On the Minimum Distance of Generalized LDPC Codes

Ayoub Otmani, Jean-Pierre Tillich, Iryna Andriyanova

On the Minimum Distance of Generalized LDPC Codes

Ayoub Otmani
ENSICAEN,
Bd. Maréchal Juin,
14050 Cedex Caen, France
Ayoub.Otmani@info.unicaen.fr

Jean-Pierre Tillich
INRIA, Projet Codes,
BP 105, Domaine de Voluceau
F-78153 Le Chesnay, France.
jean-pierre.tillich@inria.fr

Iryna Andriyanova
EPFL, School of Communications &
Computer Sciences
1015 Lausanne, Switzerland
iryna.andriyanova@epfl.ch

***Abstract*— We study necessary conditions which have to be satisfied in order to have LDPC codes with linear minimum distance. We give two conditions of this kind in this paper. These conditions are not met for several interesting code families: this shows that they are not asymptotically good. The second one concerns LDPC codes that have a Tanner graph in which there are cycles linking variable nodes of degree 2 together and provides some insight about the combinatorial structure of some low-weight codewords in such a case. When the LDPC code family is obtained from the lifts of a given protograph and if there are such cycles in the protograph, the second condition seems to capture really well the linear minimum distance character of the code. This is illustrated by a code family which is asymptotically good for which there is a cycle linking all the variable nodes of degree 2 together. Surprisingly, this family is only a slight modification of a family which does not satisfy the second condition.**

I. Introduction

Generalized LDPC codes, namely codes given by Tanner graphs where all check nodes are associated to single parity-check codes, with possibly variable nodes of degree 1 and non-transmitted (i.e. punctured) variable nodes have been shown to yield very good codes for iterative decoding for a large range of rates and lengths (see for instance [9]). If very low packet error rates are required, then great care has to be taken to design families of codes of large minimum distance. Ideally, asymptotically good families of codes are sought (that is families where the minimum distance grows linearly with the code-length). It is therefore interesting to find necessary and/or sufficient conditions for meeting this property. For well-studied code families such as turbo codes or standard (irregular) LDPC codes, upper bounds on their minimum distance can be found in [3] and [4].

The purpose of this paper is to provide sufficient conditions giving a logarithmic or sublinear upper bound on the minimum distance which are often met for **structured** families of LDPC codes. This complements the paper of Divsalar and *al.* [5] concerning LDPC codes defined by protographs. They have provided a sufficient condition for being asymptotically good which is called the *check node splitting condition*. More precisely, a protograph $\mathscr{P}$ satisfies the check node splitting condition if $\mathscr{P}$ contains only transmitted variable nodes of degree ≥ 2 and if the subgraph $\mathscr{P}_2$ of $\mathscr{P}$ that contains only degree-2 variable nodes with their attached edges and check nodes has no cycles.

This paper raises several issues:

- what happens if this subgraph contains cycles or for more general LDPC code families which are not lifts of a given protograph?
- what happens when there are non-transmitted variable nodes and variable nodes of degree 1? It should be stressed that having non-transmitted variable nodes and variable nodes of degree 1 is an important ingredient for having good iterative decoding performances, as it was put forward in [9].

This paper addresses several of these issues. First of all, for completeness, we recall a well known fact about how to handle the case of variable nodes of degree 1. If for a given check node there are more than one variable node of degree 1 attached to it, and if at least one of them is transmitted, then the minimum distance of the resulting code is at most 2. Moreover, all the variable nodes of degree 1 together with their attached check nodes can be erased from the Tanner graph without changing the linear minimum distance character of the code. The case of non-transmitted variable nodes is much more complicated however. We do not treat this case in full generality here. We do address partially however an important practical case, namely we consider the case when a certain graph denoted here by $\mathscr{G}_2$ consists of disjoint cycles. Many good LDPC codes belong to this class, see for instance [11], [9].

Definition 1: The *graph of degree-2 variable nodes* $\mathscr{G}_2$ of a generalized LDPC code is a graph whose vertex set V is formed by the check nodes involving transmitted degree-2 variable nodes. An edge connects two check nodes of V if and only if they are adjacent to a same transmitted degree 2 variable node in the Tanner graph.

