Serial List Viterbi Decoding with CRC: Managing Errors, Erasures, and Complexity

Hengjie Yang, Sudarsan V. S. Ranganathan, and Richard D. Wesel
Department of Electrical and Computer Engineering
University of California, Los Angeles, Los Angeles, CA 90095, USA
Email: {hengjie.yang, sudarsanvsr, wesel}@ucla.edu

***Abstract*—This paper analyzes the serial list Viterbi algorithm (S-LVA) used in conjunction with optimal CRC codes that minimize probability of undetected error by maximizing the minimum distance between convolutional codewords that pass the CRC check, following Lou *et al.* In particular, the paper identifies such optimal CRC codes for the 3GPP standard convolutional code (561,753). As SNR varies and the maximum list size L ranges from one to its maximum, this paper uses bounds, approximations, and simulation to characterize decoding complexity and the trade-off between erasure probability and undetected error probability. The complexity of S-LVA is captured by the expected value of the number of decoding attempts required before a CRC check passes or L codewords have been examined. For S-LVA with a degree-m CRC and maximum possible L, which is the cardinality of the set of all possible convolutional codewords, the expected value of the number of decoding attempts converges to one as SNR increases and to $2^m(1-\epsilon)$, for a small $\epsilon > 0$, as SNR decreases. For S-LVA with the maximum possible L, the erasure probability is zero. As L decreases from this maximum, the erasure probability increases and the UE probability decreases to that of $L = 1$, for which UE probability is well approximated by a nearest-neighbor bound.**

I. Introduction

Cyclic redundancy check (CRC) codes [1] are commonly used as the outer error-detection code for an inner error-correction code. At the receiver, a CRC code is used to protect against undetected errors (UEs) of the error-correction code. Koopman and Chakravarty [2] list the commonly used CRC codes up to degree 16. The designs in [2] as with most CRC designs, assume that the CRC decoder operates on a binary symmetric channel (BSC), whereas in reality the CRC decoder sees message sequences whose likelihoods depend on the codeword structure of the inner code.

For an inner convolutional code (CC), Lou *et al.* [3], for the first time, studied the design of a CRC code specifically for the inner CC. The authors presented two methods to obtain an upper bound on the UE probability of any CRC-CC pair. These methods were called the exclusion method and the construction method. A greedy CRC code search algorithm was proposed by using the fact that when the frame error rate (FER) is low, UEs with the smallest Hamming distance dominate performance. Using this search algorithm, the authors in [3] obtained CRC codes that minimize the UE probability, $\mathrm{P_{UE}}$. As an example, for a commonly used 64-state CC with 1024 information bits, the optimized CRC code typically requires 2 fewer bits to achieve a target $\mathrm{P_{UE}}$ or to reduce the $\mathrm{P_{UE}}$ by orders of magnitude (at high SNR) over the performance of standard CRC codes with the same degree.

The list Viterbi algorithm (LVA) [4] produces an ordered list of the L most likely transmitted codewords. Parallel LVA produces these L codewords all at once. Serial LVA (S-LVA) produces codewords one at a time until the CRC check passes; see Seshadri and Sundberg [5]. Several implementations of fast LVAs have appeared in literature [5]–[8]. Soong and Huang [6] proposed an efficient tree-trellis algorithm (TTA), which is a serial LVA, initially used for speech recognition. Roder and Hamzaoui [8] then improved the TTA by using several unsorted lists to eventually provide the list of L best sequences, allowing the TTA to achieve linear time complexity with respect to the list size. Wang *et al.* [9] proposed using the parity-check matrix of the CRC generator polynomial to assist decoding in a convolutionally coded system. If the soft Viterbi decoding fails, the CRC-CC pair is jointly decoded iteratively until a codeword passes the CRC check.

In this paper, we consider S-LVA combined with the optimal CRC code designed using [3] specifically for a given CC. The paper begins by presenting the optimal 8, 12, and 16-bit CRC codes for a CC in the 3GPP standard [10].

The list size L determines the maximum number of codewords that S-LVA will check. It ranges from one to $|\mathcal{C}|$, where $|\mathcal{C}|$ is the cardinality of the set $\mathcal{C}$ that includes every possible convolutional codeword. As L ranges from one to $|\mathcal{C}|$, this paper uses bounds, approximations, and simulations to characterize the trade-off between two probabilities: the erasure probability $\mathrm{P}^L_{\mathrm{NACK}}$, when no codeword passes the CRC check producing a negative acknowledgement (NACK) and the UE probability $\mathrm{P}^L_{\mathrm{UE}}$ when an incorrect codeword passes the CRC. For S-LVA with $L = |\mathcal{C}|$, $\mathrm{P}^{|\mathcal{C}|}_{\mathrm{NACK}} = 0$. As the list size L decreases, $\mathrm{P}^L_{\mathrm{NACK}}$ increases and $\mathrm{P}^L_{\mathrm{UE}}$ decreases.

The complexity of S-LVA is captured by the expected value of the number N_{LVA} of decoding attempts required before a CRC check passes or all L codewords identified by S-LVA fail the CRC check. For a degree-m CRC code and S-LVA with $L = |\mathcal{C}|$, we show that $\mathbb{E}[N_{\mathrm{LVA}}] \to 1$ as SNR increases and $\mathbb{E}[N_{\mathrm{LVA}}] \to 2^m(1-\epsilon)$, for a small $\epsilon > 0$ as SNR decreases.

