Coset graphs for LDPC codes:

Performance on the binary erasure channel

Josef Lauri

Department of Mathematics

University of Malta

Malta

C.J. Tjhai

School of Computing and Mathematics

University of Plymouth

U.K.

July 10, 2010

Abstract

We show that a popular way of constructing quasi-cyclic LDPC codes is a special case of a construction which is common in graph theory and group theory. It is shown that a generalisation of this construction as coset graphs produces $(d_v, d_c)-$regular LDPC codes which have an advantage in terms of the minimum stopping set size compared to quasi-cyclic LDPC codes. A $(d_v, d_c)-$regular quasi-cyclic LDPC code cannot have minimum stopping set size larger than $(d_v + 1)!$. However, by using coset graphs, a $(3, 5)-$regular LDPC code with minimum stopping set size of 28 and a $(3, 4)-$regular LDPC code with minimum stopping set size larger than 32 have been obtained. In addition, the idea of coset graphs also provides a compact algebraic way

of describing bipartite graph and the associated parity-check matrix of an LDPC code. Simulation results of iterative decoding of the coset graphs LDPC codes over the binary erasure channel show that some of the codes converge well and based on the truncated stopping set distributions of the codes, which are exhaustively and efficiently enumerated, the error-floor of the codes at low probability of erasure is estimated.

1 Introduction

The ultimate aim of channel coding is to construct an error-correcting code with a practical decoder that can reach the limit set by Shannon in 1948 [1]. Until the early 1990s, however, there had been no practical scheme that was deemed satisfactory in meeting Shannon's challenge. A major breakthrough was achieved in 1993 with the invention of parallel concatenated convolutional codes or turbo codes and their practical decoder–iterative decoder, by Berrou *et. al* [2]. The success of iterative decoding has brought Gallager's invention–low-density parity-check (LDPC) codes [3], back to life. These codes, which were originally invented in 1962, were rediscovered by MacKay and Neal [4] and they showed that in addition to turbo codes, LDPC codes could also approach Shannon's limit. Due to their capacity-approaching performance and their simple iterative decoder, LDPC codes have attracted a great deal of attention in the recent years–both in research community and industry.

Over the years, numerous techniques to construct LDPC codes have been proposed. These techniques can be classified into two categories: random and algebraic constructions. Random LDPC codes are constructed by placing non-zeros randomly based on a set of constraints in their parity-

check matrix, for example see [5, 6, 7]. On the other hand, for algebraic LDPC codes, the placement of non-zeros follows a certain algebraic structure. Some notable work in algebraic construction of LDPC codes are one-step majority-logic and difference-set cyclic LDPC codes [8], finite geometry LDPC codes [9] and combinatorial-based LDPC codes [10, 11]. Another class of algebraic LDPC codes are graph-theoretic codes, for example see [12, 13, 14].

In this work, we show that quasi-cyclic (QC) LDPC codes introduced by Tanner [15] are a special case of a construction which is common in graph theory and group theory. It is also shown that a generalisation of these QC LDPC codes as coset graphs may be used to construct regular LDPC codes whose upper-bound of girth is higher than that of QC LDPC codes. The rest of the paper is organised as follows. A background to LDPC codes and coset graph is presented in Section 2. Section 3 discusses the importance of stopping set of an LDPC code and its relation to the girth of the graph defined by the code. The construction of coset graph LDPC codes is presented in Section 4 and the simulation results of the constructed codes are presented in Section 5. A few concluding remarks and a discussion of future work are given in Section 6.

2 Background

A binary linear code of length n, dimension k and minimum Hamming distance d—commonly denoted as $[n, k, d]$ code, is a linear k-dimensional subspace of all binary vectors of length n $\{0, 1\}^n$. An $[n, k, d]$ code may be defined by its $m \times n$ parity-check matrix H, where $m \geq \text{Rank}(H) = n - k$. If A_w denotes the number of codewords of Hamming weight w, the Hamming

weight enumerator polynomial of the code is given by

$$A(z) = \sum_{w=0}^{n} A_w z^w,$$

where z is an indeterminate. The distribution of A_w for $0 \leq w \leq n$ is known as the Hamming weight distribution of a code.

LDPC codes form a class of linear codes whose H is sparse and is commonly associated with a bipartite or Tanner graph with left and right vertices representing the columns and rows of H respectively. Given an LDPC code, if the weight of all columns of H or the degree of all left vertices of its Tanner graph is d_v and the weight of all rows of H or the degree of the right vertices of its Tanner graph is d_c, such code is called a $(d_v, d_c)-$regular LDPC code, otherwise it is called an irregular LDPC code. For a general reference to LDPC codes, we refer the reader to [16].

In graph theory, the following construction is useful for creating edge-transitive graphs (any graph theoretic term which we use and do not define can be found in [17, 18]). Let Γ be a group and let $\mathcal{H}, \mathcal{K}$ be two subgroups of Γ. Then the graph $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$ is defined as follows: its vertices are the right cosets of $\mathcal{H}$ and $\mathcal{K}$, and two cosets $\mathcal{H}x, \mathcal{K}x$ are adjacent if and only if their intersection is non-empty. The importance of this construction stems from the following.

Theorem 2.1 *Let Γ be a finite group and $\mathcal{H}, \mathcal{K}$ subgroups of Γ whose union generates the group. Then the graph Cos$(\Gamma, \mathcal{H}, \mathcal{K})$ is a connected edge-transitive bipartite graph with vertex degrees $|\mathcal{H}|/|\mathcal{H}\cap\mathcal{K}|$ and $|\mathcal{K}|/|\mathcal{H}\cap\mathcal{K}|$ and with the two sets of cosets of $\mathcal{H}$ and $\mathcal{K}$ being the bipartition of Cos$(\Gamma, \mathcal{H}, \mathcal{K})$.*

Conversely, let G be a graph on which the group Γ acts edge-transitively but not vertex-transitively. Then G is bipartite and the vertices of G fall

into two orbits under the action of Γ. *Moreover, if* uv *is an edge of* G *and* $\mathcal{H}$ *and* $\mathcal{K}$ *are the stabilisers in* Γ *of* u *and* v, *respectively, then the union of* $\mathcal{H}$ *and* $\mathcal{K}$ *generates* Γ *and* G *is isomorphic to* $Cos(\Gamma, \mathcal{H}, \mathcal{K})$.

Checking that a coset graph does not have cycles of length 4 is easy: the condition is that there are no $h_1, h_2 \in \mathcal{H}$ and $k_1, k_2 \in \mathcal{K}$ such that $h_1 k_1 = k_2 h_2$. In [19] coset graphs were used to give a graph of order 465 whose automorphism group acted regularly on its edges and whose girth was 8. The group Γ was

$$\langle a, b, c | a^5 = b^3 = c^{31} = 1, ba = abc, ca = ac^2, cb = bc^{25} \rangle$$

with $\mathcal{H} = \langle a \rangle$ and $\mathcal{K} = \langle b \rangle$. This group is a special case of the group which we shall denote by $\Gamma(p, q, r)$ where p, q, r are primes with $r = 1 \bmod pq$ and such that

$$\Gamma(p, q, r) = \langle a, b, c | a^p = b^q = c^r = 1, ba = abc, ca = ac^s, cb = bc^t \rangle$$

and where s^p and t^q are equal to 1 mod r. Therefore the graph studied in [19] is a coset graph of $\Gamma(3, 5, 31)$.

