Information Transfer Bounds on Iterative Thresholds of Staircase Codes

Dmitri Truhachev*, Lei Zhang†, and Frank Kschischang‡

*Department of Electrical and Computer Engineering, Dalhousie University, Halifax, Canada, dmitry@dal.ca
†Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada

Abstract—We consider staircase codes and study their performance using an analysis of the decoding information flow. We derive bounds on the information transfer functions for long binary linear codes serving as component codes for staircase ensembles and compute them for several BCH codes and bounded distance decoding. The resulting expressions are used to obtain upper bounds on the iterative thresholds of staircase code ensembles used for communication over the binary symmetric channel.

I. INTRODUCTION

Staircase codes, recently introduced in [1] and proposed for ITU G.975.1 optical communication standard, form a class of spatially coupled codes on graphs designed for optical communications. A staircase code is defined via an infinite array which is filled with information and parity-check bits computed via alternating encoding of horizontal and vertical component codes. Being a class of coupled codes on graphs by their nature, staircase codes have structural similarities with coupled generalized low-density parity-check (LDPC) codes [2] and braided block codes [3]. The iterative decoding is performed by alternating hard decision syndrome-based decoding of the component codes rather than soft message-passing decoding that is typically utilized for coupled coding systems in the quest for achieving the channel capacity [4], [5], [6], [7]. Hard bit flipping decoding is demanded by high-speed operation of optical transceivers that limits the amount of computations which can be spent on the error correction decoding.

Iterative bit flipping decoding performed by alternating bounded distance decoders (BDD) of the component codes is nevertheless an efficient decoding approach [11]. At most \(t\) bits per component code has to be flipped, where \(t\) is the component code error correction radius. Due to the coupled graph structure of the staircase code error correction efforts of the component decoders result in the coupling wave propagation effect and lead to near-capacity performance at rates close to unity. The hard decision decoding, however, inevitably leads to some performance degradation compared to the soft decoding methods. The focus of the present paper is to present an approach towards quantifying the iterative limits of hard-decision based iterative processing.

Our approach focuses on the component code and attempts to quantify the amount of information which is lost in component code decoding. Syndrome-based decoding operates on the entire codeword vectors of bits opposite to soft message passing decoding which implicitly building a decoding tree for each data bit. In addition, no interleaving is performed between the component encoders in the staircase code structure. Therefore, we do not take the (extrinsic information transfer) EXIT chart bottom-up approach [8] which looks at how much information per bit is gained, approximately, at each iteration, and have the opposite, large scale, view and attempt to quantify how much information a cascade of component code encoder, channel, decoder can pass through. We then connect this knowledge to the performance of the staircase code itself. Just like in the long-standing problem of quantifying and bounding decoding error probability we make use of the codeword geometry, but compute and bound the mutual information between the transmitted codeword and the decoding decision. The information function we utilize for the component decoders is defined similar to the generalized functions [9][10]. We, however, focus on the bounded distance and minimum distance decoding rather than a posteriori probability (APP) decoding.

We start with defining the staircase code ensemble, and the proceed with formulating the problem of the component code information transfer, its exact computation and upper bounding. We show how the iterative threshold of a staircase ensemble can be upper bounded by looking at the information transfer function at the first decoding iteration. We present analytic approximations to the bounds and numerical results where we compare iterative decoding thresholds of staircase codes estimated by peeling decoding analysis to the proposed upper bounds.

II. STAIRCASE CODE ENSEMBLES

Consider an array representation of a staircase code given in Fig. 1. The encoder fills a two dimensional staircase-like array that consists of \(N/2 \times N/2\) binary matrices \(B_0^1, B_1, B_2^1, B_3, \ldots B_{2L-1}\) with information and parity-check bits. The encoding process ensures that each row and each column of the array forms a codeword of a binary \((N,K)\) code \(C\). Systematic encoder for the component code \(C\) is utilized. The matrix \(B_0^1\) is filled with zeros (or a set of predetermined bits known to the decoder). This initial “seeding” with known bits is typical for all coupled coding schemes. The information bits are then placed into the areas in the array shaded in grey (see Fig. 1). The white areas are filled by parity-check bits computed during the alternating horizontal and vertical encoding. Component encoder utilizes the first
$K$ systematic bits in each row or column to compute and fill in the remaining $N - K$ parity-check bits. The resulting design rate of the staircase code is $R_s = 2K/N - 1$.  