Notice that $\mathscr{G}_2$ is a slight modification of a graph having been considered in [4] for irregular LDPC code ensembles. We show by generalizing a result of [10], that in this case these codes are not asymptotically good when a certain condition is met (Theorem 5). This condition is not only fulfilled by several interesting families of LDPC codes, see for instance Subsections IV-A or IV-B, but also casts some light about the structure of some low-weight codewords: they basically

involve only a few variable nodes of degree greater than 2 and a sublinear number of (transmitted) variable nodes of degree 2 which are arranged around a cycle. We also provide in Section V an example of a code family which is asymptotically good but for which the Tanner graph contains a cycle of linear length joining the transmitted variable nodes of degree 2.

The case of disjoint cycles linking the transmitted variable nodes of degree 2 together in the Tanner graph of the LDPC code is in some sense the critical case. If the subgraph $\mathscr{G}_2$ is only slightly denser, i.e. if its average degree is greater than 2, then the minimum distance is only at most logarithmic in the code-length, as stated in Proposition 2. This result is probably well known, but the only place we could find the mentioning on it is [2], so we have decided to include it here. The proof we provide uses in a crucial way a rather recent proof of an old conjecture of Bollobas giving a tight upper bound on the girth of irregular graphs.

II. A LOGARITHMIC UPPER BOUND ON THE MINIMUM DISTANCE

In this section, we treat the case when $\mathcal{G}_2$ is of average degree greater than 2. In this case, it turns out that the minimum distance is at most logarithmic in the code-length, as stated in the following proposition.

Proposition 2: Let Δ be the average degree of $\mathscr{G}_2$. If $\Delta > 2$ then the minimum distance $d_{\min}$ satisfies

$$d_{\min} \leq 2\log_{\Delta-1}\left(\frac{\Delta-2}{2}m+1\right)+1 \qquad (1)$$

where m is the number of parity checks.

Proof of Proposition 2: Let g be the girth of $\mathscr{G}_2$. Note that any cycle in $\mathscr{G}_2$ is associated to a codeword whose weight is the length of the cycle (its support is given by the edges of the cycle). Therefore $d_{\min} \leq g$. To upperbound this last quantity we use the Moore bound for irregular graphs [1] which asserts that the number of vertices m_2 of $\mathscr{G}_2$ satisfies the following inequality

$$m_2 \geq 2\frac{(\Delta-1)^t-1}{\Delta-2}$$

where $t = \lfloor \frac{g}{2} \rfloor$. This implies

$$t \leq \log_{\Delta-1}\left(\frac{\Delta-2}{2}m+1\right)$$

since $m_2 \leq m$. We now conclude by:

$$d_{\min} \leq g \leq 2t+1 \leq 2\log_{\Delta-1}\left(\frac{\Delta-2}{2}m+1\right)+1.$$

■

III. A POLYNOMIAL UPPER BOUND ON THE MINIMUM DISTANCE

We have seen that when the average degree Δ of the transmitted degree-2 variable node graph $\mathscr{G}_2$ is greater than 2, then the code can not be asymptotically good. Moreover, if $\mathscr{G}_2$ contains no cycle, that is to say $\Delta < 2$, we know [5] that $\mathscr{G}_2$ is asymptotically good if it satisfies the check node splitting condition. Therefore, the critical case is $\Delta = 2$. We give in Theorem 5 sufficient conditions that enables to determine a new polynomial upper bound. It generalizes a result of [10] and some definitions are required to prove it.

Definition 3: A *dangerous cycle* is a set of parity-checks and transmitted variable nodes of degree 2 which form a single cycle in the Tanner graph. The set of parity-check nodes belonging to dangerous cycles is called the set of *dangerous* check nodes and is denoted by $\mathcal{D}$.

Definition 4: A *potentially bad set of variable nodes* X is a set of variable nodes in the Tanner graph which do no belong to dangerous cycles and such that when they are assigned to 1 and all other variable nodes are assigned to 0, the only check nodes which are not satisfied belong to dangerous cycles. The *defect* of this set is the set of unsatisfied parity-checks and is denoted by $\delta(X)$.