In [15], a class of QC LDPC codes were presented. The first code in this sequence is precisely the one whose Tanner graph is the same bipartite graph as studied in [19] for the group $\Gamma(3, 5, 31)$. The paper [15] gives some results and also simulations involving the groups $\Gamma(p, q, r)$ for $p = 3$ and $q = 5$ and $r = 61, 151$, etc up to $r = 1291$. One characteristic of these graphs is that the girth of their Tanner graph can never exceed 12. In [15] the check matrices from these graphs are described as p rows of q $r \times r$ permutation matrices, with each permutation matrix being the identity matrix shifted cyclically by an amount depending on its position in the array of matrices.

With our description of the Tanner graph as a coset graph we can also easily see that the girth cannot exceed 12, as follows.

First of all what does, in general, a path in the coset graph $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$ mean algebraically? Suppose $\mathcal{H}a$ and $\mathcal{K}b$ are adjacent. Then $h_1 a = k_1 b$ for some $h_1 \in \mathcal{H}$ and $k_1 \in \mathcal{K}$. Therefore $\mathcal{K}b = \mathcal{K}h_1 a$. Similarly, if now $\mathcal{K}h_1 a$ is adjacent to $\mathcal{H}c$, then $\mathcal{H}c = \mathcal{H}k_2 k_1 a$. Therefore, to obtain a cycle of length $2l$ we need that

$$\mathcal{H}k_1 h_2 \dots k_{2l-1} = \mathcal{H}.$$

That is, the girth g of $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$ can be defined as the length of a shortest reduced word in Γ

$$k_1 h_2 k_3 \dots k_{g-1} h_g$$

which is equal to 1, where, by a *reduced word* we mean an alternating product of elements of $\mathcal{H} - \mathcal{H} \cap \mathcal{H}$ and $\mathcal{K} - \mathcal{H} \cap \mathcal{K}$ and its length is the number of such elements.

In the case of the group $\Gamma(p, q, r)$, as sometimes happens in constructions of families of graphs with large girth [20], there is a "universal word" which stops the girth from growing beyond a certain bound for all graphs in the family. In this case we find that

$$ab^{-1}a^{-1}b^{-1}ab^2a^{-1}b^{-1}ab^{-1}a^{-1}$$

is always some power x of b, where x depends on the choice of p, q and r. Therefore

$$ab^{-1}a^{-1}b^{-1}ab^2a^{-1}b^{-1}ab^{-1}a^{-1}b^{-x}$$

is always equal to 1 and therefore the girth cannot exceed 12. The way we have described this word it appears like a rabbit out of a hat? But one

can see how this word crops up by considering the structure of part of the Tanner graph and its check matrix (this analysis is similar to that in [15] but using cosets instead of cyclically shifted identity graphs). Consider in order these three groups of vertices

$$\mathcal{H}, \mathcal{H}c, \ldots, c^{r-1}, \mathcal{H}b, \mathcal{H}bc, \ldots, \mathcal{H}bc^{r-1}, \mathcal{H}b^2, \mathcal{H}b^2c, \ldots, \mathcal{H}b^2c^{r-1}$$

calling them Groups 1, 2 and 3, and, from the other bipartition of $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$, these two groups of vertices

$$\mathcal{K}, \mathcal{K}c, \ldots, \mathcal{K}c^{r-1}, \mathcal{K}a, \mathcal{K}ac, \ldots, \mathcal{K}ac^{r-1}$$

calling them Groups A and B. The important point to observe is this: each vertex within one of the Groups 1, 2 and 3 is adjacent to exactly one vertex from each one of the Groups A and B, and vice-versa.

Now starting with the edge joining $\mathcal{K}$ to $\mathcal{H}$ (they are certainly not disjoint) move to the adjacent vertex in Group B, then to Group 2, Group A, Group 3 and Group B. Now, starting from the other end, $\mathcal{K}$ is adjacent to $\mathcal{H}b^2$ which is in Group 3. Then move to Group B, Group 2, Group A, Group 1 and Group B. By our observation, we must have traversed a cycle, and it is now not difficult to see that this cycle is represented by the above word of length 12.

A similar type of analysis done by Fossorier in [21] (but without the benefit of working with reduced words in a group) shows that some other different ways of defining check matrices by cyclically shifting the identity matrix also gives an upper bound of 12 for the girth.

3 Stopping sets and girth of Tanner graphs

We shall be presenting the results of simulating the behaviour of some LDPC codes on the binary erasure channel (BEC). For such a channel, the performance of an LDPC code under iterative decoding is dominated by its stopping set distribution. A *stopping set* of an LDPC code, which was introduced by Di *et al.* [22], is defined as follows. Let S be a set of locations in the block code, and let T be the set of check equations which involve at least one of the locations in S. Then S is said to be a stopping set if each of the locations it contains is involved in at least two of the check equations in T. In terms of the Tanner graph G of the code a stopping set can therefore be defined as follows. Let the bipartition of of the Tanner graph be $X \cup Y$ with the vertices in X representing the locations in the codewords and Y representing the check equations. Let $S \subseteq X$ and let $T \subseteq Y$ be the set of neighbours of the vertices in S. Then S is said to be a stopping set if every vertex in S has at least two neighbours in T. Clearly, the size of a minimum stopping set of a code is at equal to at most the code's minimum weight.

Like the Hamming weight distribution, we can write the stopping set distribution of a code in terms of the stopping set enumerator polynomial, that is,

$$S(z) = \sum_{w=0}^{n} S_w z^w,$$

where as before z is an indeterminate and S_w is the number of stopping sets of weight w. The importance of S_i will be become evident in the results section where it is used to accurately predict the erasure performance of an LDPC at very low error rate regions.

Stopping sets are important because most iterative methods used for

decoding cannot decode a set of erasures whose locations form a stopping set. Therefore it is desirable that the size of a stopping set of a code is large. For this reason it is also desirable that the Tanner graph of the code does not have small girth. Let $L_s(d_v, g)$ be the lower-bound of the minimum stopping set size of a $(d_v, d_c)-$ regular LDPC code of girth g. As was stated in [23], for $(d_v, d_c)-$regular LDPC codes in which $d_v \geq 2$, the girth is $g \geq 6$, we have $L_s(2, g) = g/2$, $L_s(d_v, 6) = d_v + 1$, $L_s(d_v, 8) = 2d_v$, and, for larger g,

$$L_s(d_v, g) \geq \begin{cases} 1 + \sum_{i=0}^{\frac{g-6}{4}} d_v(d_v-1)^i & \frac{g}{2} \text{ odd}, \\ 1 + \sum_{i=0}^{\frac{g-8}{4}} d_v(d_v-1)^i + (d_v-1)^{\frac{g-4}{4}} & \frac{g}{2} \text{ even}. \end{cases}$$

Therefore, for $d_v > 2$, $L_s(d_v, g)$ increases exponentially with g.

