

Guessing Random Additive Noise Decoding of Network Coded Data Transmitted over Burst Error Channels

Ioannis Chatzigeorgiou, *Senior Member, IEEE*, and Dmitry Savostyanov

Abstract—We consider a transmitter that encodes data packets using network coding and broadcasts coded packets. A receiver employing network decoding recovers the data packets if a sufficient number of error-free coded packets are gathered. The receiver does not abandon its efforts to recover the data packets if network decoding is unsuccessful; instead, it employs syndrome decoding (SD) in an effort to repair erroneous received coded packets, and then reattempts network decoding. Most decoding techniques, including SD, assume that errors are independently and identically distributed within received coded packets. Motivated by the guessing random additive noise decoding (GRAND) framework, we propose transversal GRAND (T-GRAND): an algorithm that exploits statistical dependence in the occurrence of errors, complements network decoding and recovers all data packets with a higher probability than SD. T-GRAND examines error vectors in order of their likelihood of occurring and altering the transmitted packets. Calculation and sorting of the likelihood values of all error vectors is a simple but computationally expensive process. To reduce the complexity of T-GRAND, we take advantage of the properties of the likelihood function and develop an efficient method, which identifies the most likely error vectors without computing and ordering all likelihood values.

Index Terms—Network coding, random linear coding, burst noise, network decoding, syndrome decoding, guessing random additive noise decoding (GRAND).

I. INTRODUCTION

A linear code at the physical layer (PHY) maps a sequence of K symbols onto a sequence of $N > K$ symbols, known as a *codeword*. Whereas bit-level linear codes offer forward error correction at PHY, packet-level linear codes improve the reliability of communication systems at layers above PHY. Once data have been organized into strings of symbols, called *packets*, a packet-level linear code maps a sequence of K data packets onto a sequence of $N > K$ coded packets. When packet-level *random* linear codes (RLCs) [1] are used, the N coded packets are random linear combinations of the K data packets. As a result, packet-level random linear coding over the binary field is equivalent to successive column-wise bit-level random linear coding, which means that – if the N coded

packets are stacked to form the rows of a matrix – the N elements in each column of the matrix will compose a codeword.

Packet-level RLCs, henceforth referred to as RLCs, encompass both random linear *network* codes [1] and random linear *fountain* codes [2]; in the former case, packets are combined at intermediate nodes of a multi-hop network; in the latter case, packets are combined at the input of a single-hop broadcast channel. Whereas automatic repeat request (ARQ) schemes rely extensively on delay-inducing feedback from receivers to transmitters to support reliable multicast communication, RLCs can increase throughput – as they require limited or no feedback – and can be combined with queueing methods to improve network congestion control [3]. RLC implementations are integral parts of proposed solutions that use the satellite segment to extend the coverage of terrestrial infrastructures and support passengers on vehicles, high-speed trains and aircraft; examples include the multipath transmission control protocol based on network coding (MPTCP/NC) [4], fountain code-based MPTCP (FMTCP) [5] and path-based network coding (PBNC) [6]. Furthermore, RLC-based strategies for cooperative coded caching at edge nodes of vehicular networks, e.g., roadside units, can facilitate the flexible and reliable delivery of popular content to connected vehicles [7].

Receivers that employ RLC decoding typically discard coded packets that have been corrupted by errors and attempt to reconstruct data packets from correctly received coded packets only. However, when RLC decoding is unsuccessful, correctly received coded packets can be used in conjunction with erroneous received coded packets. Packetized rateless algebraic consistency (PRAC) [8] is an error correction method that exploits the algebraic properties of RLCs to identify codewords that contain errors and iteratively search for valid codewords to replace erroneous codewords. Given the aforementioned row/column correspondence of coded packets and codewords in RLCs, correcting errors in codewords is equivalent to repairing segments in coded packets. Syndrome decoding [9] looks at the same problem from a different perspective; if errors are detected in a codeword, syndrome decoding searches for the most likely error vector that corrupted a codeword instead of looking for the most likely valid codeword. Although PRAC and syndrome decoding can improve the chances of an RLC decoder recovering all data packets, they do not account for correlations between errors, i.e., they are more suitable for memoryless channels, like the binary symmetric channel (BSC). To address this operational limitation, we look into the recently proposed guessing random additive noise decoding (GRAND) scheme.

GRAND is a maximum-likelihood decoding scheme that endeavors to identify and nullify the effect of noise on

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Part of this paper was presented at the 2022 IEEE International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT50566.2022.9834692]. Dmitry Savostyanov acknowledges the support of the Leverhulme Trust Fellowship programme (RF-2021-258). (*Corresponding author: Ioannis Chatzigeorgiou.*)

Ioannis Chatzigeorgiou is with the School of Computing and Communications, Lancaster University, LA1 4WA, United Kingdom (email: i.chatzigeorgiou@lancaster.ac.uk).

Dmitry Savostyanov is with the School of Mathematics, Statistics and Actuarial Science, University of Essex, CO4 3SQ, United Kingdom (email: d.savostyanov@essex.ac.uk).

transmitted signals [10], [11]. In low-noise entropy conditions, guessing the noise is notably faster than performing an exhaustive search through all possible codewords. GRAND has enabled *universal* decoding, as it can decode *any* linear code. Since its inception, GRAND has been extended to non-linear codes [12], hard-input iterative decoding [13], soft detection decoding [14] and soft-input soft-output decoding [15]. It has also been combined with high-order modulation techniques [16], [17], modified for fading channels [17]–[19] and implemented in hardware [20].

The extension of GRAND to channels with memory was envisioned in [11]. Soon after, GRAND Markov order (GRAND-MO) [16] was developed for guessing errors that are correlated over time. GRAND-MO focuses on coding at PHY, according to which binary codewords are transmitted over a burst error channel and, consequently, correlated errors are distributed over the bits of each codeword. By contrast, if binary RLC encoding is used to generate and transmit coded packets over a burst error channel, correlated errors will be distributed over the bits of each coded packet. Remembering that codewords in packet-level RLCs occupy the columns of a matrix formed by the transmitted coded packets as its rows, we deduce that errors *across* adjacent codewords will be correlated but errors *within* each codeword will be uncorrelated. Therefore, GRAND-MO cannot complement RLC decoding at layers above PHY.

Through this literature review, we established that currently available solutions based on GRAND, for channels with or without memory, are specifically designed for PHY. When packet-level RLC is used at layers higher than PHY, syndrome decoding [9] is the only available solution that is aligned with the principles of GRAND and complements RLC decoding, provided that the underlying channel is memoryless and, hence, received errors are uncorrelated. The objective of this paper is to propose *transversal GRAND* – a novel algorithm designed to leverage correlations in errors *across* adjacent codewords, and to work in tandem with RLC decoding at layers higher than PHY. Correlations in errors could be introduced by the communication channel and find their way to layers higher than PHY if stringent latency requirements prevent the use of interleaving [16], or introduced by the channel decoder at the PHY layer – an event known as *decoder error propagation* [21].

The basic idea behind transversal GRAND was introduced in our previous work [22], which considered single-hop transmission over a channel with memory. The contributions of this paper are as follows:

- The algorithmic details of transversal GRAND are described extensively and examples are provided to illustrate how knowledge of the statistical description of the burst error process can improve the chances of erroneous received coded packets being repaired.
- The computational complexity of a core component of transversal GRAND is studied thoroughly and its efficiency is discussed.
- A novel approach of adjustable computational complexity is developed and contrasted with the original implementation of the core component of transversal GRAND.

- Simulations of coded packet transmissions over a burst error channel verify and quantify the advantages of transversal GRAND over syndrome decoding. Numerical results also demonstrate that the computational complexity of transversal GRAND can be reduced without negatively impacting its performance.

The remainder of this paper has been organized as follows: Section II describes RLC decoding, which discards packets received in error. Section III explains how syndrome decoding can repair packets in error and assist RLC decoding. Transversal GRAND is introduced in Section IV as a means to exploit the statistical structure of errors and improve the chances of repairing packets. The implementation of a core procedure of transversal GRAND, which is responsible for sorting error vectors in order of likelihood, is discussed in Section V and revised in Section VI to reduce its complexity. Section VII dwells deeper into computational aspects of the proposed sorting procedures. Performance comparisons between syndrome decoding and transversal GRAND are presented in Section VIII and key findings are summarized in Section IX.

II. PACKET-LEVEL RANDOM LINEAR CODING AND DECODING

We consider a source node, which is about to broadcast messages to one or more destination nodes. Before transmission, each message is segmented into K source packets. Each source packet has been modeled as a sequence of B bits. The K source packets of B bits can be expressed as a matrix $\mathbf{U} \in \mathbb{F}_2^{K \times B}$, where $\mathbb{F}_2^{K \times B}$ denotes the set of all $K \times B$ matrices over $\mathbb{F}_2 = \{0, 1\}$. Random linear network coding [1] is used to encode the K source packets into $N \geq K$ coded packets. The N coded packets can also be expressed in matrix form as $\mathbf{X} \in \mathbb{F}_2^{N \times B}$. The relationship between matrices \mathbf{X} and \mathbf{U} is:

$$\mathbf{X} = \mathbf{G} \mathbf{U}, \quad (1)$$

where \mathbf{G} is known as the $N \times K$ *generator matrix* of the RLC. We employ *systematic* RLC encoding, according to which the first K of the N transmitted packets are identical to the K source packets, while the remaining $N - K$ coded packets are random linear combinations of the source packets. The generator matrix \mathbf{G} can thus be expressed as:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_K \\ \mathbf{P} \end{bmatrix}, \quad (2)$$

where \mathbf{I}_K is the $K \times K$ identity matrix. Each element of the $(N - K) \times K$ matrix \mathbf{P} is chosen uniformly and at random from \mathbb{F}_2 . The value of the seed that initializes the pseudo-random number generator and contributes to the construction of \mathbf{P} at the source node can be conveyed in the headers of the coded packets, as explained in [23]. For this reason, we assume that destination nodes have knowledge of \mathbf{P} .

Let \mathbf{Y} be an erroneous copy of \mathbf{X} that has been received by a destination node. The destination node classifies the received coded packets into error-free and erroneous coded packets, e.g., using cyclic redundancy checks (CRCs). We denote by \mathcal{R} the set of the row indices of \mathbf{Y} that correspond to correctly

received coded packets, and by N_R the number of coded packets that contain no errors, i.e., $|\mathcal{R}| = N_R \leq N$. The remaining $N - N_R$ received coded packets, which have been corrupted by errors, have indices in the set $\bar{\mathcal{R}} = \{1, \dots, N\} \setminus \mathcal{R}$. The destination node constructs matrices $\mathbf{Y}_{\mathcal{R}}$ and $\mathbf{G}_{\mathcal{R}}$ from those N_R rows of \mathbf{Y} and \mathbf{G} , respectively, with indices in \mathcal{R} . On the other hand, the indices in $\bar{\mathcal{R}}$ identify the $N - N_R$ rows of \mathbf{Y} that will be used in the construction of matrix $\mathbf{Y}_{\bar{\mathcal{R}}}$. Matrices \mathbf{Y} , $\mathbf{Y}_{\mathcal{R}}$ and $\mathbf{Y}_{\bar{\mathcal{R}}}$ can be expressed as follows:

$$\mathbf{Y} = \mathbf{X} \oplus \mathbf{E} \Leftrightarrow \begin{cases} \mathbf{Y}_{\mathcal{R}} = \mathbf{X}_{\mathcal{R}} = \mathbf{G}_{\mathcal{R}} \mathbf{U}, & (3a) \\ \mathbf{Y}_{\bar{\mathcal{R}}} = \mathbf{X}_{\bar{\mathcal{R}}} \oplus \mathbf{E}_{\bar{\mathcal{R}}}, & (3b) \end{cases}$$

where \oplus represents modulo-2 addition. The *error matrix* \mathbf{E} contains ones in positions where errors have occurred and zeros in the remaining positions. The indices in sets \mathcal{R} and $\bar{\mathcal{R}}$ can be used to decompose \mathbf{E} into $\mathbf{E}_{\mathcal{R}}$ and $\mathbf{E}_{\bar{\mathcal{R}}}$, where $\mathbf{E}_{\mathcal{R}}$ is equal to the $N_R \times B$ zero matrix.

Decoding of an RLC utilizes the N_R correctly received coded packets, represented by $\mathbf{Y}_{\mathcal{R}}$, and discards the $N - N_R$ erroneous received coded packets contained in $\mathbf{Y}_{\bar{\mathcal{R}}}$. The source message, represented by \mathbf{U} , can be obtained from $\mathbf{X}_{\mathcal{R}} = \mathbf{G}_{\mathcal{R}} \mathbf{U}$ in (3a) if the rank of $\mathbf{G}_{\mathcal{R}}$ is K . In that case, the relationship $\mathbf{X}_{\mathcal{R}} = \mathbf{G}_{\mathcal{R}} \mathbf{U}$ can be seen as a system of $N_R \geq K$ linear equations, which can be reduced to a system of K linearly independent equations with K unknowns, i.e., source packets. This $K \times K$ system will return a unique solution for \mathbf{U} . Otherwise, if $\text{rank}(\mathbf{G}_{\mathcal{R}}) < K$, the linear system is underdetermined, i.e., it has fewer linearly independent equations than unknowns, therefore a unique solution cannot be obtained.

If the RLC decoder does not discard the erroneous coded packets but attempts to repair them, the probability of recovering the source message could potentially be increased. The following section describes syndrome decoding [9], which aims to repair erroneous coded packets without taking into consideration the statistical properties of the channel.

III. GUESSING UNCORRELATED ERRORS: SYNDROME DECODING

We established in (3b) that the $N - N_R$ erroneous coded packets, which compose $\mathbf{Y}_{\bar{\mathcal{R}}}$, can be expressed as the modulo-2 sum of the corresponding transmitted coded packets, which form the rows of $\mathbf{X}_{\bar{\mathcal{R}}}$, and matrix $\mathbf{E}_{\bar{\mathcal{R}}}$. If an estimate of $\mathbf{E}_{\bar{\mathcal{R}}}$, denoted by $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$, is computed, an estimate of the transmitted coded packets can be derived as follows:

$$\hat{\mathbf{X}}_{\bar{\mathcal{R}}} = \mathbf{Y}_{\bar{\mathcal{R}}} \oplus \hat{\mathbf{E}}_{\bar{\mathcal{R}}}. \quad (4)$$

CRC verification will determine which of the estimated rows of $\hat{\mathbf{X}}_{\bar{\mathcal{R}}}$ correspond to successfully repaired coded packets. Let ν denote the number of rows in $\hat{\mathbf{X}}_{\bar{\mathcal{R}}}$ that passed CRC verification. The indices of the ν successfully repaired coded packets will be removed from set $\bar{\mathcal{R}}$ and will be added to set \mathcal{R} , while the corresponding rows of $\hat{\mathbf{X}}_{\bar{\mathcal{R}}}$ will be moved to $\mathbf{Y}_{\mathcal{R}}$ in (3a). The cardinalities of sets $\bar{\mathcal{R}}$ and \mathcal{R} will change to $N - N_R - \nu$ and $N_R + \nu$, respectively, while the dimensions of $\hat{\mathbf{X}}_{\bar{\mathcal{R}}}$ and $\mathbf{Y}_{\mathcal{R}}$ will change to $(N - N_R - \nu) \times B$ and $(N_R + \nu) \times B$, respectively. The indices of the ν repaired

packets will also be used to identify the rows of the generator matrix \mathbf{G} that should be appended to $\mathbf{G}_{\mathcal{R}}$ in (3a). If the ν repaired packets increase the rank of the enlarged $(N_R + \nu) \times K$ matrix $\mathbf{G}_{\mathcal{R}}$ to K , the process of estimating, or ‘guessing’, matrix $\mathbf{E}_{\bar{\mathcal{R}}}$ will have been successful in assisting the RLC decoder to recover the source message.

