 72, 4)$ Julin codes [10]. We shall revisit the following example later in order to show that the trellis representation of a code under a permutation that makes the code rectangular, even when this exists, is not necessarily better than that of other permutations.

Example 5.4: The ELP and the implied lower bound on the state complexity profile of the conference matrix code $\mathcal{C}_9$

i:	0	1	2	3	4	5	6	7	8	9	10	11
$h_i(\mathcal{A}_{12})$:	0	0	0	0	0	0	1/6	1/6	2/3	1.58	2.58	3.58
$\lvert V_i(\mathcal{A}_{12})\rvert$:	1	2	4	8	11	11	11	11	8	4	2	1

i:	0	1	2	3	4	5	6	7	8	9	10	11
$h_i(\mathcal{B}_{12})$:	0	0	0	0	0	1/6	1/6	2/3	1.58	2.58	3.58	4.58
$\lvert V_i(\mathcal{B}_{12})\rvert$:	1	2	4	8	16	20	20	16	8	4	2	1

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$h_i(\mathcal{N}_{16})$:	0	0	0	0	0	0	0.25	0.25	1	1.19	2.19	3	4	5
ELP bound:	1	2	4	8	16	32	**48**	**95**	64	**95**	**48**	32	16	8
CLP bound:	1	2	4	8	16	32	22	43	64	43	22	32	16	8

i:	14	15	16
$h_i(\mathcal{N}_{16})$:	6	7	8
ELP bound:	4	2	1
CLP bound:	4	2	1

under any bit-order is

i:	0	1	2	3	4	5	6	7	8	9
$h_i(\mathcal{C}_9)$:	0	0	0	0	0.3	0.8	1.41	2.34	3.33	4.32
$\lvert V_i(\mathcal{C}_9)\rvert$:	1	2	4	8	10	10	8	4	2	1

The next theorem bounds the state complexity (the maximum value of the state complexity profile) of any trellis representation of a code under any coordinate permutation. The bound is derived by the partition of the index axis $I = \{1, \cdots, n\}$ into several sections of varying length in a manner similar to [16]. This technique has been proved in the above-referenced study to improve the bound on the state complexity for some codes. Our ELP-based bound improves the corresponding CLP-based bound of [16].

Theorem 6 (State Complexity, any Coordinate Ordering): Let C be an (n, M) code over GF(q) with ELP $\boldsymbol{h}(C)$ and let $\{l_1, l_2, \cdots, l_L\}$ be any set of L positive integers provided that $l_1 + l_2 + \cdots + l_L = n$. The state complexity of any trellis representation of C under any coordinate permutation is bounded by

$$s(C) \geq \frac{\log_q M - \sum_{j=1}^{L} h_{l_j}(C)}{L-1}. \tag{5.4}$$

Proof: We partition the trellis for C into L sections of lengths $l_1, l_2, \cdots, l_L$. Let

$$f_m = \sum_{j=1}^{m} l_j, \qquad 1 \leq m \leq L$$

be the last index of the mth section with the convention $f_0 = 0$. That is, the first section covers the indices $[1, l_1]$, and the mth section includes the indices $[f_{m-1}+1, f_m]$. We denote by $\{W_m, 1 \leq m \leq L\}$ a random l_m-tuple that takes the values of the set $P_{[f_{m-1}+1, f_m]}(C)$ with probabilities that are determined by the uniform probability of the codewords. We make C into the ensemble $W \triangleq W_1 W_2 \cdots W_L$. We also denote

$$W \backslash W_j \triangleq W_1 W_2 \cdots W_{j-1} W_{j+1} \cdots W_L, \qquad 1 \leq j \leq L$$

with the conventions

$$W \backslash W_1 \triangleq W_2 W_3 \cdots W_L \qquad W \backslash W_L \triangleq W_1 W_2 \cdots W_{L-1}.$$

From (5.1) we have

$$s_{f_j}(C) \geq I(W_1 W_2 \cdots W_j; W_{j+1} W_{j+2} \cdots W_L), \qquad 1 \leq j \leq L-1.$$

Thus also

$$\begin{aligned}
\sum_{j=1}^{L-1} s_{f_j}(C) &\geq \sum_{j=1}^{L-1} I(W_1 W_2 \cdots W_j; W_{j+1} W_{j+2} \cdots W_L) \\
&= \sum_{j=1}^{L-1} \{H(W_1 W_2 \cdots W_j) \\
&\quad - H(W_1 W_2 \cdots W_j \mid W_{j+1} W_{j+2} \cdots W_L)\} \\
&= \sum_{j=1}^{L-1} H(W_1 W_2 \cdots W_j) \\
&\quad - \sum_{j=2}^{L-1} \{H(W_1 W_2 \cdots W_{j-1} \mid W_{j+1} \cdots W_L) \\
&\quad + H(W_j \mid W_1 \cdots W_{j-1} W_{j+1} \cdots W_L)\} \\
&\quad - H(W_1 \mid W_2 W_3 \cdots W_L) \\
&= H(W_1 W_2 \cdots W_{L-1}) + \sum_{j=1}^{L-2} \{H(W_1 W_2 \cdots W_j) \\
&\quad - H(W_1 W_2 \cdots W_j \mid W_{j+2} \cdots W_L)\} \\
&\quad - \sum_{j=1}^{L-1} H(W_j \mid W \backslash W_j) \\
&\geq H(W_1 W_2 \cdots W_{L-1}) - \sum_{j=1}^{L-1} H(W_j \mid W \backslash W_j) \\
&= H(W) - \sum_{j=1}^{L} H(W_j \mid W \backslash W_j).
\end{aligned}$$

We apply the inequality $H(X_J|X_{I-J}) \leq h_{|J|}(C)$ to the foregoing inequality to obtain

$$\sum_{j=1}^{L-1} s_{f_j}(C) \geq H(W_1 W_2 \cdots W_L) - \sum_{j=1}^{L} h_{l_j}(C). \qquad (5.5)$$

Consequently,

$$s(C) \geq \frac{\log_q M - \sum_{j=1}^{L} h_{l_j}(C)}{L-1}. \qquad \square$$

This bound is equal to the bound proposed in [16] for linear codes. For nonlinear codes the above bound is tighter than the corresponding CLP bound. We also note that Theorem 5 can be viewed as a special case of (5.5) for partition into two sections.