The paper is organized as follows: Sec. II briefly describes the CC-specific CRC design methodology of [3] and presents

Research supported by National Science Foundation (NSF) grant CCF-1618272. Any opinions, findings, and conclusions or recommendations are those of the author(s) and do not necessarily reflect the views of the NSF.

978-1-5386-4727-1/18/\$31.00 ©2018 IEEE

the optimal 8, 12, and 16-bit CRC codes for a CC in the 3GPP standard. Sec. III introduces S-LVA combined with the CC-optimal CRC code and analyzes its performance and complexity as SNR varies. Sec. IV studies the trade-off between $\mathrm{P}_{\mathrm{NACK}}^{L}$ and $\mathrm{P}_{\mathrm{UE}}^{L}$ for S-LVA as a function of the list size L at high SNR values. Sec. V concludes the paper.

II. Convolutional-Code-Specific CRC Design

This section presents the design methodology of [3] for designing the optimal CRC code for CCs. We use this design method to obtain, as a new result, the optimal 8, 12, and 16-bit CRC codes for the $(561, 753)$ CC in 3GPP standard.

A. System Model: CC with Soft Viterbi and CRC

A transmitter uses a CC and a CRC code to transmit an information sequence as follows: Let $f(x)$ denote a k-bit binary information sequence and $p(x)$ denote a degree-m CRC generator polynomial. Let $r(x)$ denote the remainder when $x^m f(x)$ is divided by $p(x)$. First, the CRC polynomial is used to obtain the $n = k+m$-bit sequence $x^m f(x)+r(x) = q(x)p(x)$. The transmitter then uses a feedforward, rate-$\frac{1}{N}$ CC with v memory elements and a generator polynomial $\boldsymbol{c}(x)$ to encode the n-bit sequence. The output $q(x)p(x)\boldsymbol{c}(x)$ of the convolutional encoder is transmitted over an additive white Gaussian noise (AWGN) channel using quadrature phase-shift keying (QPSK) modulation.

The receiver feeds the noisy received sequence into a soft Viterbi decoder that identifies the most likely n-bit input sequence to the convolutional encoder. The CRC decoder checks the n-bit codeword resulting from Viterbi decoding. An undetected error occurs when the Viterbi decoder identifies an incorrect codeword which then passes the CRC check. Since the CRC decoder only checks one codeword, the UE probability is exactly $\mathrm{P}_{\mathrm{UE}}^{1}$ (the "1" refers to the list size L in the context of S-LVA, which we will discuss in Sec. III).

B. Undetected Error Probability via Exclusion Method

Assume that a codeword erroneously identified by the Viterbi decoder is $q(x)p(x) + e(x)$, where $e(x) \neq 0$ and is divisible by $p(x)$. The undetected error probability $\mathrm{P}_{\mathrm{UE}}^{1}$ is the probability that such an $e(x)$ occurs with the most likely codeword identified by standard Viterbi algorithm. The exclusion method enumerates all possible error patterns of the CC and excludes those patterns that are detectable by the CRC decoder (i.e. not divisible by $p(x)$). Thus, the probability of selecting one of the remaining undetectable error patterns is equal to $\mathrm{P}_{\mathrm{UE}}^{1}$, which is upper bounded in [3] via a union bound on codeword error.

1) Undetectable Single Error: An error event occurs when the decoded trellis path leaves the encoded trellis path once and rejoins it once. The undetectable single error probability $\mathrm{P}_{\mathrm{UE},1}^{1}$ is upper bounded by the union bound $\hat{\mathrm{P}}_{\mathrm{UE},1}^{1}$,

$$\mathrm{P}_{\mathrm{UE},1}^{1} \leq \sum_{d=d_{\mathrm{free}}}^{\infty} \sum_{i=1}^{a_d} \mathbb{1}_{\{p(x)|e_{d,i}(x)\}} \cdot \max\{0, n+v-l_{d,i}+1\}\, \mathrm{P}(d), \tag{1}$$

where a_d is the number of distinct undetectable single error events with CC output Hamming distance d, $\mathbb{1}_{\{\cdot\}}$ is the indicator function, $e_{d,i}$ is the i^{th} error event with distance d and length $l_{d,i}$, and $\mathrm{P}(d)$ is the pairwise error probability of an error event with distance d. For QPSK modulation over the AWGN channel, $\mathrm{P}(d)$ can be computed using the Gaussian Q-function:

$$\mathrm{P}(d) = \mathrm{Q}(\sqrt{d\gamma_s}) \leq \mathrm{Q}(\sqrt{d_{\mathrm{free}}\,\gamma_s})e^{-(d-d_{\mathrm{free}})\gamma_s/2}, \tag{2}$$

where $\gamma_s = E_s/N_0$ is the signal-to-noise ratio (SNR) of a QPSK symbol, and E_s and $N_0/2$ denote the energy per transmitted QPSK symbol and one-dimensional noise variance, respectively[1].

A computation-friendly approximation of (1) is given as

$$\mathrm{P}_{\mathrm{UE},1}^{1} \leq \sum_{d=\tilde{d}+1}^{\infty} n a_d\, \mathrm{P}(d) + \sum_{d=d_{\mathrm{free}}}^{\tilde{d}} \sum_{i=1}^{a_d} \mathbb{1}_{\{p(x)|e_{d,i}(x)\}} \cdot (n+v-l_{d,i}+1)^{+}\, \mathrm{P}(d), \tag{3}$$

where $(\cdot)^{+}$ is the positive part and $\tilde{d}$ is a threshold selected to approximate (1). See [3] for more details.