4 Codes from the coset graphs

We shall study coset graphs as a generalisation of the QC LDPC codes studied in [15, 21]. In this section we are mainly concerned with the girth of $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$, where the union of the subgroups generates Γ, since, as we have seen, $\mathrm{Cos}(\Gamma, \mathcal{H}, \mathcal{K})$ is the Tanner graph of the corresponding LDPC code. The stopping set sizes are considered in a later section.

First we note that, in theory, there is no upper bound on the girth of such a triple. For example, let us restrict ourselves to permutations a and b of order 3 and let $\mathcal{H} = \langle a \rangle$, $\mathcal{K} = \langle b \rangle$ and $\Gamma = \langle a, b \rangle$. Consider, for example,

$$a = (1\ 2\ 3)(4\ 5\ 6)(7\ 8\ 9)$$

and

$$b = (3\ 4\ 5)(6\ 7\ 8).$$

If we follow the trajectory of the element 2 when acted upon by a permutation made up of a reduced word in a and b where the first element is a (if the first element is a^2 then consider the element 1), then we see that the shortest length of such a reduced which which maps 2 into itself is 10. But for the word to be equal to 1 all elements $1, 2 \ldots 8$ must be mapped back into themselves. Using the free software GAP [24] and its package GRAPE [25], we find that the girth of this coset graph is, in fact, 14. It is easy to see how this girth can be made larger by taking more triples in the definition of a and b, but this crude way will give a group Γ of astronomical size. So let us give here a few more reasonable examples, worked out using GAP. Unless otherwise stated, in all examples $\mathcal{H} = \langle a \rangle, \mathcal{K} = \langle b \rangle$ and $\Gamma = \langle a, b \rangle$, where a and b are the given permutations.

Example 4.1 *Let* $a = (1\ 2\ 3)$ *and* $b = (2\ 4\ 5)(3\ 6\ 7)$. *The graph* $Cos(\Gamma, \mathcal{H}, \mathcal{K})$ *also has girth* 14. *The check matrix corresponding to* $Cos(\Gamma, \mathcal{H}, \mathcal{K})$ *has size* 840×840. *Its row-rank turns out to be* 750, *therefore the resulting code corresponding to this Tanner graph is a* $(3, 3)-$*regular LDPC code with a low code-rate of* 0.10714. *(In all these examples, the computation of the check matrices from the corresponding Tanner graphs and also their ranks was carried out using GAP.)*

Example 4.2 *In order to obtain higher rate* $(d_v, d_c)-$*regular LDPC codes, the condition* $d_c > d_v$ *has to be met. Let* $a = (1\ 2\ 3\ 4)$ *and* $b = (2\ 4\ 5)(3\ 7\ 8)$. *This gives girth* 16 *and a* 45360×60480 *matrix which defines a* $(3, 4)-$*regular LDPC code of code-rate at least* 0.25.

Here are two seemingly similar small examples.

Example 4.3 *Let* $a = (1\ 3\ 5)(2\ 4\ 6)$ *and* $b = (1\ 2)(3\ 4)(5\ 6\ 7)$. *The Tanner graph has girth* 8 *with* $d_v = 3$ *and* $d_c = 6$. *The parity-check matrix has size* 420×840 *and rank* 385, *therefore the code-rate is* 0.542.

Example 4.4 *Let* $a = (1\ 2\ 3\ 4\ 5)$ *and* $b = (2\ 6\ 7)(1\ 4\ 3)$. *The Tanner graph has girth* 8 *with* $d_v = 3$ *and* $d_c = 5$. *The size of the check matrix is* 504×840 *with rank* 498, *therefore the code-rate is* 0.407.

A comparison between the last two codes is interesting. They are both of length 840 and have girth 8, and the first one has slightly better code-rate. But from our analysis we found that the first code has a minimum stopping set size of 8. In fact, there are 1365 stopping sets of weight 8 and 8064 stopping sets of weight 10, and these stopping sets are also the codewords of the code. On the other hand, the second code has a minimum stopping set size of 28. There are 120 stopping sets of that weight and they are all codewords of the code. We shall discuss these issues further in the next section.

Example 4.5 *Here we shall work backwards from a given edge-transitive graph of known girth to its representation as a coset graph. Let* T *be the graph shown in Figure 1. This is the smallest cubic graph with girth* 8. *Subdivide each edge (that is, replace each edge* $\{a, b\}$ *with two edges* $\{a, x\}$ *and* $\{x, b\}$ *where* x *is a new vertex of degree* 2*). Call the resulting graph* G. *Since* T *is arc-transitive,* G *is edge-transitive and its girth is* 16. *The order of the automorphism group of* G *is* 1440 *and the orders of the stabilisers* $\mathcal{H}$ *and* $\mathcal{K}$ *of two adjacent vertices are, respectively,* 48 *and* 32. *The order of* $\mathcal{H} \cap \mathcal{K}$ *is* 16, *confirming that the degrees of the vertices are* $48/16 = 3$ *and* $32/16 = 2$.

However, when a subgroup $\mathcal{H}'$ *of order* 3 *of* $\mathcal{H}$ *and a subgroup* $\mathcal{K}'$ *of order*

2 *of* $\mathcal{K}$ *were taken, and* Γ' *was set to be the group generated by* $\mathcal{H}' \cup \mathcal{K}'$*, the girth of* $(\Gamma', \mathcal{H}', \mathcal{K}')$ *was found to be* 4*. Therefore there is a reduced word* w *in the elements of* $\mathcal{H}'$ *and* $\mathcal{K}'$ *of length* 4*. But since* $\mathcal{H}'$ *and* $\mathcal{K}'$ *are subgroups of* $\mathcal{H}$ *and* $\mathcal{K}$*, respectively, why does not* w *give girth* 4 *in* $(\Gamma, \mathcal{H}, \mathcal{K})$*? The reason is obviously that some or all of the elements making up* w *are in* $\mathcal{H} \cap \mathcal{K}$ *so it is not a reduced word in the elements of* $\mathcal{H} - \mathcal{H} \cap \mathcal{K}$ *and* $\mathcal{K} - \mathcal{H} \cap \mathcal{K}$*. This example shows that one possible way of getting rid of short cycles could be by increasing the intersection of* $\mathcal{H}$ *and* $\mathcal{K}$ *so that the short reduced words contain elements in* $\mathcal{H} \cap \mathcal{K}$*.*