![Staircase code array](image)

Fig. 1. Staircase code array.

From horizontal encoding perspective a codeword of a staircase code is a sequence of data blocks $y_1, y_2, \ldots, y_L$, where one block  

$$y_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,N/2})$$

is composed of $N/2$ component code codewords $y_{i,j} \in \mathcal{C}$, $i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, N/2$. The first $K$ systematic bits of each $y_{i,j}$ are either the information bits or parity-check bits resulting from vertical encoding. The first block $y_1$ presents an exception since $N/2$ first bits in each vector $y_{1,j}$ are fixed to 0’s due to spatial coupling initiation.  

The codewords of the staircase code are transmitted over the BSC with crossover probability $p$. The receiver performs iterative decoding via alternating decoding of the horizontal and vertical component codes. Typically, such alternating decoding is performed for a number of iterations over a sliding window covering $W$ staircase blocks. The window slides along the code array by one block and the decoding iterations are performed again. It is important to note that the decoder operates with the code’s array directly and replaces the data by its decoded version at each iteration.  

### III. COMPONENT CODE INFORMATION TRANSFER

Consider a concatenation of a binary linear $(N, K)$ code and the binary symmetric channel (BSC) with crossover probability $p$ shown in Fig. 2. The hard output decoder shown in the figure can, for example, be bounded distance decoder (BDD), or the minimum distance i.e. maximum likelihood (ML) decoder.  

![Concatenation of an encoder, BSC, and a hard-output decoder](image)

Fig. 2. Concatenation of an encoder, BSC, and a hard-output decoder.

We focus on computing the mutual information $I(u; \hat{x})$ or $I(x; \hat{x})$, where by $u$ we denote equiprobable $K$-bit information vectors, by $x$ the corresponding $N$-bit codewords, by $r$ the corresponding $N$-bit received vectors, and by $\hat{x}$ the decoder output. The information function shows how much information the code allows to pass through for a specific crossover probability value $p$. Similar information transfer functions considered in [8] or generalized EXIT functions [9] which define as the amount of information can be passed through a channel with MAP decoding. Here we focus on the BDD and are concerned with precise commutation or tight bounding of $I(x; \hat{x})$ for long component codes, in particular primitive BCH codes with lengths of 100–1000 bits often used as component codes in staircase code ensembles.

#### A. Exact Computation of Information Transfer Function

BDD maps any sequence $r$ within error correction radius $t$ from a codeword $x_i$ into the codeword itself $\hat{x} = x_i$, $i = 0, 1, \ldots, 2^K - 1$, while any sequence outside of the error correction radius of any codeword is left unchanged. This, the decision region $\mathcal{R}_i$ of a codeword $x_i$ is a ball of radius $t$ around $x_i$. The size of the decision region $\mathcal{R}_i$ for any $i$ is the size of a ball of radius $t$ in the binary $N$-dimensional space and equals  

$$|\mathcal{R}_i| = D(N, t) = \sum_{i=0}^{t} \binom{N}{i}$$.

In case of BDD we have $2^K + 2^N - 2^K D(N, t)$ possible decision outcomes $\hat{x}$. These are the $2^K$ codewords and the set of sequences outside of the error correction area that we denote by $\bar{\mathcal{R}}$. The cardinality of this set equals  

$$|\bar{\mathcal{R}}| = 2^K (2^N - K - D(N, t))$$.

We can use the entropy expression for the mutual information  

$$I(\hat{x}; x) = H(\hat{x}) - H(\hat{x}|x)$$

and the linearity of the code that ensures the decision regions satisfy the translation property: any shift by a codeword maps the set of decision regions into themselves (the decoding is syndrome-based). The codewords $x_i$ are equiprobable, and, therefore,  

$$H(\hat{x}|x) = H(\hat{x}|x = x_j) = H(\hat{x}|0)$$

for any $j = 1, 2, \ldots, 2^K$. The probability of $r \in \bar{\mathcal{R}}$ can be evaluated as  

$$\Pr(\hat{x} = r) = 2^{-K} \sum_{i=0}^{2^K-1} p_{d_H}(r, x_i) (1 - p) N - d_H(r, x_i)$$

where $d_H(\cdot, \cdot)$ is the Hamming distance function.  

