LDPC CODES AND DECODERS FOR BURST
AND BIT-INSERTION/DELETION CORRECTION

[01]  Front matter

a.  Date April 176

b.  Abstract  We propose a coding scheme based on reverse concatenation of low-density parity check codes and spectral-null codes for correction of a single insertion and error bursts resulting from island frequency variation in bit patterned media systems. Structured codes whose Tanner graphs do not contain trapping sets on consecutive bits will be developed as they will have high robustness to error bursts. The proposed decoding algorithm combines decimation decoding, and message passing on a rewired Tanner graph which reflects the presence of insertion.

c.  Proponent(s) and affiliation(s) Bane Vasić, University of Arizona

d.  Designated contact person. Bane Vasić

[02]  Subject of research and relevance to issue(s) to be solved.

a.  Complete description of the research matter and its connection with ASTC stated goals

The proposed research pertains to development of codes and decoding algorithms tailored to channels with error burst and insertions and deletions resulting from irregular island frequency in bit patterned media (BPM) systems. The methods developed in the project will provide increased level of immunity to the effects of non uniform island spacing, written-in errors on code performance used in bi-patterned media (BPM) systems. The proposed research addresses the ASTC topic BPMR No. 7: BPM Coding and Detection for BPM Recording System.

Background

The dominant source of errors in BPM systems is irregularity of island spacing, or temporal variation of the instantaneous island frequency. In addition to insertions and deletions the BPM burst errors are also a significant impairment in the BPM systems. The writing clock frequency can be chosen to be slightly higher (or lower) than the average island frequency, resulting in repeatedly inserted (or deleted) bits. For simplicity we assume that the readback process produces insertions only. Figure 1 illustrates the write process. One notices the cumulative effect of discrepancy between island and writing frequencies – there are intervals where bits and islands are more or less aligned, as well as intervals of misalignment. A burst of errors is followed by an insertion. Recent preliminary studies indicate that the insertion frequency can be made small resulting in no more than one insertion per 4096 bits. Under these
assumptions, and assuming that the code length is less than 4096 bits, the problem reduces to designing a coding scheme that guarantees correction of a single insertion preceded by a burst of errors. This is precisely the goal of the proposed research.

![Figure 1](image1.png)

**Figure 1** The effect of non-uniform island spacing to creation of error bursts. The blue bars represent the magnetic islands. Note irregular spacing between blue bars (the variation of the instantaneous island frequency) The read bars represent the ideal islands with uniform spacing. Without perfect knowledge of time dependence of the instantaneous island frequency, a writer assumes constant island frequency i.e, that the islands are located at the position of the red bars. The figure shows an example of writing a data sequence of length 20. (a) Small timing error variance causes a burst of errors. (b) Larger timing error variance causes burst of error followed by a bit cycle slip. The bit 19 (the number at the bottom) is written on the 20th (the number at the top) island resulting in a deletion. After a deletion the bits and islands are again relatively well aligned. (c) Yellow line indicates the reliability of read back bit values. The bits in the middle are very unreliable. The bit 20 is reliable but in a wrong position.

**A Coding Scheme**

**Message passing on rewired Tanner graphs**

Let us suppose for a moment that the basis of a coding scheme is a low-density parity check (LDPC) code. Using LDPC codes in conjunction with marker codes for insertion/deletion channels has been studied in literature [6] [7]. Similar approaches are also considered for BPM and intersymbol interference channels [8], [9].

This paragraph illustrates the complications that the presence of an insertion introduces in the decoding process. **Figure 2** shows an example of writing a codeword of length $n=7$. In the absence of insertions, the actual positions of
written bits are the same as intended positions (red cells on the main diagonal in Figure 2 (a)). The insertion results of a lateral shift of the red cells Figure 2 (b)). It is clear that in order to successfully reconstruct a codeword, a decoding algorithm must have an additional feature of being able to infer the position of red cells. Let us now analyze the effect of an insertion of the Tanner graph of a code. Figure 3 shows the local neighborhood of the  \( i \)-th variable node. Insertion in a position  \( i+1 \) changes the Tanner graph, but only locally, indicating that it is conceivable that a local decoding à la sum-product algorithm may be able to reconstruct the codeword. Indeed it is not difficult to see that a naive scheme of  \( n+1 \) decoders operating in parallel on  \( n+1 \) graphs would be able to reconstruct the codeword in presence of one insertion and error patterns correctable by the decoder. Computational savings are possible by the virtue of similarity of messages passed by different decoders. This motivates the development of a message passing algorithm which operates on a Tanner graph in which nodes and edges can be duplicated. The first step is to formulate the belief propagation equations for modified Tanner graphs.