Figure 1 to come around here

5 Constructed Codes and Their Simulation Results

In this section, some results of the erasure performance over the binary erasure channel (BEC) of the binary $(d_v, d_c) - regular$ LDPC codes constructed with various coset graph parameters are presented. In one of his classic papers, Berlekamp [26] presents the error probability of a maximum-likelihood erasure decoder for a given code over the BEC. An $[n, k, d]$ linear code is guaranteed to correct $d - 1$ erasures and the error probability of a maximum-likelihood decoder, assuming that the probability of erasure is p,

is given by [26]

$$P_e \leq \underbrace{\sum_{w=n-k+1}^{n} \binom{n}{e} p^e (1-p)^{n-e}}_{\text{MDS performance}} \underbrace{\sum_{w=d}^{n-k} A_w p^w \sum_{e=w}^{n-k} \binom{n-w}{e-w} p^{e-w} (1-p)^{n-e}}_{\text{MDS shortfall}} \tag{1}$$

$$\approx \sum_{w \geq d} A_w p^w \qquad \text{for low values of } p \tag{2}$$

where A_w is the number of codewords of weight w and the second term of (1) only exists for non MDS codes. The iterative decoding of LDPC codes is sub optimal and its error performance is dictated by the stopping set rather than the weight distribution of LDPC codes. Equations (1) and (2) may be adapted by using the stopping set distribution of an LDPC code. In particular, the equation

$$P_e \approx \sum_{w \geq s_{min}} S_w p^w, \tag{3}$$

which is adapted from (2), is useful to predict the error floor of an LDPC code over the BEC. Here, s_{min} is the minimum stopping set weight of the LDPC code. The stopping set distribution of an LDPC code can be exhaustively enumerated in an efficient manner using the algorithm given by [27] and [28].

We first consider three small codes, namely Small-Cage $[9, 4, 4]$, Gen-Quad $[16, 9, 4]$ and Gray $[27, 8, 8]$. Table 1 shows the generators of these codes—the parameters a and b, the girth, the minimum stopping set size s_{min} and the lower-bound of s_{min} obtained from the formula presented in Section 3. Figure 2 shows the probability of frame error of these small LDPC

Table 1: Parameters of small LDPC codes

Name	a	b	Girth	s_{min}	$L_(s,g)$
Small-Cage $[9,4,4]$	(1 2 3)	(1 4)(2 5)(3 6)	8	4	4
Gen-Quad $[16,9,4]$	(1 2 3 4)	(1 5)(2 6) (3 7)(4 8)	8	4	4
Gray $[27,8,8]$	(1 2 3)	(1 4 7)(2 5 8) (3 6 9)	8	8	6

codes over the BEC. Approximations at lower probability of erasure obtained using (3) are also shown in the figure. The stopping set distributions are obtained using the algorithm presented in [28].

Figure 2 to come around here

The Tanner graph of Small-Cage has order 15 with $d_v = 2$ and $d_c = 3$. This is the smallest number of vertices which a bipartite graph with these degrees for the bipartition can have. It is therefore an example of what are sometimes called bi-regular cages [29]. The stopping set enumerator polynomial of Small-Cage is given by

$$S(z) = 1 + 9z^4 + 6z^5 + 18z^6 + 21z^7 + 9z^8 + z^9,$$

which contains an all-zero codeword, 9 weight codewords of weight 4 and 6 codewords of weight 6.

The Tanner graph of Gen-Quad has $d_v = 2$, $d_c = 4$, girth of 8, and diameter of 4. It is therefore a generalised quadrangle, albeit a "thin" one because of $d_v = 2$ [30]. The stopping set enumerator polynomial of Gen-

Quad is given by

$$S(z) = 1 + 36z^4 + 144z^6 + 288z^7 + 678z^8 + 1600z^9 + 2472z^{10} + \\ 2400z^{11} + 1436z^{12} + 528z^{13} + 120z^{14} + 16z^{15} + z^{16}.$$

Finally, the third Tanner graph is the well-known Gray graph which is the smallest known cubic graph (on 54 vertices) which is semisymmetric [31], that is, a regular, edge-transitive but not vertex-transitive graph. Semisymmetric graphs are not easy to find and the first examples were constructed by Folkman in [32]. The stopping set enumerator polynomial of this code is given by

$$S(z) = 1 + 27z^8 + 81z^{12} + 270z^{14} + 108z^{15} + 135z^{16} + \\ 864z^{17} + 1281z^{18} + 1620z^{19} + 2889z^{20} + 4410z^{21} + \\ 5049z^{22} + 3402z^{23} + 1305z^{24} + 270z^{25} + 27z^{26} + z^{27}.$$

Figure 3 to come around here

Figure 3 shows the performance of three $(3, 2)-$regular coset graph LDPC codes. The parameters and the truncated stopping set distributions of the codes in Figure 3 are shown in Table 2. The performance curves of these codes follow similar trend and they are not as good as other coset graph LDPC codes which have d_c of 4 and 5. The truncated stopping set distribution of these codes, indicated in the second last column of Table 2 is used to derive an approximation to the probability of decoder failure at low probability of erasure, see (3).

Figure 4 to come around here

Table 2: Parameters of $(3,2)-$regular LDPC codes in Figure 3

Name	a	b	Girth	s_{min}	S_w	$L_s(3,g)$
$[36,13,6]$	(1 2 3) (4 5 6)	(1 7)	12	6	$S_6=12$ $S_8=54$ $S_{10}=108$ $S_{11}=108$ $S_{12}=444$	3
$[90,31,9]$	(1 2 3) (4 5 6)	(1 7) (2 8)	18	9	$S_9=60$ $S_{10}=108$ $S_{11}=108$ $S_{12}=210$ $S_{13}=180$	9
$[720,241,12]$	(1 2 3) (4 5 6)	(1 7) (2 8) (4 9)	24	12	$S_{12}=720$ $S_{14}=1440$ $S_{16}=3600$ $S_{18}=13440$	12

Table 3: Parameters of $(3,4)-$regular LDPC codes in Figure 4

Name	a	b	Girth	s_{min}	S_w	$L_s(3,g)$
$[40,15,8]$	(1 2 3 4 5) (8 9 10) (11 12 13)	(1 2 3 4 5) (8 9 10 11)	8	8	$S_8=45$ $S_{12}=1200$ $S_{13}=1440$ $S_{14}=4260$ $S_{15}=7824$	6
$[720,270,8]$	(1 2 3)(4 5 6)	(3 7 8 9) (4 10)	8	8	$S_8=270$ $S_{12}=2160$ $S_{14}=2160$ $S_{15}=2160$ $S_{16}=50355$	6
$[360,117,24]$	(1 2 3)(4 5 6)	(5 6 7 8)(3 9) (2 10)(1 11)	12	24	$S_{24}=1755$	14
$[720,216,24]$	(1 2 3)(4 5 6)	(3 7 8 9)	12	24	$S_{24}=450$	14
$[2520,701,24]$	(1 2 3)(4 5 6)	(1 7) (2 8 9 10)	12	24	$S_{24}=105$	14
$[2160,591,>32]$	(1 2 3)(4 5 6)	(5 6 7 8)(3 9) (2 10)(1 11) (4 12)	12	≤ 48		14