Probabilities of codeword decisions can be expressed as  

$$\Pr(\hat{x}_i) = 2^{-K} \sum_{d=0}^{N} A_d B(d, t, p)$$

where  

$$A_d = \sum_{i=1}^{2^K} \Pr(r \in \bigcup_{i=1}^{2^K} R_i) = 2^{-K} p_{in}$$ (1)
for $i = 0, 1, \ldots, 2^K - 1$ where $A_d$ is the number of codewords of weight $d$ in the code. By

$$p_n = \sum_{d=0}^{N} A_d B(d, t, p)$$

we denote the probability that a sequence $r$ falls into a decoding region of some codeword. The probability that the received sequence $r$ falls into a ball of radius $t$ around a codeword of weight $d$ is denoted by

$$B(d, t, p) = \sum_{i, j \geq 0 \text{ s.t. } i + j \leq t} \binom{d}{i} \binom{N - d}{j} p^{d-i}(1-p)^{N-d+j}.$$  \hspace{1cm} (2)

As a result, computing the decision entropy we obtain

$$H(\hat{x}) = -2^K \text{Pr}(\hat{x} = 0) \log_2 \text{Pr}(\hat{x} = 0)
- \sum_{r \in \mathcal{R}} \text{Pr}(\hat{x} = r) \log_2 \text{Pr}(\hat{x} = r)$$

$$H(\hat{x}) = K p_n - p_n \log_2 p_n
- \sum_{r \in \mathcal{R}} \text{Pr}(\hat{x} = r) \log_2 \text{Pr}(\hat{x} = r).$$  \hspace{1cm} (3)

$$H(\hat{x}) = -K p_n + K + (1 - p_n) \log_2 (1 - p_n)$$

Exact computation of the last term in (3) is complicated. For example, such computation for triple error correcting all sequences 5 codeword (since 5 is known to be the maximum covering radius \( \Pi \)) and computing distance spectrum $A_d(r)$ for each $r$. The translation property of the decision regions can then be used to compute the probabilities for all $r \in \mathcal{R}$. This task is not computationally feasible for large $n$. Exact computation for (31,21,5) BCH code is presented in Section \( \Pi \).

Interestingly, for Hamming codes which are perfect the last term in (3) disappears and the information transfer function can be expressed analytically for any length $N$.

Computing the conditional entropy we obtain

$$H(\hat{x}|x) = -\sum_{i=0}^{2^K-1} q_i \log_2 q_i - \sum_{r \in \mathcal{R}} \text{Pr}(r|x) \log_2 \text{Pr}(r|0),$$

where $q_i = \Pr(x = x_i|0) = B(d, t, p)$ for a codeword $x_i$ of weight $d$.

Let us now denote the number of sequences $r \in \mathcal{R}$ of weight $d$ by $\bar{A}_d$. To compute $\bar{A}_d$ we consider all sequences of weight $d$ and subtract these that are already in the decision regions of the nearby codewords

$$\bar{A}_d = \binom{N}{d} - \sum_{d'=d-t}^{d+t} A_{d'} B_{t,d,d'}$$  \hspace{1cm} (5)

where $B_{t,d,d'}$ we denote the number of sequences of weight $d'$ in a ball of radius $t$ centred around a codeword of weight $d'$. This finally allows us to express the conditional entropy

$$H(\hat{x}|0) = -\sum_{d=0}^{N} (A_d B(d, t, p) \log_2 B(d, t, p)$$

$$+ \bar{A}_d p^d (1-p)^N p^d (1-p)^{N-d})$$  \hspace{1cm} (6)

The exact computation quickly becomes infeasible mainly due to the last term in $\Pi$. To overcome this difficulty we propose an upper bound presented in the next subsection.