**Figure 2** Correspondence between intended and actual bit positions in a codeword of length seven in absence of insertions (a), and when the third bit is inserted in the fourth position (b).

**Figure 3** Local rewiring of the Tanner graph around the inserted bit. (a) Original Tanner graph. (b) Tanner graph of the code corresponding to the new arrangement of bits in which the \( i \)-th bit is inserted. The red error indicates that the variable node \( i \) together with all its edges is duplicated. The duplicated edges are red.

**Reverse concatenation**
Consider a hypothetical codeword of the form 01010101 …. Since the bit values alternate, a single insertion would be readily correctable. Unfortunately, this is not an efficient code. In more useful codes, any appearance of a pair of consecutive like symbols is an indication of a possible insertion. On the other
hand, appearance of a transition is an indication that there is no insertion at this position. Thus a code which imposes a constraint on the frequency of occurrence of consecutive like symbols results would simplify the decoding algorithm by providing the information that it is not necessary to perform message passing on a particular rewiring of a Tanner graph. These codes belong to a class of constrained codes with higher-order spectral-null at zero frequency, and are well studied [5]. The connection of spectral-null codes to Levenshtein insertion/deletion codes Levenshtein is also understood [6].

The proposed coding scheme is based on a reverse concatenation scheme involving the higher-order spectral-null constrained code and a LDPC code as shown in Figure 4. The spectral-null codes have good distance properties and are naturally represented by a trellis. Thus, the BCJR algorithm may operate on a combined trellis of the ISI channel and the spectral-null code resulting in an improved performance. On the other hand, spectral-null codes have relatively low rates, and consequently a tradeoff must be found between performance and complexity benefits and rate loss. Most likely codes with first-order spectral zero will be suffice.

![Figure 4 Reverse concatenation of LDPC code and spectral-null code](image)

From the discussion above it follows that classical LDPC codes can be used to correct random errors and a single insertion. This is discussed next.

**Construction of burst error correcting codes**

Even though nonbinary codes are more suitable for this application, for the purpose of elucidating our main insights, binary codes are discussed in this section.

To construct LDPC codes with good performance it is essential to understand the failure mechanism of iterative decoders. It is now well established that failures of iterative decoders are caused by subgraphs in a Tanner graph of a code, known as trapping sets. If a trapping set is present in a Tanner graph, corruption of subset of variable nodes in this trapping set leads to decoding failure. Knowledge on trapping sets is not complete due to the complex nature of iterative decoding algorithms, but our research group has very good understanding of this phenomenon and developed codes and decoding algorithms with guaranteed and superior frame error rate performance on the binary symmetric channel (BSC).
Roughly speaking, in the context of channels with burst errors, one would like to ensure that Tanner graph has a topology such that no block of consecutive bits belongs to a small trapping set.

We will rely on codes whose parity check matrices can be represented as an array of permutation matrices. If the permutation matrices are circulant permutation matrices then the code is quasi-cyclic (QC). The decoding can be parallelized by exploiting the block structure of the parity check matrices but more importantly trapping sets of these codes typically do not involve consecutive bits (the details are omitted due to space constraints). We will develop a new class of QC LDPC codes by optimizing positions of variable nodes involved in small trapping sets.

**Decoding algorithms**

As we explained, due to complexity of the rewired Tanner graph message passing algorithm must have low-complexity. The decoding algorithm used may be modifications of the sum product algorithm or based on our recently proposed class of multi-bit message passing [1] and bit-flipping decoders [3]. These new decoders have a potential to surpass belief propagation (BP) in the error floor region with much lower complexity. They were derived by identifying potentially harmful subgraphs that could be trapping sets present in any finite-length code and designing to correct error patterns on these subgraphs in an isolated manner.

We will design novel decoding algorithms suited to bursty channels. The idea is to use the decimation as proposed in [2]. The decimation is a method originating in statistical mechanics adopted for iterative decoding. The idea is not to fix the values of reliable variables thus decimating the Tanner graph to a smaller subgraph containing erroneous variables. The main trick is to design the decimation rule so that erroneous variables are never decimated. In [2] for a specific class of seven level decoders, we proved the following Lemma.