Table 4: Parameters of $(3,5)-$regular LDPC codes in Figure 5

Name	a	b	Girth	s_{min}	S_w	$L_s(3,g)$
$[155,64,20]$	$\langle a,b,c : a^3=b^5=c^{31}$ $ba=abc,\ ca=ac^{25},$ $cb=bc^2\rangle$		8	18	$S_{18}=465$ $S_{19}=2015$ $S_{20}=9548$ $S_{21}=23715$ $S_{22}=106175$	6
$[755,334,14]$	$\langle a,b,c : a^3=b^5=c^{151}$ $ba=abc,\ ca=ac^{32},$ $cb=bc^8\rangle$		10	14	$S_{14}=755$ $S_{18}=755$ $S_{20}=3020$ $S_{22}=9815$ $S_{24}=30200$	10
$[840,342,28]$	(1 2 3 4 5)	(2 6 7) (1 4 3)	8	28	$S_{28}=120$	6
$[905,364,24]$	$\langle a,b,c : a^3=b^5=c^{181}$ $ba=abc,\ ca=ac^{48},$ $cb=bc^{42}\rangle$		12	24	$S_{24}=905$	14
$[6720,2695,24]$	(1 2 3) (4 5 6)	(3 4 5 7 8)	10	24	$S_{24}=1120$	10

Figure 4 shows the performance of several $(3,4)-$regular coset graph LDPC codes. The parameters and the truncated stopping set distribution of the codes in Figure 4 are shown in Table 3. From the figure, the codes of larger girth have higher s_{min}. Both $(3,4)-$regular coset graph LDPC codes shown in Figure 4 which have girth 8, have $s_{min}=8$. The rest of the codes in Table 3, except the one of length 2160, which have girth 12, have $s_{min}=24$. For the last code of length 2160 and dimension 591, whose Tanner graph has girth 12, exhaustive stopping set enumeration found that $32<s_{min}\leq 48$. As a consequence, unlike the other codes, an approximation to the FER performance of this code cannot be obtained yet.

Figure 5 to come around here

Figure 5 shows the performance of several $(3,5)-$regular coset graph LDPC codes whose parameters and truncated stopping set distributions are

Table 5: Parameters of $(3, 6)-$regular LDPC codes in Figure 6

Name	a	b	Girth	s_{min}	S_w	$L_s(3, g)$
$[216, 125, 8]$	(1 2 3) (4 5 7) (6 8 9)	(2 6) (3 5 8)	8	8	$S_8 = 3375$ $S_{12} = 121500$ $S_{14} = 1188000$	6
$[840, 455, 8]$	(1 3 5) (2 4 6)	(1 2) (3 4) (5 6 7)	8	8	$S_8 = 1365$ $S_{10} = 8064$ $S_{12} = 25620$	6

tabulated in Table 4. Several $(3, 5)-$regular Tanner LDPC codes are also simulated for comparison purposes. As mentioned earlier, the construction method presented in this paper is able to realise LDPC codes of the same parameters as those constructed by Tanner. From the figure, it is interesting to note that both the $[905, 364, 24]$ Tanner LDPC code and the $[6720, 2695, 24]$ coset graph LDPC code have similar error floor over the BEC. The later code has better convergence due to its larger block length. An interesting coset graph LDPC code to note is the $[840, 342, 28]$ code whose Tanner graph has girth 8. This code has a similar convergence behaviour as the $[905, 364, 24]$ Tanner LDPC code, but because of its higher s_{min}, the code has a much lower error floor.

Figure 6 to come around here

Figure 6 shows the performance of two $(3, 6)-$regular coset graph LDPC codes of girth 8: $[216, 125, 8]$ and $[840, 455, 8]$. Refer to Table 5 for the parameters and truncated stopping set distribution of these two codes. Despite having higher code-rate, the convergence of these codes are not as good as some of the $(3, 4)-$ and $(3, 5)-$regular coset graph LDPC codes already discussed.

6 Conclusion

We feel that this generalisation of earlier constructions of Tanner graphs as coset graphs is important for a number of reasons. First of all, it is conceptually a very simple idea, based only upon the notion of cosets of a group. It provides a compact algebraic way of describing the Tanner graph and the associated check matrix. Although here we have initially stressed the issue of girth, it seems that this generalisation may be useful to improve other parameters of QC LDPC codes, such as the minimum stopping set size. The encoding simplicity of quasi-cyclic LDPC codes is lost in coset graph codes, but these improvements might offset the disadvantage of losing the quasi-cyclic structure of the check matrix. In particular, a (d_v, d_c)-regular QC LDPC code has an upper-bound of s_{min} given by $(d_v+1)!$ [21] and this bound does not apply to coset graph LDPC codes. For example $(3, d_c)-$regular QC LDPC codes cannot have s_{min} larger than $(3+1)! = 24$, but we have shown two $(3, d_c)-$regular coset graph LDPC codes which have s_{min} or d larger than 24, i.e. $[840, 342, 28]$ $(3, 5)-$ and $[2160, 591, 32 \leq d \leq 48]$ $(3, 4)-$regular codes.

Secondly this generalisation brings the problem of constructing good Tanner graphs right within the heart of two well-studied areas in graph theory and group theory. The problem of finding regular or bi-regular cages [30, 29] is a well-known and well-studied problem in graph theory. In this context, finding a shortest code whose coset graph has given degrees and girth is equivalent to finding edge-transitive (possibly bi-regular) cages. Some of the examples described above are closely connected to cages.

In group theory, since the publication of the very important work by Goldschmidt [33] and Djoković and Miller (for example, [34, 35]), group amalgams have become very important not only in the study of infinite