Syndrome decoding is a method proposed in [9] for the calculation of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. This method takes into account that the $(N - K) \times K$ matrix \mathbf{P} is known to all destination nodes, and derives the $N \times (N - K)$ *parity-check matrix* \mathbf{H} as follows:

$$\mathbf{H} = [-\mathbf{P} \mid \mathbf{I}_{N-K}]^T, \quad (5)$$

so that:

$$\mathbf{H}^T \mathbf{G} = \mathbf{0}. \quad (6)$$

When operations are performed in \mathbb{F}_2 , negation has no effect on a matrix, that is, $-\mathbf{P} = \mathbf{P}$. Multiplication of \mathbf{H}^T by the received matrix \mathbf{Y} produces the $(N - K) \times B$ *syndrome matrix* \mathbf{S} , i.e., $\mathbf{S} = \mathbf{H}^T \mathbf{Y}$. Using (1), (3) and (6), we find that the relationship between the syndrome matrix \mathbf{S} and the error matrix \mathbf{E} is:

$$\mathbf{S} = \mathbf{H}^T \mathbf{Y} = \mathbf{H}^T (\mathbf{X} \oplus \mathbf{E}) = \mathbf{H}^T (\mathbf{G} \mathbf{U} \oplus \mathbf{E}) = \mathbf{H}^T \mathbf{E}. \quad (7)$$

As our focus is on the $N - N_R$ erroneous coded packets only, we use the set $\bar{\mathcal{R}}$ to isolate the $N - N_R$ of the N rows of \mathbf{H} and \mathbf{E} , and reduce the two matrices to $\mathbf{H}_{\bar{\mathcal{R}}}$ and $\mathbf{E}_{\bar{\mathcal{R}}}$, respectively. Consequently, expression (7) changes to:

$$\mathbf{S} = (\mathbf{H}_{\bar{\mathcal{R}}})^T \mathbf{E}_{\bar{\mathcal{R}}}. \quad (8)$$

If column b of \mathbf{S} and $\mathbf{E}_{\bar{\mathcal{R}}}$ is denoted by $[\mathbf{S}]_{*,b}$ and $[\mathbf{E}_{\bar{\mathcal{R}}}]_{*,b}$, respectively, expression (8) can be re-written as B independent systems of $N - K$ linear equations with $N - N_R$ unknowns per equation:

$$[\mathbf{S}]_{*,b} = (\mathbf{H}_{\bar{\mathcal{R}}})^T [\mathbf{E}_{\bar{\mathcal{R}}}]_{*,b} \quad \text{for } b = 1, \dots, B. \quad (9)$$

Mohammadi *et al.* [9] observed that erroneous received coded packets usually contain large error-free segments, thus $\mathbf{E}_{\bar{\mathcal{R}}}$ is a *sparse* matrix, that is, most of the elements in $\mathbf{E}_{\bar{\mathcal{R}}}$ are zero-valued. Based on this observation, the solution to (9) can be formulated as:

$$[\hat{\mathbf{E}}_{\bar{\mathcal{R}}}]_{*,b} = \arg \min_{\mathbf{w}^T} \|\mathbf{w}\|_0 \quad (10a)$$

$$\text{subject to } (\mathbf{H}_{\bar{\mathcal{R}}})^T \mathbf{w}^T = [\mathbf{S}]_{*,b} \quad (10b)$$

where $\mathbf{w} \in \mathbb{F}_2^{N - N_R}$ is a row vector that satisfies constraint (10b) and has the minimum possible number of non-zero elements. The norm $\|\mathbf{w}\|_0$, which counts the non-zero elements in \mathbf{w} , is defined as $\|\mathbf{w}\|_0 = |w_1|^0 + \dots + |w_{N - N_R}|^0$ assuming that $0^0 = 0$ [24]. Syndrome decoding considers (10a) and initiates an exhaustive search for a solution; the sparsity of the row vector \mathbf{w} is gradually reduced, i.e., the number of ones in \mathbf{w} increases, and the search concludes when the sparsest vector \mathbf{w} that satisfies (10b) has been identified.

Although syndrome decoding has the potential to assist RLC decoding and increase the probability of recovering the source packets [9], it does not consider the statistical properties of the channel. It selects columns for $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ that are as sparse as

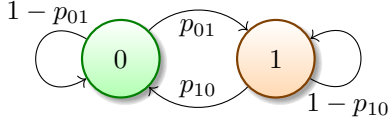


Fig. 1. The simplified Gilbert-Elliott channel model, where 0 and 1 represent the ‘good’ state and the ‘bad’ state, respectively.

possible or, equivalently, have the lowest possible Hamming weight, but it does not account for the possibility that the columns of $\hat{\mathbf{E}}_{\mathcal{R}}$ are correlated. Thus, syndrome decoding is suitable for memoryless channels only, such as the BSC. The following section describes a simplified version of the Gilbert-Elliott channel model, which is commonly used for the characterization of burst errors in channels with memory, and introduces transversal GRAND, which exploits the statistical properties of burst error channels.

IV. GUESSING BURST ERRORS: TRANSVERSAL GRAND

Coded packets at the input of the RLC decoder often contain bit errors, the distribution of which is governed by the channel type, multi-user interference, the modulation method and the forward error correction scheme at the physical layer. The wireless transmission medium and the physical layer at the source and destination nodes form a composite channel, which introduces time-dependent correlated error bursts. Gilbert [25] modeled burst error channels, and Elliott [26] proposed the more general Gilbert-Elliott channel model. The analysis of RLC-based schemes over channels with memory frequently relies on the Gilbert-Elliott channel to model transmission over wireless channels, e.g., [27]–[29].

A. The Simplified Gilbert-Elliott Channel Model

The Gilbert-Elliott channel model is a two-state Markov chain. The channel state can be in a ‘bad’ state or a ‘good’ state, and can either transition from one state to the other or remain in the same state. In the bad state, a bit error is generated with a certain high probability, while in the good state a bit error is generated with a low probability. A simplification of the model, which is often considered in the literature, e.g., [16], [30], [31], assumes that the bad state always induces a bit error, while the good state never flips the value of a bit. Therefore, the simplified channel model can be reduced to the Markov chain shown in Fig. 1, where the bad state is denoted by 1 and the good state is denoted by 0. A transition from state 0 to state 1 occurs with probability p_{01} and causes a transmitted bit to be received in error. A transition from state 1 to state 0, which occurs with probability p_{10} , implies that a transmitted bit has been received correctly. The channel may remain in state 0 or state 1 with probability $1 - p_{01}$ or $1 - p_{10}$, respectively.

The steady-state probability of being in state 1 represents the bit error probability, which is given by $\varepsilon = p_{01}/(p_{01} + p_{10})$ [27], [32]. The expected number of consecutive errors provides the average length of an error burst, and is equal to $\Lambda = 1/p_{10}$ [27]. The memory of a burst error channel is defined as $\mu = 1 - p_{10} - p_{01}$ [33] and provides a measure of the persistence

Algorithm 1: Implementation of the transversal GRAND algorithm according to which generation of the sequence of error vectors relies on the calculation and sorting of the probabilities of occurrence of the vectors in descending order of magnitude.

Inputs: $\mathbf{H}_{\mathcal{R}}, \mathbf{S}, p_{01}, p_{10}$

Output: $\hat{\mathbf{E}}_{\mathcal{R}}$

```

1 origin_vector  $\leftarrow \mathbf{0}$ 
2 for  $b \leftarrow 1$  to  $B$  do
3   Determine  $L_0$  and  $L_1$  from origin_vector
   // origin_vector changes with  $b$ 
4    $\{(\ell_0^{(\ell)}, \ell_1^{(\ell)})\}_{\ell=1}^{(L_0+1)(L_1+1)} \leftarrow$ 
     CalcProbAndSort( $p_{01}, p_{10}, L_0, L_1$ )
5   for  $\ell \leftarrow 1$  to  $(L_0+1)(L_1+1)$  do
6     err_seq  $\leftarrow$  GenErrSeq( $\ell_0^{(\ell)}, \ell_1^{(\ell)}, L_0, L_1$ )
7     forall error vectors in err_seq do
8        $\mathbf{w}^T \leftarrow$  error vector from err_seq
9       if  $(\mathbf{H}_{\mathcal{R}})^T \mathbf{w}^T = [\mathbf{S}]_{*,b}$  then
10         $[\hat{\mathbf{E}}_{\mathcal{R}}]_{*,b} \leftarrow \mathbf{w}^T$ 
11        origin_vector  $\leftarrow [\hat{\mathbf{E}}_{\mathcal{R}}]_{*,b}$ 
12        goto 16
13      end
14    end
15  end
16 end
17 return  $\hat{\mathbf{E}}_{\mathcal{R}}$ 

```

of the channel in remaining in a state. If $-1 \leq \mu < 0$, the channel has oscillatory memory. Otherwise, if $0 < \mu < 1$, the channel has persistent memory. For $\mu = 0$, the channel is memoryless.

Estimation of the bit error probability ε and the average burst length Λ facilitates computation of the transition probabilities p_{01} and p_{10} , as the latter can be expressed in terms of the former, that is, $p_{01} = \varepsilon/[\Lambda(1 - \varepsilon)]$ and $p_{10} = 1/\Lambda$. The values of p_{01} and p_{10} can then be used as inputs to our proposed algorithm and guide the derivation of matrix $\hat{\mathbf{E}}_{\mathcal{R}}$.

B. The Transversal GRAND Algorithm

Let us refer to a prospective column of $\hat{\mathbf{E}}_{\mathcal{R}}$ as an *error vector* of length $L = N - N_{\mathcal{R}}$. In syndrome decoding, error vectors that have the same number of non-zero elements, i.e., the same Hamming weight, are treated as being equally likely and can thus be queried in any order by (10). Furthermore, all error vectors of a given Hamming weight w are queried before error vectors of weight $w + 1$. The proposed method, which we call transversal GRAND, queries error vectors in descending order of likelihood, as dictated by the transition probabilities of the channel model and not by the Hamming weights of the error vectors.

As explained in Algorithm 1, transversal GRAND uses the reduced parity-check matrix $\mathbf{H}_{\mathcal{R}}$, the syndrome matrix \mathbf{S} and the transition probabilities p_{01} and p_{10} of the channel model to compute $\hat{\mathbf{E}}_{\mathcal{R}}$. Calculation of column b of $\hat{\mathbf{E}}_{\mathcal{R}}$ starts with the definition of an ‘origin’ vector of length L , which is set

equal to column $b-1$ of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ when column b is being derived, for $b = 1, \dots, B$. An all-zero origin vector is considered in the calculation of the first column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ (see line 1 of Algorithm 1). Essentially, the origin vector contains the current states of L independent Markov chains, i.e., one Markov chain for each row of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. Initializing the origin vector to the all-zero vector for $b = 1$ is equivalent to setting the initial state of each of the L chains to 0. In general, for any $b \in \{1, B\}$, the origin vector will contain L_0 zeros and L_1 ones, that is, L_0 of the chains will be in state 0 and the remaining L_1 chains will be in state 1, where $L_0 + L_1 = L$ (see line 3 of Algorithm 1).

According to the channel model, the values of ℓ_i of the L_i elements of the origin vector will change from i to j with probability p_{ij} , for $i, j \in \{0, 1\}$, $i \neq j$ and $\ell_i = 0, \dots, L_i$. The values of the remaining $L_i - \ell_i$ elements will stay the same with probability $1 - p_{ij}$. The overall probability of occurrence of the aforementioned transitions is:

$$f(\ell_0, \ell_1) = p_{01}^{\ell_0} (1 - p_{01})^{L_0 - \ell_0} p_{10}^{\ell_1} (1 - p_{10})^{L_1 - \ell_1}. \quad (11)$$

The probability of occurrence $f(\ell_0, \ell_1)$ is calculated for $\ell_0 = 0, \dots, L_0$ and $\ell_1 = 0, \dots, L_1$, and the $(L_0 + 1)(L_1 + 1)$ probability values compose the entries of the $(L_0 + 1) \times (L_1 + 1)$ matrix $\mathbf{F} = [f(\ell_0, \ell_1)]$. The entries of matrix \mathbf{F} are then sorted in descending order. The coordinates of the sorted entries constitute the set $\{(\ell_0^{(\ell)}, \ell_1^{(\ell)})\}_{\ell=1}^{(L_0+1)(L_1+1)}$, as shown in line 4 of Algorithm 1, where index ℓ indicates the position of the probability value in the sequence of sorted entries:

$$\begin{aligned} f(\ell_0^{(1)}, \ell_1^{(1)}) &\geq f(\ell_0^{(2)}, \ell_1^{(2)}) \geq \dots \\ &\dots \geq f(\ell_0^{((L_0+1)(L_1+1))}, \ell_1^{((L_0+1)(L_1+1))}). \end{aligned} \quad (12)$$

For each coordinate pair $(\ell_0^{(\ell)}, \ell_1^{(\ell)})$, the algorithm generates a sequence of $\binom{L_0}{\ell_0^{(\ell)}} \binom{L_1}{\ell_1^{(\ell)}}$ equally likely error vectors (see line 6 of Algorithm 1). The algorithm sequentially considers each error vector in the sequence, assigns it to \mathbf{w}^\top and checks whether \mathbf{w}^\top is a solution to the linear equation $(\mathbf{H}_{\bar{\mathcal{R}}})^\top \mathbf{w}^\top = [\mathbf{S}]_{*,b}$. If \mathbf{w}^\top is indeed a solution, then it is selected to be column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. Otherwise, the algorithm continues to generate and query error vectors in descending order of likelihood, until a solution is found (see lines 5–15 of Algorithm 1). As expected, the $(L_0 + 1)(L_1 + 1)$ probability values are mapped onto a total of

$$\begin{aligned} \sum_{\ell_0=0}^{L_0} \sum_{\ell_1=0}^{L_1} \binom{L_0}{\ell_0} \binom{L_1}{\ell_1} &= \sum_{\ell_0=0}^{L_0} \binom{L_0}{\ell_0} \sum_{\ell_1=0}^{L_1} \binom{L_1}{\ell_1} \\ &= 2^{L_0} 2^{L_1} = 2^{L_0+L_1} = 2^L \end{aligned}$$

error vectors. Transversal GRAND terminates when solutions for all B linear equations have been estimated and $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ has been obtained (see line 17 of Algorithm 1).

