Quasi-Cyclic LDPC Codes for Correcting Multiple Phased Bursts of Erasures

Xin Xiao and Bane Vasić
University of Arizona
Email: {7xinxiao7,vasic}@email.arizona.edu

Shu Lin and Khaled Abdel-Ghaffar
University of California, Davis
Email: {shulin, ghaffar}@ucdavis.edu

William E. Ryan
Zeta Associates Inc.
Email: bill.ryan.work@gmail.com

***Abstract*—This paper presents designs and constructions of two classes of binary quasi-cyclic LDPC codes for correcting multiple random *phased-bursts* of erasures over the binary erasure channel. The erasure correction of codes in both classes is characterized by the *cycle and adjacency structure* of their Tanner graphs. Erasure correction of these codes is a very simple process which requires only modulo-2 additions. The codes in the second class are capable of correcting *locally* and *globally* distributed phased-bursts of erasures with a *two-phase* iterative erasure-correction process.**

I. Introduction

Over a binary erasure channel (BEC) [1], erasures in a received codeword from a code may occur in random or cluster together in bursts. If erasures in a received codeword are clustered in a *span of consecutive positions (including end-around case)*, we call such an erasure pattern a *burst* of erasures. A burst of erasures is said to have length b if the erased code symbols are confined to b *consecutive positions* in which the code symbols at the first and the last positions are erased. LDPC codes for correcting erasures over a BEC were first investigated in [2], [3] and later in [4]–[8] and others.

In this paper, we consider binary quasi-cyclic (QC) LDPC code with parity-check matrices which are arrays of *circulant permutation matrices* (CPMs) of a certain size, say $l \times l$ [1], [9]. These are most widely known, studied, and used QC-LDPC codes. Let $\mathcal{C}$ be a binary QC-LDPC code whose parity check matrix $\mathbf{H}$ is an $m \times n$ array of CPMs of size $l \times l$. The array $\mathbf{H}$ consists of n columns of CPMs, called CPM column-blocks, and m rows of CPMs, called CPM row-blocks. Hence $\mathbf{H}$ is an $lm \times ln$ matrix over GF(2). Each codeword $\mathbf{v}$ in $\mathcal{C}$ consists of n sections of length l, each section consisting of l code symbols which corresponds to the l columns of a CPM column-block of $\mathbf{H}$. We call these sections of $\mathbf{v}$ *CPM-sections*.

Suppose in the transmission of a codeword $\mathbf{v}$ in $\mathcal{C}$, erasures occur and are confined in CPM-sections of $\mathbf{v}$. A burst with erasures confined to a CPM-section of $\mathbf{v}$ is called a *CPM-phased burst of erasures*. The longest length of such an erasure-burst is l. If all l code symbols in a CPM-section of $\mathbf{v}$ are erased, the burst of erasures is said to be *solid*, otherwise, it is not solid, i.e., at least one of the code symbols in the CPM-section is not erased.

The focus of this paper is in constructions of QC-LDPC codes for correcting *randomly distributed multiple* CPM-phased bursts of erasures. The construction of such a code is based on the *conventional* parity-check matrix of an Reed-Solomon (RS) code. We referred to such a code as an RS-based QC-LDPC code, simply RS-QC-LDPC code. The erasure correction of an RS-QC-LDPC code is characterized by the *cycle and adjacency structure* of its Tanner graph. Two classes of RS-QC-LDPC codes for correcting multiple CPM-phased bursts of erasures are constructed. Codes in both classes are *optimal* in the sense that the maximum number of correctable erasures of a code is equal to the number of parity-check symbols of the code, i.e., meeting the Reiger bound for erasure correction [10].

The rest of the paper is organized as follows. Section II presents the basic construction of an RS-QC-LDPC code. Section III presents a class of optimal RS-QC-LDPC codes which are capable of correcting two random CPM-phased bursts of erasures with a very simple correction process. Section IV presents a class of optimal *globally coupled* RS-QC-LDPC codes [11], [12] for correcting multiple *locally* and *globally distributed* CPM-phased bursts of erasures. Also presented in this section is a *two-phase local/global iterative erasure recovery scheme*. Section V concludes the paper with some remarks.

II. Construction of RS-QC-LDPC Codes

Let n be a *prime factor* of $2^s - 1$ and β be an element in GF(2^s) of order n. The set $\mathbf{S}_n = \{1, \beta, ..., \beta^{n-1}\}$ form a cyclic subgroup of GF(2^s) of order n. For $1 \leq d \leq n$, we form the following $d \times n$ matrix:

$$\mathbf{B}_{RS}(d,n) = \left[\beta^{ij}\right]_{1 \leq i \leq d, 0 \leq j < n}. \quad (1)$$

Label the columns of $\mathbf{B}_{RS}(d,n)$ from 0 to $n-1$. Since n is a prime, all the n entries in a row of $\mathbf{B}_{RS}(d,n)$ are distinct and form all the elements in $\mathbf{S}_n$. Except for the 0-th column, the d entries in any other column of $\mathbf{B}_{RS}(d,n)$ are distinct. The null space over GF(2^s) of $\mathbf{B}_{RS}(d,n)$ gives an $(n, n-d, d+1)$ RS code $\mathcal{C}_{RS}(d,n)$ over GF(2^s) of length n, dimension $n-d$ and minimum distance $d+1$ [1]. We call $\mathbf{B}_{RS}(d,n)$ an RS matrix.