The following theorem provides an upper bound on the state complexity profile. We claim that there exists a trellis diagram whose vertex count at any specific level is not larger than a defined measure. Furthermore, we construct a diagram that achieves the bound. The bound may not be achievable at all the trellis levels simultaneously. Thus the bound for each level of the diagram may be realized by a different construction. Clearly, when the bound is applied to a rectangular code, it bounds the state complexity of the biproper trellis of the code at all the indices simultaneously. This bound is deduced from the general bounds devised in Theorems 2 and 3. Consequently, the bound can be equivalently stated in two different frameworks. The first derivation is a graph-theoretic method that introduces the new concept of "nonconcurring codewords." The second approach rephrases the bound by information-theoretic measures similarly to Theorem 3.

Definition 5: Let C be an (n, M) code. We say that a subcode $D, D \subseteq C$ is a set of *nonconcurring codewords* at index i, $1 \leq i \leq n-1$, if $P_{i^-}(\boldsymbol{c}^l) \neq P_{i^-}(\boldsymbol{c}^m)$ and $P_{i^+}(\boldsymbol{c}^l) \neq P_{i^+}(\boldsymbol{c}^m)$ for any two codewords $\boldsymbol{c}^l, \boldsymbol{c}^m \in D$ such that $\boldsymbol{c}^l \neq \boldsymbol{c}^m$.

Definition 6: A trellis state at level i will be called a *strict-sense nonmerging state* if it has a single incoming length-i branch.

This requirement is stricter than that for a single incoming edge (length-1 branch) into the state. A nonmerging state is not necessarily a strict-sense nonmerging state.

Definition 7: A trellis state at level i will be called a *strict-sense nonexpanding state* if it has a single outgoing length-$(n-i)$ branch.

Theorem 7: There exists a (one-to-one) trellis diagram for an (n, M) code C whose vertex count at any index i is equal to N_i, the cardinality of the largest set of nonconcurring codewords (of C) at index i. Each of the states at level i in this diagram is either a strict-sense nonmerging state or a strict-sense nonexpanding state. The construction is index-dependent. That is, the above profile may not be achieved at all the trellis levels simultaneously. Furthermore, make C into an ensemble that takes one of M possible values (codewords). Assign each codeword a probability $p_m, 1 \leq m \leq M$, with $p_1 + p_2 + \cdots + p_M = 1$. Then

$$\max_{p_1, p_2, \cdots, p_M} \left\{ I(X_{i^-}; X_{i^+}) \colon \sum_{m=1}^{M} p_m = 1 \right\} = \log N_i, \quad 1 \leq i \leq n-1. \qquad (5.6)$$

Proof: The representation of the codewords of C via a past/future array is equivalent to the vertex-incident matrix of Theorem 2. A row in this array corresponds to a strict-sense nonmerging state and a column in this array corresponds to a strict-sense nonexpanding state. By Theorem 2 this array can be covered by N_i lines. We also use the construction of Theorem 2 to draw a trellis representation for C with the property that each state at level i is either a strict-sense nonexpanding state or a strict-sense nonmerging state. Finally, equality (5.6) follows from Theorem 3. $\square$

It should be noted that this form of the upper bound explicitly ensures the obvious fact that the upper bound on the state complexity profile is not smaller than the lower bound. Also, Theorem 7 immediately furnishes the following obvious bound.

Corollary 3: There exists a (one-to-one) trellis representation for an (n, M) code C whose state complexity at level i satisfies

$$s_i(C) \leq \min(i, n-i, \log_q M), \qquad 0 \leq i \leq n. \qquad (5.7)$$

It is noteworthy to comment that for linear codes we have also the following relation (under a uniform distribution of the codewords):

$$\begin{aligned} s_i(C) &= I(X_{i^-}; X_{i^+}) \\ &= H(X_{i^-}) + H(X_{i^+}) - H(X_{i^-} X_{i^+}) \\ &\leq i + (n-i) - \log_q M = n - k \end{aligned}$$

where k is the dimension of the code. This relation does not necessarily hold for nonlinear codes.

Example 5.1 (continued): Applying the upper bound of Theorem 7 to the $(11, 12, 6)$ Hadamard code $\mathcal{A}_{12}$, we obtain that this bound coincides with the lower bound of Theorem 4 under any coordinate ordering, yielding the profile shown at the bottom of this page, where X and Y, the bounds for the indices 5 and 6, are permutation-dependent and the three possible values of the bound at these indices are $12, 12$; $11, 12$; or $12, 11$; respectively. The complete profile for each permutation is achieved componentwise in this case.

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert V_i(\mathcal{A}_{12})\rvert$:	1	2	4	8	11	X	Y	11	8	4	2	1

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\lvert V_i(\mathcal{N}_{16})\rvert$:	1	2	4	8	16	32	64	96	128	96	64	32	16	8	4	2	1

Example 5.2 (continued): By Theorem 7, the least upper bound on the state complexity profile of the $(11, 24, 5)$ Hadamard code $\mathcal{B}_{12}$, over all coordinate orderings is

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert V_i(\mathcal{B}_{12})\rvert$:	1	2	4	8	16	22	22	16	8	4	2	1

This bound differs from the lower bound only at the indices 5 and 6. The lower bound at these indices is 20 states. This code is known to be a rectangular code [23]. It can be verified that the minimal biproper trellis diagram (over all bit-orders) for this code comprises 21 states at these indices.