Fig. 2 depicts an example of how transversal GRAND derives column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ from column $b-1$, when $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ consists of $L = 5$ rows and the transition probabilities of the channel model are $p_{01} = 0.1$ and $p_{10} = 0.4$. Notice in Fig. 2 that column $b-1$ of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$, which is represented by the origin vector in Algorithm 1, contains $L_0 = 2$ zeros and $L_1 = 3$ ones. Expression (11) is used to calculate the probability of an error vector occurring after column $b-1$

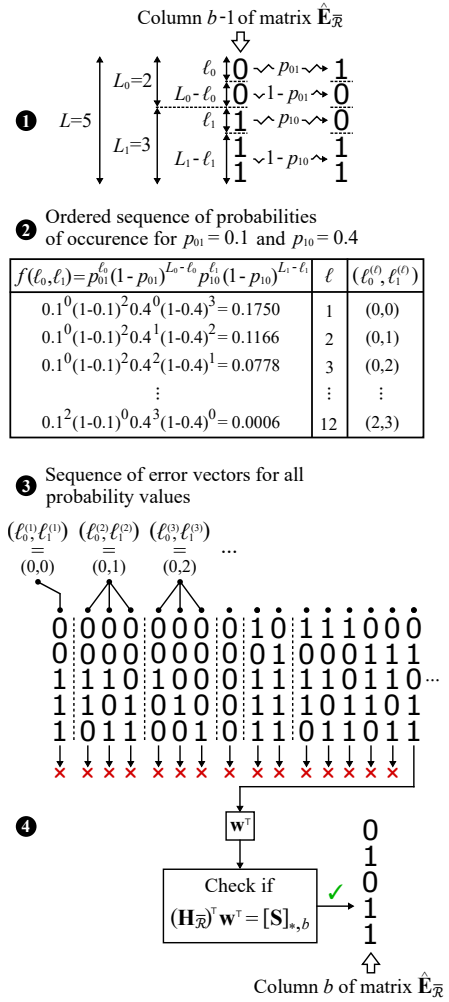


Fig. 2. Example of the derivation of column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ using Algorithm 1 when the previous column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ contains $L_0 = 2$ zeros and $L_1 = 3$ ones. The transition probabilities are set to $p_{01} = 0.1$ and $p_{10} = 0.4$.

of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ has been observed. Of the $L_0 = 2$ zeros in column $b-1$ of $\hat{\mathbf{E}}$, ℓ_0 zeros could change to 1 with probability p_{01} , while the remaining $L_0 - \ell_0$ zeros would not change value with probability $1 - p_{01}$, for $\ell_0 = 0, 1, 2$. On the other hand, ℓ_1 of the $L_1 = 3$ ones could change to 0 with probability p_{10} and the remaining $L_1 - \ell_1$ ones would stay unchanged with probability $1 - p_{10}$, for $\ell_1 = 0, 1, 2, 3$. A total of $(L_0 + 1)(L_1 + 1) = 12$ probability values are computed, stored in the 3×4 matrix \mathbf{F} , arranged in descending order and assigned an ascending index ℓ . The values of ℓ_0 and ℓ_1 that result in the ℓ -th highest probability value are denoted as $(\ell_0^{(\ell)}, \ell_1^{(\ell)})$, as shown in Fig. 2. For instance, the ordered sequence of probabilities of occurrence in Fig. 2 conveys that the probability of $\ell_0 = 0$ zeros changing to 1 and $\ell_1 = 2$ ones changing to 0 is 0.0778, which is the third highest probability value; hence, $(\ell_0^{(3)}, \ell_1^{(3)}) = (0, 2)$. This means that, when $\ell = 3$, the algorithm generates and queries $\binom{2}{0} \binom{3}{2} = 3$ error vectors – each of which occurs with probability 0.0778 – for column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. Error vectors are created in order of likelihood, as dictated by the ordered sequence of probabilities of occurrence, and each vector is assigned to \mathbf{w}^\top . The first

error vector that satisfies $(\mathbf{H}_{\bar{\mathcal{R}}})^\top \mathbf{w}^\top = [\mathbf{S}]_{*,b}$ is selected to be column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$.

In contrast to syndrome decoding, the error vectors that are generated by transversal GRAND are not necessarily of increasing Hamming weight. Furthermore, transversal GRAND does not assign by default the same probability of occurrence to error vectors that have the same Hamming weight. For example, notice that the leading weight-3 error vector in the sequence of error vectors shown in Fig. 2 is not followed by error vectors of higher weights and is not clustered together with other weight-3 error vectors.

To conclude this section, we should note that transversal GRAND – as described by Algorithm 1 – solves an optimization problem analogous to (10) for syndrome decoding, which can be formulated as follows:

$$[\hat{\mathbf{E}}_{\bar{\mathcal{R}}}]_{*,b} = \arg \max_{\mathbf{w}^\top} f(\ell_0, \ell_1) \quad (13a)$$

$$\text{where } \ell_0 = \|\overline{[\hat{\mathbf{E}}_{\bar{\mathcal{R}}}]_{*,b-1}} \odot \mathbf{w}^\top\|_0 \quad (13b)$$

$$\ell_1 = \|\overline{[\hat{\mathbf{E}}_{\bar{\mathcal{R}}}]_{*,b-1}} \odot \bar{\mathbf{w}}^\top\|_0 \quad (13c)$$

$$\text{subject to } (\mathbf{H}_{\bar{\mathcal{R}}})^\top \mathbf{w}^\top = [\mathbf{S}]_{*,b} \quad (13d)$$

where \odot represents element-wise multiplication, known as the Hadamard product, and $\bar{\mathbf{v}}$ denotes the ones' complement of a binary vector \mathbf{v} . For $b = 1$, in accordance with Algorithm 1, we set $\ell_0 = \|\mathbf{w}\|_0$ and $\ell_1 = 0$. Transversal GRAND selects a vector \mathbf{w}^\top for column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ if it maximizes the probability of occurrence $f(\ell_0, \ell_1)$, as shown in (13a), and satisfies (13d); the values of ℓ_0 and ℓ_1 can be obtained from (13b) and (13c), respectively, using the already estimated column $b - 1$ of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ and the vector under consideration.

V. IMPLEMENTATION OF THE SORTING PROCEDURE

A core element of transversal GRAND is the calculation of the probability of occurrence of every possible error vector that could form column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$, given column $b - 1$ and the transition probabilities p_{01} and p_{10} . As explained in Section IV-B, the probability of occurrence $f(\ell_0, \ell_1)$ is calculated for $\ell_0 = 0, \dots, L_0$ and $\ell_1 = 0, \dots, L_1$, where the values of L_0 and L_1 are determined by column $b - 1$. The procedure of calculating and sorting the $(L_0 + 1)(L_1 + 1)$ probability values is called `CalcProbAndSort`, as can be seen in Line 4 of Algorithm 1. This section investigates the computational complexity of `CalcProbAndSort`.

Recall from Section IV-B that the $(L_0 + 1)(L_1 + 1)$ probability values compose the entries of matrix \mathbf{F} . Let $f_0(\ell_0) = p_{01}^{\ell_0} (1 - p_{01})^{L_0 - \ell_0}$ and $f_1(\ell_1) = p_{10}^{\ell_1} (1 - p_{10})^{L_1 - \ell_1}$. Matrix \mathbf{F} is a rank-one matrix, which is generated as a trace of function $f(\ell_0, \ell_1) = f_0(\ell_0)f_1(\ell_1)$ on the two-dimensional uniform tensor product grid $\{(\ell_0, \ell_1)\}$:

$$\mathbf{F} = \begin{bmatrix} f_0(0) \\ f_0(1) \\ \vdots \\ f_0(L_0) \end{bmatrix} [f_1(0) \quad f_1(1) \quad \cdots \quad f_1(L_1)]. \quad (14)$$

When the transition probabilities take the values $p_{01} = 0$, $p_{01} = 1$, $p_{10} = 0$ or $p_{10} = 1$, the singular case of $\mathbf{F} = \mathbf{0}$ is

observed. If $p_{01} = 0$ or $p_{10} = 0$, the channel model has one state only. If $p_{01} = 1$ or $p_{10} = 1$, the channel has oscillatory memory, that is, the probability of remaining in a state is lower than the steady-state probability of being in that state [33]. These extreme cases are not of interest in the context of this paper and are ignored. We assume that $0 < p_{01} < 1$ and $0 < p_{10} < 1$ to ensure that all entries of \mathbf{F} are positive, i.e., $f(\ell_0, \ell_1) > 0$.

Since $f(\ell_0, \ell_1)$ is a product of probabilities, it could take very small values that would lead to an arithmetic underflow. A well-known approach to this problem is to operate in the log-domain and convert the products of probabilities into sums of log-probabilities¹. Noting that

$$\begin{aligned} -\log f(\ell_0, \ell_1) &= -\ell_0 \log(p_{01}) - (L_0 - \ell_0) \log(1 - p_{01}) \\ &\quad - \ell_1 \log(p_{10}) - (L_1 - \ell_1) \log(1 - p_{10}) \\ &= \underbrace{\ell_0 \log\left(\frac{1 - p_{01}}{p_{01}}\right)}_{\varphi_0(\ell_0)} + \underbrace{\ell_1 \log\left(\frac{1 - p_{10}}{p_{10}}\right)}_{\varphi_1(\ell_1)} \\ &\quad - \underbrace{(L_0 \log(1 - p_{01}) + L_1 \log(1 - p_{10}))}_{\xi} \\ &= \underbrace{\varphi_0(\ell_0) + \varphi_1(\ell_1)}_{\varphi(\ell_0, \ell_1)} - \xi, \end{aligned} \quad (15)$$

we consider the matrix $\Phi = [\varphi(\ell_0, \ell_1)]$ with elements $\varphi(\ell_0, \ell_1) = \varphi_0(\ell_0) + \varphi_1(\ell_1)$ and of the same size as \mathbf{F} . The constant term ξ in (15) contributes equally to all log-probability values and can be ignored. The negative logarithm is a monotonically decreasing function, hence the largest element of \mathbf{F} corresponds to the smallest element of Φ . Consequently, instead of arranging the elements of \mathbf{F} in decreasing order, we can now sort the elements of Φ in increasing order.

The values of

$$\alpha_0 = \log\left(\frac{1 - p_{01}}{p_{01}}\right) \quad \text{and} \quad \alpha_1 = \log\left(\frac{1 - p_{10}}{p_{10}}\right) \quad (16)$$

can be computed and stored prior to the execution of the transversal GRAND algorithm. Calculation of $\varphi_0(\ell_0) = \alpha_0 \ell_0$ requires one multiplication. Alternatively, $\varphi_0(\ell_0)$ can be written in recursive form as $\varphi_0(\ell_0) = \varphi_0(\ell_0 - 1) + \alpha_0$ and hence, calculation of $\varphi_0(\ell_0)$ for $\ell_0 = 0, \dots, L_0$ involves L_0 additions. Using the same reasoning, we deduce that L_1 additions are required for the calculation of $\varphi_1(\ell_1)$ for $\ell_1 = 0, \dots, L_1$. Having obtained the $(L_0 + 1)$ values of $\varphi_0(\ell_0)$ and the $(L_1 + 1)$ values of $\varphi_1(\ell_1)$, the procedure continues with the evaluation of the $(L_0 + 1)(L_1 + 1)$ entries of matrix Φ . One sum is evaluated for every entry of matrix Φ , according to (15). Therefore, $(L_0 + 1)(L_1 + 1) + L_0 + L_1$ additions are needed for the computation of matrix Φ .

After `CalcProbAndSort` computes the entries $\varphi(\ell_0, \ell_1)$ and inserts them in matrix Φ , it is tasked with arranging them in increasing order. Available sorting algorithms achieve different tradeoffs between runtime, stability and memory

¹Although any base can be used for the log function, numerical examples in this paper use the logarithm to the base 2.

usage. Low overall computational complexity requires a short runtime, which translates into a small number of executed comparisons. If n is the number of elements to be sorted, the worst-case number of comparisons is $O(n \log n)$, e.g., using the Heapsort [34] algorithm. The fact that Heapsort is not a stable algorithm does not affect the sorting process, as the relative order of equal entries in matrix Φ does not have to be preserved. For this reason, Heapsort is adopted by `CalcProbAndSort` to sort the $n = (L_0 + 1)(L_1 + 1)$ elements of matrix Φ .

Recall that the output of `CalcProbAndSort` is not a sequence of sorted probability values but a sequence of coordinates $\{(\ell_0^{(\ell)}, \ell_1^{(\ell)})\}$, which point to the locations of the sorted probability values in matrix Φ . The following section examines the relationship between $\varphi(\ell_0, \ell_1)$ and the structure of matrix Φ in order to develop a procedure that generates the coordinates of the ordered probability values without the need to first calculate all of the probability values.

VI. IMPLEMENTATION OF A TRACING PROCEDURE

This section introduces `TraceSortedProb`, a procedure that traces negative log-probability values, from the smallest to the largest, without first calculating them and then sorting them, and outputs their coordinates. `TraceSortedProb` is proposed as an efficient alternative to `CalcProbAndSort`. The structural properties of matrix Φ that `TraceSortedProb` takes advantage of are described in detail and the complexity of the proposed procedure is contrasted with that of `CalcProbAndSort`.

A. Fundamentals of the Tracing Procedure

Our goal is to trace the smallest elements of $\Phi = [\varphi(\ell_0, \ell_1)]$, taking in account that $\varphi(\ell_0, \ell_1) = \varphi_0(\ell_0) + \varphi_1(\ell_1)$, where $\varphi_0(\ell_0) = \alpha_0 \ell_0$ and $\varphi_1(\ell_1) = \alpha_1 \ell_1$. The value of $\varphi(\ell_0, \ell_1)$ occupies the entry in row $\ell_0 + 1$ and column $\ell_1 + 1$ of matrix Φ . We first consider $\varphi_0(\ell_0)$ and observe that, according to (16), $\alpha_0 > 0$ for $p_{01} < \frac{1}{2}$, $\alpha_0 < 0$ for $p_{01} > \frac{1}{2}$, and $\alpha_0 = 0$ for $p_{01} = \frac{1}{2}$. The sequence $\{\varphi_0(\ell_0)\}$ for $\ell_0 = 0, \dots, L_0$ is monotonically increasing for $\alpha_0 > 0$, monotonically decreasing for $\alpha_0 < 0$, and constant for $\alpha_0 = 0$. The same conclusions can be drawn for the sequence $\{\varphi_1(\ell_1)\}$ for $\ell_1 = 0, \dots, L_1$.

