# Spatially-Coupled Codes for Write-Once Memories

Santhosh Kumar†, Avinash Vem†, Krishna Narayanan†, Henry D. Pfister‡
Department of Electrical and Computer Engineering, Texas A&M University†
Department of Electrical and Computer Engineering, Duke University‡

***Abstract*—The focus of this article is on low-complexity capacity-achieving coding schemes for write-once memory (WOM) systems. The construction is based on spatially-coupled compound LDGM/LDPC codes. Both noiseless systems and systems with read errors are considered. Compound LDGM/LDPC codes are known to achieve capacity under MAP decoding for the closely related Gelfand-Pinsker problem and their coset decomposition provides an elegant way to encode the messages while simultaneously providing error protection. The application of compound codes to the WOM system is new. The main result is that spatial coupling enables these codes to achieve the capacity region of the 2-write WOM system with low-complexity message-passing encoding and decoding algorithms.**

***Index Terms*—data storage, LDGM codes, LDPC codes, message-passing algorithms, spatial coupling, write-once memory system.**

## I. Introduction

In a typical flash storage system, each cell carries an electric charge that indicates the stored bit; a higher charge denotes a 1 and a lower charge denotes a 0. While raising the charge of these cells can be done easily, lowering the charge requires resetting a large block of cells. Write-once memory (WOM) systems model such storage cells. In a binary WOM system, multiple writes are allowed but a bit with a value of 1 cannot be changed to 0. In 1982, Rivest and Shamir presented first WOM codes for two writes to store two bits in each write using only three cells [1]. Even more surprisingly, they show that only about $nt/\log(t)$ bits are required to allow $t$ writes of an $n$-bit message (for large $n$), whereas a naive scheme with no coding would require $nt$ bits. In 1985, Heegard established the capacity region of a $t$-write binary WOM system with no noise, and also the case where write errors are introduced [2].

Since their introduction, there has been a plethora of WOM-code constructions for noiseless systems [3]–[8]. These constructions are based on projective geometry [4], coset coding [5], [8], graph coverings [6], and position modulation [7]. However, the first capacity-achieving scheme for the noiseless two-write WOM system, with polynomial encoding and decoding complexity, was introduced only recently [9]. Shortly after, polar codes were constructed with an encoding and decoding complexity of $O(n\log(n))$ that achieve the capacity region of the noiseless $t$-write WOM system [10]. Constructions based on LDGM codes are also considered in [11]. In that work, a sequence of optimized irregular LDGM ensembles is used to achieve capacity for the second write of the noiseless 2-write WOM system.

The literature on WOM codes that handle errors is more limited. Error-correcting WOM codes were first constructed in [12], [13]. New constructions that are triple-error-correcting are presented in [14]. Polar codes were shown to correct a constant fraction (of blocklength) of errors [15]. Recently, polar codes were used to achieve the capacity of $t$-write WOM systems with write errors [16].

We construct low-complexity capacity-achieving codes for the binary 2-write WOM system. Our construction is a coset-type scheme based on compound LDGM/LDPC codes [17]. These are also known as Hsu-Anastasopoulos (HA) codes in the coding theory literature [18]. By spatially coupling these codes, we are able to obtain WOM codes that achieve the capacity of the noiseless 2-write WOM system under low-complexity message-passing algorithms.

While the capacity region is known for a WOM system with write errors [2], the capacity region of a WOM system with read errors is unknown to the best of our knowledge. The first write of a WOM system with read errors is an instance of the Gelfand-Pinsker problem [19], and our coding scheme achieves its capacity region. For the second write of the 2-write WOM system with read errors, our construction achieves any rate $R < 1 - \delta - h(p)$, where $\delta$ denotes the normalized weight of the state sequence after first write, and the codewords are protected against read errors from the binary symmetric channel with a bit-flip probability of $p$. In this article, the focus is mainly on the second write. For the Gelfand-Pinsker problem, we constructed practical coding schemes based on compound codes in an earlier article [20]. Empirical evidence suggests that these schemes achieve capacity for this problem. In Section III-E, we discuss this connection and the extension to the $t$-write problem for $t > 2$.

A key insight from [11] is that encoding the message for the second write can be reduced to the binary erasure quantization problem (BEQ). This reduction allows an efficient encoding algorithm with a linear computational complexity. We note that, while our encoder is also based on the reduction to the BEQ problem, our code construction differs from the one in [11] in a few important ways. In particular, it allows for low-complexity decoding and it can handle errors in the read process. Moreover, by changing the encoding algorithm, we can also handle the first write.

This material is based upon work supported in part by the National Science Foundation (NSF) under Grant No. 1320924. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these sponsors.

## II. System Model

This article deals with binary WOM systems. Consider a storage system with $n$ cells, wherein each cell can store one bit $\{0,1\}$ of information.

**Definition 1.** *In a binary WOM system, a cell with a value of* 1 *cannot be changed to* 0*. This is called the WOM constraint.*

In a single write, a $k$-bit message $s^k \in \{0,1\}^k$ is encoded into a $n$-bit sequence $x^n$ in a binary WOM system. Such a write is associated with a rate of $k/n$. We consider read errors, where it is required to decipher the message $s^k$ from a *noisy* version of $x^n$. Below, we discuss the noiseless and noisy systems separately.

