A Three-Stage Approach for Designing Non-Binary Spatially-Coupled Codes for Flash Memories

Ahmed Hareedy, Homa Esfahanizadeh, and Lara Dolecek
Electrical Eng. Department, University of California, Los Angeles, Los Angeles, CA 90095 USA
{ahareedy, hesfahanizadeh}@ucla.edu and dolecek@ee.ucla.edu

***Abstract*—Modern dense Flash memory devices operate at very low error rates, which require powerful error correcting coding techniques. Here, our focus is on spatially-coupled (SC) codes. We present a three-stage approach for the design of high performance non-binary SC (NB-SC) codes optimized for practical Flash channels; we aim at minimizing the number of detrimental general absorbing sets of type two in the graph of the designed NB-SC code. Simulation results demonstrate the effectiveness of the NB-SC codes designed using our approach.**

I. Introduction

Spatially-coupled (SC) codes are a class of graph-based codes that has recently received significant attention [1]. SC codes designed using cutting vector (CV) and minimum overlap (MO) partitioning for various applications were introduced in [2] and [3], respectively. It was recently demonstrated in [4] that general absorbing sets of type two (GASTs) are the objects that dominate the error floor of non-binary (NB) graph-based codes over practical, asymmetric Flash channels [5]. We proposed the weight consistency matrix (WCM) framework in [4] to remove GASTs from the Tanner graph of NB codes.

In this paper, we study NB-SC codes designed for practical Flash channels. The underlying block codes we focus on are circulant-based (CB) codes. Our combinatorial approach to design NB-SC codes comprises three stages:

1) ***Optimal overlap (OO) partitioning:*** We solve a discrete optimization problem to determine the optimal ***overlap parameters*** for partitioning.
2) ***Circulant power optimizer (CPO):*** We then apply a new heuristic program to optimize the ***circulant powers*** of the underlying block code.
3) ***WCM framework:*** After applying the OO-CPO technique, we perform the ***edge weight*** optimization [4].

II. Preliminaries

Let $\mathbf{H}$ be the parity-check matrix of the underlying regular non-binary CB code that has column weight γ and row weight κ. The binary image of $\mathbf{H}$, which is $\mathbf{H}^b$, has $\gamma\kappa$ circulants. Each circulant is of the form $\sigma^{f_{i,j}}$, $0 \leq i \leq \gamma-1$, $0 \leq j \leq \kappa-1$, and σ is the $p \times p$ identity matrix shifted one unit to the left. Circulant powers are $f_{i,j}$, $\forall i, j$. Array-based (AB) codes are CB codes with $f_{i,j} = ij$, $\kappa = p$, and p prime.

The NB-SC code is constructed as follows. First, $\mathbf{H}^b$ is partitioned into $m+1$ disjoint components: $\mathbf{H}_0^b, \mathbf{H}_1^b, \dots, \mathbf{H}_m^b$. Here, we focus on $\gamma = 3$ and $m = 1$. Second, $\mathbf{H}_0^b$ and $\mathbf{H}_1^b$ are coupled together L times (see [2]) to construct $\mathbf{H}_{SC}^b$. A ***replica*** is any $(L+1)\gamma p \times \kappa p$ submatrix of $\mathbf{H}_{SC}^b$ that contains $[\mathbf{H}_0^{bT} \ \mathbf{H}_1^{bT}]^T$ and zero circulants elsewhere [3]. Third, the matrix $\mathbf{H}$ is generated by replacing each 1 in $\mathbf{H}^b$ with a value $\in$ GF$(q)\backslash\{0\}$. Finally $\mathbf{H}_{SC}$, is constructed by applying the partitioning and coupling scheme above to $\mathbf{H}$.

The ***binary protograph matrix (BPM)*** of a general binary CB matrix is the matrix resulting from replacing each $p \times p$ non-zero circulant with 1, and each $p \times p$ zero circulant with 0. For example, the BPM of $\mathbf{H}_{SC}^b$ is $\mathbf{H}_{SC}^{bp}$.