For $1 \leq i \leq d$ and $0 \leq j < n$, let $l_{ij} = (ij)_n$ denotes the *smallest* nonnegative integer congruent to ij modulo n. Then,

$$\mathbf{B}_{RS}(d,n) = \left[\beta^{l_{ij}}\right]_{1 \leq i \leq d, 0 \leq j < n}. \quad (2)$$

It is easy to prove that the RS matrix $\mathbf{B}_{RS}(d,n)$ has the following structure: *any* 2×2 *submatrix of* $\mathbf{B}_{RS}(d,n)$ is *nonsingular* (NS). We referred this structure as the 2×2 submatrix (S) NS-constraint, simply 2×2 *SNS-constraint* [9].

Next, we represent each entry $\beta^{l_{ij}}$ in $\mathbf{B}_{RS}(d,n)$ by a CPM of size $n \times n$ (with rows and columns labeled from 0 to $n-1$, respectively) whose *generator* has the unit-element '1' of GF(2^s) as its *single nonzero component* at position l_{ij}. This matrix representation of $\beta^{l_{ij}}$ is referred to as the *CPM-dispersion* of $\beta^{l_{ij}}$, denoted by $CPM(\beta^{l_{ij}})$. Dispersing each entry in $\mathbf{B}_{RS}(d,n) = \left[\beta^{l_{ij}}\right]_{1 \le i \le d, 0 \le j < n}$ into an $n \times n$ CPM, we obtain a $d \times n$ array $\mathbf{H}_{RS}(d,n) = \left[CPM(\beta^{l_{ij}})\right]_{1 \le i \le d, 0 \le j < n}$ of CPMs which consists of d CPM *row-blocks* and n CPM *column-blocks*. Each CPM column-block consists of d CPMs. The array $\mathbf{H}_{RS}(d,n)$ is a $dn \times n^2$ matrix over GF(2) with column and row weights d and n, respectively. The array $\mathbf{H}_{RS}(d,n)$ is called the $n \times n$ CPM-dispersion of $\mathbf{B}_{RS}(d,n)$. As a matrix, $\mathbf{H}_{RS}(d,n)$ has the following structure: *no two rows (or two columns) have more than one location in which both have 1-entries* [1], [5], [7], [9]. We say that $\mathbf{H}_{RS}(d,n)$ satisfies the *row-column* (RC) *constraint*. This RC-constraint structure of $\mathbf{H}_{RS}(d,n)$ is ensured by the 2×2 SNS-constraint structure of the RS matrix $\mathbf{B}_{RS}(d,n)$ [9]. Following the structure of $\mathbf{B}_{RS}(d,n)$, we readily see that the n CPMs of each CPM row-block of $\mathbf{H}_{RS}(d,n)$ are *distinct*. Except for the 0-th CPM column-block, the d CPMs in any other CPM column-block of $\mathbf{H}_{RS}(d,n)$ are distinct. If $n >> 1$, $\mathbf{H}_{RS}(d,n)$ is a sparse matrix.

Using $\mathbf{H}_{RS}(d,n)$ as the parity-check matrix, the null space over GF(2) of $\mathbf{H}_{RS}(d,n)$ gives a (d,n)-regular RS-QC-LDPC code, denoted by $\mathcal{C}_{RS,ldpc}(d,n)$. Let $\mathcal{G}_{RS,ldpc}(d,n)$ denote the Tanner graph of $\mathcal{C}_{RS,ldpc}(d,n)$. The RC-constraint on $\mathbf{H}_{RS}(d,n)$ (or the 2×2 SNS-constraint on $\mathbf{B}_{RS}(d,n)$) ensures that $\mathcal{G}_{RS,ldpc}(d,n)$ has girth of at least 6 [9]. Since the parity-check matrix $\mathbf{H}_{RS}(d,n)$ and the LDPC code $\mathcal{C}_{RS,ldpc}(d,n)$ are constructed based on $\mathbf{B}_{RS}(d,n)$, we call $\mathbf{B}_{RS}(d,n)$ the *base matrix*. It is shown in [12], [13] that RS-QC-LDPC codes perform well over the AWGN channel.

III. Optimal RS-QC-LDPC Codes for Correcting Two Random CPM-Phased Bursts of Erasures

A measure of the erasure-correction performance of a code is defined as the ratio of the maximum number of erasures that the code can correct to the number of parity-check symbols of the code. We call this ratio, denoted by η, *erasure-correction efficiency* [1], [5]. An erasure-correction code is said to be *optimal* if $\eta = 1$. In this section, we present a class of optimal RS-QC-LDPC codes for correcting two *random* CPM-phased bursts of erasures. The erasure correction of a code in this class is characterized by the *cycle and adjacency structure* of its Tanner graph.