1) *Identifying the smallest entry of matrix Φ* : We established that the smallest element in sequence $\{\varphi_i(\ell_i)\}$, where $i \in \{0, 1\}$, is $\varphi_i(0)$ for $\alpha_i \geq 0$, and $\varphi_i(L_i)$ for $\alpha_i \leq 0$. Depending on the values of α_0 and α_1 , one of the corner entries of matrix Φ , which hold $\varphi(0, 0)$, $\varphi(L_0, 0)$, $\varphi(0, L_1)$ and $\varphi(L_0, L_1)$, will be the smallest entry of the matrix. Henceforth, we assume that $\alpha_0 > 0$ and $\alpha_1 > 0$. As a result, the sequences $\{\varphi_0(\ell_0)\}$ and $\{\varphi_1(\ell_1)\}$ are both monotonically increasing and, therefore, the smallest entry of matrix Φ is in row 1 and column 1, where $\varphi(0, 0)$ resides. Consequently, $(\ell_0^{(1)}, \ell_1^{(1)}) = (0, 0)$.

The cases where $\alpha_0 \leq 0$ or $\alpha_1 \leq 0$ differ only in the direction in which sequences $\{\varphi_0(\ell_0)\}$ and $\{\varphi_1(\ell_1)\}$ increase. For example, let us assume that $p_{01} < \frac{1}{2}$ and $p_{10} > \frac{1}{2}$, hence,

$\alpha_0 > 0$ and $\alpha_1 < 0$. This means that $\{\varphi_0(\ell_0)\}$ is a monotonically increasing sequence, but $\{\varphi_1(\ell_1)\}$ is a monotonically decreasing sequence. If we set $\tilde{p}_{10} = 1 - p_{10} < \frac{1}{2}$ or, equivalently, $\tilde{\alpha}_1 = -\alpha_1 > 0$, we obtain the monotonically increasing sequence $\{\tilde{\varphi}_1(\ell_1)\}$, which satisfies $\tilde{\varphi}_1(\ell_1) = \varphi_1(L_1 - \ell_1)$. The negative log-probability function $\tilde{\varphi}(\ell_0, \ell_1) = \varphi_0(\ell_0) + \tilde{\varphi}_1(\ell_1)$ is used to generate the entries of matrix $\tilde{\Phi}$. If the coordinate pair of the ℓ -th smallest entry of matrix $\tilde{\Phi}$ is $(\ell_0^{(\ell)}, \tilde{\ell}_1^{(\ell)})$, then the ℓ -th smallest entry of the original matrix Φ will simply be $(\ell_0^{(\ell)}, \ell_1^{(\ell)}) = (\ell_0^{(\ell)}, L_1 - \tilde{\ell}_1^{(\ell)})$. On the other hand, the case of $p_{10} = \frac{1}{2}$ and, thus, $\alpha_1 = 0$ leads to a constant sequence $\{\varphi_1(\ell_1)\}$. Ordering the entries of Φ becomes a trivial problem, as it depends solely on the monotonicity of sequence $\{\varphi_0(\ell_0)\}$, which is determined by the sign of α_0 . The arguments that were made for $\alpha_1 < 0$ and $\alpha_1 = 0$ can also be made for $\alpha_0 < 0$ and $\alpha_0 = 0$.

2) *Identifying the second smallest entry of matrix Φ* : Given that $\varphi(\ell_0, \ell_1) = \varphi_0(\ell_0) + \varphi_1(\ell_1)$ and that $\varphi(0, 0)$ holds the smallest entry of matrix Φ for $\alpha_0 > 0$ and $\alpha_1 > 0$, the second smallest entry will be either $\varphi(1, 0)$ or $\varphi(0, 1)$. To compare these two candidates for the second smallest entry, we evaluate the following differences:

$$\begin{aligned} \varphi(1, 0) - \varphi(0, 0) &= \varphi_0(1) - \varphi_0(0) = \alpha_0(1 - 0) = \alpha_0, \\ \varphi(0, 1) - \varphi(0, 0) &= \varphi_1(1) - \varphi_1(0) = \alpha_1(1 - 0) = \alpha_1. \end{aligned} \quad (17)$$

Therefore, the problem boils down to comparing α_0 with α_1 . If $\alpha_0 < \alpha_1$, then $\varphi(1, 0)$ is the second smallest entry of matrix Φ , hence $(\ell_0^{(2)}, \ell_1^{(2)}) = (1, 0)$; otherwise, $(\ell_0^{(2)}, \ell_1^{(2)}) = (0, 1)$. The process of pinpointing the second smallest entry of matrix Φ is illustrated in Fig. 3(a). Note that the coordinates of the smallest and second smallest entries of Φ , namely $(\ell_0^{(1)}, \ell_1^{(1)})$ and $(\ell_0^{(2)}, \ell_1^{(2)})$, have been determined by the values of α_0 and α_1 only.

3) *Generalization for the remaining entries of matrix Φ* : To explain how `TraceSortedProb` proceeds after the two smallest entries of matrix Φ have been identified, we start by making some simple observations that follow from the sum structure of $\varphi(\ell_0, \ell_1)$ and the linearity of the univariate functions $\varphi_0(\ell_0)$ and $\varphi_1(\ell_1)$. Notice that the entries of matrix Φ obey the following column-wise and row-wise properties:

$$\begin{aligned} \ell'_1 \leq \ell''_1 &\Leftrightarrow \varphi(\ell_0, \ell'_1) \leq \varphi(\ell_0, \ell''_1) \text{ for } \ell_0 = 0, \dots, L_0, \\ \ell'_0 \leq \ell''_0 &\Leftrightarrow \varphi(\ell'_0, \ell_1) \leq \varphi(\ell''_0, \ell_1) \text{ for } \ell_1 = 0, \dots, L_1. \end{aligned} \quad (18)$$

To extend this observation, we introduce a weak partial order relation and a strong partial order relation for the coordinate pairs, denoted by \preceq and \prec , respectively:

$$\begin{aligned} (\ell'_0, \ell'_1) \preceq (\ell''_0, \ell''_1) &\Leftrightarrow \ell'_0 \leq \ell''_0 \text{ and } \ell'_1 \leq \ell''_1, \\ (\ell'_0, \ell'_1) \prec (\ell''_0, \ell''_1) &\Leftrightarrow (\ell'_0, \ell'_1) \preceq (\ell''_0, \ell''_1) \text{ and } (\ell'_0, \ell'_1) \neq (\ell''_0, \ell''_1). \end{aligned}$$

Then, it follows from (18) that

$$(\ell'_0, \ell'_1) \preceq (\ell''_0, \ell''_1) \Rightarrow \varphi(\ell'_0, \ell'_1) \leq \varphi(\ell''_0, \ell''_1). \quad (19)$$

Suppose that the order of the ℓ smallest entries of matrix Φ has been determined, that is:

$$\varphi(\ell_0^{(1)}, \ell_1^{(1)}) \leq \varphi(\ell_0^{(2)}, \ell_1^{(2)}) \leq \dots \leq \varphi(\ell_0^{(\ell)}, \ell_1^{(\ell)}), \quad (20)$$

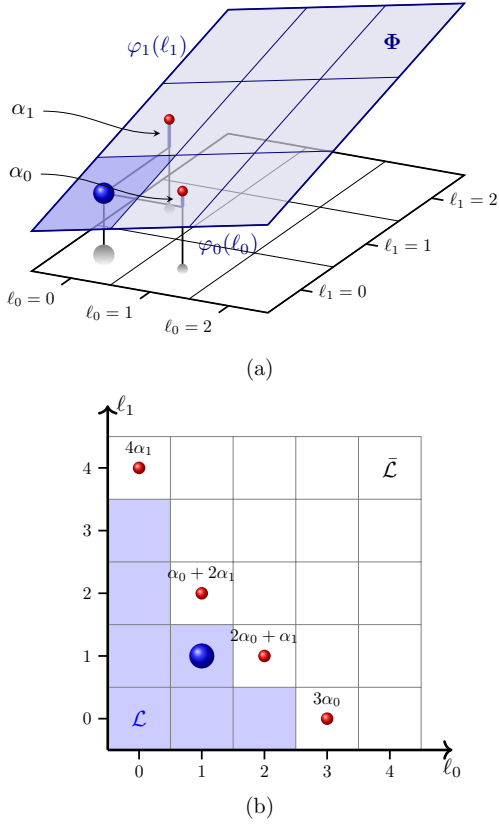


Fig. 3. Pictorial representation of how `TraceSortedProb` pinpoints the entries of matrix Φ , from the smallest to the largest, for $p_{01} < \frac{1}{2}$, $p_{10} < \frac{1}{2}$ and $p_{01} > p_{10}$. Equivalently, $\alpha_0 > 0$, $\alpha_1 > 0$ and $\alpha_0 < \alpha_1$. (a) A fragment of the matrix Φ with the plane above visualizing the value of its elements. The smallest entry of Φ , which corresponds to the largest probability value, is at $(\ell_0^{(1)}, \ell_1^{(1)}) = (0, 0)$ and is depicted by a large blue ball. Two candidates for the second smallest value are shown as small red balls at $(\ell_0, \ell_1) = (1, 0)$ and $(\ell_0, \ell_1) = (0, 1)$, where $\varphi(\ell_0, \ell_1)$ takes the values α_0 and α_1 , respectively. Since $\alpha_0 < \alpha_1$, the second smallest entry of Φ is at $(\ell_0^{(2)}, \ell_1^{(2)}) = (1, 0)$. (b) A larger fragment of matrix Φ . The shaded blue region represents the set \mathcal{L} of the ℓ smallest entries of Φ . The largest element in \mathcal{L} is depicted by a large blue ball. The candidates for the $(\ell+1)$ -th smallest entry are the corners of \mathcal{L} , shown as small red balls. Labels on candidate elements represent the values $\varphi(\ell_0, \ell_1) = \alpha_0 \ell_0 + \alpha_1 \ell_1$.

and the ℓ coordinate pairs form the set $\mathcal{L} = \{(\ell_0^{(\ell')}, \ell_1^{(\ell')})\}_{\ell'=1}^{\ell}$, which has a particular geometric structure, as shown in Fig. 3(b). If $(\ell_0, \ell_1) \in \mathcal{L}$ and $(\ell_0', \ell_1') \preccurlyeq (\ell_0, \ell_1)$, then from (19) we have $\varphi(\ell_0', \ell_1') \leq \varphi(\ell_0, \ell_1)$, thus $(\ell_0', \ell_1') \in \mathcal{L}$. This means that every element $(\ell_0, \ell_1) \in \mathcal{L}$ defines a set of indices $\mathcal{S}(\ell_0, \ell_1) = \{(\ell_0', \ell_1') : (\ell_0', \ell_1') \preccurlyeq (\ell_0, \ell_1)\}$ that make up a rectangle with coordinates $(0, 0)$, $(0, \ell_1)$, (ℓ_0, ℓ_1) and $(\ell_0, 0)$ in the grid specified by matrix Φ . The elements of $\mathcal{S}(\ell_0, \ell_1)$ are also members of \mathcal{L} , i.e.,

$$(\ell_0, \ell_1) \in \mathcal{L} \Rightarrow \mathcal{S}(\ell_0, \ell_1) \subseteq \mathcal{L}. \quad (21)$$

Hence, $\mathcal{L} = \bigcup_{\ell'=1, \dots, \ell} \mathcal{S}(\ell_0^{(\ell')}, \ell_1^{(\ell')})$. Many of these rectangles are fully contained in others, and can be omitted from the union of sets that compose \mathcal{L} without changing the result. Rectangles that cannot be omitted are those that enclose elements, denoted as $(c_0, c_1) \in \mathcal{L}$, that are not contained in any other rectangular set:

$$(c_0, c_1) : \nexists (\ell_0, \ell_1) \in \mathcal{L} : (c_0, c_1) \prec (\ell_0, \ell_1). \quad (22)$$

We call such elements the *corners* of \mathcal{L} , and denote the set of all corners of \mathcal{L} as $\mathcal{C}_{\mathcal{L}}$. For example, if \mathcal{L} is defined as in Fig. 3(b), the set of all corners of \mathcal{L} is given by $\mathcal{C}_{\mathcal{L}} = \{(0, 3), (1, 1), (2, 0)\}$. By tracking $\mathcal{C}_{\mathcal{L}}$, we maintain the minimal representation of the coordinate set \mathcal{L} at each step of the procedure, as $\mathcal{L} = \bigcup_{(c_0, c_1) \in \mathcal{C}_{\mathcal{L}}} \mathcal{S}(c_0, c_1)$. Furthermore, the element that was added last to \mathcal{L} , that is, $(\ell_0^{(\ell)}, \ell_1^{(\ell)})$, is always a corner of \mathcal{L} . Indeed, if $(\ell_0^{(\ell)}, \ell_1^{(\ell)}) \notin \mathcal{C}_{\mathcal{L}}$, then $\exists \ell' < \ell$ such that $(\ell_0^{(\ell)}, \ell_1^{(\ell)}) \prec (\ell_0^{(\ell')}, \ell_1^{(\ell')})$, which leads to $\varphi(\ell_0^{(\ell)}, \ell_1^{(\ell)}) < \varphi(\ell_0^{(\ell')}, \ell_1^{(\ell')})$ according to (19). The latter inequality contradicts (20) for $\ell' < \ell$, therefore the original assumption $(\ell_0^{(\ell)}, \ell_1^{(\ell)}) \notin \mathcal{C}_{\mathcal{L}}$ must be false, hence $(\ell_0^{(\ell)}, \ell_1^{(\ell)}) \in \mathcal{C}_{\mathcal{L}}$ must be true.