Example 5.3 (continued): Applying Theorem 7 to the $(16, 256, 6)$ Nordstrom–Robinson code $\mathcal{N}_{16}$, we obtain that there are coordinate permutations whereby this bound yields the value of 96 states at the indices 7 and 9. The least upper bound on the complete profile is shown at the top of this page.

This result disproves the conjecture of Lafourcade and Vardy [16] that the state complexity profile of the known construction of the code in [4] and [24] is optimal componentwise. The construction in the aforementioned papers includes 128 states at the levels 7 and 9.

The Nordstrom–Robinson code can be expressed as a union of eight cosets of the first order $(16, 5, 8)$ Reed–Muller code. A twisted squaring construction of the code $|(8,7)/(8,4)/(8,1)|^2$ was given by Forney in [4]. Vardy [24] showed that the code can be expressed as the union of binary images of two isomorphic linear $(4, 2, 3)$ codes over $\mathrm{GF}(4)$. The trellises of the two constructions coincide, yielding the same state complexity profile. This profile can be formulated by

$$s_i(\mathcal{N}_{16}) = \min\{i, 16 - i, s_i[\mathrm{RM}(1,4)] + 3\}$$

where $s_i[\mathrm{RM}(1,4)]$ is the state complexity profile of the $(16, 5, 8)$ Reed–Muller code in its standard bit-order which is optimal componentwise and derived in [17]. The expression $s_i[\mathrm{RM}(1,4)] + 3$ is due to the construction of the code as eight cosets of the $\mathrm{RM}(1,4)$ code. The vertex count of these constructions is therefore as shown in the first expression at the bottom of this page.

This diagram comprises a single type of states (expanding, merging, etc.) at each level. Using our upper bound we obtained a new construction of the code. Our construction utilizes the asymmetry of the code and the interconnections between the constituent cosets. The new (biproper) trellis incorporates different types of states at each level, unlike the known construction, and it also reduces the vertex count at levels 6 and 10 to 48 vertices in comparison to 64 states in the known construction. Recalling that the lower bound at these indices is 48 states we realize that the lower bound on the state complexity profile of this code is quite tight. The trellis representation of the new construction includes, however, 128 states at level 8, whereas the known construction includes only 64 states at this level. The total number of edges in both constructions is identical (764 edges). The novel construction consists of four disjoint parallel subtrellises without cross-connections between the indices 6–10. These four components are structurally identical. Each one of the four parallel parts incorporates paths that describe eight codewords of each coset and 64 codewords altogether. The state complexity profile of the diagram of our new construction is as listed at the bottom of the page.

We shall henceforth denote the four possible types of trellis states (in the constructed diagram) as follows:

- —<: Expanding state—a single incoming edge and two outgoing edges.
- >—: Merging state—two incoming edges and a single outgoing one.
- —•—: Simple extension state—a single incoming and a single outgoing edge.
- ><: Butterfly state—two incoming edges and two outgoing edges.

In the new construction, all the states in the levels 0–5 are expanding states, and all the states in the levels 11–16 are merging states. The states at the other levels are delineated at the top of the following page.

Interconnections Caption: The butterfly states at level 6 are connected to the simple extension states at level 7, and thus the expanding states at level 6 are connected to the same type of states at the next level. All the simple extension states at level 7 are connected to simple extension states at level 8.

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\lvert V_i(\mathcal{N}_{16})\rvert$:	1	2	4	8	16	32	64	128	64	128	64	32	16	8	4	2	1

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\lvert V_i(\mathcal{N}_{16})\rvert$:	1	2	4	8	16	32	48	96	128	96	48	32	16	8	4	2	1

i:	6		7		8		9		10
States/Count:	$>\!< - 16$		$-\!\bullet\!- - 32$		$\succ\!- - 32$		$\succ\!- - 64$		$\succ\!- - 32$
	$-\!\prec - 32$		$-\!\prec - 64$		$-\!\bullet\!- - 64$		$-\!\bullet\!- - 32$		$>\!< - 16$
					$-\!\prec - 32$				
Total no. of states:	48		96		128		96		48
Total no. of edges:		96		160		160		96	

Thus the expanding states at level 7 are connected to all three types of states at level 8. All the merging and expanding states at level 8 are connected to merging states at the next level. All the merging states at level 9 are connected to merging states at level 10. The diagram is symmetrical about the center level $i = 8$.

There are numerous coordinate permutations that yield this structure. For example, we list the codewords as the union of eight cosets of the $(16,5,8)$ Reed–Muller code. We take the following eight coset representatives:

0000	0000	0000	0000	0001	1110	1000	1000
0110	1010	1100	0000	1011	1000	0100	1000
0011	1001	1010	0000	1000	1101	0010	1000
1010	0011	0110	0000	0010	1011	0001	1000

We use the standard ordering of the $(16,5,8)$ Reed–Muller code, i.e., we build the code by the generator matrix

$$\begin{bmatrix} 1111111111111111 \\ 0000000011111111 \\ 0000111100001111 \\ 0011001100110011 \\ 0101010101010101 \end{bmatrix}$$

If we denote by $c_1, c_2, \cdots, c_{16}$ the coordinates of the foregoing ordering, we apply the coordinate permutation

$$(c_3, c_4, c_5, c_8, c_9, c_{11}, c_{10}, c_{16}, c_{14}, c_{12}, c_{15}, c_{13}, c_6, c_7, c_1, c_2)$$

to achieve the new trellis structure. It should be noted, however, that this is just an example for a coordinate ordering that provides the new construction, and there are many other permutations that result in the same trellis.