2) Undetectable Double Error: An undetectable double error involves two disjoint error events. The undetectable double error probability $\mathrm{P}_{\mathrm{UE},2}^{1}$ can be upper bounded by the union bound $\hat{\mathrm{P}}_{\mathrm{UE},2}^{1}$ derived in [3], equation (4), as well. We refer the reader to [3] for more details.

3) Undetected Error Probability: In general, the UE probability $\mathrm{P}_{\mathrm{UE}}^{1}$ is upper bounded by the sum of the probability of arbitrary number of disjoint error event combinations, which can be generalized as follows:

$$\mathrm{P}_{\mathrm{UE}}^{1} \leq \sum_{s=1}^{\infty} \hat{\mathrm{P}}_{\mathrm{UE},s}^{1}\,, \tag{4}$$

where $\hat{\mathrm{P}}_{\mathrm{UE},s}^{1}$ refers to the union bound on the s-tuple undetectable error. The corresponding approximation of (4) given in [3] uses a maximum search depth threshold of $\tilde{d}$. Due to this limitation, the exclusion method is only useful when undetectable errors with distance $d \leq \tilde{d}$ dominate.

In this paper, all truncated union bounds on UE probability are plotted by considering only the undetectable errors with $d \leq \tilde{d}$ and neglecting the remainder terms.

C. Convolutional-Code-Specific CRC Code Search Algorithm

When FER is low, UE probability is dominated by the UEs with the smallest distance. Thus, the algorithm to find an optimal CRC code for a given CC focuses on UEs with the smallest distance. Since the coefficients of x^m and x^0 are both 1, there are 2^{m-1} candidate degree-m CRC generator polynomials. The CRC code search starts with $d = d_{\mathrm{free}}$ and updates the candidate list by keeping only the CRC generator polynomials with the fewest undetectable errors at

[1] In [3], there is a typo in the expression for equation (2) that includes erroneously a factor of two in the square root.

TABLE I
STANDARD CRC CODES VERSUS OPTIMAL CRC CODES FOR CC $G = (561, 753)$ WITH $n = 504$ BITS

Name	Gen. Poly.	Undetected Error Distance Spectrum A_d					
	d	16	18	20	22	24	26
Standard-8	0x19B	983	4387	19909	105000	672724	3972970
Optimal-8	0x19D	0	979	22349	111304	686314	3830340
Standard-12	0x180F	0	0	969	5815	42893	245211
Optimal-12	0x108B	0	0	0	4793	45795	246729
Standard-16	0x11021	0	0	484	0	1765	14752
Optimal-16	0x1F8FD	0	0	0	0	0	13240

that distance. The search then progressively increases d and repeats the same procedure until only one polynomial remains in the candidate list, the polynomial with the best "truncated" distance spectrum only considering distances up to $\tilde{d}$. Note that when $d \geq 2\, d_{\text{free}}$, undetectable double errors also have to be considered. In [3], the authors point out that the algorithm usually terminates before $d = 3\, d_{\text{free}}$.

D. Optimal CRC Codes for 3GPP Standard

We now present the optimal CRC codes designed for the $(561, 753)$ CC that is specified in the 3GPP TS 25.212 version 7.0.0 Release 7 document [10]. The technical specification document provides 8, 12, and 16-bit CRC codes. These CRC codes and the corresponding newly designed optimal CRC codes are given in Table I. The table presents the UE distance spectrum A_d (the number of distinct UEs at distance d with positions taken into account) for each CRC. The CRC code generator polynomials are represented in hexadecimal and d represents the CC output Hamming distance. Fig. 1 shows the histogram of UE distance spectrum in Table I. For a degree-m CRC code in the 3GPP standard, the minimum distance d_{CRC} at which UEs first occur is always smaller than the corresponding degree-m optimal CRC code.

Fig. 2 shows the nearest neighbor approximations (NNAs) on P_{UE}^{1} in (13) for each CRC code as well as the NNA of (11) on $\text{P}_{\text{NACK}}^{1}$. Also shown is the NNA of the FER of soft Viterbi decoding with no CRC code. As shown in Sec. III, these NNAs are tight at high SNR. The figure shows that considerable improvement in P_{UE}^{1} can be obtained with CRC codes designed specifically for the 3GPP $(561, 753)$ CC.

III. S-LVA PERFORMANCE VS. SNR

S-LVA begins by finding the closest codeword c_1 to the received sequence and passing it to the CRC code for verification. If the CRC check fails, S-LVA outputs the next closest codeword c_2 and repeats the above procedure until the CRC check is successful or the best L codewords $c_1, \ldots c_L$ all fail the CRC check, in which case the decoder declares erasure and a NACK is generated. The failure rate of S-LVA is

$$\text{P}_{\text{F}}^{L} = \text{P}_{\text{UE}}^{L} + \text{P}_{\text{NACK}}^{L}\,. \tag{5}$$

This section examines S-LVA performance as a function of SNR (E_s/N_0), where performance metrics include P_{F}^{L}, P_{UE}^{L}, $\text{P}_{\text{NACK}}^{L}$, and $\mathbb{E}[N_{\text{LVA}}]$. The extreme cases of SNR (very low and very high) and list size ($L = 1$ and $L = |\mathcal{C}|$) are given

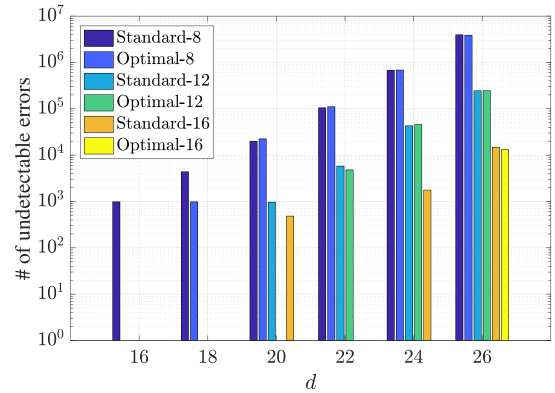

Fig. 1. Undetected error distance spectrum of standard CRC codes versus optimal CRC codes for $(561, 753)$ CC with $n = 504$ bits.