Theorem 5: Let n be the length of a generalized LDPC code. If there exist $K_1 n^{\alpha}$ disjoint potentially bad sets of variable nodes which are all of cardinality less than K_2 and defect size less than K_3, then the minimum distance $d_{\min}$ of the code satisfies

$$d_{\min} \leq K_4 n^{1-\frac{\alpha}{K_3}} + 2K_2 + 1,$$

where $K_4 = 2\frac{K_3^2}{K_1}$.

To prove this theorem we will also need the following

Definition 6: An *annihilating configuration* of a subset $A \subset \mathcal{D}$ of dangerous check nodes is a set of (transmitted) variable nodes of degree 2 belonging to dangerous cycles, such that when they are assigned to 1 and all other variable nodes are assigned to 0, the set of unsatisfied parity-check nodes is precisely A. The set of annihilating configurations for A is called the *annihilating set* of A and is denoted by $\mathrm{Ann}(A)$. The distance between two subsets A and B of parity-check nodes belonging to $\mathcal{D}$ is defined by the minimum size of an annihilating configuration of the symmetric difference of A and B denoted by $A \oplus B$. When A and B are disjoint and if the annihilating set $\mathrm{Ann}(A \oplus B)$ is empty, then the distance is infinite. The corresponding quantity is denoted by $\Delta(A, B)$. We set by definition $\Delta(A, A) = 0$.

It should be noted that $\Delta(,)$ is indeed a distance. This is a consequence of the following fact. If x is an annihilating configuration for $A \oplus B$ and y is an annihilating configuration for $B \oplus C$, then by linearity of parity-checks $x \oplus y$ is an annihilating configuration for $A \oplus C$. This implies the triangular inequality for $\Delta(,)$.

Since for any pair of disjoint potentially bad sets of variable nodes A and B, and any $x \in \mathrm{Ann}(\delta(A) \oplus \delta(B))$, $A \cup B \cup x$ is the support of a codeword, we have

$$d_{\min} \leq \Delta(\delta(A), \delta(B)) + |A| + |B|. \qquad (2)$$

This is the crucial inequality. What we are going to prove now is that there are two potentially bad sets of variable nodes of small size (less than K_2) A and B such that there is an annihilating configuration for $\delta(A) \oplus \delta(B)$ of sublinear size. In other words we are going to prove the existence of a codeword

which involve at most $2K_2$ variable nodes which are not in dangerous cycles and a sublinear number of positions which belong to dangerous cycles.

Proof of Theorem 5: This is essentially a packing argument over the set of all possible defects of a given size. Let t be the number in $\{1,\dots,K_3\}$ for which the number of potentially bad sets of size at most K_2 and defect size t is the largest. We know that this number is at least $\frac{K_1}{K_3}n^\alpha$ and denote by $\mathcal{B}$ the corresponding set of potentially bad sets. For a subset $A\subset\mathcal{D}$, we denote by $\mathscr{B}_r(A)$ the set of subsets of $\mathcal{D}$ of size t which are at distance at most r of A. From Inequality (2) we know that all the sets $\mathscr{B}_{\lfloor\frac{d_{\min}-2K_2-1}{2}\rfloor}(\delta(X))$ are disjoint for $X\in\mathcal{B}$. Therefore

$$\sum_{X\in\mathcal{B}}\left|\mathscr{B}_{\lfloor\frac{d_{\min}-2K_2-1}{2}\rfloor}(\delta(X))\right|\leq\binom{m}{t}\leq\frac{m^t}{t!},\qquad(3)$$

where $m\stackrel{\text{def}}{=}|\mathcal{D}|$. On the other hand there is a simple lower bound on the sizes of such balls when all the dangerous cycles are large enough. There are two cases to consider : all dangerous cycles are of size at least $2M$ where $M\stackrel{\text{def}}{=}K_4n^{\frac{K_3-\alpha}{K_3}}+2K_2+1$ or there is one dangerous cycle whose size is smaller. In the latter case there is a non-zero codeword of weight less than M with support the variable nodes of degree 2 belonging to the aforementioned cycle and we are done.