Example 5.4 (continued): Applying the upper bound of Theorem 7 to the conference matrix code $\mathcal{C}_9$, we obtain the following (least) upper bound which is indeed achieved for some permutations componentwise.

i:	0	1	2	3	4	5	6	7	8	9
$\lvert V_i(\mathcal{C}_9)\rvert$:	1	2	4	8	12	12	8	4	2	1

Notably, the code is not rectangular under any of the optimal bit permutations that attain this profile. The vertex count of the code, under the permutations that make it rectangular, is 16 at indices 4 and 5 [23]. The following codeword list is an example of an optimal bit-order:

000000000	110101010	111001001	110011100
011100101	001111100	101100110	000011111
100110011	011010011	101010101	010110110
001101011	101011010	110000111	010111001
111110000	011001110	100101101	111111111

VI. Bounds on the Branch Complexity Profile of (Nonlinear) Block Codes

The computational complexity of the Viterbi algorithm is proportional to the number of edges in the trellis diagram. Thus this number is a more accurate measure of the decoding complexity. The following theorems bound the total number of branches between any two levels of the trellis diagram. These two levels need not be adjacent. When we apply the bound to adjacent levels we deduce a bound on the edge complexity profile as a special case. Among all trellis representations of linear codes, the BCJR trellis has the fewest edges [20]. This trellis representation, and actually any proper trellis for both linear and nonlinear codes satisfies $s_i(C) \leq b_{i,i+1}(C) \leq s_i(C)+1$. Biproper trellises also satisfy $s_{i+1}(C) \leq b_{i,i+1}(C) \leq s_{i+1}(C)+1$. Hereafter, we shall show that the lower bounds on the state and the branch complexity (and not the quantities themselves) satisfy these relations.

In this section we dwell on the study of trellis representation of nonlinear block codes. We shall be concerned with the problem of minimizing the branch count between the indices i and $j > i$ in a trellis representation for C. We do not confine our interest only to the case $j = i+1$. We elaborate on the observation of [15] that a vertex in a trellis for a code C at level i represents a product-form subset of C. The minimization of the vertex count at each level has been shown to be equivalent to the problem of covering the past/future Cartesian array by rectangles. We shall show that the branch set between any two levels also represents a product-form relation induced by C.

A useful representation of the codewords of C for the purpose of bounding the branch complexity between the levels i and $j > i$ is by the pair $[\boldsymbol{a}, \boldsymbol{b}]$, where $\boldsymbol{a} = P_{j^-}(\boldsymbol{c})$ and $\boldsymbol{b} = P_{i^+}(\boldsymbol{c})$ and $\boldsymbol{c} \in C$. Clearly, the concatenation of $\boldsymbol{a}$ and $\boldsymbol{b}$ does not provide a codeword of C since these two portions of a codeword have $(j-i)$ common symbols of the codeword. Thus we use the square brackets when we use this representation of the codewords. We denote

$$\tilde{C} \triangleq \{[P_{j^-}(\boldsymbol{c}), P_{i^+}(\boldsymbol{c})]\colon \boldsymbol{c} \in C\}.$$

Any (n, M) block code C induces a relation between the set of the past projection of the codewords of $C, P_{j^-}(C)$, and the future projection set $P_{i^+}(C)$. Clearly,

$$\tilde{C} \subset P_{j^-}(C) \times P_{i^+}(C), \forall i, j > i.$$

Similarly to the partitioning of C into past/future portion, we consider the modified code $\tilde{C}$ as a code of length two over the alphabet $P_{j^-}(C) \times P_{i^+}(C)$. The problem of minimizing the branch count between levels i and $j > i$ may also be visualized via a Cartesian array. Each row of the prescribed array is identified with a distinct value of the set $P_{j^-}(C)$ and each column represents a different value of the set $P_{i^+}(C)$. Consequently, the minimization problem addressed in this section is identical to finding a minimal covering of the array for $\tilde{C}$ by product-form subsets. These subsets are represented by complete rectangles in the array. Using the above model, we can utilize the general results of Section IV and devise bounds on the branch complexity profile easily. We hereby evade proving these results directly, though the latter approach is also not a too intricate task.

Theorem 8 (Branch Complexity Profile Under a Given Coordinate Permutation): The branch complexity between the indices i and $j > i$ of any trellis representation for an (n, M) block code C is bounded by

$$\begin{aligned} b_{i,j}(C) &\geq I_u(X_{j^-}; X_{i^+}) \\ &= H(X_{[i+1,j]}) + I_u(X_{i^-}; X_{j^+} | X_{[i+1,j]}) \\ &= H(X_{i^-} X_{i^+}) - H(X_{i^-} | X_{i^+}) - H(X_{j^+} | X_{j^-}), \\ &\qquad 0 \leq i < j \leq n. \end{aligned} \tag{6.1}$$

Particularly, we have for codes over a fixed alphabet of size q (e.g., GF(q))

$$b_{i,j}(C) \geq I_u(X_{j^-}; X_{i^+}) = g'_j(C) - g_i(C), \qquad 0 \leq i < j \leq n. \tag{6.2}$$

The bound is met with equality for group codes.

The proof of this theorem follows from Theorem 1 and the model presented at the beginning of this section.

For $j = i+1$ we obtain $b_{i,i+1}(C) \geq g'_{i+1}(C) - g_i(C)$. From Lemma 3 we also have $g'_i(C) \leq g'_{i+1}(C) \leq g'_i(C) + 1$ and $g_i(C) \leq g_{i+1}(C) \leq g_i(C) + 1$. Thus the lower bound on the branch complexity between the indices i and $i+1$ is neither smaller than the bound on the state complexity at levels i and $i+1$ nor larger than this bound by more than 1. For linear codes, this relation exists between the state and branch complexities.

Definition 8: Let k be an integer, $1 \leq k \leq n$. Two length-k branches through the trellis that correspond to the same k-tuple, start at the same state and terminate in the same state will be called *congruent branches*.

It is readily verified that Theorem 8 can be stated in a stronger way. It actually bounds the minimal number of noncongruent branches between the levels i and $j > i$.