A. Codes

Set $d = 2$ in (1), we obtain a $2 \times n$ RS matrix $\mathbf{B}_{RS}(2,n) = \left[\beta^{ij}\right]_{1 \le i \le 2, 0 \le j < n}$. The $n \times n$ CPM-dispersion of $\mathbf{B}_{RS}(2,n)$ gives a $2 \times n$ array $\mathbf{H}_{RS}(2,n) = \left[CPM(\beta^{l_{ij}})\right]_{1 \le i \le 2, 0 \le j < n}$ of CPMs of size $n \times n$ which consists of two CPM row-blocks and n CPM column-blocks. Each CPM column-block consists of two CPMs. The array $\mathbf{H}_{RS}(2,n)$ is a $2n \times n^2$ matrix over GF(2) with column and row weights 2 and n, respectively. Since the sum of n rows in each CPM row-block results in a row all ones, the sum of all rows of $\mathbf{H}_{RS}(2,n)$ gives a row of all zeros. Based on the composition of CPMs in $\mathbf{H}_{RS}(2,n)$, it can be proved that the rank of $\mathbf{H}_{RS}(2,n)$ is $2n-1$. The matrix $\mathbf{H}_{RS}(2,n)$ contains one redundant row. The null space over GF(2) of $\mathbf{H}_{RS}(2,n)$ gives a $(2,n)$-regular (n^2, n^2-2n+1) RS-QC-LDPC code $\mathcal{C}_{RS,ldpc}(2,n)$.

Label the CPM row-blocks and column-blocks of $\mathbf{H}_{RS}(2,n)$ from 1 to 2 and from 0 to $n-1$, respectively. Label the n columns of each CPM column-block from 0 to $n-1$. Each codeword $\mathbf{v} = (\mathbf{v}_0, \mathbf{v}_1, ..., , \mathbf{v}_{n-1})$ in $\mathcal{C}_{RS,ldpc}(2,n)$ consists of n CPM-sections $\mathbf{v}_j, 0 \le j < n$, each consisting of n consecutive code symbols. The n code symbols in the j-th section $\mathbf{v}_j$ of $\mathbf{v}$ correspond to n columns of the j-th CPM column-block of $\mathbf{H}_{RS}(2,n)$.

B. Cycle Structure of the Tanner Graph of $\mathcal{C}_{RS,ldpc}(2,n)$

The Tanner graph, denoted by $\mathcal{G}_{RS,ldpc}(2,n)$, of the $(2,n)$-regular RS-QC-LDPC code $\mathcal{C}_{RS,ldpc}(2,n)$ consists of n^2 variable nodes (VNs) and $2n$ check-nodes (CNs). Each VN v is connected to two CNs and each CN c is connected to n VNs. Each VN v is connected to $n-1$ other VNs through the same CN c by *paths of length* 2 which are called *neighbor* VNs of v. Hence, each VN in $\mathcal{G}_{RS,ldpc}(2,n)$ has $2(n-1)$ neighbor VNs. Since $\mathbf{H}_{RS}(2,n)$ satisfies the RC-constraint, $\mathcal{G}_{RS,ldpc}(2,n)$ contains no cycle of length 4. In the following, we analyze *local cycle structure* of $\mathcal{G}_{RS,ldpc}(2,n)$ which is pertinent for $\mathcal{C}_{RS,ldpc}(2,n)$ to correct CPM-phased bursts of erasures.

Let $CPM(l)$ denote an $n \times n$ CPM in which the *generator* has its single 1-enty in position l, $0 \le l < n$. Then, $CPM(l)$ has n 1-entries in positions $(i,(i+l)_n), 0 \le i < n$, where $(x)_n$ denotes the least nonnegative integer equal to x modulo n. For $0 \le i,j < n$, a 1-entry resides in position (i,j) of $CPM(l)$ if and only if $j - i \equiv l (\mathrm{mod} n)$.

For $0 \le i,j < n$ with $i < j$, consider the following 2×2 subarray of $\mathbf{H}_{RS}(2,n)$:

$$\mathbf{H}(i,j) = \begin{bmatrix} CPM(\beta^i) & CPM(\beta^j) \\ CPM(\beta^{2i}) & CPM(\beta^{2j}) \end{bmatrix} \quad (3)$$

which consists of the i-th and j-th CPM column-blocks of $\mathbf{H}_{RS}(2,n)$. Let $a = (i)_n, b = (2i)_n, c = (2j)_n$, and $d = (j)_n$. Then, $0 \le a,b,c,d < n$. It is clear $a = i$ and $d = j$. The numbers, a, b, c, and d, specify the locations of the single 1-entries of the generators of the 4 CPMs in $\mathbf{H}(i,j)$. Replacing the notations of the 4 CPMs in $\mathbf{H}(i,j)$ by $CPM(a), CPM(b), CPM(c)$, and $CPM(d)$, respectively. Then, the 2×2 array $\mathbf{H}(i,j)$ given by (3) is put in the following form:

$$\mathbf{H}(i,j) = \begin{bmatrix} CPM(a) & CPM(d) \\ CPM(b) & CPM(c) \end{bmatrix} \quad (4)$$