Consider a noiseless binary WOM system initialized with zeros, and a successive representation of $t$ messages with rates $R_1, \ldots, R_t$. The capacity region of the $t$-write system is given by [2]

$$\Big\{(R_1,\ldots,R_t) \mid 0 \le R_i \le h(\delta_i)\prod_{j=1}^{i-1}(1-\delta_j), 1 \le i \le t-1,$$
$$0 \le R_t \le \prod_{j=1}^{t-1}(1-\delta_j), 0 \le \delta_i \le 1\Big\}, \tag{1}$$

where $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$.

The capacity region in (1) has an intuitive description. Define the normalized weight of a binary sequence as the number of ones in the sequence divided by its length. Consider the first write and suppose that the normalized weight of the sequence in the first write is $0 \le \delta_1 \le 1$. Then, the rate of the first write must satisfy $R_1 < h(\delta_1)$. Now, a fraction $1-\delta_1$ of the cells that are zeros can be utilized for subsequent writes. For the second write, if a further fraction $0 \le \delta_2 \le 1$ of these cells are allowed to set to 1, then the rate of the second write satisfies $R_2 < (1-\delta_1)h(\delta_2)$. Generalizing this argument gives the upper bound in (1) on the achievable rates.

The primary focus here is on the second write of the 2-write WOM system. Suppose the normalized weight of the sequence after the first write is $\delta$. Then, we construct a coding scheme for the second write of the 2-write WOM system that achieves any rate

$$R < 1-\delta. \tag{2}$$

Now, suppose there are read errors. Say a message $s^k$ is encoded into a sequence $x^n$ in a WOM system. However, the message can only be decoded from $y^n \in \{0,1\}^n$, a noisy version of $x^n$. This models the bit-flips caused by read errors in a storage system. We assume that the errors are caused by a BSC with bit-flip probability $p$. That is, $y_i = x_i \oplus \mathsf{Ber}(p)$, where $\oplus$ denotes the modulo 2 addition and $\mathsf{Ber}(p)$ denotes the Bernoulli random variable with parameter $p$.

It is important to note that this error model is different from the write error model. In write-error model, the error occurs during the write and the reads occur without error. In this case, the encoder can achieve optimal performance by biasing the input distribution. In the read-error model,

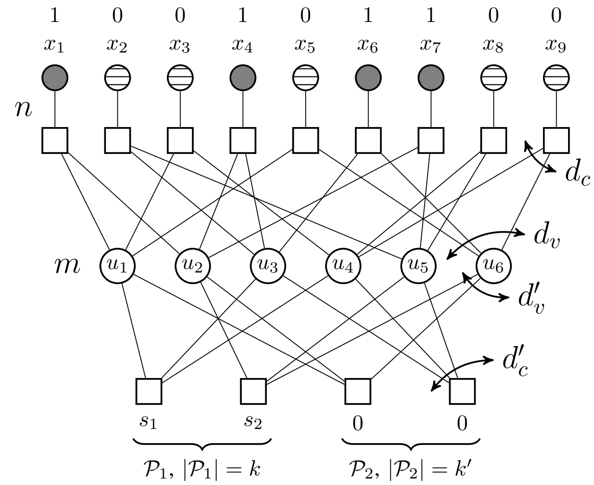


Fig. 1. A Tanner graph representation of a compound LDGM/LDPC code. The top part represents the LDGM code, and the bottom part represents the LDPC code. The parities in $\mathcal{P}_1$ carry the message, and the parities in $\mathcal{P}_2$ provide the error correction.

the write occurs without errors but the each read experiences independent errors. If sufficient error correction is used, then a single read can be used to recover the correct codeword with high probability and the system will have perfect knowledge of the memory state during rewrites. To the best of the authors' knowledge, the capacity region of the WOM system with read errors is unknown.

For the second write of the two-write WOM system with read errors, we can construct codes that achieve any rate

$$R < 1-\delta-h(p), \tag{3}$$

where $\delta$ is the normalized weight of the state sequence after first write.

## III. Coding Scheme

### A. Compound LDGM/LDPC Codes

Our coding scheme for the WOM system is based on compound LDGM/LDPC codes. We refer the reader to [17], [18] for a detailed description. Below, we give a brief description and state some important properties of these codes.

A Tanner graph representation of a compound LDGM/LDPC code is shown in Fig. 1. The relationship between a codeword $x^n = (x_1, \cdots, x_n)$ and the information bits $u^m = (u_1, \cdots, u_m)$ is described by the upper portion of the graph, which resembles an LDGM code. For example, in Fig. 1, we have $x_1 = u_1 \oplus u_2$, $x_2 = u_3 \oplus u_5$, etc., where $\oplus$ denotes the modulo 2 addition. Moreover, the information bits $u^m$ are required to satisfy parity constraints in the bottom part of the graph, which resembles an LDPC code. The code-bits $x^n$ are called LDGM bit-nodes and the information bits $u^m$ are called LDPC bit-nodes. We also note the distinction between LDGM (top check-nodes) and LDPC check-nodes (bottom check-nodes). In Fig. 1, we have $u_1 \oplus u_3 \oplus u_4 = s_1$, $u_2 \oplus u_5 \oplus u_6 = s_2$, $u_1 \oplus u_2 \oplus u_6 = 0$, $u_3 \oplus u_4 \oplus u_5 = 0$. LDPC check-nodes are split into two disjoint groups $\mathcal{P}_1$ and $\mathcal{P}_2$ with sizes $|\mathcal{P}_1| = k$ and $|\mathcal{P}_2| = k'$, respectively. The parities in $\mathcal{P}_1$ specify a fixed constraint given by $s^k$, while the parities in $\mathcal{P}_2$ are all set to 0. The codebook $\mathcal{C}(s^k)$, parameterized by the parity sequence $s^k$, is the set of all $x^n$ sequences that satisfy the above constraints.