Observe in Fig. 3(b) that, as ℓ increases, more entries of Φ are traced, their coordinate pairs are added to \mathcal{L} , and \mathcal{L} grows in size. The $(L_0 + 1)(L_1 + 1) - \ell$ coordinate pairs of the entries of Φ that have not been considered yet in the tracing process compose the complementary set $\bar{\mathcal{L}} = \{(\ell_0, \ell_1) \notin \mathcal{L}\}$, which has structural properties that mirror those of \mathcal{L} . Whereas \mathcal{L} holds the coordinates of the ℓ smallest entries of Φ , $\bar{\mathcal{L}}$ consists of the coordinates of the $(L_0 + 1)(L_1 + 1) - \ell$ largest entries of Φ . Whereas the coordinate pair that corresponds to the largest of the ℓ smallest entries of Φ is one of the corners of \mathcal{L} , the coordinate pair that corresponds to the smallest of the $(L_0 + 1)(L_1 + 1) - \ell$ largest entries of Φ is one of the corners of $\bar{\mathcal{L}}$. Let the corners of \mathcal{L} be arranged in $\mathcal{C}_{\mathcal{L}} = \{(c_0^{(1)}, c_1^{(1)}), \dots, (c_0^{(C)}, c_1^{(C)})\}$ such that $c_0^{(1)} < \dots < c_0^{(C)}$ and $c_1^{(1)} > \dots > c_1^{(C)}$, where $C = |\mathcal{C}_{\mathcal{L}}|$. The corners of $\bar{\mathcal{L}}$ can then be calculated from the corners of \mathcal{L} as follows:

$$\begin{aligned} \mathcal{C}_{\bar{\mathcal{L}}} = \{ & \underbrace{(0, c_1^{(1)} + 1), (c_0^{(1)} + 1, c_1^{(2)} + 1), \dots}_{\text{omit if } c_1^{(1)} \geq L_1} \\ & \dots, (c_0^{(C-1)} + 1, c_1^{(C)} + 1), \underbrace{(c_0^{(C)} + 1, 0)}_{\text{omit if } c_0^{(C)} \geq L_0} \}. \end{aligned} \quad (23)$$

In Fig. 3(b), we mentioned that $\mathcal{C}_{\mathcal{L}} = \{(0, 3), (1, 1), (2, 0)\}$, therefore $\mathcal{C}_{\bar{\mathcal{L}}} = \{(0, 4), (1, 2), (2, 1), (3, 0)\}$ based on (23).

To determine the $(\ell+1)$ -th smallest entry of Φ and obtain the pair $(\ell_0^{(\ell+1)}, \ell_1^{(\ell+1)})$, we need to find the smallest entry among the candidates $(\ell_0, \ell_1) \in \mathcal{C}_{\bar{\mathcal{L}}}$, i.e.,

$$\begin{aligned} (\ell_0^{(\ell+1)}, \ell_1^{(\ell+1)}) &= \arg \min_{(\ell_0, \ell_1) \in \mathcal{C}_{\bar{\mathcal{L}}}} \varphi(\ell_0, \ell_1) \\ &= \arg \min_{(\ell_0, \ell_1) \in \mathcal{C}_{\bar{\mathcal{L}}}} (\alpha_0 \ell_0 + \alpha_1 \ell_1). \end{aligned} \quad (24)$$

The bivariate function $\varphi(\ell_0, \ell_1) = \alpha_0 \ell_0 + \alpha_1 \ell_1$ can be seen as a ‘penalty’ that is minimized by `TraceSortedProb`. Fig. 3(b) provides a visualization of the penalty for each of the four elements of $\mathcal{C}_{\bar{\mathcal{L}}}$.

The step-by-step process of `TraceSortedProb` to identify the six smallest entries of matrix Φ for $\alpha_0 = 0.59$, $\alpha_1 = 1.22$, $L_0 = 3$ and $L_1 = 3$ is illustrated in Fig. 4. As explained in Section VI-A1 and shown in Fig. 4(b), the smallest entry is at $(0, 0)$ given that $\alpha_0 > 0$ and $\alpha_1 > 0$. Hence, $\mathcal{C}_{\mathcal{L}} = \{(0, 0)\}$ for $\ell = 1$, and $\mathcal{C}_{\bar{\mathcal{L}}} = \{(0, 1), (1, 0)\}$

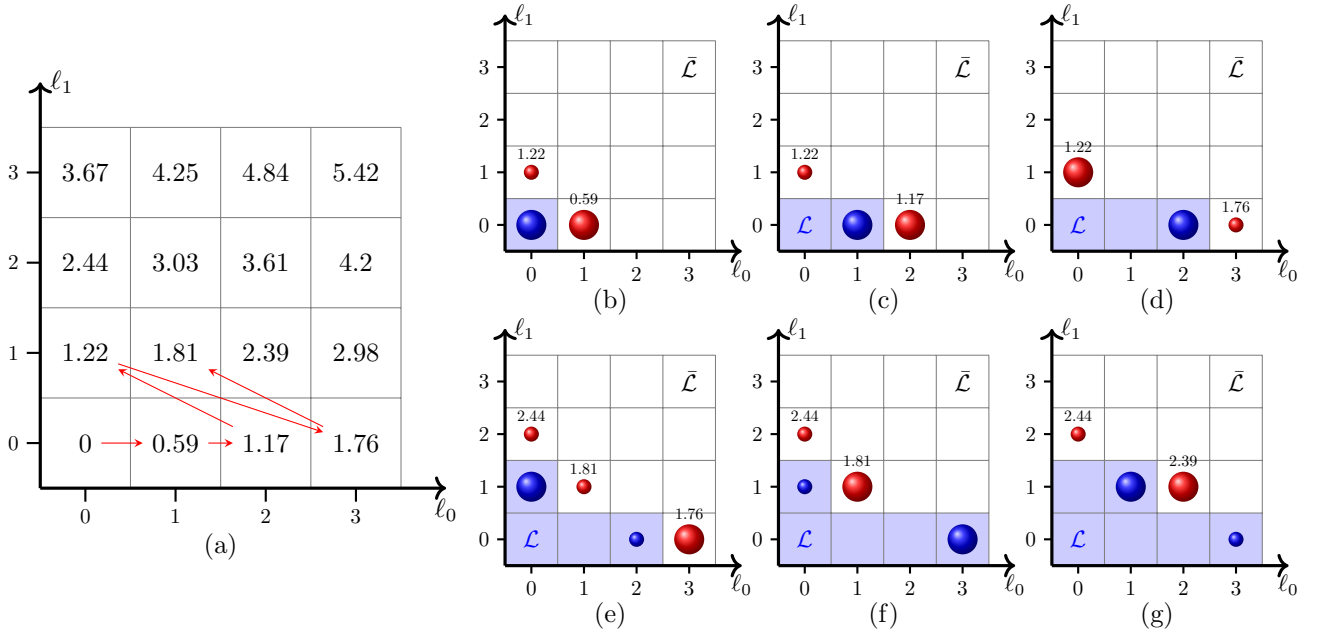


Fig. 4. Example demonstrating the operation of `TraceSortedProb` for $p_{01} = 0.4$, $p_{10} = 0.3$, $L_0 = 3$ and $L_1 = 3$. Based on these values, $\alpha_0 = 0.59$ and $\alpha_1 = 1.22$ (displayed values have been rounded up to two decimal places). The entries of Φ as well as a path that begins at the smallest entry and ends at the sixth smallest entry of Φ are presented in (a) for reference. The process of `TraceSortedProb` for identifying the six smallest entries of Φ , without first calculating all entries and then ordering them, is described in (b)-(g). Blue balls of any size identify corners of \mathcal{L} , while red balls of any size represent corners of $\bar{\mathcal{L}}$. A large blue ball marks the coordinate pair that has been added last to \mathcal{L} , as it points to the ℓ -th smallest value of Φ , for $\ell = 1$ in (b) to $\ell = 6$ in (g). Labels on red balls represent the values of $\varphi(\ell_0, \ell_1)$ at these points. A large red ball pinpoints the corner of $\bar{\mathcal{L}}$ that is chosen for a value of ℓ and changes to a large blue ball after it is added to \mathcal{L} for the next value of ℓ . Observe that the large blue balls in (b)-(g) occupy the same coordinates and in the same order as the coordinates of the six smallest entries in (a).

based on (23). Using (24), we find that transitioning from $(0,0)$ to $(1,0)$ carries a lower penalty than transitioning to $(0,1)$. Therefore, the second smallest entry, i.e., for $\ell = 2$, is at $(1,0)$. The pair $(1,0)$ is added to \mathcal{L} and $\mathcal{C}_{\mathcal{L}}$, previous members of $\mathcal{C}_{\mathcal{L}}$ that do not comply with (22) are removed from $\mathcal{C}_{\mathcal{L}}$, the complementary set $\mathcal{C}_{\bar{\mathcal{L}}}$ is obtained from $\mathcal{C}_{\mathcal{L}}$ using (23), and the coordinate pair in $\mathcal{C}_{\bar{\mathcal{L}}}$ that minimizes (24) points to the location of the third smallest entry of Φ . This iterative process could continue until all entries of Φ have been considered, but it could also terminate when the coordinates of a desired proportion of the smallest – and, hence, most significant – entries of Φ have been determined.

B. Complexity Analysis

As in `CalcProbAndSort`, the terms $\alpha_0 = \log(\frac{1-p_{01}}{p_{01}})$ and $\alpha_1 = \log(\frac{1-p_{10}}{p_{10}})$ are computed prior to the execution of the algorithm. Part of the running time of `TraceSortedProb` is dedicated to the evaluation of the products $\alpha_0 \ell_0$ and $\alpha_1 \ell_1$ in (24) for all values of ℓ_0 and ℓ_1 . Notice that as ℓ_0 increases in steps of 1, the value of $\alpha_0 \ell_0$ increases in steps of α_0 (i.e., one addition per step). Calculation of the products $\alpha_0 \ell_0$ and $\alpha_1 \ell_1$, for $\ell_0 = 0, \dots, L_0$ and $\ell_1 = 0, \dots, L_1$, involves L_0 increments of ℓ_0 and L_1 increments of ℓ_1 , which result in $L_0 + L_1$ sequential additions. An extra addition is required to evaluate the penalty $\alpha_0 \ell_0 + \alpha_1 \ell_1$ whenever a new coordinate pair (ℓ_0, ℓ_1) is inserted in $\mathcal{C}_{\bar{\mathcal{L}}}$.

As explained in Section VI-A3, `TraceSortedProb` maintains and updates $\mathcal{C}_{\mathcal{L}}$ and $\mathcal{C}_{\bar{\mathcal{L}}}$, which contain the corners

of \mathcal{L} and $\bar{\mathcal{L}}$, respectively. As the procedure traverses through the entries of matrix Φ and the size of \mathcal{L} increases in each step, maintaining and updating $\mathcal{C}_{\mathcal{L}}$ and $\mathcal{C}_{\bar{\mathcal{L}}}$ requires no more than $|\mathcal{C}_{\mathcal{L}}|$ operations per step. These operations between integer-valued indices can be omitted from the analysis, given that computational complexity is primarily driven by floating-point operations. At each step, the complexity of `TraceSortedProb` is dominated by the process of finding the lowest penalty among $|\mathcal{C}_{\bar{\mathcal{L}}}|$ penalties associated with the $|\mathcal{C}_{\bar{\mathcal{L}}}|$ corners of $\mathcal{C}_{\bar{\mathcal{L}}}$, as described by (24) and depicted in Fig. 4. Let the $|\mathcal{C}_{\bar{\mathcal{L}}}|$ penalties be stored in the form of a linked list data structure, the elements of which are organized in ascending order. Therefore, the first element of the list always corresponds to the lowest penalty. With the aid of Fig. 4, one can observe that, at each step, a new coordinate pair – on average – is inserted in $\mathcal{C}_{\bar{\mathcal{L}}}$ and the corresponding penalty is computed and inserted in the list. Identification of the position in the sorted linked list where the new penalty value should be inserted requires $O(\log |\mathcal{C}_{\bar{\mathcal{L}}}|)$ comparisons. The lowest penalty value is subsequently removed from the linked list and the corresponding corner in $\mathcal{C}_{\bar{\mathcal{L}}}$ is moved to $\mathcal{C}_{\mathcal{L}}$. After $(L_0 + 1)(L_1 + 1)$ steps, when all entries of matrix Φ have been considered, `TraceSortedProb` has computed and inserted $(L_0 + 1)(L_1 + 1)$ penalty values in the linked list, raising the total number of additions to $(L_0 + 1)(L_1 + 1) + L_0 + L_1$ and the total number of comparisons to $O((L_0 + 1)(L_1 + 1) \log |\mathcal{C}_{\bar{\mathcal{L}}}|)$. As can be inferred from (23), the structure of \mathcal{L} ensures that $|\mathcal{C}_{\bar{\mathcal{L}}}| \leq \min(L_0 + 1, L_1 + 1)$. The number of comparisons performed by `TraceSortedProb`

TABLE I
NUMBER OF OPERATIONS REQUIRED BY TWO COMPETING PROCEDURES
FOR THE EVALUATION OF A SINGLE COLUMN OF THE $L \times B$ MATRIX $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$,
WHERE $L = L_0 + L_1$.

Operations	CalcProbAndSort
Additions	$(L_0 + 1)(L_1 + 1) + L_0 + L_1$
Worst-case number of comparisons	$O(n \log n)$ for $n = (L_0 + 1)(L_1 + 1)$

Operations	TraceSortedProb
Additions	$\ell_{\text{th}} + L_0 + L_1$ for $1 \leq \ell_{\text{th}} \leq (L_0 + 1)(L_1 + 1)$
Worst-case number of comparisons	$O(\ell_{\text{th}} \log \min(L_0 + 1, L_1 + 1))$ for $1 \leq \ell_{\text{th}} \leq (L_0 + 1)(L_1 + 1)$

is thus $O((L_0 + 1)(L_1 + 1) \log \min(L_0 + 1, L_1 + 1))$.

We established that both CalcProbAndSort and TraceSortedProb carry out $(L_0 + 1)(L_1 + 1) + L_0 + L_1$ additions. If we set $n = (L_0 + 1)(L_1 + 1)$ and $L_0 = L_1$ to simplify expressions, CalcProbAndSort performs $O(n \log n)$ comparisons, whereas TraceSortedProb performs $O(n \log(\sqrt{n} - 1))$ comparisons. If we take into account that $n \log(\sqrt{n} - 1) < \frac{n}{2} \log n$, we conclude that the worst-case number of comparisons in both procedures is $O(n \log n)$.

Although both procedures have computational complexities of the same order, the quick convergence of GRAND-based methods to optimal solutions [11] gives TraceSortedProb an edge on CalcProbAndSort. The error vector that is selected to be a column of matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ is found among error vectors that have been born out of the first few elements of the output sequence of coordinate pairs. Assume that transversal GRAND introduces a threshold ℓ_{th} , which limits the number of coordinate pairs that are considered. CalcProbAndSort still has to sort the $(L_0 + 1)(L_1 + 1)$ entries of matrix Φ and then output the coordinates of only the ℓ_{th} smallest entries. Therefore, the complexity of CalcProbAndSort remains constant and independent of ℓ_{th} . In contrast, TraceSortedProb will generate the sequence of ℓ_{th} coordinate pairs after the first ℓ_{th} steps and will then terminate execution, thus ignoring the remaining $(L_0 + 1)(L_1 + 1) - \ell_{\text{th}}$ entries of matrix Φ . This means that the product $(L_0 + 1)(L_1 + 1)$ can be replaced by ℓ_{th} in the complexity expressions derived in this section. A summary of the total number of operations performed by CalcProbAndSort and TraceSortedProb for the evaluation of a single column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ is provided in Table I. Note that for $n = (L_0 + 1)(L_1 + 1)$ and $L_0 = L_1$, the number of additions carried out by TraceSortedProb reduces from $O(n)$ to $O(\sqrt{n})$, while the worst-case number of comparisons drops from $O(n \log n)$ to $O(\log n)$, provided that ℓ_{th} is independent of n .