Let $\mathcal{G}(i,j)$ be the Tanner graph associated with the 2×2 array $\mathbf{H}(i,j)$ of CPMs. Then, the Tanner graph $\mathcal{G}(i,j)$ has $2n$ VNs and $2n$ CNs. Consider a cycle in $\mathcal{G}(i,j)$. Then, its length should be a *multiple* of 4, say $4t, 2 \leq t \leq n$. Such a cycle corresponds to a *sequence* of 1-entires in $4t$ positions in $\mathbf{H}(i,j)$ of the following form (*traced in counter clock-wise column-row order*):

$$(i_0,j_0),(i_1,j_0),(i_1,j_1),(i_2,j_1),(i_2,j_2),(i_3,j_2),...,$$
$$(i_{2t-2},j_{2t-2}),(i_{2t-1},j_{2t-2}),(i_{2t-1},j_{2t-1}),(i_0,j_{2t-1}),(i_0,j_0) \quad (5)$$

where: (1) $i_0, i_1, ..., i_{2t-1}$ are distinct and $j_0, j_1, ..., j_{2t-1}$ are distinct; (2) $(i_0,j_0),(i_2,j_2),...,(i_{2t-2},j_{2t-2})$ are the positions of n 1-entries in $CPM(a)$; (3) $(i_1,j_0),(i_3,j_2),...,(i_{2t-1},j_{2t-2})$ are the positions of n 1-entries in $CPM(b)$; (4) $(i_1,j_1),(i_3,j_3),...,(i_{2t-1},j_{2t-1})$ are the positions of n 1-entries in $CPM(c)$; and (5) $(i_0,j_1),(i_2,j_3),...,(i_{2t-2},j_{2t-1})$ are positions of n 1-entries in $CPM(d)$.

The positions of 1-entries in the $CPM(a), CPM(b), CPM(c)$ and $CPM(d)$, must satisfy the following conditions, respectively: (1) $j_r - i_r \equiv a \pmod{n}$ for $r = 0, 2, ..., 2t-2$; (2) $j_{r+1} - i_r \equiv b \pmod{n}$ for $r = 1, 3, ..., 2t-1$; (3) $j_r - i_r \equiv c \pmod{n}$ for $r = 1, 3, ..., 2t-1$; and (4) $j_{r+1} - i_r \equiv d \pmod{n}$ for $r = 0, 2, ..., 2t-2$. Following the constraints on the positions of 1-entries in $CPM(a), CPM(b), CPM(c)$ and $CPM(d)$, it can be proved readily with some algebraic manipulations that

$$\sum_{r=0}^{2t-1} j_r - \sum_{r=0}^{2t-1} i_r \equiv at + ct (\mathrm{mod} n) \quad (6)$$

$$\sum_{r=0}^{2t-1} j_r - \sum_{r=0}^{2t-1} i_r \equiv bt + dt (\mathrm{mod} n). \quad (7)$$

It follows from (6) and (7) that we have $at + ct \equiv bt + dt$ modulo n. Hence, $t(a-b+c-d)$ is divisible by n, i.e., $(t(a-b+c-d))_n \equiv 0$. From this, we readily see that the *shortest cycle* in $\mathcal{G}(a,b,c,d)$ has length 4λ where $\lambda = n/\gcd(n,(a-b+c-d)_n)$ and $\gcd(n,(a-b+c-d)_n)$ denotes the greatest common divisor of n and $(a-b+c-d)_n$. Since n is prime, $\gcd(n,(a-b+c-d)_n) = 1$. Hence, $\lambda = n$ and $\mathcal{G}(i,j)$ is composed of a *single cycle* of length $4n$ which includes the $2n$ VNs, $2n$ CNs and $4n$ edges of $\mathcal{G}(i,j)$. Since $\mathcal{G}(i,j)$ is a subgraph of the Tanner graph $\mathcal{G}_{RS,ldpc}(2,n)$ of $\mathcal{C}_{RS,ldpc}(2,n)$, the cycle structure of $\mathcal{G}(i,j)$ is referred to as the *local cycle structure* of $\mathcal{G}_{RS,ldpc}(2,n)$.