**Lemma 1:** The decimation-enhanced decoding algorithm will never decimate a node initially correct to a wrong value, and a node initially wrong to a correct value.

The proposed decimation algorithm serves as a guide to help the multi-bit decoding algorithm to converge faster on a small number of errors. The main insight is that the role of decimation should not necessarily be to correct errors, but to ensure that more variable nodes in the graph that initially receive right values from the channel are shielded from the erroneous messages emanating from the error nodes by decimating those correct variable nodes. The burst errors can be efficiently corrected by not allowing decimation of variable nodes between pairs of variable nodes whose reliability exhibits fluctuation.
To prove this concept we simulated decoding of a short Tanner code on a simple bursty channel model in which the beginning of a burst is uniformly distributed over the length of a code, and the burst error length, $L$, has the exponential distribution with parameter $\lambda$, that is $Pr\{L = k\} = \lambda e^{-\lambda k}$, $k \geq 0$. The results for a simple decimation strategy which does not change the variable between “suspicious” nodes are shown in Figure 5. The algorithm is able to correct burst error patterns up to length 11 which is promising. In the proposed research the interplay between the code structure and decimation will be thoroughly studied.

![Figure 5](image.png)

**Figure 5** The frame error rate performance of decimation decoders on a channel with burst of errors.

b. Proposed research approach(es)

i. Experimental NA

ii. Computational

- Burst error and insertion/deletion statistics will be gathered from sponsor companies to match the models used in study. For example, burst errors may be modeled as Gilbert-Eliot channel.
- Simulations of a scheme composed of $n+1$ parallel decoders will be initially preformed to assess the ballpark of potential benefits.
- Message passing algorithms for the rewired Tanner graph will be developed and implemented in software.
- Reverse concatenation scheme will be implemented in software. This involves identification of parameters of the spectral-null code and LDPC code. Off-shelf constrained codes will be used.
If time permits, the BCJR algorithm operating on a combined ISI channel-constrained code trellis will be simulated. Otherwise, BCJR will operate on the ISI-channel only.

- A tradeoffs between constrained code rate and complexity will be studied.
- QC LDPC codes with a special trapping set arrangement suitable for burst errors will be developed. The construction will be based on Latin squares, and will also include removing small trapping sets. Trapping sets will be obtained from our work on BSC.
- Capacity results for insertion channel will be obtained from literature. This will aid code construction. Capacity of the “burst-and-insertion” channel is not known, thus an experimental study will be performed to find appropriate code parameters (rate, length, minimum distance, Levenshtein distance).
- Decimation decoder will be developed for correcting error burst.
- Decimation decoder will be combined with message passing on the rewired Tanner graph to correct both error bursts and insertions.
- The interdependence between code structure and decimation algorithm will be studied and optimal code-decoder pairs will be developed.

References


c. Likely outcome of research
- Message passing on rewired Tanner graphs capable of correcting single insertion and random errors.
- Optimized parameters and performance of the reverse concatenation scheme involving LDPC codes and spectral-null codes.
- Good quasi-cyclic burst error correction codes and their performance and trapping set structure.
- Decimation decoding algorithms for burst error correction
- Decoding algorithm for joint burst error and insertion correction

[03] Resources required to perform project

a. Personnel, students, etc.
   i. ½ time – research assistant
   ii. Faculty – ½ month summer

b. Equipment, lab, etc.

c. Computational

[04] Resources other than ASTC funding dedicated to perform project

a. Grants None

b. Contracts None

c. Other None

[05] Resources requested from ASTC and how they will be utilized

a. Funding
   i. Overhead $30,738
   ii. Direct project cost $43,293
   iii. Facility use fees $1,000
   iv. Materials $2,000
   v. Student stipends $18,074
   vi. Faculty support $7,414
   vii. Travel $5,000
viii. Justification

The budget requests support for personnel and non-personnel expenses for one year. The detailed justification of expenses is given below.