**Definition 2.** *A $(d_v, d_c, d'_v, d'_c)$ compound code is defined as a compound LDGM/LDPC code with the following Tanner graph. Each of the LDPC bit-nodes have $d_v$, $d'_v$ edges to the LDGM and LDPC check-nodes, respectively. Moreover, each of the LDGM and LDPC check-nodes have $d_c$ and $d'_c$ edges to the LDPC bit-nodes, respectively.*

The rate of the codebook $\mathcal{C}(s^k)$ is $(m-k-k')/n$, since $m-k-k'$ independent information bits can be selected to generate a codeword of length $n$. Also, the parities in $\mathcal{P}_1$ provide an elegant coset decomposition. More precisely, let $\mathcal{C}$ be the codebook generated by the information bits $u^m$ by ignoring the parities in $\mathcal{P}_1$. Then,

$$\mathcal{C} = \bigcup_{s^k \in \{0,1\}^k} \mathcal{C}(s^k),$$

and $\mathcal{C}(s^k)$ are cosets of $\mathcal{C}$.

### B. Write-Once Memory Codes

We now describe our coding scheme for the second write of the 2-write WOM system based on compound codes. Let $s^k$ be the $k$-bit message for the second write and $n$ be the size of the WOM system. Consider Fig. 1, where the gray LDGM bit-nodes represent the indices that are set to 1 and the horizontally patterned LDGM bit-nodes represent the zeroes after the first write. The parities $s^k$ in group $\mathcal{P}_1$ carry the message.

**Construction 3.** *The $k$-bit message $s^k$ is encoded into a $n$-bit sequence $x^n$ in the WOM system by finding a codeword $x^n$ in the codebook $\mathcal{C}(s^k)$ with the constraint that the indices that are 1 after the first write remain the same in $x^n$.*

For such a construction, the message $s^k$ can then be retrieved from $x^n$ (or a noisy version of $x^n$), which gives a rate of $R = k/n$. It remains to find codes $\mathcal{C}(s^k)$ and algorithms that allow Construction 3 and also achieve the rates in (2) and (3).

A crucial step in Construction 3 is finding a codeword in a code with specified values in certain bit positions. This is an instance of the *erasure quantization* problem. In the binary erasure quantization (BEQ) problem, it is required to quantize a source sequence over the alphabet $\{0,1,*\}$ to some codeword in a given codebook over the binary field $\{0,1\}$. The requirement is that 0s and 1s in the source sequence cannot be changed but the erasures $*$ can be set to either 0 or 1 in the codeword. To map the WOM problem to the BEQ problem, we can take the source sequence as the state of the WOM system after first write and set all zeroes to erasures.

A key observation in [21] is that the BEQ problem is closely related to the channel coding over the binary erasure channel (BEC). In particular, [21, Theorem 2] states that an erasure quantizer for a code can successfully quantize every source sequence with erasure pattern $e^n \in \{0,1\}^n$ if and only if the channel decoder for the dual code can correct all received sequences with the erasure pattern $1^n \oplus e^n$.

**Remark 4.** *It is well known that the compound codes achieve the capacity on erasure channels under bit-MAP decoding [18]. However, under bit-MAP decoding, linear codes achieve capacity if and only if their dual codes achieve capacity. Thus, the compound codes also achieve the capacity region of the BEQ problem.*

Another important property of compound codes is that they are good channel codes under MAP decoding for the BSC [22]. Roughly speaking, this means that if a compound code (of large enough degrees and blocklength) of rate $1-h(p)-\varepsilon$ is used for the channel coding problem over a BSC with parameter $p$, then the message can be reliably estimated under optimal decoding with probability of error at most $\varepsilon$.

**Theorem 5.** *Compound codes achieve the rate regions in* (2) *and* (3) *under optimal encoding and decoding.*

*Proof:* The following provides heuristic arguments that can be made rigorous using standard techniques. Moreover, we only discuss the achievability of (3) since (2) is a special of (3) with $p = 0$.

Choose a compound code $\mathcal{C}(s^k)$ with parameters $m = n$ and $(n-k')/n = 1-h(p)-\varepsilon$. The code $\mathcal{C}$ is also a compound code with rate $(n-k)/n = 1-h(p)-\varepsilon$. Since $\mathcal{C}$ is a good channel code for the BSC, the codeword $x^n$ (and subsequently $s^k$) can be recovered from the BSC channel with parameter $p$.

Since the dual code $\mathcal{C}(s^k)^\perp$ (with rate $1-(n-k-k')/n$) can correct up to a fraction $(n-k-k')/n$ of erasures, the code $\mathcal{C}(s^k)$ can quantize all source sequences with at least a fraction $1-(n-k-k')/n$ of erasures. Thus, for a fraction of at most $(n-k-k')/n$ ones in the WOM system after first write, we can find a codeword in $\mathcal{C}(s^k)$ according to Construction 3. Thus, any $\delta < (n-k-k')/n = 1-h(p)-R-\varepsilon$ is achievable, which is the rate region in (2). ■

While the above construction achieves the desired rate regions under optimal encoding and decoding, our interest is in practical encoding and decoding techniques. But, the compound LDGM/LDPC codes do not achieve capacity under the message-passing algorithms, and so the above construction cannot achieve the desired rate regions under the low-complexity message-passing algorithms. To alleviate this issue, we consider spatially-coupled codes because these codes have linear time message-passing algorithms that achieve capacity.