Assume now that all dangerous cycles are of size at least $2M$. For $A\subset\mathcal{D}$ of size t and $r\leq M$ we have

$$|\mathscr{B}_r(A)|\geq\frac{\binom{t+r}{t}}{t!}\geq\frac{r^t}{(t!)^2}.\qquad(4)$$

To check this point let us choose some order on the dangerous checks and let us observe that the aforementioned ball contains all subsets B of t dangerous check nodes such that for all $i\in\{1,\dots,t\}$ the i-th check node in B is in the same cycle as the i-th check node of A and at distance l_i such that

$$\sum_i l_i\leq r.$$

The number of such subsets is clearly lower bounded by the number of non-negative t-tuples $(l_1,\dots,l_t)$ such that $l_1+\cdots+l_t\leq r$ divided by $t!$. The number of such t-tuples is equal to $\binom{t+r}{t}$ and this implies Equation (4). Combining Equations (3) and (4) and letting $r=\lfloor\frac{d_{\min}-2K_2-1}{2}\rfloor$, we obtain

$$\frac{K_1}{K_3}n^\alpha\frac{r^t}{(t!)^2}\leq\frac{m^t}{t!},$$

which implies

$$r\leq\left(t!\frac{K_3}{K_1}\right)^{\frac{1}{t}}n^{-\frac{\alpha}{t}}m,$$

and since $m\leq n$ we deduce that

$$r\leq\left(t!\frac{K_3}{K_1}\right)^{\frac{1}{t}}n^{1-\frac{\alpha}{t}},$$

so we have that

$$d_{\min}\leq 2\left(t!\frac{K_3}{K_1}\right)^{\frac{1}{t}}n^{1-\frac{\alpha}{t}}+2K_2+1.$$

Using the fact that $(t!)^{\frac{1}{t}}\leq K_3$ for $1\leq t\leq K_3$, we obtain

$$d_{\min}\leq\frac{2K_3^2}{K_1}n^{1-\frac{\alpha}{K_3}}+2K_2+1.$$

■

IV. Examples

A. LDPC codes with two variable nodes of degree 2 per parity-check equation

An important class of LDPC codes is obtained by choosing structured LDPC codes where each parity check involves exactly two variable nodes of degree 2. They display many interesting features which make them quite attractive for standardisation: they can be linearly encoded [8], [11], the minimum distance is typically some power of the code-length [10] and they can be decoded in a repeat-accumulate way which generally decreases drastically the number of decoding iterations. It is known that "regular" codes of this kind can not be asymptotically good. By "regular", we mean LDPC codes where all parity checks involve exactly 2 variable nodes of degree 2 and some constant number c of variable nodes of degree d. It is namely proved in [10] that the minimum distance of such a code with length n is always upper-bounded by a term of order $O(n^{\frac{d-1}{d}})$ where d is the degree of the variable nodes of degree greater than 2. This result is a special case of Theorem 5. Indeed, in this case there are $\Theta(n)$ variable nodes of degree d. Any such variable node forms a potentially bad set of variable node of size 1 and defect size d. More generally, by the same kind of argument we obtain

Proposition 7: Any LDPC code of length n with two variable nodes of degree 2 per parity-check equation with $\Theta(n)$ variable nodes of degree d is of distance at most $O(n^{\frac{d-1}{d}})$.

B. A Multi-edge example

In [9] a few generalized LDPC codes are presented, and some of them have quite good iterative decoding performances. In particular, the multi-edge code of rate $\frac{1}{2}$ defined in table VIII displays one of the best known performances for low-complexity iterative decoding for lengths in the range $1000-10000$ and target block error rates in the range $10^{-5}-10^{-2}$. This code family presents however an error floor which begins in the range $10^{-5}-10^{-3}$, depending on the code length. One might wonder whether or not such a code can be asymptotically good. Actually, this is not the case.

To see this, recall that the Tanner graph of such a code of length n is given by the following figure, where the transmitted nodes are in black and the non-transmitted variable nodes are in white. The edges are obtained by matching together the sockets associated with the variable nodes and the sockets associated with the check nodes which are of the same type. The type is given by the color (blue,red,black) and the fact that it is represented by a solid or a dashed line. There are 6 types of sockets here.