- **Graduate Student Support:** The support of one graduate student for the academic year is requested. The graduate student will participate in theoretical research as well as the extensive development and application of computer simulation tools required to carry out the proposed investigations. Graduate students will participate in summer internship programs at sponsor companies.
- **PI Support:** A summer salary support of total $7,414k for the PI is requested.
- **Operations and Supplies:** Research support in the amount of $3,000 per year for each of three years is requested for conference registrations, and project specific supplies. Supplies include minor computer materials and upgrades necessary for simulation and theoretical studies. A Facility Usage fee of $1,000 is requested for use of the Information Theory and Coding Laboratory.
  - **Travel:** Travel support in the amount of $5,000 is requested. This will support one international conference trip and two domestic conference trips per year to be taken by the PIs or graduate student to disseminate and discuss results of the supported research.

b. Expected technical cooperation with sponsor(s): materials to be provided by sponsor(s) (e.g., targets, devices, engineering support, etc.)

c. Sponsors’ facility utilization NA

d. Expected students’ internships: A student internship will take place during Summer 2012. Internship at Western Digital, Seagate, LSI or Marvell.

[06] Time line

a. First quarter: Develop burst error models, decoders for burst errors channels.

b. Second quarter: Develop binary quasi-cyclic LDPC codes for burst error channels.

c. Third quarter: Develop generalized LDPC codes for burst error channels. Develop decoders for insertion channels

d. Fourth quarter: Develop detectors for burst and insertion channel.
Information Theory and Coding Laboratory at the University of Arizona has the following computational resources available for this project:

<table>
<thead>
<tr>
<th>Machine type</th>
<th>Number of machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel Pentium 4 CPU 1.70 GHz</td>
<td>3</td>
</tr>
<tr>
<td>Intel Pentium 4 CPU 1.80 GHz</td>
<td>10</td>
</tr>
<tr>
<td>Intel Pentium 4 CPU 2.40 GHz</td>
<td>1</td>
</tr>
<tr>
<td>Intel Pentium 4 CPU 3.06 GHz</td>
<td>5</td>
</tr>
<tr>
<td>Dual Intel Xeon CPU 2.20 GHz</td>
<td>2</td>
</tr>
<tr>
<td>Dual Intel XeonCPU 2.40 GHz</td>
<td>2</td>
</tr>
<tr>
<td>Dual Intel XeonCPU 2.80 GHz</td>
<td>10</td>
</tr>
<tr>
<td>Dual Intel Xeon CPU 3.00 GHz</td>
<td>5</td>
</tr>
<tr>
<td>Dual Intel Quad-Core Xeon CPU 2.00 GHz</td>
<td>2</td>
</tr>
<tr>
<td>Sun Blade 1000 Workstation</td>
<td>1</td>
</tr>
</tbody>
</table>

Contact Information

**Technical:**
Bane Vasic  
The University of Arizona  
Department of Electrical and Computer Engineering  
1230 East Speedway Boulevard  
Tucson, AZ 85721  
Tel: 520-626-5550  
Fax: 520-626-3144  
E-mail: vasic@ece.arizona.edu

**Business:**
Angie Gomez (ampgomez@email.arizona.edu)  
Engineering Proposal Services  
Room 303  
Civil Engineering Building  
1209 E. 2nd Street  
PO Box 210072  
Tucson, AZ 85721-0072  
Tel: 520-621-6595  
E-mail: ENGR-EPSdept@email.arizona.edu  
Fax: 520-621-9291
Biographical Sketch

**Dr. Bane Vasić** is a Professor of Electrical and Computer Engineering and Mathematics at the University of Arizona. Prior to this appointment he was at Bell Laboratories, where he was involved in the development of the soft error-event decoding algorithm, and read channel architecture for Bell Labs magnetic recording read channel chips which were regarded as the best in industry. His pioneering work on structured low-density parity check (LDPC) error correcting codes has contributed to enabling low-complexity iterative decoder implementations. Structured LDPC codes are today adopted in a number of communications standards, and play an important role in extremely high-density magnetic recording systems. Dr. Vasić was the first to propose using LDPC codes in magnetic recording systems. He was also involved in a Digital Versatile Disc (DVD) Copy Protection Standardization Group sponsored by DVD Industry Consortium and the movie studios. Dr. Vasić is also known for his theoretical work in error correction coding theory and codes on graphs which has led to analytical characterization of the hard decision iterative decoders of low-density parity check (LDPC) codes, and design of codes with best error-floor performance on BSC known today.

**Professional Preparation**
- Undergraduate Institution: University of Nis, Serbia, B. S. (Electrical Engineering), 1989.
- Graduate Institution: University of Nis, Serbia, M. S. (Electrical Engineering), December 1991, University of Nis, Serbia, Ph.D. (Electrical Engineering), 1994.

**Appointments**
- Professor, Electrical and Computer Engineering, University of Arizona, 2000-present
- Member of Technical Staff Bell Laboratories, 1998-2000