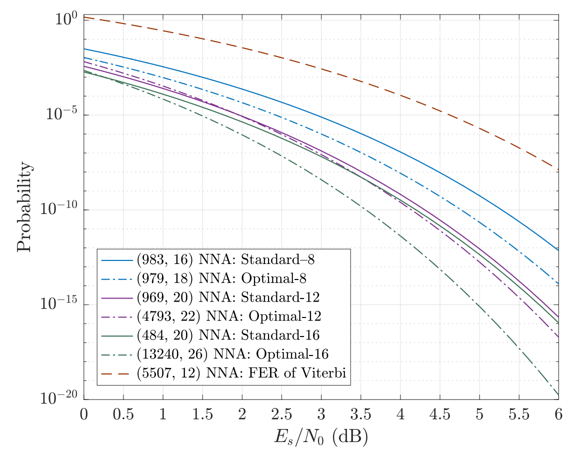

Fig. 2. $(A_{d_{\text{CRC}}},\ d_{\text{CRC}})$ nearest neighbor approximation (NNA) on P_{UE}^{1} of standard CRC codes versus optimal CRC codes for $(561, 753)$ CC with $n = 504$ bits. $(A_{d_{\text{free}}},\ d_{\text{free}})$ NNA on FER of soft Viterbi decoding with no CRC code for the same CC is also given as a reference.

particular attention as they frame the overall performance landscape.

In the discussion below, certain sets of codewords are important to consider. First, $\mathcal{C}$ is the set of all convolutional codewords. Since we consider a finite blocklength system where there are n message bits and v termination bits (completely determined by the $n = k + m$ message bits) fed into the convolutional encoder, the size of $\mathcal{C}$ is

$$|\mathcal{C}| = 2^{n} = 2^{k+m}. \tag{6}$$

Let c^* denote the transmitted codeword. A superscript of $-$ indicates a set that excludes c^*. For example $\mathcal{C}^-$ is the set of all convolutional codewords except the transmitted codeword c^*. The set $\mathcal{C}_{\text{CRC}}$ is the set of all convolutional codewords whose corresponding input sequences pass the CRC check. The size of $\mathcal{C}_{\text{CRC}}$ is

$$|\mathcal{C}_{\text{CRC}}| = 2^{n-m} = 2^{k}. \tag{7}$$

The set $\mathcal{C}_{\overline{\text{CRC}}}$ is the set of all convolutional codewords whose corresponding input sequences *do not* pass the CRC check.

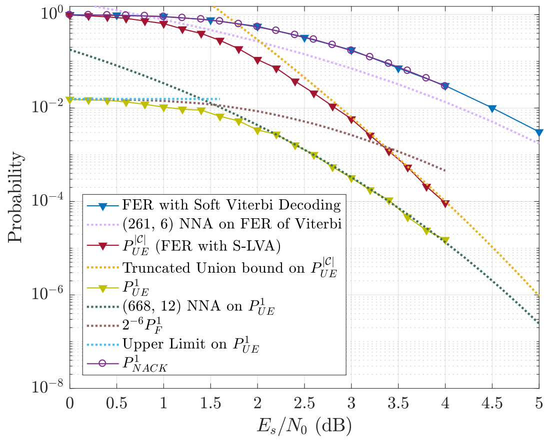

Fig. 3. Comparison of FER between S-LVA combined with the optimal degree-6 CRC code 0x43 and soft Viterbi decoding (without a CRC code) for $(13, 17)$ CC when $n = 256 + 6$ bits. $(261,\ 6)$ NNA on soft Viterbi decoding, truncated union bound at $\tilde{d} = 24$ on $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$, conjecture of $2^{-6}\,\mathrm{P}_{\mathrm{F}}^1$, upper limit of 2^{-6}, and $(668,\ 12)$ NNA on $\mathrm{P}_{\mathrm{UE}}^1$ are also provided as a reference.

The size of this set is

$$|\mathcal{C}_{\overline{\mathrm{CRC}}}| = 2^n - 2^k. \tag{8}$$

A. The Case of $L = |\mathcal{C}|$

Consider S-LVA with the largest possible list size $L = |\mathcal{C}|$. Regardless of SNR, $\mathrm{P}_{\mathrm{NACK}}^{|\mathcal{C}|} = 0$ always holds because S-LVA with $L = |\mathcal{C}|$ will always find a codeword that passes the CRC check. Let A_d be the number of distinct UEs of distance d with positions taken into account. The UE probability $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ is upper bounded by the union bound that some codeword in $\mathcal{C}_{\mathrm{CRC}}^{-}$ is pairwise more likely than c^*:

$$\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|} \leq \sum_{c \in \mathcal{C}_{\mathrm{CRC}}^{-}} \mathrm{P}(d(c, c^*)), \tag{9}$$

where $d(c, c^*)$ is the distance between c and c^*, and $\mathrm{P}(d(c, c^*))$ is defined in (2). This is precisely the union bound of [3] given as an upper bound on $\mathrm{P}_{\mathrm{UE}}^1$, which was presented in Sec. II-B. That it is also a valid upper bound for $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ indicates that, at least at low SNR, this bound will be loose for $L = 1$. At very low SNR, $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ converges to $\frac{|\mathcal{C}_{\mathrm{CRC}}^{-}|}{|\mathcal{C}_{\mathrm{CRC}}|} \approx 1$.