B. An Upper Bound on the Information Transfer Function

Consider the last term in (3). To make an upper bound on $H(\hat{x})$ assume that all sequences $r \in \mathcal{R}$ are equiprobable. Note that in this case we can compute the probability of $r \in \mathcal{R}$ as

$$\text{Pr}(\hat{x} = r) \approx (1 - 2^K \text{Pr}(\hat{x} = 0)) / |\mathcal{R}|.$$  \hspace{1cm} (7)

Applying this expression to $\Pi$ and using the concavity of the log function we obtain

$$H(\hat{x}) \leq h(p_n) + K + (1 - p_n) \log_2 (|\mathcal{R}|)$$

where $h(\cdot)$ is the binary entropy function. The entropy $H(\hat{x}|0)$ can be computed exactly as in $\Pi$ leading to the upper bound

$$I(\hat{x}:x) \leq I(\hat{x}:\hat{x})$$

$$= h(p_n) + K + (1 - p_n) \log_2 (|\mathcal{R}|)$$

$$- \sum_{d=0}^{N} (A_d B(d, t, p) \log_2 B(d, t, p)$$

$$+ \bar{A}_d p^d (1-p)^N p^d (1-p)^{N-d})$$  \hspace{1cm} (8)

We can use binomial spectrum approximation

$$A_d \approx \binom{N}{d} 2^{-N+K}$$

which is in fact very accurate for the information transfer function already for $N = 31$ as demonstrated by our numerical results. Interestingly, in this case, we can show by simple combinatorics that

$$\bar{A}_d \approx \binom{N}{d} (1 - D(N,t)2^{-N+K})$$  \hspace{1cm} (9)

C. Application to Staircase Codes

Let us consider a staircase code with $2L$ blocks. We denote the vectors consisting of the first $K$ systematic bits of horizontal component codewords $y_{i,j}$ by $u_{i,j}$, for $i = 1, 2, \ldots, L$ and $j = 1, 2, \ldots, N/2$. We pass each $u_{i,j}$ through the channel depicted in Fig. 2 where the encoder is the (systematic) encoder for component code $C$ reconstructing the vector $y_{i,j}$, $p$ is the BSC crossover probability, and the hard decision decoder is the BDD with error correction radius $t$. Thus, passing all vectors $u_{i,j}$ in parallel through such cascade channels is equivalent to passing a codewords of the staircase code through the BSC with crossover probability $p$ and performing horizontal decoding (utilizing parallel decoding schedule) or the first decoding iteration. Here, for
convenience, we assume a flooding decoding schedule in which the decoding window size \( W = 2L \).

For large \( L \) the total information content in all vectors \( u_{i,j} \) approximately equals \( LN^2/2R_0 \) or \( NR_0 \) bits per vector. From channel coding theorem we know that

\[
NR_0 \leq C = \max_{p(x)} I(x; \hat{x})
\]

where \( C \) is the capacity of the cascade channel and \( p(x) \) is the input distribution. The information transfer value we compute assumes equiprobable distribution of the information tuples at the input of the cascade channel. We can show that the upper bound on the information transfer function \( \bar{I}(x; \hat{x}) \) computed in Subsection [II-B] is, in fact, an upper bound on the capacity of the cascade channel.

**Proposition 1.**

\[
C \leq \bar{I}(x; \hat{x}) .
\]

Proof: Consider an arbitrary distribution \( p(x) \) of codewords \( x_0, x_1, \cdots, x_{2^K-1} \) of the codeword \( C \). Note that \( H(\hat{x}|x) = H(\hat{x}|0) \) does not depend on \( p(x) \). Thus, our focus is on upper bounding \( H(\hat{x}) \).

First we show that the probability that a sequence \( r \) falls inside one of the balls around codewords \( p_{in} \) does not depend on \( p(x) \).

\[
p_{in} = \sum_{j=0}^{2^{K-1}} \text{Pr}(\hat{x} = x_j) = \sum_{j=0}^{2^{K-1}} \sum_{i=0}^{2^{K-1}} \text{Pr}(\hat{x} = x_j| x = x_i) p(x = x_i)
\]

\[
= \sum_{i=0}^{2^{K-1}} p(x = x_i) \sum_{j=0}^{2^{K-1}} \text{Pr}(\hat{x} = x_j| x = x_i)
\]

\[
= \sum_{i=0}^{2^{K-1}} p(x = x_i) \sum_{d=0}^{N} A_d B(d, t, p) = \sum_{d=0}^{N} A_d B(d, t, p)
\]

Then for any \( p(x) \)

\[
H(\hat{x}) = -\sum_{j=0}^{2^{K-1}} \text{Pr}(\hat{x} = x_j) \log_2 \text{Pr}(\hat{x} = x_j)
\]

\[
- \sum_{r \in \mathcal{R}} \text{Pr}(\hat{x} = r) \log_2 \text{Pr}(\hat{x} = r)
\]

\[
H(\hat{x}) \leq -p_{in} \log_2(p_{in}2^{-K}) - \sum_{r \in \mathcal{R}} \text{Pr}(\hat{x} = r) \log_2 \text{Pr}(\hat{x} = r)
\]

where the equality is achieved for uniform distribution of codewords (see [4]).