VII. COMPLEXITY CHARACTERIZATION BASED ON SYSTEM PARAMETERS

The analysis in Sections V and VI-B used the number of zeros and ones, denoted by L_0 and L_1 respectively, in column $b-1$ of the $L \times B$ matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ to characterize the computational

complexity associated with the generation of ordered coordinate pairs $(\ell_0^{(\ell)}, \ell_1^{(\ell)})$ for the estimation of column b of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$, where $\ell = 1, \dots, (L_0 + 1)(L_1 + 1)$ in CalcProbAndSort, and $\ell = 1, \dots, \ell_{\text{th}}$ in TraceSortedProb. This section expresses the complexity of each procedure for the estimation of the full matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ in terms of primary system parameters.

Both syndrome decoding and transversal GRAND rely on RLC decoding, which has complexity $O(K^3)$ if Gaussian elimination is employed [35]. As explained in Section III, syndrome decoding considers candidate error vectors for each column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ in the same order always, regardless of the channel parameters. In transversal GRAND, the ordering of the candidate vectors is informed by the channel parameters. As a result, transversal GRAND incurs additional complexity, which is further discussed in this section, but has the potential to obtain better estimates of the error matrix and, thus, stands a higher chance of recovering the source message, as demonstrated in Section VIII.

A. Distribution of the elements of the error matrix

Section VI-B established that the complexity of the ordering procedures used in transversal GRAND depends on the number of zeros and ones in each column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. Recall that matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ has dimensions $L \times B$, where $L = N - N_{\mathcal{R}}$ is the number of received coded packets that contain errors. The worst-case scenario for transversal GRAND occurs when errors are spread out over all N coded packets, in which case the dimensions of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ are $N \times B$. Let X_b be a discrete random variable that represents the number of zeros in column b of the $N \times B$ matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$. In other words, X_b represents the number of independent Markov chains that are in state 0 in step b , for $b = 1, \dots, B$. The number of ones in column b is given by $N - X_b$. We denote the expected value of X_b by $E(X_b)$. The following recurrence relation can be obtained from the Markov chain in Fig. 1:

$$\begin{aligned} E(X_b) &= (N - E(X_{b-1}))p_{10} + E(X_{b-1})(1 - p_{01}) \\ &= Np_{10} + E(X_{b-1})(1 - p_{10} - p_{01}), \end{aligned} \quad (25)$$

which states that the expected number of zeros in column b is the sum of the expected number of ones in column $b-1$ that changed to zeros with probability p_{10} and the expected number of zeros in column $b-1$ that did not change value with probability $1 - p_{01}$. As mentioned in Section IV-B, the N Markov chains start from state 0, therefore $E(X_0) = X_0 = N$.

Iteration of (25) for an increasing value of b reveals that $E(X_b)$ can be written via geometric series as follows:

$$\begin{aligned} E(X_b) &= N \left(1 - p_{01} \sum_{i=0}^{b-1} (1 - p_{10} - p_{01})^i \right) \\ &= N \left(1 - p_{01} \left(\frac{1 - (1 - p_{10} - p_{01})^b}{p_{10} + p_{01}} \right) \right) \\ &= N (1 - \varepsilon (1 - \mu^b)), \end{aligned} \quad (26)$$

where both the bit error probability $\varepsilon = p_{01}/(p_{01} + p_{10})$ and the channel memory $\mu = 1 - p_{10} - p_{01}$ have been defined in Section IV-A. For increasing b , the contribution of μ to the

value of $W(X_b)$ diminishes as the Markov chains approach the steady state. At the steady state, (26) reduces to:

$$E(X) = N(1 - \varepsilon), \quad (27)$$

which is independent of b . Although elements in adjacent columns of $\hat{\mathbf{E}}_{\bar{r}}$ remain correlated at the steady state, we infer from (27) that the number of zeros in each column of $\hat{\mathbf{E}}_{\bar{r}}$ follows a binomial distribution with parameters N and $1 - \varepsilon$. Hence, the number of zeros in any column of $\hat{\mathbf{E}}_{\bar{r}}$ can be represented by a single random variable $X \sim \text{Binomial}(N, 1 - \varepsilon)$ with $E(X)$ given by (27).

Knowledge of the distribution of X can be used to estimate the number of a particular type of operations, e.g., additions or comparisons, performed by each of the ordering procedures presented in Sections V and VI-B.

B. Computational complexity of additions

Let $g_{\text{add}}^{\text{sort}}(B, N, \varepsilon)$ denote the average number of additions required for B runs of `CalcProbAndSort` to obtain the B columns of $\hat{\mathbf{E}}_{\bar{r}}$. From Table I, for $L_0 \leftarrow X$ and $L_1 \leftarrow N - X$, we obtain:

$$g_{\text{add}}^{\text{sort}}(B, N, \varepsilon) = E \left(\sum_{b=1}^B ((X+1)(N-X+1) + N) \right) \quad (28)$$

$$= B (NE(X) - E(X^2) + 2N + 1),$$

where

$$E(X^2) = \text{Var}(X) + (E(X))^2. \quad (29)$$

Since $X \sim \text{Binomial}(N, 1 - \varepsilon)$, the variance of X is given by $\text{Var}(X) = N\varepsilon(1 - \varepsilon)$ and the mean of X is given by (27). Substituting $\text{Var}(X)$ and $E(X)$ into (28) results in:

$$g_{\text{add}}^{\text{sort}}(B, N, \varepsilon) = B(N(N-1)\varepsilon(1-\varepsilon) + 2N + 1) \quad (30)$$

$$\leq BN^2\varepsilon + 2BN + B,$$

which leads us to the conclusion that:

$$g_{\text{add}}^{\text{sort}}(B, N, \varepsilon) = O(BN^2), \quad (31)$$

for a given $\varepsilon > 0$ as $B, N \rightarrow \infty$. Positive values, which are not too close to zero, are considered for the bit error probability ε because transversal GRAND is triggered only when RLC decoding cannot cope with the large number of erroneous received coded packets.

The average number of additions required for B runs of `TraceSortedProb` can be obtained following a similar line of reasoning as before and the help of Table I:

$$g_{\text{add}}^{\text{trace}}(B, N, \varepsilon) = E \left(\sum_{b=1}^B (\ell_{\text{th}} + N) \right) \quad (32)$$

$$= BE(\ell_{\text{th}}) + BN.$$

As we show in Section VIII, ℓ_{th} can be kept small, therefore

$$g_{\text{add}}^{\text{trace}}(B, N, \varepsilon) = O(BN). \quad (33)$$

C. Computational complexity of comparisons

Let $g_{\text{comp}}^{\text{sort}}(B, N, \varepsilon)$ denote the average number of comparisons required for B runs of `CalcProbAndSort` to obtain the B columns of $\hat{\mathbf{E}}_{\bar{r}}$ and $g_{\text{comp}}^{\text{trace}}(B, N, \varepsilon)$ denote the number of comparisons required by `TraceSortedProb` for the same task.

The $O(n \log n)$ worst-case estimates in Table I are for a comparison sort algorithm, such as heapsort [34]. The average number of operations in heapsort is $n \log_2 n$, i.e. the leading term constant is one. Hence, the average complexity of one `CalcProbAndSort` run is

$$n \log_2 n = (L_0 + 1)(L_1 + 1) (\log_2(L_0 + 1) + \log_2(L_1 + 1)),$$

while the complexity of one `TraceSortedProb` run is

$$\ell_{\text{th}} \log_2 \min(L_0 + 1, L_1 + 1) \leq \frac{1}{2} \ell_{\text{th}} (\log_2(L_0 + 1) + \log_2(L_1 + 1)),$$

thus `CalcProbAndSort` is at least $2(L_0 + 1)(L_1 + 1)/\ell_{\text{th}}$ times more complex than `TraceSortedProb`.

As before, we plug $L_0 \leftarrow X$ and $L_1 \leftarrow N - X$, and average over the binomial random variable X . Since \log is a concave function, we can bound $\log_2(L_0 + 1) + \log_2(L_1 + 1)$ from above as follows:

$$\log_2(L_0 + 1) + \log_2(L_1 + 1) \leq 2 \log_2(L_0 + L_1 + 2) = 2 \log_2(N + 2).$$

Substituting this result in the estimates above, and proceeding in a similar way to (30), we obtain:

$$g_{\text{comp}}^{\text{sort}}(B, N, \varepsilon) \leq 2BE((X+1)(N-X+1)) \log_2(N+2) = 2B(NE(X) - E(X^2) + N + 1) \log_2(N+2) \leq 2BN^2 \log_2(N+2)\varepsilon + 2B(N+1) \log_2(N+2) = O(BN^2 \log N). \quad (34)$$

Following a similar line of reasoning for the number of comparisons in `TraceSortedProb`, we get:

$$g_{\text{comp}}^{\text{trace}}(B, N, \varepsilon) \leq BE(\ell_{\text{th}}) \log_2(N+2). \quad (35)$$

As we show in numerical experiments in Section VIII, ℓ_{th} can be kept small, therefore the tracing procedure is asymptotically faster – with respect to the number of comparisons – than the sorting procedure.

D. Relative complexity of additions and comparisons

Additions are, on average, more expensive than comparisons. To explain this fact, consider a pair of integers $a = \overline{a_s \dots a_1 a_0} = \sum_{k=0}^s a_k 2^k$ and $b = \overline{b_s \dots b_1 b_0} = \sum_{k=0}^s b_k 2^k$, written in big-endian notation, with s bits $a_k, b_k \in \mathbb{F}_2$ for $k = 0, \dots, s$. Computing their sum $c = a + b$ in the same form $c = \overline{c_{s+1} c_s \dots c_1 c_0}$ would require $2s$ bit operations, $c_k = a_k + b_k + r_{k-1} \pmod 2$ for $k = 1, \dots, s$ where r_{k-1} is the carry-over from the previous bit.

In contrast, a comparison of a and b would require to look at the most significant bits a_s and b_s first. If $a_s > b_s$, we can instantly conclude that $a > b$. If $a_s < b_s$, then $a < b$, and only if $a_s = b_s$ we need to consider the subsequent digits.

In binary arithmetic, $a_s = b_s$ happens with probability $1/2$. Hence, the comparison of a and b requires one bit-operation with probability $1/2$, two bit-operations with probability $1/2^2$, and so on. Overall, the average complexity of a comparison for numbers of length s bits is equal to:

$$\sum_{k=1}^s \frac{k}{2^k} = 2 - \frac{2+s}{2^s} < 2, \quad (36)$$

i.e., a comparison requires, on average, less than two bit-operations.

To summarise, comparisons are on average s times faster than additions, where $s = 32$ for single-precision computations and $s = 64$ for double-precision computations. We can therefore consider additions to dominate the complexity of the whole computation, and neglect the effect of comparisons.

VIII. RESULTS AND DISCUSSION

In this section, the bit error probability, ε , and the average length of an error burst, Λ , are used as the channel parameters; they can be derived from the transition probabilities, p_{10} and p_{01} , as explained in Section IV-A. Recall that if a destination node receives $L = N - N_R$ erroneous coded packets of B bits, transversal GRAND attempts to repair them by estimating the $L \times B$ matrix $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ column by column. Estimation of a column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ relies on the entries of the previous column. If the previous column contains L_0 zeros and L_1 ones, where $L_0 + L_1 = L$, transversal GRAND classifies all 2^L possible vectors that are candidates for the next column of $\hat{\mathbf{E}}_{\bar{\mathcal{R}}}$ into $(L_0 + 1)(L_1 + 1)$ groups according to their probability of occurrence, as explained in Section IV-B. The TraceSortedProb procedure, introduced in Section VI, provides transversal GRAND with the option to identify only the ℓ_{th} groups that are most likely to occur, instead of all $(L_0 + 1)(L_1 + 1)$ groups, thus reducing computational complexity without necessarily degrading the estimation accuracy of the algorithm.

The authors of [9] considered memoryless channels and demonstrated the performance gain achieved by syndrome decoding when it is combined with RLC decoding, in terms of the overall *decoding probability*, that is, the probability of a destination node recovering the K source packets when N coded packets have been transmitted. Since the proposed transversal GRAND algorithm is an extension of syndrome decoding for burst error channels, the objective of the following section is to compare the decoding probabilities of stand-alone RLC decoding, RLC decoding combined with syndrome decoding (RLC with SD) and RLC decoding combined with transversal GRAND (RLC with T-GRAND) for various channel configurations.

A. Impact of T-GRAND on the decoding probability

The decoding probability was measured through simulations² for $K = 10$ source packets and a range of values for the

²Software simulations were implemented in MATLAB. The decoding probability shown in the plots for a tuple $(K, N, \varepsilon, \Lambda, B)$ has been averaged over 6×10^4 channel realizations. For each realization, the generator matrix of the RLC at the transmitter was randomly generated, as described in Section II.

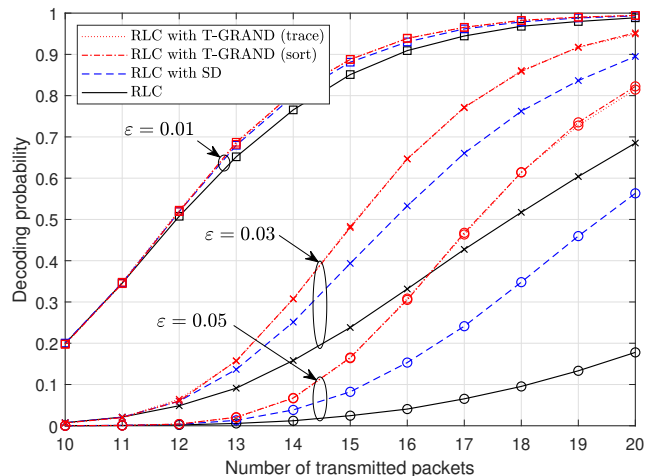


Fig. 5. Comparison of the decoding probabilities of RLC decoding, RLC decoding with SD, and RLC decoding with T-GRAND for $K = 10$ source packets, N transmitted coded packets, where $N = 10, \dots, 20$, bit error probability $\varepsilon \in \{0.01, 0.03, 0.05\}$, average length of error bursts $\Lambda = 4$ and packet length $B = 64$.

number of transmitted packets ($N = 10, 11, \dots, 20$), the bit error probability ($\varepsilon \in \{0.01, 0.03, 0.05\}$), the average length of burst errors ($\Lambda \in \{3, 5, 7\}$), and the packet length in bits ($B \in \{16, 32, 96\}$). Measurements of the decoding probability are presented in Fig. 5, Fig. 6 and Fig. 7. In all three figures, T-GRAND implementations using CalcProbAndSort and TraceSortedProb have been considered. They have been abbreviated to “T-GRAND (sort)” and “T-GRAND (trace)”, respectively. In the latter case, we set $\ell_{\text{th}} = 8$. The figures establish that the threshold ℓ_{th} used by TraceSortedProb can be set to a small value to reduce the overall complexity of T-GRAND without compromising the decoding probability achieved by T-GRAND based on CalcProbAndSort.