C. Erasure Correction

Let $\mathbf{v} = (\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{n-1})$ be a codeword in $\mathcal{C}_{RS,ldpc}(2,n)$ which consists of n CPM-sections, each consisting of n code symbols. Two code symbols in $\mathbf{v}$ are said to be *neighbors* if they correspond to two VNs in $\mathcal{G}_{RS,ldpc}(2,n)$ which are neighbors (connected by a path of length 2). It follows from the local cycle structure of the Tanner graph of the code $\mathcal{C}_{RS,ldpc}(2,n)$ that each code symbol in a CPM-section $\mathbf{v}_i$ has exactly two neighbors in a different CPM-section $\mathbf{v}_j$ of $\mathbf{v}$. This neighbor structure is referred to as *adjacency structure* of code symbols of a codeword in $\mathcal{C}_{RS,ldpc}(2,n)$.

Suppose, during the transmission of a codeword $\mathbf{v} = (\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{n-1})$ in $\mathcal{C}_{RS,ldpc}(2,n)$, two CPM-phased bursts of erasures occur, one in the i-th CPM-section $\mathbf{v}_i$ and the other in the j-th CPM-section $\mathbf{v}_j$ of $\mathbf{v}$. A CPM-phased burst of erasures may contain up to n erasures. If the number of erasures is n, then we say that the CPM-phased burst of erasures is *solid*. Two CPM-phased bursts of erasures are said to be *mutually semi-solid* if the total number of erasures is $2n-1$, i.e., one CPM-phased burst is solid and the other contains $n-1$ erasures (not solid). In the following, we show that based on the local cycle structure of its Tanner graph, the code $\mathcal{C}_{RS,ldpc}(2,n)$ is capable of correcting two *random* CPM-phased-bursts with the number of erasures up to $2n-1$ (i, e., two mutually semi-solid CPM-phased bursts of erasures), which is equal to number parity-check symbols of the code $\mathcal{C}_{RS,ldpc}(2,n)$.

It follows from the adjacency structure of $\mathcal{C}_{RS,ldpc}(2,n)$ that there exists at least one erased symbol v in the i-th (or j-th) CPM-section which is the neighbor of an *un-erased* code symbol in the j-th (or i-th) CPM-section of $\mathbf{v}$. Hence, there is row $\mathbf{h}$ in $\mathbf{H}_{RS}(2,n)$ which checks v as the only erased symbol. Take the inner product of $\mathbf{h}$ and $\mathbf{v}$ to form a check-sum $\sum$. Then, $\sum$ contains v as the only *unknown*. From $\sum$, we recover v by taking the modulo-2 sum of the un-erased symbols in $\sum$. After recovering v, we find another erased code symbol which is the neighbor of an un-erased code symbol or the recovered code symbol v in the two CPM-phased bursts of erasures confined in the CPM-sections, $\mathbf{v}_i$ and $\mathbf{v}_j$, of $\mathbf{v}$. We recover this erased code symbol in exactly the same manner as the recovery of the first erased code symbol v. Continue this recovery process until all the erased code symbols in the CPM-sections $\mathbf{v}_i$ and $\mathbf{v}_j$ of $\mathbf{v}$ are recovered. Since all the erased code symbols in $\mathbf{v}_i$ and $\mathbf{v}_j$ correspond to the $2n$ VNs on the single cycle of length $4n$ in $\mathcal{G}(i,j)$, there is always an un-erased or recovered code symbol which is the neighbor of an erased code symbol. Hence, every erased code symbol in the CPM-sections $\mathbf{v}_i$ and $\mathbf{v}_j$ of $\mathbf{v}$ can be recovered. The erasure recovery process will never fail as long as there is at least one un-erased code symbol in two CPM-phased bursts of erasures which are confined in two CPM-sections of a transmitted codeword $\mathbf{v}$ in $\mathcal{C}_{RS,ldpc}(2,n)$. Erasure correction requires only modulo-2 addition operations. The above erasure recovering process is referred to as the *peeling algorithm* [1], [14].

Since the maximum number of correctable erasures confined to two CPM-sections of a received codeword $\mathbf{v}$ in the code $\mathcal{C}_{RS,ldpc}(2,n)$ is $2n-1$ which is equal to the number of parity-check symbols of the code, hence the code $\mathcal{C}_{RS,ldpc}(2,n)$ is optimal i.e., the erasure-correction efficiency $\eta = 1$.

Example 1: Let GF(2^9) be the field for code construction. Note that $2^9 - 1 = 511$ can be factored as the product of primes 7 and 73. Set $n = 73$. Let β an element in GF(2^9) of order 73. We first construct a 2×73 RS matrix $\mathbf{B}_{RS}(2,73)$ over GF(2^9) in the form of (2). Dispersing each entry in $\mathbf{B}_{RS}(2,73)$ into a 73×73 CPM, we obtain a 2×73 array

$\mathbf{H}_{RS}(2, 73)$ of CPM of size 73×73 which is a 146×5329 matrix. The null space over GF(2) of $\mathbf{H}_{RS}(2, 73)$ gives a (5329, 5184) RS-QC-LDPC code $\mathcal{C}_{RS,ldpc}(2, 73)$ with rate 0.9737. Each codeword in $\mathcal{C}_{RS,ldpc}(2, 73)$ consists of 73 CPM-sections, each consisting 73 code symbols. The number of parity-check symbols in each codeword of $\mathcal{C}_{RS,ldpc}(2, 73)$ is 145 parity-check symbols. The code $\mathcal{C}_{RS,ldpc}(2, 73)$ is capable of correcting any two random CPM-phased bursts of erasures, one not solid, with a maximum number of 145 correctable erasures. The code $\mathcal{C}_{RS,ldpc}(2, 73)$ is optimal in correction of two random CPM-phased bursts of erasures, one not solid.