First, we describe an iterative encoding procedure that finds a codeword $x^n$ in the coset $\mathcal{C}(s^k)$ with the WOM constraint. Once such a codeword $x^n$ is found, retrieving $s^k$ from $x^n$ (or a noisy version of $x^n$) can be done efficiently by message-passing algorithms for channel coding on the codebook $\mathcal{C}$. Message-passing algorithms for channel coding are well-known in the coding theory literature. See [23] for their description. In the following, we focus only on the encoding.

**Algorithm 1** Iterative Quantization Algorithm [21]

**Input:** Seq. $z^n \in \{0,1\}^n$, Msg. $s^k$, Comp. code $G(U,V)$.
**Output:** Seq. $x^n$ in the $\mathcal{C}(s^k)$ that satisfies WOM const.
  Associate $z^n$ with LDGM check-nodes as in Fig. 1.
  Set all 0s in $z^n$ to erasures $*$.
  Set parities in $\mathcal{P}_1$ to $s^k$.
  **while** $\exists$ non-erasures in $V$ **do**
    **if** $\exists$ non-erased $u \in U$ such that only one of its neighbors $v \in V$ is not erased **then**
      Pair $(u,v)$.
      Erase $u$ and $v$.
    **else**
      FAIL.
      **break**.
    **end if**
  **end while**
  **if** pairing did not FAIL **then**
    Set non-erased $u \in U$ to $0$.
    In the inverse order in which bit-nodes in $U$ are erased, set them to satisfy the paired parity.
    Evaluate the codeword $x^n$ from the graph $G(U,V)$ and information bits $u^m$.
  **end if**

### C. Iterative Quantization Algorithm

The encoding algorithm that is described below is introduced in the context of the BEQ problem as the iterative quantization algorithm [21]. In contrast to the BEQ problem, the additional challenge in WOM problem is for the quantized sequence to represent a given message $s^k$. This can be overcome by running the iterative quantization algorithm on the coset $\mathcal{C}(s^k)$.

The duality between the BEQ problem and the channel decoding over the BEC is further illustrated by the similarity between the iterative quantization algorithm and the peeling decoder for the BEC. In the iterative quantization algorithm, degree-one bit-nodes are peeled off the graph rather than the degree-one check-nodes. Suppose $G(U,V)$ is a Tanner graph representation of the compound LDGM/LDPC, where $U$ denotes the LDPC bit-nodes (nodes to be peeled off from the graph) and $V$ denotes the union of check-nodes (analogous to the code-bits in peeling decoder). The details are presented in Algorithm 1.

Let $z^n$ denote the state of the WOM system after first write. That is, 1's in $z^n$ represent the WOM constraints. First, associate the sequence $z^n$ with the LDGM check-nodes and set all 0's in $z^n$ to erasures $*$. Now, LDPC bit-nodes which have a single non-erased check-node as its neighbor are paired with that unique check-node. Then, both those LDPC bit-nodes and the paired check-nodes are erased from the graph and this process is repeated. This is akin to the peeling decoder which assigns values to erased bit-nodes that have a degree-one check-node as its neighbor and then erases those bit-nodes. Once this pairing is done, the values of the LDPC bit-nodes are set to satisfy the paired check-node in the *reverse order that the LDPC bit-nodes were paired.*

For this algorithm to succeed in finding a desired codeword, there should be a sufficient number of degree-one LDPC bit-nodes to begin with. In fact, when this algorithm is used for the WOM problem with compound LDGM/LDPC codes and *regular* degrees, the process never succeeded. Therefore, it is required to design irregular degree distributions to have sufficient degree-one bit-nodes and achieve the optimal performance. However, this is not an issue in spatially-coupled codes since the boundary termination in these codes provides the necessary degree-one LDPC bit-nodes to get the encoding process started.

### D. Spatial Coupling

For a comprehensive introduction to the spatially-coupled codes, see [24]–[27]. Below, we give a brief description of the construction of the spatially-coupled compound LDGM/ LDPC codes. Two important reasons we use spatially-coupled compound codes are:

a) The boundary condition provides a natural way to start the iterative quantization procedure. Thus, this obviates the need for designing irregular distributions with optimal number of degree-one bit-nodes.
b) When these codes are used for the WOM system with read errors, the decoding thresholds under the belief-propagation decoder saturate to the MAP threshold and achieve the rate region in (3).

We describe the construction with regular degrees (for the single system) for both bit- and check-nodes. As in Fig. 1, let $d_v$ and $d'_v$ denote the LDPC bit-node degrees to the LDGM and LDPC check-nodes, respectively. Also, let $d_c$ and $d'_c$ denote the LDGM and LDPC check-node degrees to the LDPC bit-nodes, respectively. For simplicity, we assume that the parities in $\mathcal{P}_1$ have the same degree as the parities in $\mathcal{P}_2$.