The dangerous cycles are formed by variable nodes of degree 2 alternating with check nodes of degree 5 (which might belong either to the first group of check nodes of degree

5 or to the second one). The potentially bad set of variable nodes are given by single (transmitted) vertices of degree 3. Their defect is of size 3. There are therefore $\frac{3n}{10}$ potentially bad sets of variable nodes. By using Theorem 5, we know that the minimum distance of this code is at most of order $O(n^{\frac{2}{3}})$. It should be added that in this case, by using the same kind of proof technique as in [10], it could be proved that by taking random matchings of sockets of the same kind the typical minimum distance would be smaller: it would be of order $O(n^{\frac{1}{3}})$.

C. A protograph example

In all the previous examples, the potentially bad sets of variables nodes were formed by single vertices of degree greater than 2. We will now give a more complicated example of a code of designed rate $\frac{1}{3}$ and of sublinear minimum distance where the potentially bad sets of variable nodes have a more complicated structure. This code is defined by a Tanner graph which is a lift of the protograph of Figure 1. There are three kinds of variable nodes, white, red and blue variable nodes. The dangerous cycles of the Tanner graph are formed here by cycles with alternating red check nodes and blue variable nodes. The structure of the potentially bad sets of variable nodes is now given by the subgraph of the Tanner graph induced by a white vertex and the three neighboring red vertices as shown in Figure 2. If the code is of length n, there are $\frac{n}{3}$ white variable nodes and therefore also $\Theta(n)$ disjoint potentially bad sets of variable nodes of size 4 and defect size 6. The minimum distance of such codes is therefore at most of order $O(n^{\frac{5}{6}})$ by Theorem 5.

V. An asymptotically good family of codes with a cycle of degree 2 vertices

One might wonder whether codes where $\mathscr{G}_2$ consists of cycles may have linear minimum distance. This is of course possible just by taking a Tanner graph of a code of some length n with linear minimum distance and by adding a new cycle of length $2n$ consisting of n additional degree-2 variable nodes alternating with check nodes of degree 2. This defines a new code of length $2n$ with the same minimum distance as the former code. But there are far more interesting examples of codes with linear minimum distance for which the Tanner graph contains dangerous cycles of linear size. Consider the slight modification of the code of Subsection IV-C which consists in changing just one edge of the protograph as indicated in Figure 3.

We consider codes of length $3n$ consisting in n-lifts of this protograph where the n blue vertices of degree 2 form a cycle of length $2n$ with the check nodes of degree 3. Note that in this case, Theorem 5 does not apply : there are no potentially bad sets of variable nodes of constant size.

It turns out that a constant fraction of codes of this kind have linear minimum distance. The proof uses considerations on the average weight distribution (the average being taken over all codes of this kind of the same length). Let $\bar{a}_{s,t,u}$ be the average number of codewords of such a code consisting of s blue variables being equal to 1, t red variable nodes being equal to 1, u white variable nodes being equal to 1 and all remaining variable nodes being equal to 0. Let $d_{\min}$ be the minimum distance of our code. We will use that

$$\mathbf{Prob}(d_{\min} \leq v) \leq \sum_{s,t,u:0<s+t+u\leq v} \bar{a}_{s,t,u}. \tag{5}$$

Let $b_{s,t}$ be the number of codewords of the code of length $2n$ given by the Tanner graph given by Figure 4, where there are exactly s blue vertices assigned to 1 and t red vertices assigned to 1. One might check, following [7], that $b_{s,t} = 0$ if t is odd, and if t is even that

$$b_{s,t} = \frac{2n\binom{s-1}{t/2-1}\binom{n-s-1}{t/2-1}}{t}. \tag{6}$$

We also let

$$C(y,z) \stackrel{\text{def}}{=} \sum_{t,u} c_{t,u} y^t z^u \stackrel{\text{def}}{=} (1+y^2+6yz+3z^2+2yz^3+3y^2z^2)^n. \tag{7}$$

It can be checked that $c_{t,u}$ is the number of codewords of the code of length $5n$ given by the Tanner graph given by Figure 5, where there are exactly t red vertices assigned to 1 and u white vertices which are assigned to 1.