groups but also in finite group theory (see, for example [36, 37]). Basically a (finite group) amalgam consists of two finite groups $\mathcal{H}$ and $\mathcal{K}$ and a subgroup $\mathcal{B}$ common to both, and such that $\phi_1\mathcal{B} \to \mathcal{H}$ and $\phi_2\mathcal{B} \to \mathcal{K}$ are injective homomorphisms (the amalgams arising in the work of Goldschmidt and Djoković and Miller are also simple, that is, there is no subgroup of $\mathcal{B}$ which is normal in both $\mathcal{H}$ and $\mathcal{K}$). Roughly speaking, ϕ_1 and ϕ_2 determine how $\mathcal{H}$ and $\mathcal{K}$ are amalgamated together along $\mathcal{B}$. If it is assumed that there are no relations between elements of $\mathcal{H}$ and $\mathcal{K}$ except for the elements in $\phi_1(\mathcal{B})$ and $\phi_2(\mathcal{B})$ then we get the free product with amalgamation $\Gamma' = \mathcal{H} *_{\mathcal{B}} \mathcal{K}$, which is an infinite group. The corresponding coset graph is an the infinite tree $T_{p,q}$ with degrees $p = |\mathcal{H}|/|\mathcal{B}|$ and $q = |\mathcal{K}|/|\mathcal{B}|$ [38]. A finite group generated by $\mathcal{H} \cup \mathcal{K}$ is said to be a *completion* of Γ'. One can think of a completion Γ' as being obtained by adding relations to $\mathcal{H} *_{\mathcal{B}} \mathcal{K}$ so that its coset graph is obtained as the corresponding quotient of $T_{p,q}$; more precisely, a completion is a pair of homomorphisms (ψ_1, ψ_2) from $\mathcal{H}$ and $\mathcal{K}$, respectively, to some group Γ such that $\phi_1\psi_1 = \phi_2\psi_2$. The free amalgamated product $\Gamma' = \mathcal{H} *_{\mathcal{B}} \mathcal{K}$ itself is often called the universal completion of the amalgam.

In Goldschmidt's work, the two subgroups making up the amalgam were the stabilisers of two adjacent vertices while Djoković and Miller used the stabiliser of a vertex and the stabiliser of an edge incident to the vertex. The remarkable achievement of these authors was to show that, for particular symmetric graphs, most notably cubic s-transitive graphs, the corresponding amalgam can only be one of exactly a finite number of well-described types (the type of an amalgam being the isomorphism types of $\mathcal{H}, \mathcal{K}$ and $\mathcal{B}$ and ϕ_1 and ϕ_2). The connection between finite group amalgams and the problem of finding minimal graphs with given degree and girth comes out very well in [39] and in [40]. In the latter paper, the smallest known cubic graphs of

girth 12 and of girth 20 are obtained as coset graphs of particular completions of amalgams first described by Goldscmidt. One wonders whether it is possible, using the sophisticated group theoretic techniques developed by these authors, to determine what types of amalgams are determined by imposing certain pre-conditions on the coset graphs required to get an LDPC code with good parameters, such as the appropriate degrees, girth, minimum weight and minimum stopping set size.

Acknowledgements

The first author would like to thank the GAP-forum community for help in the use of GAP, especially to David Hobby for suggesting Example 4.5 and to Leonard Soicher for help with the programmes used to construct the coset graphs using GRAPE, and also to Janice Rapa and Victor Buttigieg for bringing [15] to his notice. The second author would like to thank Martin Tomlinson and Marcel Ambroze of the University of Plymouth for discussions on stopping set enumeration algorithm.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding: Turbo codes," in *Proc. IEEE International Conference on Communications*, (Geneva, Switzerland), pp. 1064–1070, 23–26 May 1993.

[3] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 21–28, Jan. 1962.

[4] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low-density parity-check codes," *Electron. Lett.*, vol. 32, no. 18, pp. 1645–1646, 1996.

[5] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, pp. 399–431, Mar. 1999.

[6] Y. Mao and A. Banihashemi, "A heuristic search for good low-density parity-check codes at short block lengths," in *Proc. IEEE Int. Conf. Communications (ICC)*, (Helsinki, Finland), pp. 41–44, Jun. 2001.

[7] X. Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and Irregular Progressive Edge-Growth Tanner Graphs," *IEEE Trans. Inf. Theory*, vol. 51, pp. 386–398, Jan. 2005.

[8] R. Lucas, M. P. C. Fossorier, Y. Kou, and S. Lin, "Iterative decoding of one-step majority logic decodable codes based on belief propagation," *IEEE Trans. Commun.*, vol. 46, pp. 931–937, June 2000.

[9] Y. Kou, S. Lin, and M. Fossorier, "Low density parity check codes based on finite geometries: A rediscovery and new results," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2711–2736, Nov. 2001.

[10] S. J. Johnson and S. R. Weller, "Construction of Low-Density Parity-Check Codes from Kirkman Triple Systems," *Proc. IEEE Inform. Theory Workshop,* Cairns, Australia, 2-7 Sept., pp. 90–92, 2001.

[11] B. Vasic and M. Milenkovic, "Combinatorial Constructions of Low-Density Parity-Check Codes for Iterative Decoding," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1156–1176, June 2004.

[12] J. Rosenthal and P. Vontobel, "Construction of LDPC codes based on Ramanujan graphs and ideas from Margulis," in *Proc. 38th Annu. Allerton Conf. Communications, Control, and Computing*, (Monticello, IL, USA), pp. 248–257, Oct. 2000.

[13] P. Vontobel and R. Tanner, "Construction of codes based on finite generalized quadrangles for iterative decoding," in *Proc. IEEE International Symposium on Information Theory*, (Washington, USA), p. 223, 24 Jun.–29 Jun. 2001.

[14] Z. Liu and D. Pados, "LDPC codes from generalized polygons," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3890–3898, Nov. 2005.

[15] R. Tanner, D. Sridhara, and T.E.Fuja, "A class of group-structured LDPC codes," in *Proceedings of the International Symposium on Communication Theory and Applications*, (Ambleside, UK), pp. 365–370, 2001.

[16] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*. Cambridge University Press, 2009.

[17] A. Bondy and U. Murty, *Graph Theory (Graduate Texts in Mathematics)*. Springer, 2008.

[18] J. Lauri and R. Scapellato, *Topics in Graph Automorphisms and Reconstruction*. Cambridge University Press, 2003.

[19] J. Lauri, "Constructing graphs with several pseudosimilar vertices or edges," *Discrete Maths*, to appear.

[20] N. Biggs, "Constructions for cubic graphs with large girth," *The Electronic Journal of Combinatorics*, vol. 5, 1998.

[21] M. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1788–1793, August 2004.

[22] C. Di, D. Proietti, I. Telatar, T. Richardson, and R. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1570–1579, Jun. 2002.

[23] A. Orlitsky, R. Urbanke, K. Viswanathan, and J. Zhang, "Stopping sets and the girth of tanner graphs," in *Proc. IEEE International Symposium in Information Theory*, (Lausanne, Switzerland), p. 2, June 30 – July 5 2002.

[24] The GAP Group, Aachen, St Adrews, *GAP — Groups, Algorithms, and Programming Version*, version 4.4 ed.

[25] L. H. Soicher, "GRAPE: a system for computing with graphs and groups," in *Groups and Computation* (L. Finkelstein and W. Kantor, eds.), vol. 11 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science, pp. 287–291, American Mathematical Society, 1993.

[26] E. R. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE*, vol. 68, pp. 564–593, 1980.

[27] E. Rosnes and Ø. Ytrehus, "An efficient algorithm to find all small-size stopping sets of low-density parity-check matrices," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4167–4178, Sep. 2009.

