Constructing High-rate Scale-free LDPC Codes

X. Zheng, F. C. M. Lau and C. K. Tse
Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hunghom, Hong Kong
Email: {encmlau,encktse}@polyu.edu.hk.

***Abstract*— Low-density parity-check (LDPC) codes with scale-free (SF) symbol-node degree distribution have been shown to provide very good error performance. When the code rate becomes high, however, there will be a lot of degree-2 symbol nodes in the "pure" SF-LDPC codes. As a consequence, when the codes are constructed by connecting the symbol nodes with the check nodes, many small-size cycles will be formed. Such small-cycles will degrade the error performance of the codes. In this paper, we address the issue by imposing a new constraint on the design of high-rate SF-LDPC codes. We will compare the error rates of the constrained SF-LDPC codes and other optimized LDPC codes.**

I. Introduction

Complex networks have been studied across many disciplines, including computer science, social sciences, engineering and biology [14], [1], [2], [6]. Some basic properties such as average path length (APL), clustering coefficient, degree distribution and many other properties such as betweenness centrality, assortative mixing and dis-assortative mixing have been analyzed for different types of networks. Such properties are very valuable and effective in characterizing many different physical networks. The most commonly known networks include the random network, lattice, small-world network and scale-free network [14], [4]. Recently, there has been a growing interest in applying features of complex networks to solving engineering problems [6], [16], [17], [12], [13]. One of the applications is the design of short-length low-density parity-check (LDPC) codes [16], [17].

Low-density parity-check (LDPC) codes [5] form a class of linear block codes with error performance approaching the Shannon limit [11]. An LDPC code can be represented by a Tanner graph characterized by a set of nodes V and a set of connections (edges) E among the nodes. We define k and m, respectively, as the number of message bits and the number of parity bits in a codeword which has a length of $n = k + m$. Thus, there are n symbol nodes and m check nodes in an LDPC code, and the code rate R_{code} is given by $k/n = 1 - m/n$. The node set V is further divided into the set of symbol nodes V_s and the set of check nodes V_c. The sets of symbol nodes and check nodes can be expressed by $V_s = \{v_1, v_2, \ldots, v_n\}$ and $V_c = \{c_1, c_2, \ldots, c_m\}$, respectively, where v_j represents the j-th symbol node and c_i denotes the i-th check node. Fig. 1 shows a Tanner graph representation of an LDPC code in which there are 10 symbol nodes and 5 check nodes.

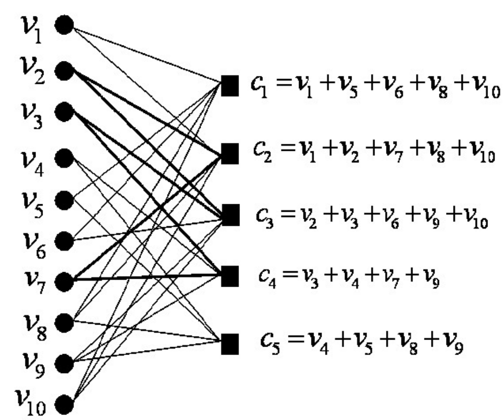


Fig. 1. A Tanner graph representation of an LDPC code with 10 symbol nodes and 5 check nodes.

Suppose the channel is an additive white Gaussian noise (AWGN) channel and the typical belief propagation (BP) iterative decoder [10] is employed at the receiving end. As the corrupted signals are received, messages regarding the probability that the individual received symbol should be a "0" or a "1" will be passed from the symbol nodes $v_1, v_2, \ldots, v_n$ to their neighboring check nodes $c_1, c_2, \ldots, c_m$ based on the set of connections E. Then, each check node will return messages to its connected symbol nodes. Moreover, the message from a symbol node v to a check node c is calculated based on the received messages from both the channel and the neighboring check nodes except c. Similarly, the message from a check node c to a symbol node v is computed based on the incoming messages from the neighboring symbol nodes except v. The iterative process continues until a valid codeword is found, or the maximum number of iterations has been reached.

In [16], [17], the "shortest-APL" feature of scale-free networks has been applied to the design of short-length LDPC) codes, forming the scale-free LDPC (SF-LDPC) codes. Compared with the LDPC codes optimized by the DE-algorithm, the short-length SF-LDPC codes, which are characterized with scale-free symbol-node-degree distributions, can produce similar error rates in the waterfall region and lower error rates at the high signal-to-noise ratio (SNR) region. One issue arises, however, when the code rate increases. In the SF-LDPC codes, a large proportion of the symbol nodes has a degree of 2 and the proportion increases with the code rate. When the ratio of degree-2 symbol nodes is too high, short-cycles containing only degree-2 symbol nodes will be produced in the Tanner graph, resulting in a degradation of the error performance. In this paper, we aim to tackle this issue by imposing a restriction

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

on the proportion of degree-2 symbol nodes in SF-LDPC codes.