For $k = 256$ bits, Fig. 3 shows $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ as a function of E_s/N_0 for the $(13, 17)$ CC using soft Viterbi decoding without a CRC code and S-LVA with $L = |\mathcal{C}|$ combined with the optimal degree-6 CRC code 0x43. The truncated union bound at $\tilde{d} = 24$ on $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ of (9) derived via exclusion method in Sec. II-B is also shown. It can be seen that the union bound on $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$ becomes tight as SNR increases.

B. The Case of $L = 1$

For $L = 1$, with the same blocklength n, $\mathrm{P}_{\mathrm{F}}^1$ is exactly the FER of the CC under soft Viterbi decoding with no CRC code. The addition of the CRC code separates the failures into erasures and UEs, with probabilities $\mathrm{P}_{\mathrm{NACK}}^1$ and $\mathrm{P}_{\mathrm{UE}}^1$, respectively. Thus we have union bounds, nearest neighbor approximation (NNA), and a low-SNR upper limit as follows:

$$\mathrm{P}_{\mathrm{NACK}}^1 \leq \sum_{c \in \mathcal{C}_{\overline{\mathrm{CRC}}}} \mathrm{P}(d(c, c^*)) \tag{10}$$

$$\approx A_{d_{\mathrm{free}}} \mathrm{P}(d_{\mathrm{free}}), \tag{11}$$

$$\mathrm{P}_{\mathrm{UE}}^1 \leq \sum_{c \in \mathcal{C}_{\mathrm{CRC}}^{-}} \mathrm{P}(d(c, c^*)) \tag{12}$$

$$\approx A_{d_{\mathrm{CRC}}} \mathrm{P}(d_{\mathrm{CRC}}), \tag{13}$$

$$\lim_{\gamma_s \to -\infty} \mathrm{P}_{\mathrm{UE}}^1 = 2^{-m}. \tag{14}$$

Note that (12) is identical to (9), but $\mathrm{P}_{\mathrm{UE}}^1$ should be significantly smaller than $\mathrm{P}_{\mathrm{UE}}^{|\mathcal{C}|}$. Thus we propose an improved bound on $\mathrm{P}_{\mathrm{UE}}^1$ as follows: for a randomly chosen degree-m CRC code and $L = 1$ we expect an incorrectly chosen convolutional codeword to pass the CRC check with probability 2^{-m}. This should be an upper bound on the performance of CRCs optimized according to [3]. Thus we conjecture that

$$\mathrm{P}_{\mathrm{UE}}^1 \leq 2^{-m}\,\mathrm{P}_{\mathrm{F}}^1\ . \tag{15}$$

This upper bound should be loose for well-designed CRCs at high SNR. However, at very low SNR we expect this bound to be tight based on the fact that the upper limit of $\mathrm{P}_{\mathrm{UE}}^1$ satisfies (14). Fig. 3 shows that (15) is accurate at very low SNR and the NNA of $\mathrm{P}_{\mathrm{UE}}^1$ in (13) is quite accurate at high SNR. The parameters of the NNA are $A_{d_{\mathrm{CRC}}} = 668$ and $d_{\mathrm{CRC}} = 12$.

C. Complexity Analysis: $\mathbb{E}[N_{\mathrm{LVA}}]$ at High and Low SNR

In [8], the authors present tables that compare the time and space complexity for different implementations of the LVA. For a fixed blocklength and a specified CC-CRC pair, the decoding complexity of LVA depends mainly on the number of decoding trials performed. Denote by N_{LVA} the random variable indicating the number of decoding trials of S-LVA for a received codeword randomly drawn according to the noise distribution. We show that with list size $|\mathcal{C}|$, the expected value of N_{LVA}, $\mathbb{E}[N_{\mathrm{LVA}}]$, converges to 1 as SNR increases and converges to $2^m(1-\epsilon)$, for a small $\epsilon > 0$ as SNR decreases.

Theorem 1: The expected number of decoding trials $\mathbb{E}[N_{\mathrm{LVA}}]$ for S-LVA with list size $|\mathcal{C}|$, used with a degree-m CRC code, satisfies (i) $\lim_{\gamma_s \to \infty} \mathbb{E}[N_{\mathrm{LVA}}] = 1$; (ii) $\lim_{\gamma_s \to -\infty} \mathbb{E}[N_{\mathrm{LVA}}] = 2^m(1-\epsilon)$, where $\epsilon \to 0$ as $n \to \infty$.

Proof: Let $\tilde{x}_i^n$ denote the i^{th} output of the S-LVA, which is the codeword at position i in the list of all possible codewords sorted according to increasing soft Viterbi metric (typically Hamming or Euclidean distance) with respect to the received noisy codeword.

(i) Consider the event $A_i \triangleq \cap_{j=1}^{i-1}\{p(x) \nmid \tilde{x}_j^n\} \cap \{p(x) \mid \tilde{x}_i^n\}$, where $p(x)$ is the CRC polynomial. Because of the existence of codewords that have $p(x)$ as a factor (i.e. that pass the CRC check), there exists a maximum decoding depth $\tilde{N} < \infty$ such that $\Pr\{A_j\} = 0, \forall j > \tilde{N}$.