[28] M. Tomlinson, M. Ambroze, and L. Yang, "Exhaustive weight spectrum analysis of LDPC codes," in *Proc. 10^{th} International Symposium Com-*

munication Theory and Applications, (St. Martin College, Ambleside, UK), H. W. Comms. Ltd., 13–17 Jul. 2009.

[29] G. Araujo-Pardo, C. Balbuena, P. García-Vázquez, X. Marcote, and J. Valenzuela, "On the order of $(\{r, m\}; g)$-cages of even girth," *Discrete Math.*, 2007.

[30] C. Godsil and G. Royle, *Algebraic Graph Theory*. Springer-Verlag, 2001.

[31] D. M. T. Pisanski, "The Gray graph revisited," *J. Graph Theory*, vol. 35, pp. 1–7, 2000.

[32] J. Folkman, "Regular line-symmetric graphs," *J. Combin. Theory*, vol. 3, pp. 215–232, 1967.

[33] D. Goldschmidt, "Automorphisms of trivalent graphs," *Ann. Math.*, vol. 111, pp. 377–406, 1980.

[34] D. Djoković, "Automorphisms of regular graphs and finite simple group-amalgams," in *Algebraic Methods in Graph Theory* (L. Lovász and V. Sós, eds.), vol. 25 of *Colloquia Mathematica Societatis János Bolyai*, (Szeged, Hungary), pp. 95–118, 1978.

[35] D. Djoković and G. Miller, "Regular groups of automorphisms of cubic graphs," *J. Combin. Theory (Ser. B)*, vol. 29, pp. 195–230, 1980.

[36] A. Delgado, D. Goldschmidt, and B. Stellmacher, *Groups and Graphs: New Results and Methods*. Birkhäuser-Verlag, 1985.

[37] H. Kurzweil and B. Stellmacher, *The Theory of Finite Groups : An Introduction*. Springer-Verlag, 2004.

[38] J. Serre, *Trees (Corrected Second Printing)*. Springer-Verlag, 2003.

[39] N. Biggs, "Graphs with large girth," *Ars Combin.*, vol. 25C, pp. 73–80, 1988.

[40] C. Parker and P. Rowley, "Completions of Goldschmidt amalgams of type G_4 in dimension 3," *J. Algebraic Combin.*, vol. 13, pp. 77–82, 2001.

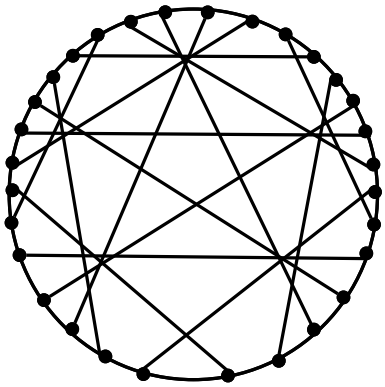

Figure 1: Tutte's cage which is the smallest cubic graph of girth 8.

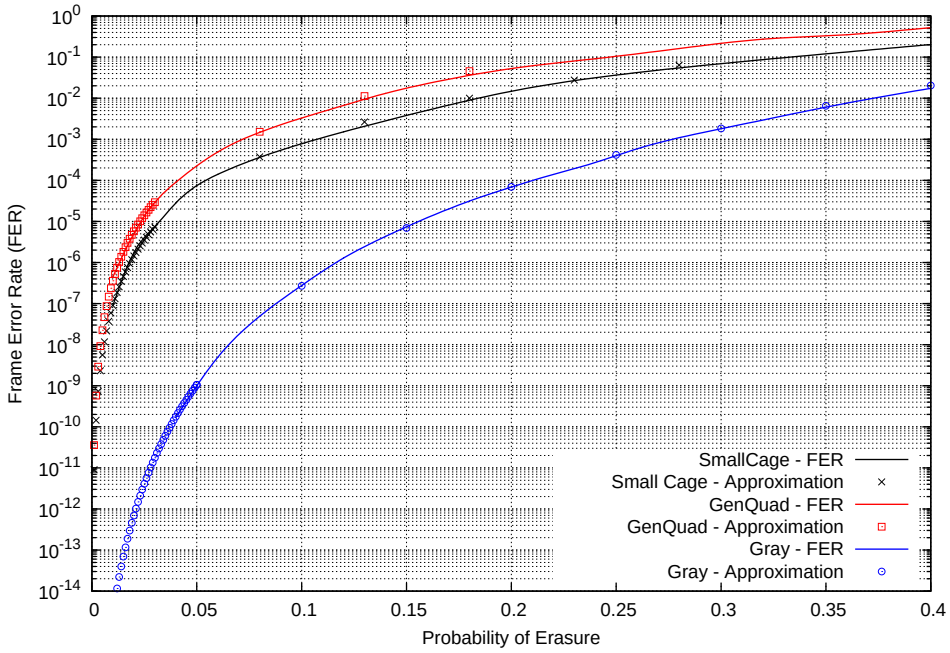

Figure 2: Frame Error Probability of Small LDPC Codes over the BEC

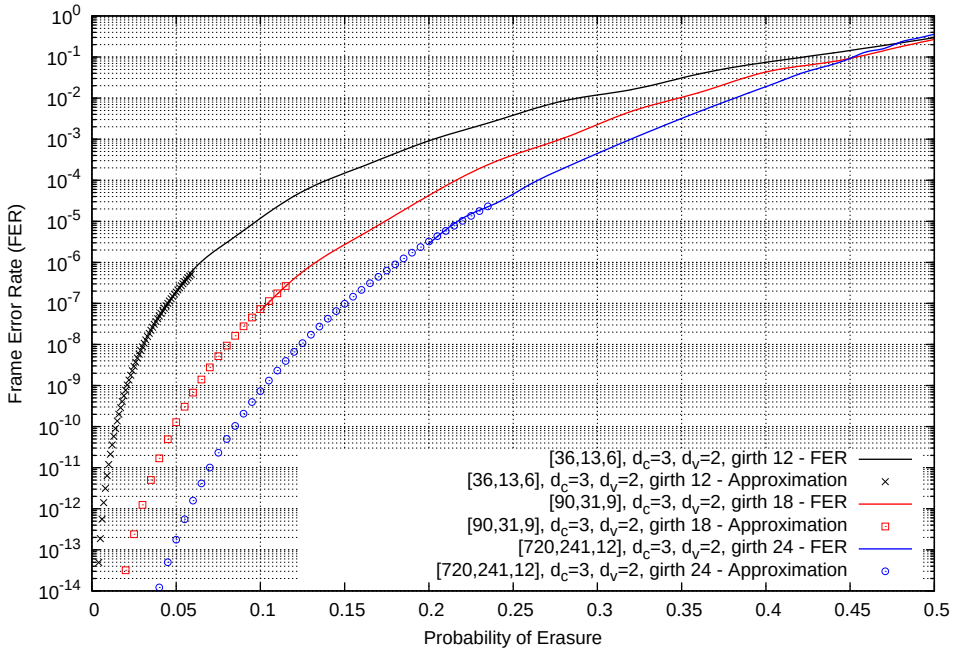

Figure 3: Frame Error Probability of $(3, 2)-$regular coset graph LDPC Codes over the BEC