Let $L$ be the chain length of the spatially-coupled system and $w$ be the coupling window of the spatially-coupled code. Consider groups of LDPC bit-nodes at sections indexed by $\mathcal{N}_v = \{1, \cdots, L+w-1\}$, LDGM check-nodes at $\mathcal{N}_c = \{1, \cdots, L\}$, and LDPC check-nodes at $\mathcal{N}'_c = \{w, \cdots, L+w-1\}$. We describe the coupling structure in the LDGM part; similarly, this can be repeated for the LDPC part, both for the parities in $\mathcal{P}_1$ and $\mathcal{P}_2$. To begin with, let $N$ denote the number of LDPC bit-nodes in a given section $i \in \mathcal{N}_v$, and pick $N$ so that

- $Nd_v/d_c$, $Nd'_v/d'_c$ are integers; this is to ensure that all LDGM and LDPC check-nodes have regular degrees,
- $Nd_v/w$, $Nd'_v/w$ are integers; this ensures an appropriate partition of the edges in a given section for the coupling.

Place $d_v$ edge sockets to each LDPC bit-node. Place $Nd_v/d_c$ LDGM check-nodes at each section in $\mathcal{N}_c$, each check-node with $d_c$ edge sockets. For both LDPC bit-nodes and LDGM check-nodes, this ensures each a total of $Nd_v$ edge sockets. Now, partition the $Nd_v$ sockets at both LDPC bit-nodes and LDGM check-nodes into $w$ groups using a uniform random permutation. Denote these partitions by $\mathcal{N}^v_{i,\ell}$, $\mathcal{N}^c_{j,\ell}$, where $1 \le i \le L+w-1$, $1 \le j \le L$, $1 \le \ell \le w$. The coupled LDGM component is constructed by connecting the sockets

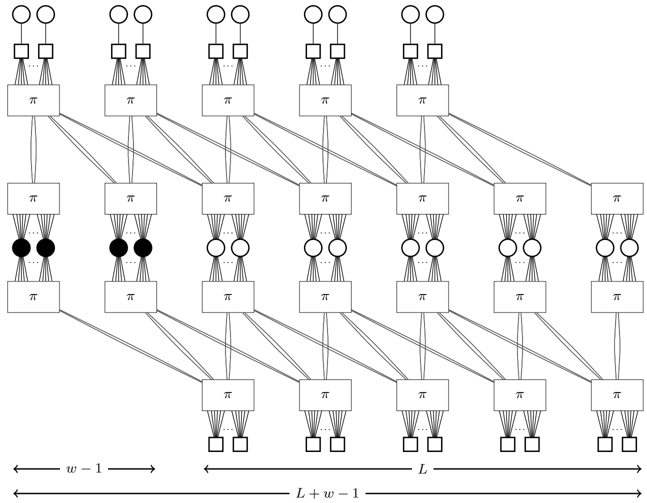


Fig. 2. Illustration of connections in a spatially-coupled compound LDGM/LDPC code. The top part denotes the coupling in the LDGM part, and the bottom part denotes the coupling in the LDPC part. The LDPC bit-nodes in the first $w-1$ sections (black bit-nodes in the middle) are set to 0.

in $\mathcal{N}^c_{j,\ell}$ to $\mathcal{N}^v_{j+\ell-1,\ell}$. See Fig. 2 for an illustration of these connections.

To construct the coupled LDPC component, one can start by placing $d'_v$ edge sockets at each LDPC bit-node and repeating the above process. Moreover, the coupling is done separately for parities in $\mathcal{P}_1$ and $\mathcal{P}_2$. We ensure that each LDPC bit-node has connections to check-nodes in both $\mathcal{P}_1$ and $\mathcal{P}_2$. When decoding the message $s^k$ from $x^n$, the channel decoder does not use the parities in $\mathcal{P}_1$. Thus, if there are LDPC bit-nodes with no connections to the parities in $\mathcal{P}_2$ (but has connections only to $\mathcal{P}_1$), this causes a small error floor in the decoding.

This construction leaves some edge sockets of the LDPC bit-nodes at the boundary unconnected. Also, the LDPC bit-nodes in the first $w-1$ sections are shortened to 0. The shortening and the boundary connections are necessary for the spatially-coupled "encoding/decoding wave" to get started and the threshold saturation phenomenon to take effect in spatial-coupling [24].

If we run the iterative quantization procedure in Algorithm 1 on the spatially-coupled code without imposing any boundary condition, the resulting assignment of the LDPC bit-nodes in the first $w-1$ sections will not necessarily be all zeroes. But it is required for these LDPC bit-nodes to be 0 so that the decoding wave gets started and the threshold saturation phenomenon to take effect.

**Remark 6.** *We initialize the LDPC bit-nodes in the first $w-1$ sections to erasures to make sure that the iterative quantization algorithm does not pair these bit-nodes with any check-nodes.*

The LDPC bit-nodes in the first $w-1$ sections can subsequently be set to zero. After this modification, the encoding wave starts from the other end of the spatially-coupled code. As such, we cannot place message bits (parities in group $\mathcal{P}_1$) in the first $w-1$ sections of the LDPC check-nodes, since the zeroes in the first $w-1$ sections of the LDPC bit-nodes cannot be consistent with odd parities in the check-nodes. For the same reason, we cannot have WOM constraints in the first $w-1$ sections of the LDGM check-nodes. This results in a rate-loss due to termination which is a standard side-effect of spatially-coupled constructions.

### E. Multi-Write WOM System

For a 2-write WOM system, the first write differs from the second in two key aspects. The first aspect is that there are no WOM constraints for the first write unlike in the second. That is, the state of the system before first write is all zeroes. The second aspect is that the normalized weight of the sequence after first write must be within a given $\delta$. The second write has no such restrictions.