We now recall the GAST definition. VNs (resp., CNs) refer to variable (resp., check) nodes.

Definition 1. *(cf. [4]) Consider a subset $\mathcal{V}$ of VNs in the Tanner graph of an NB code. Set all the VNs in $\mathcal{V}$ to values $\in$ GF(q)$\backslash\{0\}$ and set all other VNs to 0. The set $\mathcal{V}$ is said to be an (a, b, d_1, d_2, d_3)* ***GAST*** *over GF(q) if the size of $\mathcal{V}$ is a, the number of unsatisfied CNs connected to $\mathcal{V}$ is b, the number of degree-1 (resp., 2 and > 2) CNs connected to $\mathcal{V}$ is d_1 (resp., d_2 and d_3), $d_2 > d_3$, all the unsatisfied CNs connected to $\mathcal{V}$ (if any) have either degree 1 or degree 2, and each VN in $\mathcal{V}$ is connected to strictly more satisfied than unsatisfied neighboring CNs (for some set of given VN values).*

We also recall the **unlabeled GAST (UGAST)** definition [4]. The WCM framework removes a GAST by careful processing of its edge weights [4]. See [6] for more details.

III. OO Partitioning: Theoretical Analysis

We determine a common substructure in multiple GASTs, then minimize the number of instances of this substructure in the graph of $\mathbf{H}_{SC}^b$ (the unlabeled graph of the SC code) [2]. For the dominant GASTs we have encountered in NB codes with $\gamma = 3$ simulated over Flash channels, the $(3, 3, 3, 0)$ UGAST occurs as a common substructure most frequently [6].

A cycle of length 6 (cycle 6) in the graph of $\mathbf{H}_{SC}^{bp}$ (the binary protograph), which is defined by the non-zero entries $\{(h_1, \ell_1), (h_2, \ell_2), \dots, (h_6, \ell_6)\}$ in $\mathbf{H}_{SC}^{bp}$, results in p cycles 6 in the graph of $\mathbf{H}_{SC}^b$ if and only if [6]:

$$\sum_{e=1}^{3} f_{h_{2e-1}, \ell_{2e-1}} \equiv \sum_{e=1}^{3} f_{h_{2e}, \ell_{2e}} \mod p, \qquad (1)$$

where $f_{h,\ell}$ is the power of the circulant indexed by (h, ℓ) in $\mathbf{H}_{SC}^b$. The $(3, 3, 3, 0)$ UGAST is a cycle 6. Thus, our OO partitioning aims at deriving the overlap parameters of $\mathbf{H}^{bp}$ that result in the minimum number of cycles 6 in the graph of $\mathbf{H}_{SC}^{bp}$. Then, we run the CPO to further reduce the number of $(3, 3, 3, 0)$ UGASTs in the graph of $\mathbf{H}_{SC}^b$ by breaking the condition in (1) for as many cycles as possible in the optimized graph of $\mathbf{H}_{SC}^{bp}$ (see also [6]).

We establish a discrete optimization problem by expressing the number of cycles 6 in the graph of $\mathbf{H}_{SC}^{bp}$ as a function of the overlap parameters and standard code parameters, then solve for the optimal overlap parameters.

Lemma 1. *In the graph of an SC code with $\gamma = 3$, κ, $p = 1$, $m = 1$, and L, the number of cycles 6 is given by:*

$$F = LF_s + (L-1)F_d, \qquad (2)$$

where F_s (resp., F_d) is the number of cycles 6 having their VNs spanning one (resp., two consecutive) particular replica(s).

Let the ***overlapping set*** of x rows of a binary matrix be the set of positions in which all the x rows have 1's simultaneously (overlap). Seven independent parameters, referred to as t_0, t_1, t_2, $t_{0,1}$, $t_{0,2}$, $t_{1,2}$, and $t_{0,1,2}$, govern the overlapping sets [6].