With the help of these quantities it is readily checked that

Lemma 8:

$$\bar{a}_{s,t,u} = \frac{b_{s,t} c_{2t,3u} \binom{n}{u}}{\binom{2n}{2t}\binom{3n}{3u}}.$$

It will be convenient to bring in the quantities

$$\begin{aligned} \sigma &\stackrel{\text{def}}{=} \frac{s}{n} \\ \tau &\stackrel{\text{def}}{=} \frac{t}{n} \\ \nu &\stackrel{\text{def}}{=} \frac{u}{n} \\ \alpha(\sigma,\tau,\nu) &\stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{\ln(\bar{a}_{s,t,u})}{n} \end{aligned}$$

To evaluate $\alpha(\sigma,\tau,\nu)$, we first handle the $c_{2t,3u}$ terms with

Lemma 9: Let $\gamma(\tau,\mu) \stackrel{\text{def}}{=} \limsup \frac{1}{n} \ln c_{t,u}$. Then

$$\gamma(\tau,\nu) \leq \tau + \frac{3}{2}\nu - \tau \ln \frac{\tau}{1+3\mu} - \frac{3\nu}{2} \ln \frac{\nu}{2+2/\mu}, \tag{8}$$

where $\mu \stackrel{\text{def}}{=} \frac{3\nu-2\tau+\sqrt{(2\tau-3\nu)^2+8\tau\nu}}{4\tau}$.

We will not give the details of the calculations here, but we just indicate that this follows from the trivial upper bound

$$c_{2t,3u} \leq \inf_{y>0,z>0} \frac{C(y,z)}{y^{2t} z^{3u}},$$

from which it follows that

$$\begin{aligned} \gamma(\tau,\nu) \quad \leq \quad & \inf_{y>0,z>0} \ln(1+y^2+6yz+3z^2+2yz^3+3y^2z^2) \\ & -2\tau \ln y - 3\nu \ln z. \end{aligned}$$

From this upper bound, we deduce the handier expression

Lemma 10: For $\tau + \nu \leq 0.009$,

$$\gamma(\tau,\nu) \leq 2\tau + 2\nu - \tau \ln \tau - \frac{15}{8}\nu \ln \nu.$$

For the other terms, which involve binomial coefficients, we use the following inequalities which are quite sharp for small t and which can be deduced from Stirling's approximation, [6, §II.9].

Fact 11: There exist two constants K and K' such that

$$K' e^{-\frac{t^2}{n-t}} \sqrt{\frac{n}{t(n-t)}} \left(\frac{ne}{t}\right)^t \leq \binom{n}{t} \leq K \sqrt{\frac{n}{t(n-t)}} \left(\frac{ne}{t}\right)^t.$$

For the term $b_{s,t}$ we proceed by bringing in $\beta(\sigma,\tau) \stackrel{\text{def}}{=} \frac{1}{n} \ln b_{s,t}$ which we upper-bound as follows by using Fact 11 :

$$\beta(\sigma,\tau) \leq -\tau \ln \tau + \tau + \frac{\tau}{2} \ln\left(4\sigma(1-\sigma)\right) + \frac{\ln(\tau/\sigma)}{n} \quad (9)$$

On the other hand the term $\frac{\binom{n}{u}}{\binom{2n}{2t}\binom{3n}{3u}}$ is upper-bounded by Fact 11 as follows

$$\frac{1}{n} \ln\left(\frac{\binom{n}{u}}{\binom{2n}{2t}\binom{3n}{3u}}\right) \leq -2\tau - 2\nu + 2\tau \ln \tau + 2\nu \ln \nu + K(\tau+\nu)^2, \quad (10)$$

for some constant $K > 0$. Putting all these upper bounds together we obtain

Lemma 12: For $\tau + \nu \leq 0.009$, there exists some constant $K > 0$ such that:

$$\alpha(\sigma,\tau,\nu) \leq \frac{\tau}{2} \ln\left(4e^2\sigma(1-\sigma)\right) + \frac{\ln(\tau/\sigma)}{n} + \frac{1}{8}\nu \ln \nu + K(\tau+\nu)^2.$$

From the last lemma and $\bar{a}_{s,0,0} = 0$ for $s \neq 0, s \neq n$ we deduce that

Lemma 13: There exists $\delta > 0$ such that

$$\sum_{0<s+t+u<\delta n} \bar{a}_{s,t,u} \leq \frac{1}{2}.$$

By using this lemma and Inequality (5) we obtain

Proposition 14: At least half of the codes defined in this section have their minimum distance greater than δn.