Fig. 5 shows measurements for an increasing probability of the channel flipping bits in transmitted packets, when each packet contains $B = 64$ bits and errors occur in bursts of length $\Lambda = 4$ on average. For $\varepsilon = 0.01$ and a growing number of transmitted packets, the destination node is likely to receive K error-free linearly independent packets among the received packets and recover the K source packets using RLC decoding; SD and T-GRAND offer only a marginal improvement in decoding probability. If the value of ε is raised to 0.05, the proportion of erroneous received packets increases, the RLC decoder experiences a notable drop in decoding performance, and SD and T-GRAND become instrumental in improving the chances of recovering the source packets. For example, when $\varepsilon = 0.05$ and $N = 20$, SD increases the decoding probability of RLC from 0.18 to 0.56 whereas T-GRAND, which considers correlations in errors when repairing coded packets, boosts the decoding probability to 0.82.

Fig. 6 depicts the impact of the average length of error bursts on the decoding probability, when $\varepsilon = 0.03$. When the value of ε is constant, the same number of errors – on average – impair the transmitted packets, but errors aggregate in fewer packets as the average length of error bursts increases. Consequently, destination nodes receive an increasingly larger

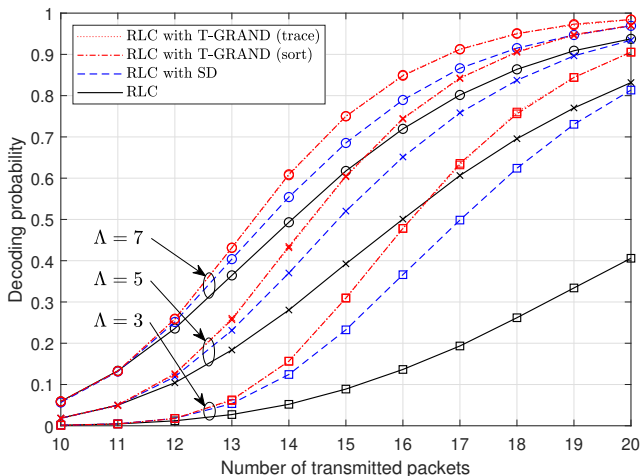


Fig. 6. Comparison of the decoding probabilities of RLC decoding, RLC decoding with SD, and RLC decoding with T-GRAND for $K = 10$ source packets, N transmitted coded packets, where $N = 10, \dots, 20$, bit error probability $\varepsilon = 0.03$, average length of error bursts $\Lambda \in \{3, 5, 7\}$ and packet length $B = 64$.

proportion of error-free packets as Λ grows, and RLC decoding stands a greater chance of success. Although fewer packets are received in error for higher values of Λ , erroneous packets are more severely damaged by longer error bursts. Nevertheless, SD and especially T-GRAND can still improve the decoding probability of RLC. For example, the decoding probability of RLC, RLC with SD and RLC with T-GRAND is 0.72, 0.79 and 0.85, respectively, for $\Lambda = 7$ and $N = 16$. On the other hand, low values of Λ have a negative impact on RLC decoding; for a decreasing length of error bursts and a fixed average number of errors, the proportion of correctly received packets reduces. Packets corrupted by errors may dominate but errors are more sparsely distributed, and both SD and T-GRAND are more effective in repairing corrupted packets. Notice in Fig. 6 that the decoding probability of RLC, RLC with SD and RLC with T-GRAND is 0.41, 0.81 and 0.91, respectively, for $\Lambda = 3$ and $N = 20$.

The effect of the packet length B on the decoding probability is shown in Fig. 7. For fixed values of ε and Λ , the likelihood that a packet will be corrupted by errors increases with B and, thus, the performance of RLC decoding deteriorates. When it comes to SD and T-GRAND, the value of B also signifies the number of times that the algorithms have run in order to estimate the errors in the B positions of the corrupted packets. The advantage of T-GRAND over SD becomes apparent as B increases. For $B = 96$ and $N = 20$, SD lifts the decoding probability of RLC from 0.08 to 0.62; T-GRAND further increases the decoding probability to 0.82 as is more successful than SD in making consecutive accurate guesses of error occurrences in the course of the $B = 96$ runs.

B. Impact of T-GRAND on the number of transmitted packets

Whereas figures 5 to 7 depict the decoding probability as a function of the number of transmitted packets, Fig. 8 shows the relationship between the *average* number of transmitted packets, denoted by $E(N)$, and the bit error probability for a

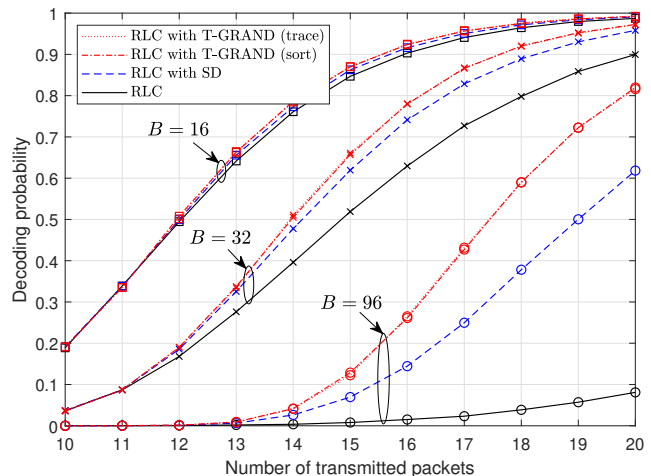


Fig. 7. Comparison of the decoding probabilities of RLC decoding, RLC decoding with SD, and RLC decoding with T-GRAND for $K = 10$ source packets, N transmitted coded packets, where $N = 10, \dots, 20$, bit error probability $\varepsilon = 0.03$, average length of error bursts $\Lambda = 3$ and packet length $B \in \{16, 32, 96\}$.

decoding probability equal to 1. To obtain Fig. 8, simulations allowed the number of transmitted packets N to increase until all K source packets had been recovered by the destination node. The average number of packet transmissions for the successful delivery of the K source packets influences the *average completion delay* [36], given by $E(N) - K$. In a broadcast scenario, a single destination node can be the major contributor to the average completion delay for the system, if it experiences the highest bit error probability compared to other destination nodes and, thus, frequently imposes the highest completion delay among all destination nodes.

We observe in Fig. 8 that, for $\varepsilon = 0.05$, $E(N)$ increases from 20.8 (for $K = 10$) to 41.3 (for $K = 20$) to 62.2 (for $K = 30$) in the case of RLC decoding. This is equivalent to $E(N) \approx 2.07K$ for $K \in \{10, 20, 30\}$ and $\varepsilon = 0.05$. Therefore, the ratio $E(N)/K$ for a given bit error probability remains largely unaffected by K when stand-alone RLC decoding is employed. When RLC decoding is combined with SD for $\varepsilon = 0.05$, the ratio $E(N)/K$ is smaller than 2.07, and drops from 1.77 (for $K = 10$) to 1.57 (for $K = 30$). T-GRAND further reduces the requirement for a drastic increase in the value of $E(N)$ to compensate for the increase in the bit error probability; when $\varepsilon = 0.05$, the ratio $E(N)/K$ decreases from 1.67 (for $K = 10$) to 1.41 (for $K = 30$). The gap in the average number of transmitted packets between T-GRAND and SD grows as the bit error probability increases. For instance, as can be seen in Fig. 8, $E(N) = 28.5$ for T-GRAND while $E(N) = 30.4$ for SD when $K = 20$ and $\varepsilon = 0.04$, i.e., T-GRAND requires 1.9 fewer packets to be transmitted – on average – than SD to achieve a decoding probability of 1. This difference increases to 3.1 for $\varepsilon = 0.05$, and to 4.3 for $\varepsilon = 0.06$. This example is used in Fig. 9 to further explore the potential of T-GRAND and SD to converge to a correct estimate of the error matrix.

Fig. 9 considers both T-GRAND and SD and depicts the probability that each method will correctly estimate the $L \times B$ matrix $\mathbf{E}_{\bar{R}}$ if $N - N_R$ or fewer packets are received in error,

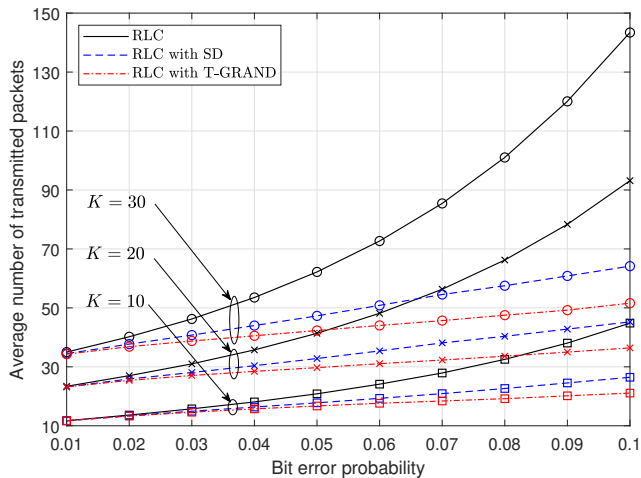


Fig. 8. Comparison of the average number of transmitted packets required by RLC decoding, RLC decoding with SD, and RLC decoding with T-GRAND to recover $K \in \{10, 20, 30\}$ source packets with decoding probability 100% for a varying bit error probability ε . The average length of error bursts and the packet length have been set to $\Lambda = 4$ and $B = 64$, respectively.

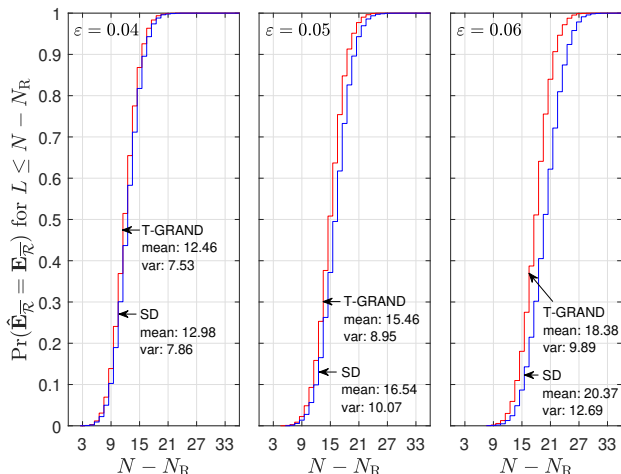


Fig. 9. Probability that the $L \times B$ estimated error matrix matches the actual error matrix when $L \leq N - N_R$, where $N - N_R$ is the number of packets received in error. The bit error probability ε takes the values 0.04 (left), 0.05 (middle) and 0.06 (right). The remaining parameters have been set as follows: $K = 20$, $\Lambda = 4$ and $B = 64$.

i.e., $L \leq N - N_R$, for $K = 20$ and $\varepsilon \in \{0.04, 0.05, 0.06\}$. The three subfigures illustrate that T-GRAND can accurately estimate $\mathbf{E}_{\bar{\mathcal{R}}}$ with a progressively higher chance than SD for a given number of erroneous received packets, as the value of ε increases from 0.04 to 0.06. Fig. 9 also establishes that T-GRAND needs fewer received packets than SD to compute an accurate estimate of $\mathbf{E}_{\bar{\mathcal{R}}}$ with the same probability as SD. Therefore, the results in Fig. 9 tally with the results in Fig. 8, which exhibited that T-GRAND requires fewer packet transmissions than SD to recover the source packets.

C. Impact of T-GRAND on the completion time

Recall that SD assumes that the channel is memoryless, even when it has memory, as is the case of this paper. As explained in Section III, the order with which SD generates and

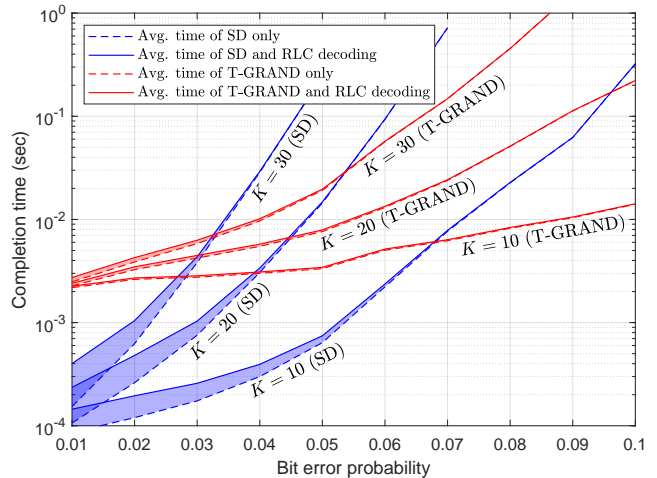


Fig. 10. Completion time of T-GRAND and SD, and overall completion time of a decoder that combines RLC decoding with T-GRAND or SD. Shaded regions represent the time added by RLC decoding to the overall completion time. The system configuration is identical to that considered in Fig. 8.

queries error vectors for the column-wise estimation of $\mathbf{E}_{\bar{\mathcal{R}}}$ is fixed and independent of the channel parameters. T-GRAND incorporates an additional stage, which uses knowledge of the channel model to inform the ordering process, thus making the estimation of $\mathbf{E}_{\bar{\mathcal{R}}}$ dependent on the channel parameters. This stage has given T-GRAND an advantage over SD in terms of the decoding probability (Figs. 5-7) and the average number of transmitted packets (Fig. 8). This section investigates the impact of this additional stage on the overall time complexity of T-GRAND and contrasts it to that of SD. The T-GRAND implementation that employs CalcProbAndSort has been considered to ascertain if a complexity gain over SD can be achieved, even when T-GRAND uses the more computational expensive of the two proposed ordering algorithms.

Fig. 10 compares the average completion time³ of the two error correction methods, that is, T-GRAND and SD. Solid lines represent the overall completion time, which includes the time taken by the initial attempt of RLC decoding to recover the source packets, the execution of the error correction method, and the subsequent successful RLC decoding of the error-free and repaired received packets. Dashed lines depict the average completion time of each error correction method, while shaded areas draw attention to the time added by the two RLC decoding operations to the overall completion time.