In Sect. II, we describe our proposed algorithm for constructing LDPC codes with high rates. The characteristics of some codes constructed, together with their error performance, are given in Sect. III. Finally, we make some remarks in Sect. IV.

II. Construction Algorithm for LDPC Codes

A. Degree Distributions

Given a code rate R_{code} and a maximum symbol-node degree d_v, we first derive the degree distributions of the scale-free LDPC codes under a certain proportion of degree-2 symbol nodes. We define

$$\lambda(x) := \sum_{k=2}^{d_v} \lambda_k x^{k-1} \quad \text{and} \quad \rho(x) := \sum_{k=2}^{d_c} \rho_k x^{k-1} \tag{1}$$

as the symbol-node degree distribution and the check-node degree distribution, respectively. In (1), λ_k denotes the fraction of edges connected to degree-k symbol nodes and ρ_k denotes the fraction of edges connected to degree-k check nodes. Also, d_v and d_c denote the maximum symbol-node degree and maximum check-node degree, respectively. In the following, we show the algorithm in deriving $\lambda(x)$ and $\rho(x)$.

1) Denote the probability that a symbol node has k connections by $P_\lambda(k)$.
2) Suppose $P_\lambda(2)$ has been set to F_2, which is equal to or slightly larger than $1 - R_{\text{code}}$. The total number of edges connecting to degree-2 symbol nodes is thus $2nF_2$, where n represents the number of symbol nodes.
3) For $k = 3, 4, \ldots, d_v$, we assume that the fraction of symbol nodes with degree k follows a scale-free distribution, i.e., $P_\lambda(k) = Ak^{-\gamma}$ with A being the normalizing coefficient and γ being the characteristic exponent.
4) Since
$$\sum_{k=2}^{d_v} P_\lambda(k) = 1, \tag{2}$$
we have
$$A = \frac{1 - F_2}{\sum_{k=3}^{d_v} k^{-\gamma}}. \tag{3}$$
5) The total number of edges connecting to symbol nodes with degree k ($k = 3, 4, \ldots, d_v$) equals
$$knP_\lambda(k) = \frac{n(1 - F_2)}{\sum_{l=3}^{d_v} l^{-\gamma}} \, k^{1-\gamma}. \tag{4}$$
6) Thus, when $k = 2$,
$$\lambda_2 = \frac{2F_2}{2F_2 + \frac{\sum_{u=3}^{d_v} u^{1-\gamma}(1-F_2)}{\sum_{l=3}^{d_v} l^{-\gamma}}}; \tag{5}$$
when $k = 3, 4, \ldots, d_v$,
$$\lambda_k = \frac{k^{1-\gamma}(1 - F_2)/\sum_{l=3}^{d_v} l^{-\gamma}}{2F_2 + \frac{\sum_{u=3}^{d_v} u^{1-\gamma}(1-F_2)}{\sum_{l=3}^{d_v} l^{-\gamma}}}. \tag{6}$$
7) The average symbol-node degree, denoted by $< k_v >$, is computed using
$$< k_v > = \sum_{k=2}^{d_v} kP_\lambda(k) = 2F_2 + \sum_{k=3}^{d_v} \frac{(1 - F_2)k^{1-\gamma}}{\sum_{l=3}^{d_v} l^{-\gamma}}. \tag{7}$$
8) We assume that a check node can only have d_c-2, d_c-1, or d_c connections and that the check-node degree $k \in \{d_c - 2, d_c - 1, d_c\}$ follows a Poisson distribution with parameter ν, i.e.,
$$P_\rho(k) = \frac{\nu^k \exp(-\nu)}{k!}. \tag{8}$$
9) The fraction of edges connecting to check nodes with degree $k \in \{d_c - 2, d_c - 1, d_c\}$ therefore equals
$$\rho_k = \frac{\frac{\nu^k \exp(-\nu)}{(k-1)!}}{\sum_{l=d_c-2}^{d_c} \frac{\nu^l \exp(-\nu)}{(l-1)!}}. \tag{9}$$
10) The check-node degree distribution in (1) can be further written as
$$\rho(x) = \sum_{k=d_c-2}^{d_c} \frac{\frac{\nu^k \exp(-\nu)}{(k-1)!}}{\sum_{l=d_c-2}^{d_c} \frac{\nu^l \exp(-\nu)}{(l-1)!}} x^{k-1}. \tag{10}$$
11) Combining the aforementioned results gives
$$\frac{< k_v >}{1 - R_{\text{code}}} = \frac{(d_c-2)(d_c-1)d_c + (d_c-1)d_c\nu + d_c\nu^2}{(d_c-1)d_c + d_c\nu + \nu^2}. \tag{11}$$
12) Since d_c is an integer greater than 2, we have
$$d_c - 2 < \frac{< k_v >}{1 - R_{\text{code}}} < d_c \tag{12}$$
and hence d_c equals $\left\lceil \frac{<k_v>}{1-R_{\text{code}}} \right\rceil$ or $\left\lceil \frac{<k_v>}{1-R_{\text{code}}} \right\rceil + 1$, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x.
13) Once d_c is selected, the corresponding ν can be found using (11).