Recently it has been proved ([20], [22], and [26]) that the known construction of trellises (for linear codes) of Bahl *et al.* [1] and Forney [4] minimizes a wide variety of complexity measures. This construction minimizes the vertex count and the edge count at each index. It also minimizes the *expansion index* [26] (maximizes the *Euler characteristic* [22]), and thus it minimizes the overall Viterbi decoding complexity. Our lower bounds (Theorems 4 and 8) can be considered as an alternative proof to the first two claims. Moreover, using Theorem 8 we establish an additional property of the unique minimal biproper trellis for group codes and in particular of the BCJR trellis [1].

Corollary 4: The total number of branches between any two indices in the minimal trellis for a *group* code C is not larger than that of any other trellis representation for C.

Proof: The branch complexity of the minimal trellis for group codes is given in [5] and [6]. Evidently, this complexity coincides with the right-hand side of (6.1). □

Theorem 9 (Branch Complexity Profile, any Coordinate Ordering): The branch complexity profile of any trellis representation for an (n, M) block code C over GF(q), under any coordinate ordering ΠC, is bounded by

$$\begin{aligned} b_{i,j}(\Pi C) &\geq h'_j(C) - h_i(C) \\ &= \log_q M - h_i(C) - h_{n-j}(C) \\ &= h'_j(C) + h'_{n-i}(C) - \log_q M. \end{aligned} \tag{6.3}$$

Proof: We replace the ordered profiles in (6.2) with the corresponding unordered profiles, recalling that $\boldsymbol{h}'(C) \leq \boldsymbol{g}'(C)$ and $\boldsymbol{h}(C) \geq \boldsymbol{g}(C)$, and the theorem follows. □

It is evident that the foregoing bounds are not smaller (and usually larger) than the corresponding CLP bounds of [16].

Example 6.1: Using the conditional ELP of the $(11, 12, 6)$ Hadamard code $\mathcal{A}_{12}$ as evaluated in Example 5.1, Theorem 9 provides the following lower bound on the edge (length-1 branches) count of any trellis representation of the code under any coordinate ordering:

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert B_{i,i+1}(\mathcal{A}_{12})\rvert$:	2	4	8	11	11	12	11	11	8	4	2	

Furthermore, applying Theorem 8 to every bit-order, we achieve the above bound, except for the bound on the edge count between the indices 4–5 and 6–7. Actually, there are three typical permutations providing the bounds $12, 12$; $11, 12$; and $12, 11$ between the above indices, respectively. This indicates that the bound of 11 edges between the indices 4–5 and 6–7 cannot be achieved simultaneously. The corresponding CLP bound is six edges between the indices 3–4, 4–5, 6–7, and 7–8.

Example 6.2: Using the conditional ELP of the $(11, 24, 5)$ Hadamard code $\mathcal{B}_{12}$, as found in Example 5.2, in conjunction with Theorem 9, we obtain the following lower bound on the edge count of any trellis representation of the code under any bit-order:

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert B_{i,i+1}(\mathcal{B}_{12})\rvert$	2	4	8	16	22	20	22	16	8	4	2	

Applying Theorem 8 to every coordinate permutation, we achieve the above bound except for the bound on the edge count between the indices 5–6. It follows from Theorem 8 that the number of edges between these indices cannot be smaller than 22. The corresponding CLP bound is 12 edges between the indices 3–4, 4–5, 6–7, and 7–8. Likewise, it provides the bound of six edges between the levels 5 and 6. It is readily verified that the bound of 22 edges between the levels 4–5,

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\lvert B_{i,i+1}(\mathcal{N}_{16})\rvert$:	2	4	8	16	32	57	95	108	108	95	57	32	16	8	4	2	

5–6, and 6–7 is realizable. Obviously, the bound for the other levels is achievable as well.

Example 6.3: Applying Theorem 9 to the $(16, 256, 6)$ Nordstrom–Robinson code, along with the conditional ELP that was derived in Example 5.3, we obtain the following lower bound on the edge count between adjacent levels of any trellis representation of the code (see the top of the following page).

The corresponding CLP bound, without the use of code-dependent combinatorial considerations, is much looser than the above bound. The list at the bottom of this page depicts the CLP bound on the branch complexity profile of [16], along with the improved bound (also of [16]). The latter bound employs combinatorial properties of this specific code.

The improved bound also implies that the sum of the total number of edges between the levels 7–8 and 8–9 is at least 256 edges.

Theorem 10 (Length-k Branch Complexity, any Coordinate Ordering): Let C be an (n, M) code over GF(q) with conditional ELP $\boldsymbol{h}(C)$, and let $l_1, l_2, \cdots, l_L$ be any L positive integers such that $l_1 + l_2 + \cdots + l_L = n - k(L-1)$, then the length-$k$ branch complexity of any trellis representation of C, under any coordinate ordering, is bounded by

$$b^k(C) \geq \frac{\log_q M - \sum_{j=1}^{L} h_{l_j}(C)}{L-1}. \tag{6.4}$$

Proof: We partition the trellis for C into L sections of lengths $l_1, l_2, \cdots, l_L$. We denote by f_m the last index of the mth section,

$$f_m = \sum_{j=1}^{m} l_j + k(m-1), \qquad 1 \leq m \leq L$$

with the convention $f_0 = -k$. We denote by $W_m, 1 \leq m \leq L$ a random l_m-tuple that takes on the values of the set $P_{[f_{m-1}+k+1, f_m]}(C)$, respectively, with probabilities that are determined by the uniform distribution of the codewords. Likewise, we define another set of random k-tuples, $\{U_i, 1 \leq i \leq L-1\}$ that take the values of the indices that are not covered by $\{W_m\}$. That is, each U_i takes the values of the set $P_{[f_i+1, f_i+k]}(C)$. We make C into the ensemble W and we use the following notations:

$$\begin{aligned} W &\triangleq W_1U_1W_2U_2 \cdots W_{L-1}U_{L-1}W_L, \\ W \backslash W_j &\triangleq W_1U_1W_2U_2 \cdots W_{j-1}U_{j-1}U_jW_{j+1}U_{j+1} \cdots \\ &\quad W_{L-1}U_{L-1}W_L, \qquad 1 \leq j \leq L \end{aligned}$$

with

$$W \backslash W_1 \triangleq U_1W_2U_2 \cdots W_{L-1}U_{L-1}W_L$$

and

$$W \backslash W_L \triangleq W_1U_1W_2U_2 \cdots W_{L-1}U_{L-1}.$$

In the proposed partition, each ensemble $U_i, 1 \leq i \leq L-1$, is a random k-tuple. We sum the branch complexity between the indices f_j and $f_j + k$ for the $L-1$ possible values of $j, 1 \leq j \leq L-1$.