If Fig. 10 is looked at in combination with Fig. 8, we observe that, for small values of K and ε , SD and T-GRAND have a marginal difference in the average number of transmitted packets required to recover the source packets. This is because the pre-set ordering of SD and the adaptive ordering of T-GRAND generate similar outputs, although the adaptive capability of the latter method incurs a complexity cost. As the values of K and ε increase, adaptive ordering starts to exhibit a clearer advantage over pre-set ordering, as T-GRAND requires

³The simulations were carried out on a 64-bit Microsoft Windows 10 system using an Intel i7-1370P processor running at 1.9 GHz and 32 GB of RAM. Standard sequential programming in MATLAB was used for both SD and T-GRAND, that is, parallelization was not explored in this paper.

fewer packet transmissions than SD. Although this advantage of T-GRAND comes at the expense of added complexity, Fig. 10 shows that the added computational cost is offset by the quicker convergence to a solution. Indeed, the pre-set ordering of SD results in the generation and querying of more error vectors than T-GRAND before an appropriate vector is identified for each column of $\hat{\mathbf{E}}_{\overline{R}}$. Therefore, for increasing values of K and ε , T-GRAND has the potential to also achieve a shorter completion time than SD, in addition to the improved decoding probability and the reduced packet transmissions.

D. Discussion

The results presented in this section established that T-GRAND has the potential to enhance the performance of an RLC decoder, that is, improve the probability of a destination node recovering the source packets, or reduce the required number of packet transmissions. The proliferation of RLC implementations, such as fountain coding and network coding, in vehicular networks was mentioned in Section I. RLC has been promoted as a way to combat intermittent connections, e.g., due to the high-speed movement of cars, or the sparse and highly dynamic traffic. For example, RLC has been proposed for comfort applications (e.g., multimedia streaming [37]), safety applications, (e.g., broadcast of post-crash warning messages [38]), and cooperative content distribution (e.g., sharing information about hazards [39]). In vehicle-to-vehicle scenarios, T-GRAND could improve the reliability of applications that rely on RLC decoding in adverse channel conditions. When RLC is used on vehicle-to-infrastructure links, T-GRAND could be integrated into the RLC decoder of roadside units to allow vehicles to transmit fewer packets without compromising reliability; therefore, T-GRAND could contribute to the improvement of the energy efficiency of vehicular communications and the reduction of the packet backlog at roadside units.

IX. CONCLUSIONS AND FUTURE WORK

In this paper, we built on the ‘guessing random additive noise decoding’ (GRAND) concept and developed transversal GRAND, a hard detection decoder that leverages correlations between channel errors and works in tandem with the decoder of a random linear code (RLC) at layers higher than the physical layer. Transversal GRAND generates and queries error vectors in order of likelihood, and attempts to correct bits that occupy the same position in erroneous received packets. After all bit positions have been considered, RLC decoding utilizes both repaired and correctly received packets to recover the original information packets. Calculation and sorting of the likelihood values of all error vectors for each bit position is a simple but computationally expensive procedure. For this reason, we also presented a computationally efficient method, which identifies the most likely error vectors without computing and ordering their likelihoods. Simulation results demonstrated that transversal GRAND has the potential to markedly improve the performance of RLC decoding when the channel introduces error bursts in transmitted packets.

The work presented in this paper has paved the way for new research directions, which we shall briefly review here. The first direction is the derivation of an expression for the probability of an RLC decoder recovering the source packets, when it is assisted by syndrome decoding or transversal GRAND. The decoding probability is directly related to the probability that the estimated error matrix is correct or, equivalently, each of the columns of the estimated error matrix has been identified correctly. The work by Galligan *et al.* [40], which was inspired by Forney’s work on error exponents [41], could be used to obtain an approximation of the likelihood that a column of the estimated error matrix is correct. The chain rule can then be applied to obtain an approximation of the joint probability that all columns of the error matrix have been estimated correctly. Of course, a product of approximated likelihoods will amplify the overall approximation error, but the product could still provide a useful bound on the probability that the estimated error matrix is accurate.

Another direction is the extension of transversal GRAND to RLCs over $\mathbb{F}_{2^q} = \{0, 1, \dots, 2^q - 1\}$ for $q \geq 1$. This would involve the replacement of the binary Gilbert-Elliott channel model by a non-binary model, e.g., [42], and the derivation of a multidimensional function that describes the probability of occurrence of each error vector, instead of the two-dimensional function given in (11). The procedure for identifying the most likely error vectors would then have to be configured to trace a path of ordered negative log-probability values, not in a two-dimensional matrix as in Fig. 4, but in a multidimensional matrix. The generalization of transversal GRAND for RLCs over \mathbb{F}_{2^q} will allow comparison of its performance with that of other non-binary schemes, e.g., PRAC [8].

A third, and equally interesting direction, is the investigation of transversal GRAND from an implementation perspective. Abbas *et al.* [43] compiled a treatise on hardware architectures of GRAND variants for the physical layer. Although [43] could inform aspects of the design of transversal GRAND, random linear coding has been standardized as application-layer forward error correction (AL-FEC) for multimedia broadcast and multicast services (MBMS) [44]. Thus, the efficiency of a software-based implementation will depend on the underlying hardware architecture, e.g., ARM or x86, and on the chosen programming language, that is, its capability to directly access hardware resources and its flexibility to explore parallelization.

REFERENCES

- [1] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, “A random linear network coding approach to multicast,” *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [2] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, “A digital fountain approach to reliable distribution of bulk data,” in *Proc. ACM SIGCOMM*, Vancouver, Canada, Sep. 1998.
- [3] Q. Huang, J. Li, Y. Zhu, and D. Wu, “Trickle irrigation: Congestion relief for communication with network coding,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8695–8708, Sep. 2018.
- [4] J. Cloud, F. D. P. Calmon, W. Zeng, G. Pau, L. M. Zeger, and M. Médard, “Multipath-TCP with network coding for mobile devices in heterogeneous networks,” in *Proc. IEEE 78th Veh. Technol. Conf.*, Las Vegas, NV, USA, Sep. 2013.
- [5] Y. Cui, L. Wang, X. Wang, H. Wang, and Y. Wang, “FMTC: A fountain code-based multipath transmission control protocol,” *IEEE/ACM Trans. Netw.*, vol. 23, no. 2, pp. 465–478, Apr. 2015.

- [6] G. Giambene, D. K. Luong, T. de Cola, V. A. Le, and M. Muhammad, "Analysis of a packet-level block coding approach for terrestrial-satellite mobile systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8117–8130, Aug. 2019.
- [7] J. Chen, H. Wu, P. Yang, F. Lyu, and X. Shen, "Cooperative edge caching with location-based and popular contents for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10291–10305, Sep. 2020.
- [8] G. Angelopoulos, A. P. Chandrakasan, and M. Médard, "PRAC: Exploiting partial packets without cross-layer or feedback information," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Sydney, Australia, Jun. 2014.
- [9] M. S. Mohammadi, Q. Zhang, and E. Dutkiewicz, "Reading damaged scripts: Partial packet recovery based on compressive sensing for efficient random linear coded transmission," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3296–3310, Aug. 2016.
- [10] K. R. Duffy, J. Li, and M. Médard, "Guessing noise, not code-words," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018.
- [11] —, "Capacity-achieving guessing random additive noise decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4023–4040, Jul. 2019.
- [12] A. Cohen, R. G. L. D'Oliveira, K. R. Duffy, J. Woo, and M. Médard, "AES as error correction: Cryptosystems for reliable communication," Sep. 2022. [Online]. Available: <https://arxiv.org/pdf/2203.12047>
- [13] K. Galligan, A. Solomon, A. Riaz, M. Médard, R. T. Yazicigil, and K. R. Duffy, "IGRAND: Decode any product code," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021.
- [14] K. R. Duffy, "Ordered reliability bits guessing random additive noise decoding," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, Jun. 2021.
- [15] C. Condo, "Iterative soft-input soft-output decoding with ordered reliability bits GRAND," Jul. 2022. [Online]. Available: <https://arxiv.org/abs/2207.06691>
- [16] W. An, M. Médard, and K. R. Duffy, "Keep the bursts and ditch the interleavers," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 3655–3667, Jun. 2022.
- [17] I. Chatzigeorgiou and F. A. Monteiro, "Symbol-level GRAND for high-order modulation over flat fading channels," Jul. 2022. [Online]. Available: <https://arxiv.org/abs/2207.07748>
- [18] S. M. Abbas, M. Jalaeddine, and W. J. Gross, "GRAND for Rayleigh fading channels," in *Proc. IEEE Global Commun. Conf. Workshops (GC Wkshps)*, Rio de Janeiro, Brazil, Dec. 2022.
- [19] H. Srieddeen, M. Médard, and K. R. Duffy, "GRAND for fading channels using pseudo-soft information," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Rio de Janeiro, Brazil, Dec. 2022.
- [20] S. M. Abbas, T. Tonnellier, F. Ercan, M. Jalaeddine, and W. J. Gross, "High-throughput and energy-efficient VLSI architecture for ordered reliability bits GRAND," *IEEE Trans. VLSI Syst.*, vol. 30, no. 6, pp. 681–693, Jun. 2022.
- [21] M. Zhu, D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, "Decoder error propagation mitigation for spatially coupled ldpc codes," in *Proc. Int. Symp. on Inform. Theory and its Applications (ISITA)*, Kapolei, USA, Oct. 2020.
- [22] I. Chatzigeorgiou, "Transversal GRAND for network coded data," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Espoo, Finland, Jun. 2022.
- [23] J. Jin, B. Li, and T. Kong, "Is random network coding helpful in WiMAX?" in *Proc. IEEE 27th Conf. on Comp. Commun. (INFOCOM)*, Phoenix, AZ, USA, Apr. 2008.
- [24] D. L. Donoho, "Sparse components of images and optimal atomic decompositions," *Constr. Approx.*, vol. 17, no. 1, pp. 353–382, Jan. 2001.
- [25] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, no. 5, pp. 1253–1266, Sep. 1960.
- [26] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 5, pp. 1977–1997, Sep. 1963.
- [27] J. Cloud and M. Médard, "Network coding over SATCOM: Lessons learned," in *Proc. Int. Conf. on Wireless and Satellite Systems (WiSATS)*, Bradford, UK, Jul. 2015.
- [28] B. Jun, P. Yang, J.-S. No, and H. Park, "New fountain codes with improved intermediate recovery based on batched zigzag coding," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 23–36, Jan. 2017.
- [29] J. Huang, Z. Fei, C. C. No, M. Xiao, and J. Yuan, "Weighted online fountain codes with limited buffer size and feedback transmissions," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 7960–7973, Dec. 2021.
- [30] N. Aboutorab, P. Sadeghi, and S. Sorour, "Enabling a tradeoff between completion time and decoding delay in instantly decodable network coded systems," *IEEE Trans. Commun.*, vol. 62, no. 4, pp. 1296–1309, Apr. 2014.
- [31] R. D. Cideciyan, S. Furrer, and M. A. Lantz, "Performance of interleaved block codes with burst errors," *IEEE Trans. Magn.*, vol. 55, no. 3, pp. 1–5, Mar. 2019.
- [32] T. Rancurel, D. Roviras, and F. Castanié, "Expression of the capacity for the Gilbert channel in presence of interleaving," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, USA, May 2001.
- [33] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.
- [34] J. W. J. Williams, "Algorithm 232 – Heapsort," *Communications of the ACM*, vol. 7, no. 6, pp. 347–348, Jun. 1964.
- [35] W. Chen, F. Lu, and Y. Dong, "The rank distribution of sparse random linear network coding," *IEEE Access*, vol. 7, pp. 43806–43819, Mar. 2019.
- [36] R. Su, Q. T. Sun, Z. Zhang, and Z. Li, "Completion delay of random linear network coding in full-duplex relay networks," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 7843–7857, Dec. 2022.
- [37] S. Yousefi, T. Chahed, S. M. M. Langari, and K. Zayer, "Comfot applications in vehicular ad hoc networks based on fountain coding," in *Proc. IEEE 71st Veh. Technol. Conf.*, Taipei, Taiwan, May 2010.
- [38] N. F. Abdullah, A. Doufexi, and R. J. Piechocki, "Car-to-car safety broadcast with interference using raptor codes," in *Proc. IEEE 73rd Veh. Technol. Conf.*, Budapest, Hungary, May 2011.
- [39] S. Ahmed and S. S. Kanhere, "VANETCODE: Network coding to enhance cooperative downloading in vehicular ad-hoc networks," in *Proc. Int. Conf. on Wireless Comm. and Mobile Comp. (IWCMC)*, Vancouver, Canada, Jul. 2006.
- [40] K. Galligan, P. Yuan, and K. R. D. Muriel Médard, "Upgrade error detection to prediction with GRAND," May 2023. [Online]. Available: <https://arxiv.org/abs/2305.05777>
- [41] G. Forney, "Exponential error bounds for erasure, list, and decision feedback channels," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.
- [42] W. Griffiths, H.-J. Zepernick, and M. Caldera, "APP decoding of non-binary block codes on Gilbert-Elliott channels using generalized weight polynomials," in *Proc. IEEE Int. Symp. on Inf. Theory and its Appl. (ISITA)*, Auckland, New Zealand, Dec. 2008.
- [43] S. M. Abbas, M. Jalaeddine, and W. J. Gross, *Guessing random additive noise decoding: A hardware perspective*. Switzerland: Springer, 2023.
- [44] *Technical specification group services and system aspects, multimedia broadcast/multicast services, architecture and functional description, TS 23.246, V6.9.0, 3GPP Std.*, Dec. 2005.



Ioannis Chatzigeorgiou (M'05-SM'15) received the Dipl.Ing. degree in Electrical Engineering from Democritus University of Thrace, Greece, the M.Sc. degree in Satellite Communication Engineering from the University of Surrey, UK, and the Ph.D. degree from the University of Cambridge, UK. He is currently a Senior Lecturer at Lancaster University, UK. Prior to his appointment, he held postdoctoral positions at the University of Cambridge and the Norwegian University of Science and Technology (NTNU). His research interests include signal processing, coding theory and performance analysis of communication systems.



Dmitry Savostyanov received B.Sc. and M.Sc. degrees in Applied Physics and Mathematics from the Moscow Institute of Physics and Technology, and the Ph.D. degree from the Institute of Numerical Mathematics of the Russian Academy of Sciences. He is currently a Lecturer in Applied Mathematics with the School of Mathematics, Statistics and Actuarial Science at the University of Essex, UK. His research interests include tensor product algorithms for high-dimensional problems, with applications to quantum control, epidemics on networks, and

fractional differential equations.