Note that when $\gamma_s \rightarrow \infty$, $\Pr\{A_1\} \rightarrow 1$ and $\sum_{i=2}^{\tilde{N}} \Pr\{A_i\} \rightarrow 0$. Thus,

$$\begin{aligned}\lim_{\gamma_s\to\infty} \mathbb{E}[N_{\text{LVA}}] &= \lim_{\gamma_s\to\infty} \left[1\cdot \Pr\{A_1\} + \sum_{i=2}^{\infty} i \Pr\{A_i\}\right] \\ &= \lim_{\gamma_s\to\infty} \left[1\cdot \Pr\{A_1\} + \sum_{i=2}^{\tilde{N}} i \Pr\{A_i\}\right] \\ &\leq \lim_{\gamma_s\to\infty} \left[1\cdot \Pr\{A_1\} + \tilde{N}\sum_{i=2}^{\tilde{N}} \Pr\{A_i\}\right] \\ &= 1. \end{aligned} \tag{16}$$

Since $N_{\text{LVA}} \geq 1$, $\mathbb{E}[N_{\text{LVA}}] \geq 1$. It follows that $\lim_{\gamma_s\to\infty} \mathbb{E}[N_{\text{LVA}}] = 1$.

(ii) When $\gamma_s \rightarrow -\infty$, the SNR is low enough such that with high probability the received sequence $\boldsymbol{y}$ is far away from the entire constellation of all possible sequences that can be transmitted in $\mathbb{R}^n$. This implies that with very high probability $\boldsymbol{y}$ is almost equidistant from all possible convolutional codewords that can be transmitted. For those received sequences almost equidistant from all convolutional codewords, the S-LVA decoding process can be modeled as follows: In a basket of "blue" balls (codewords that pass the CRC check) and "red" balls (codewords that do not pass the CRC check), the S-LVA chooses balls at random without replacement with the objective of stopping when it successfully picks a blue ball. Thus, $\mathbb{E}[N_{\text{LVA}}]$ can be computed using a standard result in combinatorics as follows. For a decoded sequence with n message and parity-check bits and v trailing zero bits, the total number of balls in the basket is $N = 2^n$ and the number of blue balls in the basket is $M = 2^{n-m}$:

$$\begin{aligned}\lim_{\gamma_s\to-\infty} \mathbb{E}[N_{\text{LVA}}] &= 1 + \frac{N-M}{M+1} \\ &= \frac{N+1}{M+1} \\ &= 2^m \left[1 - \frac{2^m-1}{2^m+2^n}\right] \\ &= 2^m(1-\epsilon), \end{aligned} \tag{17}$$

where $\epsilon = \frac{2^m-1}{2^m+2^n} > 0$. When m is fixed, $\lim_{n\to\infty} \mathbb{E}[N_{\text{LVA}}] = 2^m$. ■

Fig. 4 shows empirical $\mathbb{E}[N_{\text{LVA}}]$ for the $(13, 17)$ CC with the optimal CRC codes with degrees ranging from 1 to 6 when $k = 256$ bits. The curves verify Theorem 1; $\mathbb{E}[N_{\text{LVA}}] \rightarrow 1$ as the SNR increases and $\mathbb{E}[N_{\text{LVA}}] \approx 2^m$ as the SNR decreases to very low values. While the result we have obtained in Theorem 1 for the case of $\gamma_s \rightarrow -\infty$ requires very low SNR values for the arguments made to hold, it is interesting to see from the figure that S-LVA behaves similar to random guessing as soon as the SNR value is below the Shannon limit, shown as a vertical line for $m = 1$. (The limits for the other values of m are very close to the limit for $m = 1$).

Theorem 1 studies the limit of $\mathbb{E}[N_{\text{LVA}}]$ in the limit of extremely high and low SNR regimes. In practice, SNRs

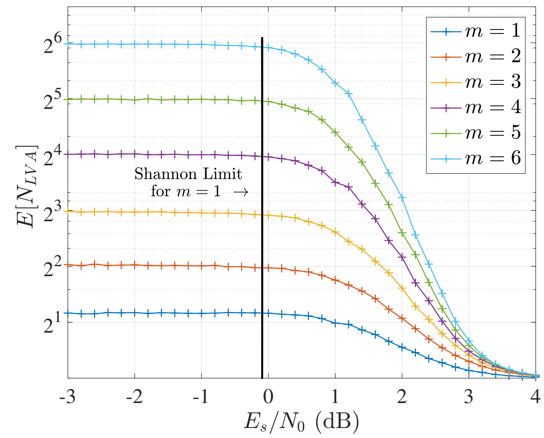

Fig. 4. $\mathbb{E}[N_{\text{LVA}}]$ vs. E_s/N_0 of degree $1-6$ optimal CRC codes for $(13, 17)$ CC, with $k = 256$.

ranging between 0.5 dB and 4 dB above the Shannon limit are of particular interest. As shown in Fig. 4, $\mathbb{E}[N_{\text{LVA}}]$ traverses its full range from $\approx 2^m$ to 1 in this range of practical interest.

IV. S-LVA Performance vs. L

As we learned in Sec. III-A, the "complete" S-LVA algorithm with $L = |\mathcal{C}|$ achieves $\text{P}_{\text{NACK}}^{|\mathcal{C}|} = 0$ and $\text{P}_{\text{UE}}^{|\mathcal{C}|}$ is well approximated by truncating the union bound of (9) at a reasonable $\tilde{d}$. In the context of a feedback communication system, it is often preferable to retransmit a codeword or to lower the rate of the transmission through incremental redundancy rather than to accept undetectable errors. Thus the full complexity $L = |\mathcal{C}|$ may actually lead to detrimental results in certain cases, especially at very low SNRs where $\text{P}_{\text{UE}}^{|\mathcal{C}|}$ approaches 1.