IV. Globally Coupled RS-QC-LDPC Codes for Correcting Multiple CPM-Phased Bursts of Erasures

The RS matrix $\mathbf{B}_{RS}(2, n)$ given in form of (2) can be used as a *building block* to construct codes to correct multiple CPM-phased bursts of erasures. One such construction is presented in this section.

For an integer $k \geq 2$, form the following $(2k + n) \times kn$ matrix over GF(2^s):

$$\mathbf{B}_{GC,RS}(2, n, k) = \begin{bmatrix} \mathbf{B}_{RS}(2, n) & & & \\ & \mathbf{B}_{RS}(2, n) & & \\ & & \ddots & \\ & & & \mathbf{B}_{RS}(2, n) \\ \mathbf{I}(n, n) & \mathbf{I}(n, n) & \cdots & \mathbf{I}(n, n) \end{bmatrix} \quad (8)$$

This matrix consists of two submatrices. The upper submatrix is a $k \times k$ array with k copies of $\mathbf{B}_{RS}(2, n)$ lying on its main diagonal and zeros elsewhere. The lower submatrix is $1 \times k$ array of $n \times n$ identify matrices $\mathbf{I}(n, n)$. The $n \times n$ CPM-dispersion of $\mathbf{B}_{GC,RS}(2, n, k)$ gives an array $\mathbf{H}_{GC,RS}(2, n, k)$ which is a $(2k + n)n \times kn^2$ matrix over GF(2) with constant column weight 3. Note that the $n \times n$ CPM-dispersion of $\mathbf{B}_{RS}(2, n)$ is $\mathbf{H}_{RS}(2, n)$ and the CPM-dispersion of the $n \times n$ identity $\mathbf{I}(n, n)$ is an $n^2 \times n^2$ identify matrix $\mathbf{I}(n^2, n^2)$. The matrix $\mathbf{H}_{GC,RS}(2, n, k)$ consists of two submatrices, the upper one and the lower one. The upper submatrix of $\mathbf{H}_{GC,RS}(2, n, k)$ is a $k \times k$ diagonal array with k copies of the $2 \times n$ array $\mathbf{H}_{RS}(2, n)$ lying on it main diagonal. The lower submatrix is a row of k copies of the $n^2 \times n^2$ identity matrix $\mathbf{I}(n^2, n^2)$.

It can be proved that the rank of the CPM-dispersion $\mathbf{H}_{GC,RS}(2, n, k)$ of $\mathbf{B}_{GC,RS}(2, n, k)$ is

$$rank(\mathbf{H}_{GC,RS}(2, n, k)) = (2n - 1)(k - 1) + n^2. \quad (9)$$

(Proof is skipped due to page limitation.) Hence, the null space over GF(2) of $\mathbf{H}_{GC,RS}(2, n, k)$ gives a $(kn^2, (k-1)(n-1)^2)$ QC-LDPC code $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$.

Let $\mathcal{G}_{GC,RS,ldpc}(2, n, k)$ be the Tanner graph of $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$. From (8), we see that $\mathcal{G}_{GC,RS,ldpc}(2, n, k)$ has a *global structure*. It consists of k disjoint copies of the Tanner graph $\mathcal{G}_{RS,ldpc}(2, n)$ of $\mathcal{C}_{RS,ldpc}(2, n)$ connected by a set of n^2 global CNs which correspond to the n^2 rows of the lower submatrix $\mathbf{G}(n^2, k) = [\mathbf{I}(n^2, n^2), \mathbf{I}(n^2, n^2), ..., \mathbf{I}(n^2, n^2)]$ of $\mathbf{H}_{GC,RS}(2, n, k)$. With this globally coupling structure of $\mathcal{G}_{GC,RS,ldpc}(2, n, k)$, the code $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$ is referred to as a *CN-based globally coupled* (GC) RS-based QC-LDPC code, denoted by CN-GC-RS-QC-LDPC code. The code $\mathcal{C}_{RS,ldpc}(2, n)$ and its Tanner graph $\mathcal{G}_{RS,ldpc}(2, n)$ are called *local code* and *local graph*, respectively. The matrix $\mathbf{G}(n^2, k)$ is called the *global connection matrix*. Note that between two adjacent 1-entries in a row of $\mathbf{G}(n^2, k)$, there is span of n^2-zeros. Such a span of zeros is called a *zero-span*. Global coupling was first introduced in [11], [12].