The first write is an instance of the Gelfand-Pinsker problem [19], where the encoder side-information is the all-zero sequence. For this problem, we have constructed low-complexity schemes based on spatial coupling of compound codes that were empirically shown to achieve capacity in an earlier article [20]. To find a desired codeword within a normalized distance $\delta$ from $0^n$, the standard belief-propagation (BP) algorithms are not sufficient. The difficulty comes from the fact that there are several codewords that are within the desired distance. As such, the BP algorithm cannot converge to any particular codeword. It is necessary to decimate variable-nodes (set them to 0 or 1 based on current log-likelihood ratio) every few iterations. It is worth noting that in the iterative quantization procedure presented in Algorithm 1, there is no way to restrict the weight of the output sequence. As such, we cannot use Algorithm 1 for the first write.

Although the belief-propagation with guided decimation (BPGD) algorithm is crucial for the first write, the question remains whether one can leverage the this to restrict the output weight after the second write. Our current simulations of the BPGD algorithm for the second write never succeeded in finding a codeword that satisfies both the WOM constraints and the message parities $s^k$. The circumstance is different in the first write since there are no WOM constraints. Getting the BPGD algorithm to succeed for the second write immediately enables one to achieve the capacity region of the t-write WOM system.

## IV. Numerical Results

Consider a $(d_v, d_c, d'_v, d'_c)$ compound code for the second write of the noiseless 2-write WOM problem as shown in Fig. 1. We can assume $k' = 0$, since there is no need for error correction, and can have empty $\mathcal{P}_2$. Then,

$$m = n\frac{d_c}{d_v}, \qquad k = m\frac{d'_v}{d'_c} = n\frac{d_c d'_v}{d_v d'_c}.$$

For the second write, this gives a rate of

$$R = \frac{k}{n} = \frac{d_c d'_v}{d_v d'_c}.$$

TABLE I

ACHIEVABLE THRESHOLD $\delta$ FOR THE NOISELESS WOM SYSTEM WITH SPATIALLY-COUPLED COMPOUND LDGM/LDPC CODES WITH $L = 30$ AND A SINGLE SYSTEM BLOCKLENGTH OF $\approx 24000$

| LDGM/LDPC $(d_v, d_c, d'_v, d'_c)$ | $\delta^*$ | $\delta$ $w=2$ | $\delta$ $w=3$ | $\delta$ $w=4$ |
|---|---|---|---|---|
| $(3,3,3,6)$ | 0.500 | 0.477 | 0.492 | 0.494 |
| $(3,3,4,6)$ | 0.333 | 0.294 | 0.324 | 0.326 |
| $(3,3,5,6)$ | 0.167 | 0.095 | 0.156 | 0.158 |
| $(4,4,3,6)$ | 0.500 | 0.461 | 0.491 | 0.492 |
| $(4,4,4,6)$ | 0.333 | 0.278 | 0.323 | 0.325 |
| $(4,4,5,6)$ | 0.167 | 0.086 | 0.155 | 0.159 |
| $(5,5,3,6)$ | 0.500 | 0.436 | 0.488 | 0.491 |
| $(5,5,4,6)$ | 0.333 | 0.260 | 0.320 | 0.324 |
| $(5,5,5,6)$ | 0.167 | 0.079 | 0.154 | 0.159 |

Thus, from the capacity region in (2), the maximum normalized weight $\delta^*$ of the state sequence after first write below which we can successfully encode the second message is

$$\delta^* = 1 - R = 1 - \frac{d_c d'_v}{d_v d'_c}.$$

In Table I, we list the achievable thresholds $\delta$ for the spatially-coupled compound LDGM/LDPC codes. For the spatially-coupled codes, a chain length of $L = 30$ is used and the value of the threshold $\delta$ is shown for different coupling-window sizes. Note that the rate loss in the spatially-coupled codes is not included when reporting the optimal threshold $\delta^*$. It is implicit that current constructions based on spatially-coupled codes suffer a rate loss of $O(w/L)$. While there has been some progress in mitigating this loss [28]–[30], minimizing this loss is an important open problem in the spatially-coupled codes. Also, for a fixed coupling window length $w$, the rate loss can be reduced arbitrarily by increasing $L$ without changing the achievable threshold $\delta$. As such, the thresholds shown in Table I are not sensitive to the parameter $L$. The blocklength of the single system is roughly of size 24000, which gives an effective blocklength for the coupled-system of about 720000. Such enormous blocklengths are required for the spatially-coupled codes to mitigate the rate loss and also operate close to capacity.

For the simulations, we have tested the encoding of 10 sequences with the iterative quantization algorithm, and the reported thresholds are the maximum values for which the majority of the 10 sequences are successfully encoded. A few observations are in order. From Table I, it is clear that one requires a coupling window lengths of at least $w = 3$ for the thresholds to have small gap to the optimal values. The decrease in the gap to the optimal threshold from changing $w = 3$ to $w = 4$ is minimal. To close the gap further, one needs to increase the blocklength of the single system beyond 24000.