Sec. III-B showed how the other extreme of $L = 1$ significantly lowers the UE probability with P_{UE}^{1} well approximated by the minimum between the upper bound of (15) and the NNA of (13). The reduction in P_{UE} comes at the cost of a significantly increased $\text{P}_{\text{NACK}}^{1}$, which is approximately the FER of the CC decoded by soft Viterbi without a CRC code.

We expect the best choice of L for many systems to be in between these two extremes. The rest of this section explores how P_{UE}^{L} and $\text{P}_{\text{NACK}}^{L}$ vary with L. In general, with SNR fixed, $\text{P}_{\text{NACK}}^{L}$ and P_{UE}^{L} have the following properties: $\text{P}_{\text{NACK}}^{L}$ is a decreasing function of L with $\lim_{L\to|\mathcal{C}|} \text{P}_{\text{NACK}}^{L} = 0$, and P_{UE}^{L} is an increasing function of L with $\lim_{L\to|\mathcal{C}|} \text{P}_{\text{UE}}^{L} = \text{P}_{\text{UE}}^{|\mathcal{C}|}$, which is well approximated by (9).

Therefore, one could ask what the optimal list size L^* is such that, for example, $\text{P}_{\text{NACK}}^{L} \leq \text{P}_{\text{NACK}}^{*}$ and $\text{P}_{\text{UE}}^{L} \leq \text{P}_{\text{UE}}^{*}$, where $\text{P}_{\text{NACK}}^{*}$ and P_{UE}^{*} are target erasure and UE probabilities, respectively. We present useful bounds on $\text{P}_{\text{NACK}}^{L}$ and P_{UE}^{L} to further explore the concept of an optimal list size L^*.

Corollary 1 (Markov bound on $\text{P}_{\text{NACK}}^{L}$): The erasure probability $\text{P}_{\text{NACK}}^{L}$ satisfies $\text{P}_{\text{NACK}}^{L} \leq \frac{1}{L}$ if $\gamma_s \rightarrow \infty$.

Proof: The result is a direct consequence of Markov inequality. The erasure probability with a list size L is given as $\text{P}_{\text{NACK}}^{L} = \Pr\{N_{\text{LVA}} > L\}$, where N_{LVA} is the

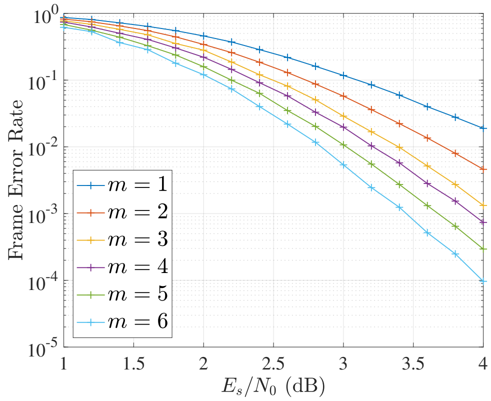

Fig. 5. FER vs. E_s/N_0 of degree $1-6$ optimal CRC codes for $(13, 17)$ CC with $k = 256$.

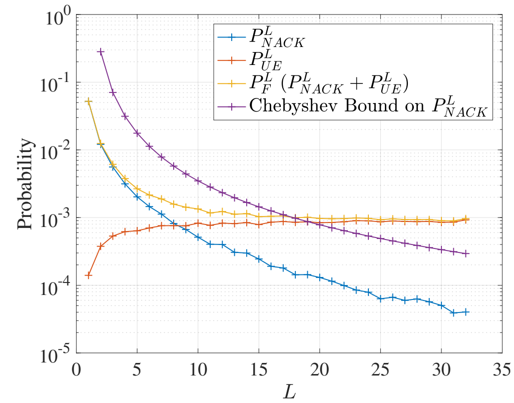

Fig. 6. Trade-off between $\mathrm{P}^L_{\mathrm{NACK}}$ and $\mathrm{P}^L_{\mathrm{UE}}$ for the optimal degree-5 CRC code 0x2D and $(13, 17)$ CC when $k = 256$, $\gamma_s = 3.7$ dB.

random variable representing the decoding trial at which the CRC check first passes. By applying Markov inequality for $\gamma_s \to \infty$, we have

$$\mathrm{P}^L_{\mathrm{NACK}} = \Pr\{N_{\mathrm{LVA}} > L\} \leq \frac{\mathbb{E}[N_{\mathrm{LVA}}]}{L} = \frac{1}{L}. \tag{18}$$

■

A more useful Chebyshev bound on $\mathrm{P}^L_{\mathrm{NACK}}$ could be obtained if one knows the variance $\mathrm{var}(N_{\mathrm{LVA}})$ at high SNR.

Corollary 2 (Chebyshev bound on $\mathrm{P}^L_{\mathrm{NACK}}$*):* Given $\mathrm{var}(N_{\mathrm{LVA}})$ at $\gamma_s \gg 0$, $\mathrm{P}^L_{\mathrm{NACK}}$ satisfies $\mathrm{P}^L_{\mathrm{NACK}} \leq \frac{\mathrm{var}(N_{\mathrm{LVA}})}{(L-1)^2}$, where $L \geq 2$.