The code $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$ has local and global CPM-phased erasure-burst correction characteristics. A codeword $\mathbf{u}$ in $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$ contains of k sub-codewords, denoted by $\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{k-1}$, each consisting of n^2 consecutive code symbols of $\mathbf{u}$. From the structure of $\mathbf{B}_{GC,RS}(2, n, k)$ displayed by (8), we see that each sub-codeword $\mathbf{u}_i$ of $\mathbf{u}$ is a codeword in the local code $\mathcal{C}_{RS,ldpc}(2, n)$. These sub-codewords are called *local codewords*. The k local codewords, $\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{k-1}$, in cascade must satisfies the n^2 global parity-check constraints specified by the n^2 rows of the global connection submatrix $\mathbf{G}(n^2, k)$ of $\mathbf{H}_{GC,RS}(2, n, k)$.

The erasure-correction process consists of two phases, the *local phase* and the *global phase*. When a transmitted codeword $\mathbf{u} = (\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{k-1})$ in $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$ is received. The local erasure-correction phase is activated. We decode each local codeword $\mathbf{u}_i$ of $\mathbf{u}$ as soon as it is received by a *local erasure (LE) decoder*. If each received local codeword $\mathbf{u}_i$ contains no more than two CPM-phased bursts with no more than $2n - 1$ erasures, regardless of their locations, the erased code symbols can be recovered. If every local received codeword is contaminated by two or fewer correctable CPM-phased bursts of erasures, all the erasures in the received codeword $\mathbf{u}$ will be corrected in the local erasure-correction phase. The local erasure correction process is exactly the same as describe in Section II.B. In the local phase of erasure-correction process, each received local codeword is decoded by the same *LE-decoder*.

At the completion of local phase of erasure-correction process, if all the local received codewords are successfully decoded, we stop the erasure-correction process. Otherwise, the global phase of erasure-correction process is activated. In this case, the erasure correction is performed based on the global connection matrix $\mathbf{G}(n^2, k)$. Let $\mathbf{u}^{(1)}$ be decoded word at the output of the LE-decoder. For an erasure e in $\mathbf{u}^{(1)}$, if there is a row $\mathbf{g}$ in $\mathbf{G}(n^2, k)$ which checks the erasure e only and no other erasures in $\mathbf{u}^{(1)}$, then the erasure e can be recovered from the check-sum $\sum$ formed by the inner product of $\mathbf{g}$ and $\mathbf{u}^{(1)}$. Such an erasure e is said to be *globally recoverable*. If all the erasures in $\mathbf{u}^{(1)}$ are globally recoverable, then the *global erasure* (GE) decoder will recover all the erasures in $\mathbf{u}^{(1)}$. In this case, we stop the erasure correction process.

Based on the zero-span structure of the global connection matrix $\mathbf{G}(n^2, k)$, we readily see that all the erasures in $\mathbf{u}^{(1)}$ are recoverable if the locations of CPM-phased bursts of erasures in the fail decoded local codewords in $\mathbf{u}^{(1)}$ are *disjoint*. In this case, each erasure in $\mathbf{u}^{(1)}$ is checked alone by a row in $\mathbf{G}(n^2, k)$ and hence it is recoverable. A very special case is that all the fail decoded CPM-phased bursts of erasures in $\mathbf{u}^{(1)}$ are confined in n consecutive CPM-sections of $\mathbf{u}^{(1)}$ (including the end-around case). In this case, the GE-decoder will correct all the erasures in $\mathbf{u}^{(1)}$. This means that, during the transmission of a global codeword $\mathbf{u}$ in $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$, if a single local codeword in $\mathbf{u}$ is completely erased, it can be recovered by the GE-decoder.

If the GE-decoder fail to correct all the CPM-phased bursts of erasures in $\mathbf{u}^{(1)}$, the erasure correction is switched back to local phase. Let $\mathbf{u}^{(2)}$ be the decoded word at the output of GE-decoder. If each fail decoded local codeword in $\mathbf{u}^{(2)}$ contains no more than two CPM-phased bursts of erasures, then all the erasures in $\mathbf{u}^{(2)}$ will be recovered by the LE-decoder. Then, the erasure correction process stops. If there are still uncorrected CPM-phased bursts of erasures, the erasure correction is switched to the global phase. The local and global erasure corrections continue iteratively until either all the erasures in the received global codeword are recovered, or the remaining CPM-phased bursts of erasures in a decoded word are un-recoverable, or a preset maximum number of local/global iterations of erasure correction is reached.

With the above two-phase iterative erasure-correction process, the maximum number of recoverable erased symbols is $n^2 + (k-1)(2n-1)$ (a single local codeword is completely erased and two mutually semi-solid CPM-phased bursts of erasures occur in each of the other $k-1$ local codewords) which is equal to the number of parity-check symbols of the code. Hence, the code $\mathcal{C}_{GC,RS,ldpc}(2, n, k)$ is also optimal ($\eta = 1$).