REFERENCES

[1] N. Alon, S. Hoory, and N. Linial. The Moore bound for irregular graphs. *Graphs Combin.*, 18:53–57, 2002.

[2] I. Andriyanova. *"Analysis and design of a certain family of graph-based codes: TLDPC codes"*. PhD thesis, National Superior School of Telecommunications, Paris, France, December 2006.

[3] M. Breiling. A logarithmic upper bound on the minimum distance of turbo codes. *IEEE Transactions on Information Theory*, 50(8):1692–1710, August 2004.

[4] C. Di, T. Richardson, and R. Urbanke. Weight distribution of low-density parity-check codes. *IEEE Transactions on Information Theory*, 52(11):4839–4855, November 2006.

[5] D. Divsalar, S. Dolinar, and C. Jones. Construction of protograph LDPC codes with linear minimum distance. In *Proceedings ISIT 2006*, July 2006.

[6] W. Feller. *An introduction to probability theory and its applications*, volume 1. John Wiley and Sons Inc., New York, 3rd edition, 1968.

[7] R. Ikegaya, K. Kasai, Y. Shimoyama, T. Shibuya, and K. Sakaniwa. Stopping set distributions of two-edge type LDPC code ensembles. In *Proceedings ISIT 2005*, pages 985–989, September 2005.

[8] D.J.C. Mackay. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.

[9] T. Richardson and R. Urbanke. Multi-edge LDPC codes. Available at:http://lthcwww.ep.ch/papers/multiedge.ps.

[10] J.P. Tillich and G. Zémor. On the minimum distance of structured LDPC codes with two variable nodes of degree-2 per parity-check equation. In *Proceedings of ISIT 2006*, Seattle, USA, 2006.

[11] M. Yang, W. E. Ryan, and Y. Li. Design of efficiently encodable moderate-length high-rate irregular LDPC codes. *IEEE Transactions on Communications*, 52(4):564–571, April 2004.

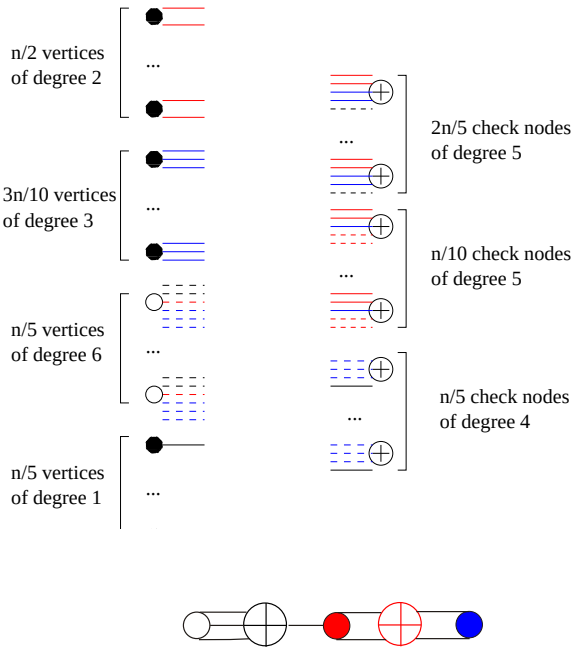

Fig. 1. A protograph defining a code of designed rate $\frac{1}{3}$

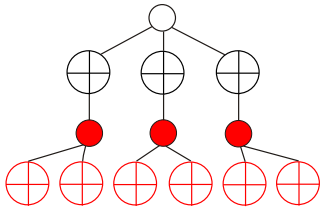

Fig. 2. A potentially bad set of size 4 and defect 6

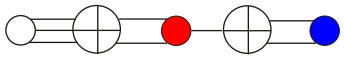

Fig. 3. A slight modification of the protograph given in Figure 1

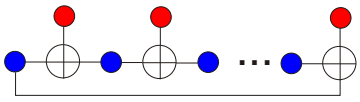

Fig. 4. A cyclic Tanner graph

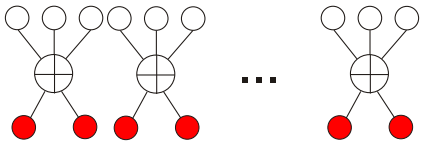

Fig. 5. The Tanner graph with weight distribution $C(y, z)$