Next, we present the simulations for a smaller blocklength. In Fig. 3, we show the encoding failure probability for the $(3,3,3,6)$ spatially-coupled compound code with $(L, w) = (30, 3)$ and a single system blocklength of 1200, which gives an effective blocklength of 36000. A total of $10^5$ messages were attempted to encode and no failures were observed for $\delta < 0.43$. These results appear to be better than the implementations based on polar codes. For example, in [11, Figure 2] for a polar code of length 16000 and rate $1/2$, at $\delta = 0.42$ the encoding failure probability is more than $5\times10^{-2}$. However, note that our construction based on spatial coupling with parameters $(L, w) = (30, 3)$ suffers a rate loss of 10%. While the precise trade-offs between constructions based on spatially-coupled codes and polar codes are not completely clear, this is worth pursuing.

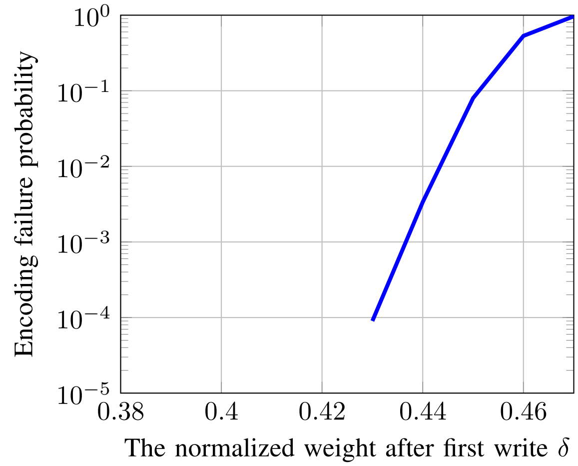


Fig. 3. Encoding failure probability for the second write as a function of the normalized weight after first write, for the spatially-coupled compound code with parameters $(d_v, d_c, d'_v, d'_c) = (3,3,3,6)$, $(L, w) = (30, 3)$ and a single system block length of 1200. A total of $10^5$ messages were attempted to encode, and no failures were observed for $\delta$ less than 0.43.

TABLE II

ACHIEVABLE THRESHOLDS $(\delta, p)$ FOR THE WOM SYSTEM WITH READ ERRORS AND SPATIALLY-COUPLED COMPOUND CODES WITH $L = 30$ AND A SINGLE SYSTEM BLOCKLENGTH OF $\approx 32000$

| LDGM/LDPC $(d_v, d_c, d'_v, d'_c)$ | $w$ | $(\delta^*, p^*)$ | $(\delta, p)$ |
|---|---|---|---|
| $(3,3,4,6)$ | 3 | $(0.333, 0.0615)$ | $(0.321, 0.0585)$ |
| $(3,3,4,8)$ | 3 | $(0.500, 0.0417)$ | $(0.490, 0.0387)$ |
| $(3,3,6,8)$ | 4 | $(0.250, 0.0724)$ | $(0.239, 0.0684)$ |
| $(4,4,4,6)$ | 4 | $(0.333, 0.0615)$ | $(0.324, 0.0585)$ |
| $(4,4,4,8)$ | 4 | $(0.500, 0.0417)$ | $(0.492, 0.0387)$ |
| $(4,4,6,8)$ | 4 | $(0.250, 0.0724)$ | $(0.241, 0.0694)$ |

Now, lets consider the error-correcting WOM codes. For simplicity, we assume the parities in groups $\mathcal{P}_1$ (size $k$) and $\mathcal{P}_2$ (size $k'$) have equal size and equal degrees. Recall that when decoding message $s^k$ from a noisy version of $x^n$, the codebook of interest is $\mathcal{C}$ (where parities $\mathcal{P}_1$ are not present). In the construction of the compound code, we ensure that all LDPC bit-nodes have equal degrees (except at the boundary in a spatially-coupled code) in both $\mathcal{C}$ and $\mathcal{C}(s^k)$. Thus, if the code $\mathcal{C}(s^k)$ has the degree profile $(d_v, d_c, d'_v, d'_c)$, due to the equality assumptions about $\mathcal{P}_1$ and $\mathcal{P}_2$, the code $\mathcal{C}$ has the degree profile $(d_v, d_c, d'_v/2, d'_c)$.

Consider a compound code $\mathcal{C}(s^k)$ with the degree profile $(d_v, d_c, d'_v, d'_c)$. With the above assumptions,

$$m = n\frac{d_c}{d_v}, \qquad k = k' = \frac{m}{2}\frac{d'_v}{d'_c} = \frac{n}{2}\frac{d_c d'_v}{d_v d'_c}.$$

Thus, the rate of the second-write is given by

$$R = \frac{k}{n} = \frac{1}{2}\frac{d_c d_v'}{d_v d_c'}.$$

Since the degree profile of $\mathcal{C}$ is given by $(d_v, d_c, d_v'/2, d_c')$, its code-rate is given by

$$\frac{m-k'}{n} = \frac{d_c}{d_v} - \frac{1}{2}\frac{d_c d_v'}{d_v d_c'} = \frac{d_c}{d_v}\left(1 - \frac{d_v'}{2d_c'}\right).$$

Therefore, the maximum BSC parameter $p^*$ below which the code $\mathcal{C}$ can correct errors with high probability is

$$p^* = h^{-1}(1 - \text{code-rate}) = h^{-1}\left(1 - \frac{d_c}{d_v}\left(1 - \frac{d_v'}{2d_c'}\right)\right).$$

From the rate region in (3), the maximum normalized weight $\delta^*$ of the state sequence after first write below which we can successfully encode the second message is

$$\delta^* = 1 - h(p^*) - R = \frac{d_c}{d_v} - \frac{d_c d_v'}{d_v d_c'}.$$

In Table II, the achievable thresholds $(\delta, p)$ are shown for the spatially-coupled compound codes. For these codes, a chain length of $L = 30$ is used and the thresholds are shown for different coupling window sizes. Again, we have not included the rate loss when reporting the optimal thresholds $(\delta^*, p^*)$. The blocklength of the single system is roughly of size 32000.