Example 2: Let GF(2^5) be the field for code construction. Set $n = 2^5 - 1 = 31$ (a prime) and $k = 16$. First we construct a 2×31 $\mathbf{B}_{RS}(2, 31)$ over GF(2^5) in the form of (2). Next we take 16 copies of $\mathbf{B}_{RS}(2, 31)$ and form a 63×496 matrix $\mathbf{B}_{GC,RS}(2, 31, 16)$ in the form of (8). The 31×31 CPM-dispersion of $\mathbf{B}_{GC,RS}(2, 31, 16)$ gives a 1953×15376 globally coupled matrix $\mathbf{H}_{GC,RS}(2, 31, 16)$ of rank 1876.

The null space over GF(2) of $\mathbf{H}_{GC,RS}(2, 31, 16)$ gives a (15376, 13500) CN-QC-RS-QC-LDPC code with rate 0.878. The code can recover a single erased local codeword of 961 code symbols and correct up to two mutually semi-solid CPM-phased bursts of erasures in each of the other 15 local received codewords. The maximum number of recoverable erased symbols is 1876. The erasure-correction efficiency of the code is $\eta = 1$. The code is optimal.

V. Conclusion

In this paper, we presented two classes of *binary* QC-LDPC codes for correcting multiple random CPM-phased bursts of erasures over the BEC. The codes in both classes are optimal. The erasure correction of these codes is characterized by the *cycle and adjacency structure* of their Tanner graphs. The constructions of binary codes in these two classes can be generalized for constructing nonbinary codes over fields of characteristic 2 if we disperse each entry of the base matrix of a code into a nonbinary CPM [2].

An interesting question is that whether the codes presented in this paper can be applied to distributed storage systems.

Acknowledgment

This work was partially supported by the NSF grants ECCS-1500170 and SaTC-1813401, the AppoTech and NuFront gift grants.

References

[1] W. E. Ryan and S. Lin, *Channel Codes: Classical and Modern*. New York, NY: Cambridge Univ. Press, 2009.

[2] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inf. theory*, vo. 47, no. 2, pp.569-584, Feb. 2001.

[3] J. Ha and S. w. McLaughlin, "Low-density parity-check codes over Gaussian channels with erasures," *IEEE Trans. Inf. Theory*, vol.49, no. 7, pp.1801-1809, Jul. 2003.

[4] M. Yang and B. E. Ryan, "Performance of efficiently encodable low-density-parity codes in noise bursts on the RPR4 channel," *IEEE Trans. Magnetics*, vol. 40, no. 2, pp. 507-512, Mar. 2004.

[5] Y. Y Tai, L. Lan, L. -Q. Zeng, S. Lin, and K. Abdel-Ghaffar, "Algebraic construction of construction of quasi-cyclic codes for the AWGN and erasure channel," *IEEE Trans. Commun*, vol. 54, no. 10, pp. 1765-1774, Oct. 2006.

[6] G. Hosoya, H. Yagi, T. Matsushima, and S. Hirosawa, "A modification method for constructing low-density parity-check code burst erasures," *ICICE Trans. Fundamentals*, vol. E89-A, no.10. pp. 1501- 2509, Oct. 2006.

[7] L. Lan, L. -Q. Zeng, Y. Y. Tai, L. Chen, S. Lin, and K. Abdel-Ghaffar, "Construction of quasi-cyclic LDPC codes for AWGN and binary erasure channels: a finite field approach," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2429-2458, Jul. 2007.

[8] S. J. Johnson, "Burst erasure correcting LDPC codes," *IEEE Trans. Commun.*, vol. 57, n0. 3, pp. 641-652, Mar. 2009.

[9] Q. Diao, Q. Huang, S. Lin, and K. Abdel-Ghaffar, "A matrix-theoretic approach for analyzing quasi-cyclic low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 4030 - 4048, Jun. 2012.

[10] S. H. Reiger, "Codes for the correction of clustered errors", *IRE Trans. Inf. Theory*, vol. IT-6, pp. 16-21, 1960.

[11] J. Li, S. Lin, K. Abdel-Ghaffar, W. E. Ryan, and D. J. Costello, Jr., "Globally coupled LDPC Codes," *Proc. Inform. Theory and Applications (ITA)*, San Diego, CA., Jan. 31 ? Feb. 5, 2016.

[12] J. Li, S. Lin, K. Abdel-Ghaffar, W.E. Ryan, and D. J. Costello, Jr., "LDPC Code Designs, Constructions and Unification", Cambridge University Press, Cambridge, UK , 2017.

[13] J. Li, K. Liu, S. Lin, and K. Abdel-Ghaffar, "Reed-Solomon based nonbinary LDPC code," *Proc. Int. Symp. Inf. Theory and Its Applic. (ISITA)*, Monterey, CA, USA, Oct. 30 - Nov. 2, 2016.

[14] V. Savin, "LDPC decoder," in *Channel Coding: Theory, Algorithms, and Applications*, D. Declercq, M. Fossorier, and E. Biglieri Eds., Oxford, UC: Academic Press, 2014.