From (6.1) we have

$$\begin{aligned} &\sum_{j=1}^{L-1} b_{f_j, f_j+k}(C) \\ &\quad \geq \sum_{j=1}^{L-1} \{H(W_1U_1W_2U_2 \cdots W_jU_j) \\ &\qquad - H(W_1U_1W_2U_2 \cdots W_{j-1}U_{j-1}W_j | U_jW_{j+1}U_{j+1} \cdots \\ &\qquad\quad W_{L-1}U_{L-1}W_L)\}. \end{aligned}$$

After some algebra we obtain that the right-hand term of the above inequality is not smaller than

$$H(W) - \sum_{j=1}^{L} H(W_j | W \backslash W_j). \tag{6.5}$$

To complete the proof we recall that

$$H(W_j | W \backslash W_j) \leq h_{l_j}(C).$$

We replace the ordered ELP quantities of inequality (6.5) with the unordered quantities to obtain

$$\sum_{j=1}^{L-1} b_{f_j, f_j+k}(C) \geq H(W) - \sum_{j=1}^{L} h_{l_j}(C) \tag{6.6}$$

which establishes the theorem. □

The bound proposed in [16] for linear codes is identical to the above bound for the special case $j = i + 1$. For other values of k, (6.4) is a new bound. For nonlinear codes the

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CLP bound:	2	4	8	16	32	43	43	64	64	43	43	32	16	8	4	2	
Improved bound:	2	4	8	16	32	64	96	96	96	96	64	32	16	8	4	2	

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert B_{i,i+1}(\mathcal{A}_{12})\rvert$:	2	4	8	11	U	12	V	11	8	4	2	

foregoing bound is tighter than the corresponding CLP bound. Theorem 9 can be viewed as a special case of inequality (6.4) for partition into two sections.

The notion of nonconcurring codewords and the equivalent information-theoretic bound of Theorem 7 can be extended to the branch complexity profile. Similarly to Theorem 7, the following upper bound is constructive. We construct a trellis diagram whose branch count between two levels is equal to a defined measure. The minimum achievable branch count between the two levels is therefore not larger than that of the devised diagram. The bound may not be attainable at all the trellis levels simultaneously. This bound completes the analogy between the state complexity profile and the branch complexity profile. Actually, both the lower bounds on the state complexity profile and the upper bound on this profile can be considered as special cases of the branch complexity bounds evaluated for $j = i$.

Definition 9: Let C be an (n, M) code. We say that a subcode D, $D \subseteq C$ is a set of *wide-sense nonconcurring codewords at indices* i *and* $j > i$ if $P_{j^-}(\boldsymbol{c}^l) \neq P_{j^-}(\boldsymbol{c}^m)$ and $P_{i^+}(\boldsymbol{c}^l) \neq P_{i^+}(\boldsymbol{c}^m)$ for any two codewords $\boldsymbol{c}^l, \boldsymbol{c}^m \in D$ such that $\boldsymbol{c}^l \neq \boldsymbol{c}^m$ and $0 \leq i < j \leq n$. Two codewords in such a set need not necessarily satisfy $P_{i^-}(\boldsymbol{c}^l) \neq P_{i^-}(\boldsymbol{c}^m)$ or $P_{j^+}(\boldsymbol{c}^l) \neq P_{j^+}(\boldsymbol{c}^m)$.

Theorem 11: There exists a (one-to-one) trellis diagram for an (n, M) code C whose branch count between the indices i and j is equal to $N_{i,j}$, the cardinality of the largest set of wide-sense nonconcurring codewords (of C) at indices i and j. The two states at the two ends of each branch (at levels i and j) are either strict-sense nonexpanding states or strict-sense nonmerging states. This profile may be realized by a different construction for each level pair (i, j). Using information-theoretic concepts, we consider C as an ensemble that takes one of M possible values (codewords). We assign each codeword a probability p_m, $1 \leq m \leq M$, with

$$p_1 + p_2 + \cdots + p_M = 1.$$

Then

$$\max_{p_1, p_2, \cdots, p_M} \left\{ I(X_{j^-}; X_{i^+}) \colon \sum_{m=1}^{M} p_m = 1 \right\} = \log N_{i,j}, \quad 0 \leq i < j \leq n. \tag{6.7}$$

Proof: The representation of the codewords of C via an array as described in the introduction of the section is equivalent to the vertex-incident matrix of Theorem 2. A row in this array corresponds to a strict-sense nonmerging state at level j, and a column in this array corresponds to a strict-sense nonexpanding state at level i. By Theorem 2 this array can be covered by $N_{i,j}$ lines. Again, (6.7) follows from Theorem 3. □

Example 6.1 (continued): Applying Theorem 11 to the $(11, 12, 6)$ Hadamard code $\mathcal{A}_{12}$, we obtain that the upper bound on the edge count coincides with the lower bound (Theorem 8) componentwise, under every bit order. The resulting bound is the one shown at the top of this page, where U and V are permutation-dependent. Three typical coordinate permutations provide the following values of U and V: $12, 12$; $11, 12$; and $12, 11$; respectively. Thus for every coordinate ordering the lower bound (Theorem 8) coincides with the upper bound (Theorem 11) evaluated for the same permutation, yielding the edge complexity of the minimal biproper trellis which is the minimal trellis for this code thereupon.