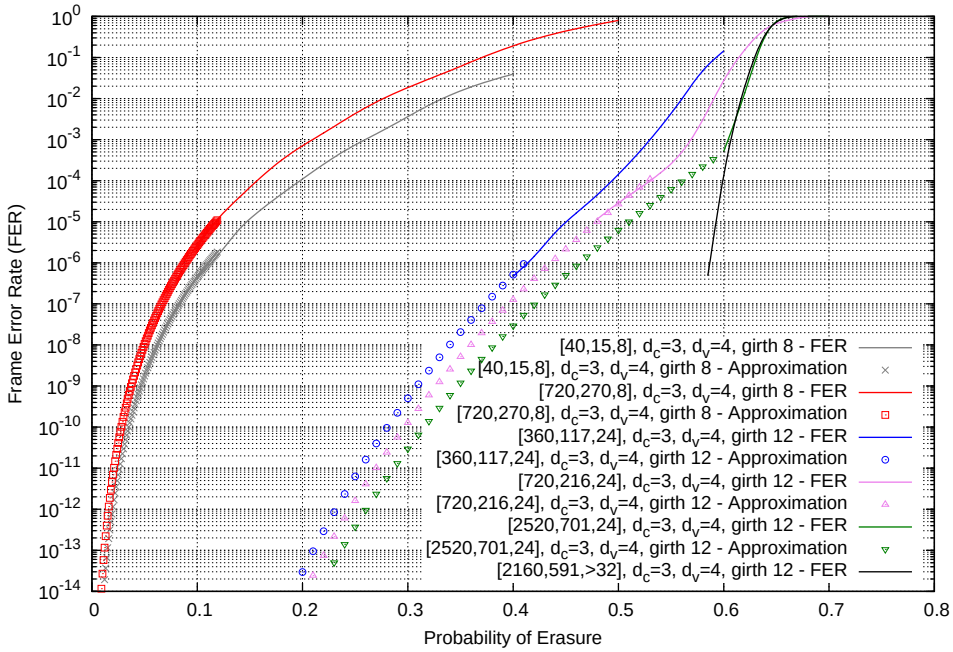

Figure 4: Frame Error Probability of $(3, 4)-$regular coset graph LDPC Codes over the BEC

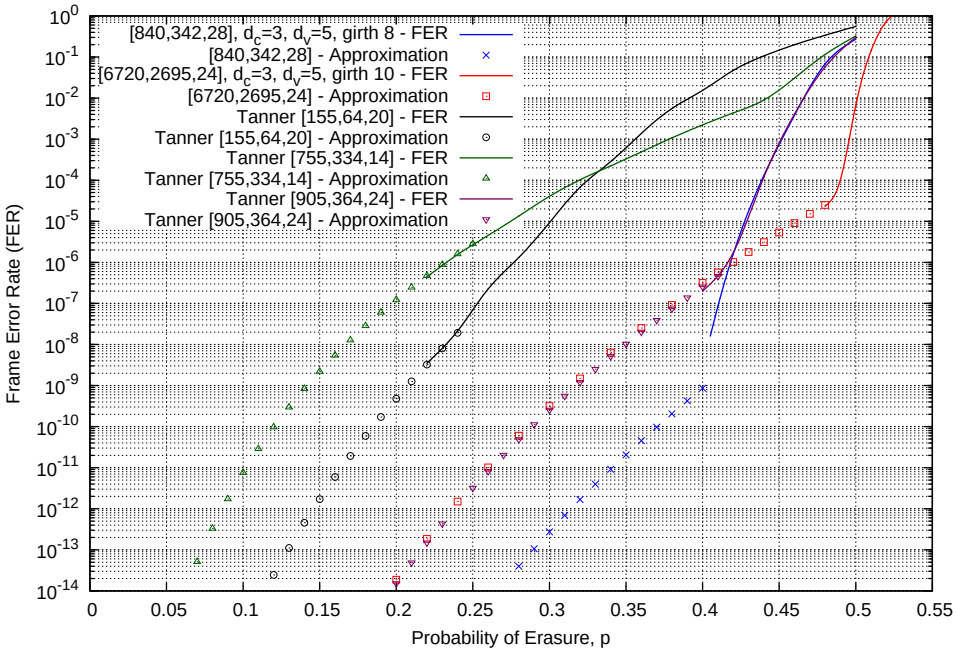

Figure 5: Frame Error Probability of $(3, 5)-$regular coset graph LDPC Codes over the BEC

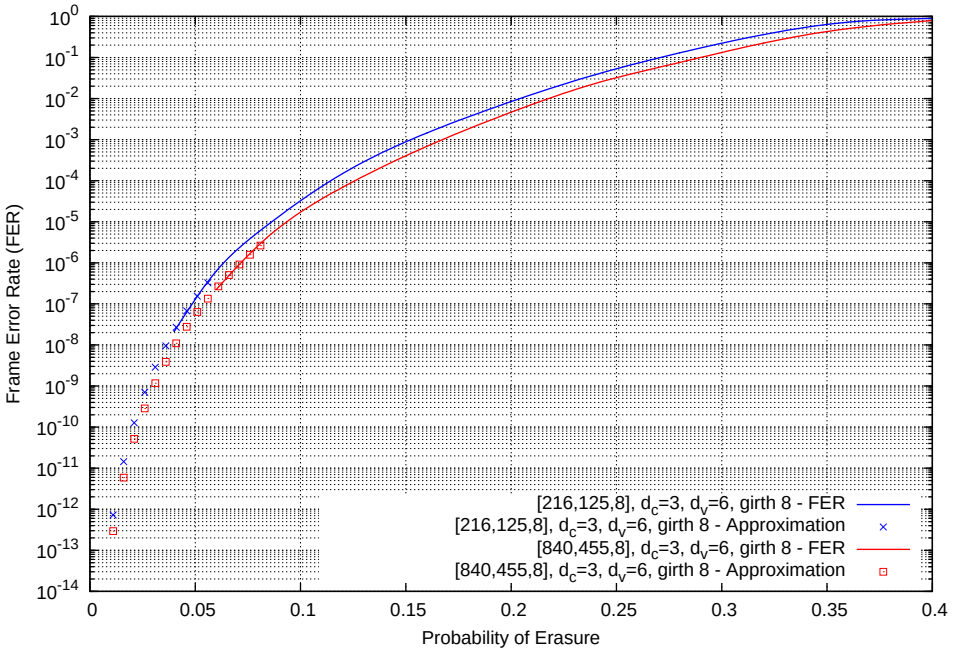

Figure 6: Frame Error Probability of $(3, 6)-$regular coset graph LDPC Codes over the BEC