We define the functions $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ to be used in Theorem 1. The definitions of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are given in [6]. Theorem 1 uses combinatorics to give the exact expressions for F_s and F_d in terms of the above overlap parameters.

Theorem 1. *In the graph of an SC code with $\gamma = 3$, κ, $p = 1$, $m = 1$, and L, F_s and F_d are computed as follows:*

$$F_s = F_{s,0} + F_{s,1} + F_{s,2} + F_{s,3}, \text{ and} \quad (3)$$

$$F_d = F_{d,0} + F_{d,1} + F_{d,2} + F_{d,3}, \quad (4)$$

where $F_{s,0}$, $F_{s,1}$, $F_{s,2}$, $F_{s,3}$, $F_{d,0}$, $F_{d,1}$, $F_{d,2}$, and $F_{d,3}$ are functions of the overlap parameters, and are given in [6].

The idea of Theorem 1 is that F_s and F_d can be computed by decomposing each into four more tractable terms.

Now, define F^* to be the minimum number of cycles 6 in the graph of $\mathbf{H}_{SC}^{bp}$ (the binary protograph). Thus, our ***discrete optimization problem*** is formulated as follows:

$$F^* = \min_{t_0, t_1, t_2, t_{0,1}, t_{0,2}, t_{1,2}, t_{0,1,2}} F. \quad (5)$$

The constraints are linear constraints capturing interval constraints and the balanced partitioning constraint.

Since all the solutions result in the same number of OO partitioning choices and the same F^*, we work with one of these solutions, and call it an optimal vector, $\mathbf{t}^*$.

IV. Circulant Power Optimization

The steps of the CPO are (more details are in [6]):

1) Assign AB circulant powers to the $\gamma\kappa$ 1's in $\mathbf{H}^{bp}$.
2) Design $\mathbf{H}_{SC2}^{bp}$ using $\mathbf{H}^{bp}$ and $\mathbf{t}^*$ such that $\mathbf{H}_{SC2}^{bp}$ contains only two replicas, $\mathbf{R_1}$ and $\mathbf{R_2}$.
3) Locate all the cycles of lengths 4 and 6 in $\mathbf{H}_{SC2}^{bp}$.
4) Specify the cycles 6 in $\mathbf{H}_{SC2}^{bp}$ that have (1) satisfied, and call them ***active cycles***. Let $2F_s^a$ (resp., F_d^a) be the number of active cycles having their VNs spanning only $\mathbf{R_1}$ or only $\mathbf{R_2}$ (resp., both $\mathbf{R_1}$ and $\mathbf{R_2}$).
5) Compute the number of $(3, 3, 3, 0)$ UGASTs in $\mathbf{H}_{SC}^{b}$:
$$F_{SC} = (LF_s^a + (L-1)F_d^a)\, p. \quad (6)$$
6) Count the weighted number of active cycles [6] each 1 in $\mathbf{H}_{SC2}^{bp}$ is involved in.
7) Map the counts from step 6 to the 1's in $\mathbf{H}^{bp}$, and sort these 1's in a list descendingly according to the counts.
8) Pick a subset of 1's from the top of this list, and change the circulant powers associated with them.
9) Using these interim new powers, do steps 4 and 5.
10) If F_{SC} is reduced with no cycles of length 4 in $\mathbf{H}_{SC}^{b}$, update F_{SC} and the circulant powers, then go to step 6. Otherwise, return to step 8.
11) Iterate until the target F_{SC} is achieved.

V. Simulation Results

We start our experimental results with a table comparing the number of $(3, 3, 3, 0)$ UGASTs in SC codes designed using various techniques. Table I demonstrates reductions in the number of $(3, 3, 3, 0)$ UGASTs achieved by the OO-CPO technique over other techniques.

