CCIT REPORT #849 January 2014

LDPC Codes for Two-Dimensional Arrays

Yuval Cassuto, *Member, IEEE,* and Amin Shokrollahi, *Fellow, IEEE*

Abstract

Binary codes over two-dimensional arrays are very useful in data storage, where each array column represents a storage device or unit that may suffer failure. In this paper we propose a new framework for probabilistic construction of codes on two-dimensional arrays. Instead of a pure combinatorial erasure model used in traditional array codes, we propose a mixed combinatorial-probabilistic model of limiting the number of column failures, and assuming a binary erasure channel in each failing column. For this model we give code constructions and detailed analysis that allow sustaining a large number of column failures with graceful degradation in the fraction of erasures correctable in failing columns. Another advantage of the new framework is that it uses low complexity iterative decoding. The key component in the analysis of the new codes is to analyze the decoding graphs induced by the failed columns, and infer the decoding performance as a function of the code design parameters, as well as the array size and failure parameters. A particularly interesting class of codes, called probabilistically MDS array codes, gives fault-tolerance that is equivalent to traditional MDS array codes. The results also include a proof that the two-dimensional codes outperform standard one-dimensional LDPC codes.

I. Introduction

Linear codes constructed from low density matrices are the keystone of contemporary coding theory, with huge impact on both theoretical research and practical applications. The sparsity of the code matrices offers a compelling complexity advantage in implementation, and achieving this low complexity with optimality in redundancy is the great achievement of a large body of deep research. Interestingly, two separate coding-theory fields aim at the above objective of developing low-density codes with good information efficiency.

One is the field of *array codes* [3]. Array codes, defined over two-dimensional binary arrays, use low-density parity-check matrices to construct codes for erasures or errors of full columns. The low-density property offers complexity benefits in encoding and decoding operations. Hence the task of array codes is to combat column-level erasures/errors, while being defined by operations over smaller information units (bits or small groups of bits). The ultimate goal of array-code research is code families that are Maximum Distance Separable (MDS) from the column perspective, thus having optimal redundancy for a given column correction requirement. Array codes are widely used in practice, being employed as a central element of RAID (Redundant Arrays of Inexpensive Disks) [14] storage systems. Each array column then represents an individual storage device, whose failure is modeled as a column erasure.

The second, and better-known low-density coding field, is *low-density parity-check (LDPC) codes under iterative decoding* [15]. Without need for detailed introduction, in this area the objective is to construct low-density (one dimensional) codes that will decode well under sub-optimal iterative decoding algorithms.

So far, despite the structural similarities, the two low-density coding fields evolved in essential separation. Array codes have concentrated on algebraic constructions and decoding algorithms that guarantee fixed numbers of column erasures over small array dimensions. In contrast, the theory of iteratively decoded low-density parity-check codes has sought probabilistic code constructions that with high probability have good iterative decoding performance in the limit of large block lengths. Previous work exists where iterative decoding of known array codes over one-dimensional channels is experimentally examined [9], [2], but no attempt has been made (to the best of our knowledge) to construct new array codes for iterative decoding over two-dimensional channels. Such constructions are highly motivated by the current state of matters in array-code theory and practice. Algebraic array codes rigidly assume that columns are erased at full, while in practice many storage devices have failure modes that render only part of their data inaccessible. This strong error model introduces significant complexity penalties, with encoding and decoding complexities that steeply grow with the number of column erasures. Moreover, algebraic array codes guarantee correction of a certain number of column erasures, with a sharp transition to failure if the specified number of column erasures is exceeded. These issues have triggered a recent line of work constructing array codes for combinations of full-column and individual-symbol erasures (and errors) [4], [16].

For the same purpose of targeting more realistic erasures in storage arrays, the study in this paper proposes the first framework for construction and analysis of two-dimensional array codes under iterative decoding. The new framework merges the fields of array coding and LDPC coding in the sense that it applies the deep construction and analysis tools of LDPC codes to an error model very natural for array codes. The outcome from this framework are codes that can extend the fault tolerance of storage arrays with complexities significantly lower than what algebraic array codes can achieve. The new framework comprises three main components

1) An erasure model that combines an integer bound on the number of failing columns with an erasure probability within failing columns.

Yuval Cassuto is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa Israel (email: ycassuto@ee.technion.ac.il). Part of the research was done while at the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL).

Amin Shokrollahi is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station 14, CH-1015 Lausanne, Switzerland (e-mail: amin.shokrollahi@epfl.ch). Part of the results of this paper appeared in the proceedings of the 2011 International Symposium on Information Theory [8].