## V. Conclusion

We have constructed spatially-coupled compound LDGM/LDPC codes that achieve the capacity region of the write-once memory (WOM) systems. The focus is on the second write of the 2-write WOM system. The first write is an instance of the Gelfand-Pinsker problem, for which we constructed capacity-achieving practical coding schemes from compound codes in an earlier article [20]. Encoding for the second write is done by the iterative quantization algorithm from a reduction to the binary erasure quantization problem. The performance of these codes appears to be better than some current implementations of the polar codes. Finally, this construction appears to be the only non-polar coding scheme that can correct a constant fraction of errors with high probability.

## References


[1] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Information and Control*, vol. 55, no. 13, pp. 1 – 19, 1982.

[2] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inf. Theory*, vol. 31, pp. 34–42, Jan 1985.

[3] J. K. Wolf, A. D. Wyner, J. Ziv, and J. Krner, "Coding for a write-once memory," *AT&T Bell Laboratories Technical Journal*, vol. 63, no. 6, pp. 1089–1112, 1984.

[4] F. Merkx, "Womcodes constructed with projective geometries," *Traitement du signal*, vol. 1, pp. 227–231, 1984.

[5] G. Cohen, P. Godlewski, and F. Merkx, "Linear binary code for write-once memories (corresp.)," *IEEE Trans. Inf. Theory*, vol. 32, pp. 697–700, Sep 1986.

[6] Y. Wu, "Low complexity codes for writing a write-once memory twice," in *IEEE Int. Symp. Inf. Th.*, pp. 1928–1932, June 2010.

[7] Y. Wu and A. Jiang, "Position modulation code for rewriting write-once memories," *IEEE Trans. Inf. Theory*, vol. 57, pp. 3692–3697, June 2011.

[8] E. Yaakobi, S. Kayser, P. Siegel, A. Vardy, and J. Wolf, "Codes for write-once memories," *IEEE Trans. Inf. Theory*, vol. 58, pp. 5985–5999, Sept 2012.

[9] A. Shpilka, "New constructions of wom codes using the Wozencraft ensemble," *IEEE Trans. Inf. Theory*, vol. 59, pp. 4520–4529, July 2013.

[10] D. Burshtein and A. Strugatski, "Polar write once memory codes," *IEEE Trans. Inf. Theory*, vol. 59, pp. 5088–5101, Aug 2013.

[11] E. E. Gad, W. Huang, Y. Li, and J. Bruck, "Rewriting flash memories by message passing," in *IEEE Int. Symp. on Inf. Th.*, (Hong Kong, China), June 2015.

[12] G. Zémor, *Problemes combinatoires lies a l'ecriture sur des memoires*. PhD thesis, ENST, Paris, France, 1989.

[13] G. Zémor and G. Cohen, "Error-correcting wom-codes," *IEEE Trans. Inf. Theory*, vol. 37, pp. 730–734, May 1991.

[14] E. Yaakobi, P. Siegel, A. Vardy, and J. Wolf, "Multiple error-correcting wom-codes," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2220–2230, April 2012.

[15] A. Jiang, Y. Li, E. Gad, M. Langberg, and J. Bruck, "Joint rewriting and error correction in write-once memories," in *IEEE Int. Symp. Inf. Theory*, pp. 1067–1071, July 2013.

[16] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," *CoRR*, vol. abs/1410.3542, 2014.

[17] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inform. Theory*, vol. 55, pp. 1061–1079, March 2009.

[18] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving codes with bounded graphical complexity and maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 992–1006, 2010.

[19] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Inform. Transm.*, vol. 9, pp. 19–31, 1980.

[20] S. Kumar, A. Vem, K. Narayanan, and H. D. Pfister, "Spatially-coupled codes for side-information problems," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 516–520, 2014.

[21] E. Martinian and J. S. Yedidia, "Iterative quantization using codes on graphs," in *in Proc. Allerton Conf. Comm., Cont., and Comp.*, 2003.

[22] M. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1351–1368, 2010.

[23] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. New York, NY: Cambridge University Press, 2008.

[24] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inform. Theory*, vol. 57, pp. 803–834, Feb. 2011.

[25] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5274–5289, Oct. 2010.

[26] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inform. Theory*, vol. 59, pp. 7761–7813, Dec. 2013.

[27] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially-coupled LDPC and LDGM codes on BMS channels," *IEEE Trans. Inform. Theory*, vol. 60, pp. 7389–7415, Dec. 2014.

[28] K. Tazoe, K. Kasai, and K. Sakaniwa, "Efficient termination of spatially-coupled codes," in *IEEE Inf. Th. Workshop (ITW)*, pp. 30–34, Sept 2012.

[29] A. Iyengar, P. Siegel, R. Urbanke, and J. Wolf, "Windowed decoding of spatially coupled codes," *IEEE Tran. on Inf. Th.*, vol. 59, pp. 2277–2292, April 2013.

[30] P. Olmos and R. Urbanke, "A scaling law to predict the finite-length performance of spatially-coupled ldpc codes," *IEEE Tran. on Inf. Th.*, vol. 61, pp. 3164–3184, June 2015.