The upper bound \( I(\hat{x}; x) \) is obtained under assumption that \( \text{Pr}(\hat{x} = r) \) are the same for any \( r \in \mathcal{R} \). Since \( \text{Pr}(r \in \mathcal{R}) = 1 - p_{in} \) irrespectively of \( p(x) \). Again using the concavity of the log function we find that [7] is valid irrespectively of \( p(x) \). \( \square \)

The sum of the capacities of parallel cascade channels gives \( \bar{I}(x; x)/N \) as an upper bound on the achievable staircase code rate. Proposition [1] allows us to use the upper bound on the information transfer \( \bar{I}(x; x)/N \) to find an upper bound on \( p^* \), the crossover probability of the BSC for which a staircase code can operate reliably \( p^* \leq \bar{p} \) where

\[
\bar{p} = \arg \max_{p} \text{ such that } \bar{I}(x; x)/N \geq R_0 .
\]

**IV. Numerical Results**

Fig. 3 demonstrates the exact information transfer function for \((31, 21, 5)\) BCH (green circles) and the upper bound (8) (blue curve). The information transfer function is normalized per bit, i.e., it is divided by \( N \). Binary primitive BCH \((2^n - 1, 2^n - 2n - 1, 5)\) codes are known to have only four distinct coset weight distributions [12] for odd \( n > 4 \) - one for each coset leader of weights 0, 1, 2, 3. Thus, all outcomes \( r \in \mathcal{R} \) correspond to cosets of weight 3, have the same weight distribution, and are, therefore, equiprobable. As a result, the exact computation of the information transfer coincides with the upper bound (blue curve). The slight difference between the blue and green curves is due to the binomial approximation for the spectrum used in the upper bound computation. The BSC channel capacity is given by the red curve.

The achievable \( p^* \) for the corresponding classic staircase code (black star) with BDD is estimated using a peeling decoding analysis. Magenta line indicates the rate of the staircase code on the x-axis. The peeling decoding, which is outside of the scope of the present paper, considers a graph of the staircase code chain in which horizontal and vertical component codes are vertices. All-zero codeword transmission is considered and the two vertices are connected by edges if and only if the bit at their intersection equals 1. The decoder then peels off the edges from the graph and the average number of edges per block can be computed at each decoding iteration. The highest value of BSC crossover probability \( p \) for which all edges are eventually peeled off is denoted by \( p^* \). The peeling analysis gives an approximation for the iterative decoding threshold since it does not account for miss-corrections.

Upper bound (8) computed for \((1023, 993, 7)\) BCH code with \( t = 3 \) is shown in Fig. 4 (blue curve). The bound is useful at the staircase code rate, indicated by magenta bar. The gap between the \( p^* \) point and the upper bound is due to the assumption that sequences \( r \) outside of the decoding balls are equiprobable and due to the fact that consecutive decoding iterations contribute to the information loss due to processing, and therefore, consideration of the first iteration alone is suboptimal.

Upper bounds (9) on the iterative thresholds of staircase codes are given in Fig. 5 for two code classes. The first class is based on binary primitive BCH \((2^n - 1, 2^n - nt - 1, 2t + 1)\) for \( t = 2 \) and \( n = 6, 7, 8, \cdots, 11 \) (black dots), while for the second \( t = 3 \) and \( n = 7, 8, \cdots, 11 \) (blue dots). The peeling decoding approximations for the thresholds \( p^* \) are shown by black stars for \( t = 2 \) and blue stars for \( t = 3 \).
The bound and the approximation are very close for $t = 2$ codes starting from $n = 7$. The bound also gets tighter as $n$ and the respective staircase code rate increases.

V. CONCLUSION

We have derived an upper bound on the information transfer function of long linear codes for transmission over BSC used with bounded distance decoding. We then use the resulting bound to upper bound thresholds of staircase codes. The information transfer bounds provided a computational tool to bound iterative thresholds of code ensembles used with vector-based hard-decision decoding.

REFERENCES