Next, we provide simulation results verifying the performance gains achieved by our NB-SC code design approach for Flash memories. The Flash channel we use is a practical Flash channel, which is the normal-Laplace mixture (NLM) Flash channel [5]. Here, we use 3 reads, and the sector size is 512 bytes. RBER is the raw BER (no error correction) and UBER is the uncorrectable BER (after error correction) [4].

All the NB-SC codes we simulated are defined over GF(4), and have $\gamma = 3$, $\kappa = p = 19$, $m = 1$, and $L = 20$ (block length $= 14440$ bits and rate ≈ 0.834). Code 1 is uncoupled (AB). Code 2 is designed using the CV technique. Code 3 is designed using the OO technique (with no CPO applied). The underlying block codes of Codes 2 and 3 are AB codes. Code 4 is designed using the OO-CPO technique. The edge weights of Codes 1, 2, 3, and 4 are selected randomly. Code 5 (resp., Code 6) is the result of applying the WCM framework to Code 1 (resp., Code 4) to optimize the edge weights.

Fig. 1 demonstrates the performance gains achieved by each stage of our NB-SC code design approach. The gain of the first stage (OO) is about 0.6 of an order of magnitude. The gain of the second stage (CPO) is about 0.7 of an order of magnitude. The gain of the third stage (WCM) is about 1.2 orders of magnitude. Moreover, the figure shows that the NB-SC code designed using our approach, which is Code 6, achieves about 200% RBER gain compared to Code 2.

TABLE I

Number of $(3,3,3,0)$ UGASTs in SC codes with $\gamma = 3$, $m = 1$, and $L = 30$ designed using different techniques.

Design technique	Number of $(3,3,3,0)$ UGASTs			
	$\kappa = p = 7$	$\kappa = p = 11$	$\kappa = p = 13$	$\kappa = p = 17$
Uncoupled with AB	8820	36300	60840	138720
SC CV with AB	3290	14872	25233	59024
SC MO with AB	609	3850	6851	15997
SC best with AB	609	3520		
SC OO-CPO with CB	203	2596	5356	14960

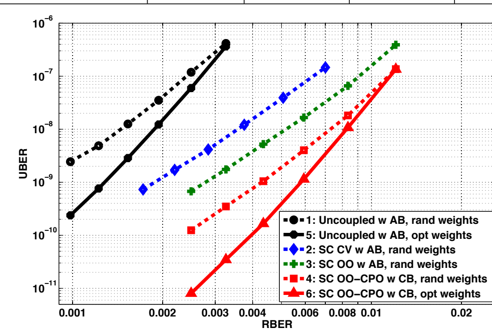

Fig. 1. Simulation results over the NLM Flash channel for SC codes with $\gamma = 3$, $m = 1$, and $L = 20$ designed using different techniques.

VI. Conclusion

We proposed a combinatorial approach for the design of NB-SC codes optimized for practical Flash channels. The OO-CPO technique efficiently optimizes the underlying topology of the NB-SC code, then the WCM framework optimizes the edge weights. NB-SC codes designed using our approach outperform existing NB-SC codes over Flash channels.

References

[1] A. R. Iyengar *et al.*, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, 2012.

[2] H. Esfahanizadeh *et al.*, "Spatially-coupled codes optimized for magnetic recording applications," *IEEE Trans. Magn.*, vol. 53, no. 2, 2016.

[3] H. Esfahanizadeh *et al.*, "A novel combinatorial framework to construct spatially-coupled codes: minimum overlap partitioning," in *Proc. IEEE ISIT*, Aachen, Germany, 2017.

[4] A. Hareedy *et al.*, "A general non-binary LDPC code optimization framework suitable for dense Flash memory and magnetic storage," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 9, 2016.

[5] T. Parnell *et al.*, "Modelling of the threshold voltage distributions of sub-20nm NAND flash memory," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, 2014.

[6] A. Hareedy *et al.*, "High performance non-binary spatially-coupled codes for Flash memories," 2017. [Online]. Available: http://arxiv.org/abs/1706.07531