Proof: The result is a direct consequence of Chebyshev inequality. Since $\gamma_s \gg 0$, $\mathbb{E}[N_{\mathrm{LVA}}] \to 1$. From Chebyshev inequality, we have

$$\begin{aligned}
\mathrm{P}^L_{\mathrm{NACK}} &= \Pr\{N_{\mathrm{LVA}} > L\} \\
&= \Pr\{N_{\mathrm{LVA}} \geq L + 1\} \\
&\leq \Pr\{|N_{\mathrm{LVA}} - \mathbb{E}[N_{\mathrm{LVA}}]| \geq L - \mathbb{E}[N_{\mathrm{LVA}}] + 1\} \\
&\leq \frac{\mathrm{var}(N_{\mathrm{LVA}})}{(L - (\mathbb{E}[N_{\mathrm{LVA}}] - 1))^2} \\
&\leq \frac{\mathrm{var}(N_{\mathrm{LVA}})}{(L-1)^2}.
\end{aligned} \tag{19}$$

■

We study the trade-off between $\mathrm{P}^L_{\mathrm{NACK}}$ and $\mathrm{P}^L_{\mathrm{UE}}$ for the $(13, 17)$ CC. Assume at $\gamma_s = 3.7$ dB, $\mathrm{P}^*_{\mathrm{NACK}} = 10^{-3}$ and $\mathrm{P}^*_{\mathrm{UE}} = 8 \times 10^{-4}$. In Fig. 5, the FER of degree $1-6$ optimal CRC codes is plotted. Here we use the optimal degree-5 CRC code with the $(13, 17)$ CC to illustrate how to find the optimal list size L^*. Fig. 6 shows the trade-off between $\mathrm{P}^L_{\mathrm{NACK}}$ and $\mathrm{P}^L_{\mathrm{UE}}$ when $k = 256$ at 3.7 dB. It can be seen that $L^* = 8$ satisfies $\mathrm{P}^L_{\mathrm{NACK}} \leq \mathrm{P}^*_{\mathrm{NACK}}$ and $\mathrm{P}^L_{\mathrm{UE}} \leq \mathrm{P}^*_{\mathrm{UE}}$.

If $\mathrm{P}^*_{\mathrm{NACK}} = 10^{-3}$, $\mathrm{P}^*_{\mathrm{UE}} = 10^{-3}$ and empirical $\mathrm{var}(N_{\mathrm{LVA}}) = 0.2823$ is known, since $\mathrm{P}^L_{\mathrm{UE}} \leq \mathrm{P}^*_{\mathrm{UE}}$ always holds, one can directly apply the empirical Chebyshev bound to obtain $L^* \geq 18$ without knowing the true $\mathrm{P}^L_{\mathrm{NACK}}$ curve.

V. Conclusion

This paper first applied Lou *et al.*'s CC-specific CRC code design algorithm to identify the optimal CRC codes for the 3GPP CC $(561, 753)$. The performance and the analysis of the optimal CRC codes showed that the current 3GPP standard CRC codes still leave much room for improvement in terms of their UE probability. We then studied S-LVA with a list size L and a degree-m CRC code as SNR varies and as L ranges from one to its maximum possible value. We characterized the decoding complexity of S-LVA and the trade-off between the erasure probability $\mathrm{P}^L_{\mathrm{NACK}}$ and the undetected error probability $\mathrm{P}^L_{\mathrm{UE}}$. In particular, the paper showed that the expected number of decoding attempts of S-LVA converges to 1 as SNR increases and to $2^m(1-\epsilon)$, for a small $\epsilon > 0$ as SNR decreases. For S-LVA with $L = |\mathcal{C}|$, $\mathrm{P}^{|\mathcal{C}|}_{\mathrm{NACK}} = 0$. Here, $\mathrm{P}^{|\mathcal{C}|}_{\mathrm{UE}}$ is minimized by Lou *et al.*'s CRC design which focuses on the union bound. As the list size L decreases, $\mathrm{P}^L_{\mathrm{NACK}}$ increases and $\mathrm{P}^L_{\mathrm{UE}}$ decreases to $\mathrm{P}^1_{\mathrm{UE}}$, which is well approximated by a nearest-neighbor bound.

References

[1] R. E. Blahut, *Algebraic Codes for Data Transmission.* Cambridge, UK: Cambridge University Press, 2003.

[2] P. Koopman and T. Chakravarty, "Cyclic redundancy code (CRC) polynomial selection for embedded networks," in *Int. Conf. Dependable Systems and Networks*, Jun. 2004, pp. 145–154.

[3] C. Y. Lou, B. Daneshrad, and R. D. Wesel, "Convolutional-code-specific CRC code design," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3459–3470, Oct. 2015.

[4] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, J. B. Anderson, Ed. New Jersey, USA: IEEE Press, 1999.

[5] N. Seshadri and C. E. W. Sundberg, "List viterbi decoding algorithms with applications," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 313–323, Feb. 1994.

[6] F. K. Soong and E. F. Huang, "A tree-trellis based fast search for finding the n-best sentence hypotheses in continuous speech recognition," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 1991, pp. 705–708 vol.1.

[7] C. Nill and C. E. W. Sundberg, "List and soft symbol output viterbi algorithms: extensions and comparisons," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 277–287, Feb. 1995.

[8] M. Roder and R. Hamzaoui, "Fast tree-trellis list viterbi decoding," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 453–461, Mar. 2006.

[9] R. Wang, W. Zhao, and G. B. Giannakis, "CRC-assisted error correction in a convolutionally coded system," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1807–1815, Nov. 2008.

[10] E. T. S. Institute, "3gpp ts 25.212 version 7.0.0 release 7."