Example 6.2 (continued): Applying Theorem 11 to the $(11, 24, 5)$ Hadamard code $\mathcal{B}_{12}$, we obtain that the upper bound on the edge count coincides with the lower bound (Theorem 8) componentwise, under every bit-order. The least upper bound is given in the first expression at the bottom of this page.

Theorem 11 provides this same bound profile under all coordinate permutations, excluding the bound on the total number of edges between the indices 5 and 6. The upper bound on this quantity is 24 for some permutations. The lower bound for a given bit-order (Theorem 8) coincides with the upper bound under every given order.

Example 6.3 (continued): The ELP upper bound (Theorem 11) on the edge count of the $(16, 256, 6)$ Nordstrom–Robinson code $\mathcal{N}_{16}$ is as shown in the second expression at the bottom of this page.

The trellis diagram of the constructions of the code due to Forney [4] and Vardy [24] meets this bound except for the number of edges between the levels 6–7 and 9–10. These constructions comprise 128 edges between these levels. Our

i:	0	1	2	3	4	5	6	7	8	9	10	11
$\lvert B_{i,i+1}(\mathcal{B}_{12})\rvert$:	2	4	8	16	22	22	22	16	8	4	2	

i:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\lvert B_{i,i+1}(\mathcal{N}_{16})\rvert$:	2	4	8	16	32	64	96	128	128	96	64	32	16	8	4	2	

upper bound implies that this number of 128 edges is not optimal at the corresponding levels and actually our new construction indeed includes 96 edges at the above levels. Thus our lower bound on the edge count between these indices is tight (95 edges). We also note that the above profile is pretty tight and in each level it coincides with the least edge count among Forney's and Vardy's construction and our new construction.

VII. Conclusion

A considerable advance in the study of trellis representation for linear block codes has been achieved in recent studies. Yet, less is known about the trellis structure of nonlinear codes. The problem of minimizing the vertex count of nonlinear codes may have several solutions, unlike the corresponding problem for linear codes. Furthermore, minimal trellis representations for nonlinear codes may not be proper or one-to-one. In this paper, we have defined entropy/length profiles. These profiles seem useful to the study of nonlinear codes. Additionally, we have established bounds on the minimal covering of a bipartite graph by complete subgraphs. The minimization of the vertex set of a trellis representation for the code is known to be equivalent to the minimal covering problem. Furthermore, we have derived a model under which the minimization of the branch count between any two levels in the trellis can also be considered as a minimal covering problem. Equipped with these models (analogies) and the entropy/length profiles, we have derived both lower and upper bounds on the trellis state and branch complexity profiles of (nonlinear) block codes. The new approach, which is also applicable to improper and non-one-to-one trellis structures, leads to appreciably tight bounds on the complexity measures of nonlinear codes. It is argued that the state complexity at any given level cannot be smaller than the mutual information between the past and the future portions of the code at this level under a uniform distribution of the codewords. The prescribed lower bounds are tighter than other known bounds. Apparently, they also establish the validity of the CLP-based bound to non-one-to-one trellises. We have also devised a particular trellis construction and an appropriate probabilistic model to prove that the minimum achievable state complexity over all possible trellises is not larger than the maximum value of the above mutual information over all possible probability distributions of the codewords. Moreover, we have shown how to construct a diagram that meets the bound. This construction yields a one-to-one trellis diagram endowed with special properties. Similar lower and upper bounds on the branch complexity profile and on the branch complexity are also derived.

A legitimate question that may arise at this stage is what is the computational complexity of our information-theoretic bounds? For brevity we give some final conclusions regarding this issue in gross. The total number of operations required to compute the lower bound on the entire state (or branch) complexity profile of an (n,M) code C over GF(q) under a given permutation is upper-bounded by $O(n\cdot M)$. This is actually the complexity of the computation of the ordered ELP. The computational complexity of the CLP bound is the same. The derivation of the actual trellis representation of the code (also under a given permutation) can be carried out in an $O(n\cdot M\cdot \log M)$ time algorithm [22], [26]. This algorithm applies to rectangular codes only. The computation of our upper bound is equivalent to the construction of a maximum matching in a bipartite graph. This search is known as the *Hungarian Method* due to Kuhn (e.g., [27]). One can think of the following efficient algorithm to implement this search: put the codewords into a two-dimensional array, scan a part of the array, and exchange rows and columns (when necessary). The computational complexity of each of the constituent steps of this algorithm is bounded above by $O(M^2)$ per a given index. Hence the derivation of a bound for all the trellis levels requires $O(n\cdot M^2)$ operations. We see that the construction of the trellis (for rectangular codes) entails larger computational complexity than the proposed lower bound but it necessitates less operations than the proposed upper bound.

We end this section with a brief summary of the bounds on the state and branch complexity and the state and branch complexity profiles with the relations between the different bounds.

- Minimum achievable state complexity profile under a given coordinate permutation $s_i(C)$
$$\begin{aligned} &h'_i(C) - h_i(C) \\ &\quad \le \{I(X_{i^-}; X_{i^+})\colon\, p_m = 1/M, \forall m\} \\ &\quad \le s_i(C) \le \max_{p_1,p_2,\cdots,p_M} \left\{ I(X_{i^-}; X_{i^+})\colon\, \sum_{m=1}^{M} p_m = 1 \right\}. \end{aligned}$$
- Minimum branch complexity profile under a given coordinate ordering $b_{i,j}(C)$
$$\begin{aligned} &h'_j(C) - h_i(C) \\ &\le \{I(X_{j^-}; X_{i^+})\colon\, p_m = 1/M, \forall m\} \\ &\le b_{i,j}(C) \le \max_{p_1,p_2,\cdots,p_M} \left\{ I(X_{j^-}; X_{i^+})\colon\, \sum_{m=1}^{M} p_m = 1 \right\}. \end{aligned}$$
- Minimum achievable state complexity profile under any coordinate ordering $s_i(\Pi C)$
$$s_i(\Pi C) \ge h'_i(C) - h_i(C).$$
The minimum value of
$$\max_{p_1 p_2,\cdots,p_M} \{I(X_{i^-}; X_{i^+})\colon\, \sum_{m=1}^{M} p_m = 1\}$$
over all coordinate orderings is the (least) upper bound on $s_i(\Pi C)$.
- Minimum achievable branch complexity profile under any coordinate ordering $b_{i,j}(\Pi C)$
$$b_{i,j}(\Pi C) \ge h'_j(C) - h_i(C).$$
The minimum value of
$$\max_{p_1,p_2,\cdots,p_M} \{I(X_{j^-}; X_{i^+})\colon\, \sum_{m=1}^{M} p_m = 1\}$$
over all coordinate orderings upper bounds $b_{i,j}(\Pi C)$.

- State complexity

$$s(C) \geq \max_{L,\{l_j\}:\, l_1+l_2+\cdots+l_L=n} \frac{\log_q M - \sum_{j=1}^{L} h_{l_j}(C)}{L-1} \geq \max_{i:\, 0\leq i\leq n} \{h_i'(C) - h_i(C)\}.$$

- Length-k branch complexity (any coordinate permutation):

$$b^k(C) \geq \max_{L,\{l_j\}:\, l_1+l_2+\cdots l_L=n-k(L-1)} \frac{\log_q M - \sum_{j=1}^{L} h_{l_j}(C)}{L-1} \geq \max_{i:\, 0\leq i\leq n-k} \{h_{i+k}'(C) - h_i(C)\}.$$

ACKNOWLEDGMENT

The authors wish to thank V. Sidorenko for his review in which he pointed out the application of the results of this paper to graph theory as presented in Section IV. This viewpoint enhanced the exposition of the results. The authors are also indebted to him for simplifying the proof of Theorem 2 and for providing copies of [22] and [23]. His constructive comments and those of the other anonymous reviewer improved the generality and the presentation of this paper.

REFERENCES

[1] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[2] Y. Berger and Y. Be'ery, "The twisted squaring construction, trellis complexity, and generalized weights of BCH and QR codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1817–1827, Nov. 1996.

[3] M. R. Best, "Binary codes with a minimum distance of four," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 738–742, Nov. 1980.

[4] G. D. Forney, Jr., "Coset codes—Part II: Binary lattices and related codes)," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, Sept. 1988.

[5] ——, "Dimension/length profiles and trellis complexity of linear block codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1741–1752, Nov. 1994.

[6] G. D. Forney, Jr., and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams, and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New-York: Wiley, 1968.

[8] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: Freeman, 1979.

[9] M. Hall, Jr., *Combinatorial Theory*. Waltham, MA: Blaisdell, 1967.

[10] D. Julin, "Two improved block codes," *IEEE Trans. Inform. Theory*, vol. IT-11, p. 459, July 1965.

[11] T. Kasami, T. Takata, T. Fujiwara, and S. Lin, "On the optimum bit orders with respect to the state complexity of trellis diagrams for binary linear codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 242–245, Jan. 1993.

[12] ——, "On complexity of trellis structure of linear block codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1057–1064, May 1993.

[13] A. D. Kot and C. Leung, "On the construction and dimensionality of linear block code trellises," in *Proc. 1993 IEEE Int. Symp. on Information Theory* (San Antonio, TX, Jan. 17–22, 1993), p. 291.

[14] F. R. Kschischang, "The trellis structure of maximal fixed-cost codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1828–1838, Nov. 1996.

[15] F. R. Kschischang and V. Sorokine, "On the trellis structure of block codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1924–1937, Nov. 1995.

[16] A. Lafourcade and A. Vardy, "Lower bounds on trellis complexity of block codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1938–1954, Nov. 1995.

[17] C.-C. Lu and S.-H. Huang, "On bit-level trellis of Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2061–2064, Nov. 1995.

[18] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. New York: North-Holland, 1977.

[19] J. L. Massey, "Foundation and methods of channel encoding," in *Proc. Int. Conf. on Information Theory and Systems, NTG-Fachberichte* (Berlin, Germany, Sept. 1978), vol. 65, pp. 148–157.

[20] R. J. McEliece, "On the BCJR trellis for linear block codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1072–1092, July 1996.

[21] D. J. Muder, "Minimal trellises for block codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1049–1053, Sept. 1988.

[22] V. R. Sidorenko, "The Euler characteristic of the minimal code trellis is maximum," *Probl. Inform. Transm.*, vol. 33, no. 1, pp. 72–77, Mar. 1997.

[23] V. Sidorenko, I. Martin, and B. Honary, "On separability of nonlinear block codes," preprint.

[24] A. Vardy, "The Nordstrom–Robinson code: Representation over GF(4) and efficient decoding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1686–1693, Sept. 1994.

[25] A. Vardy and Y. Be'ery, "Maximum-likelihood soft decision decoding of BCH codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 546–554, Mar. 1994.

[26] A. Vardy and F. R. Kschischang, "Proof of a conjecture of McEliece regarding the expansion index of the minimal trellis," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2027–2034, Nov. 1996.

[27] D. B. West, *Introduction to Graph Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[28] J. K. Wolf, "Efficient maximum likelihood decoding of linear block codes using a trellis," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 76–80, Jan. 1978.

[29] V. V. Zyablov and V. R. Sidorenko, "Bounds on complexity of trellis decoding of linear block codes," *Probl. Pered. Inform.*, vol. 29, pp. 3–9, 1993.