B. Optimized Degree Distributions

In the previous section, we derive systematically the degree distributions of the symbol nodes and check nodes. However, the degree distributions have not been optimized. In order to obtain the largest achievable threshold, denoted by σ^*[1], we further perform the following iterations.

- Set F_2 to $1 - R_{\text{code}}$.
- For $\gamma = 1.90$ to 2.50 in steps of 0.01, do
  - Based on the method in Sect. II-A, derive the degree distributions of the symbol nodes and check nodes.
  - With the degree distributions, evaluate the threshold σ^* using the density-evolution algorithm [10], [9].

[1]The threshold σ^* can be regarded as the theoretical largest noise standard deviation of an AWGN channel that allows the LDPC codes to perform error-free communications.

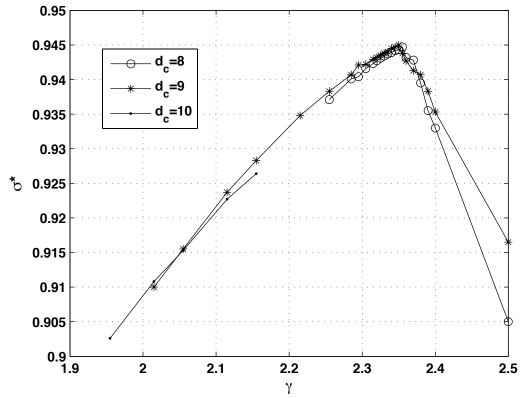


Fig. 2. A typical plot of achievable error performance σ^* versus γ.

- End ($\gamma = 1.90$ to 2.50 in steps of 0.01)
- If $F_2 == 1.05 - R_{\text{code}}$, stop. Otherwise, increment F_2 by 0.001 and repeat the previous "for" loop.

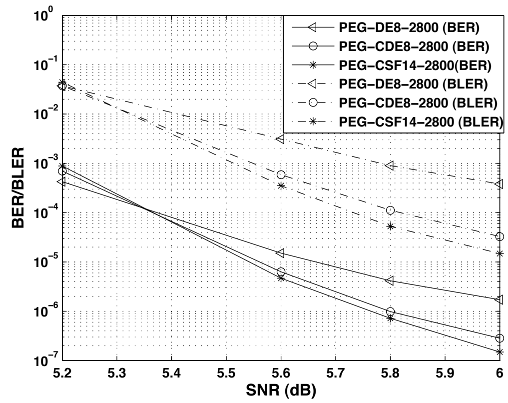
Finally, we record the largest achievable threshold and the associated parameters such as degree distributions, average symbol-node degree and characteristic exponent. Fig. 2 shows a typical plot of the achievable threshold as the characteristic exponent γ varies. It shows that the threshold peaks at a certain value of γ before it decreases.

C. Connecting Symbol Nodes with Check Nodes

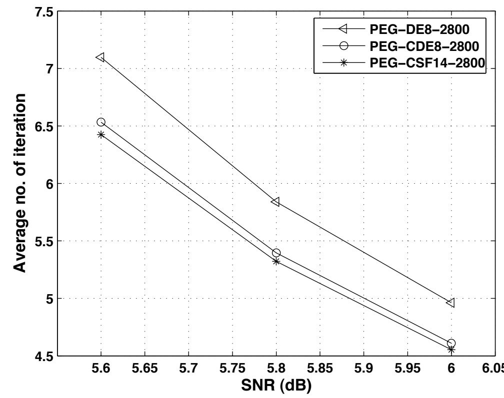
All the parameters regarding the LDPC code have been found. The next procedure is to connect the symbol nodes with the check nodes. We define the "girth" of a symbol node as the length of the smallest cycle in the Tanner graph that is originated from the symbol node. The "girth average" is then the mean of the girths over all the symbol nodes. In [8], the authors have related the error performance of LDPC codes to the girth average in the associated Tanner graph. The results have concluded that short-length codes with good error-correcting performance usually have a large girth average. To construct short-length LDPC codes with large girth, a broad class of methods have been proposed [8], [7]. Among the methods, progressive edge growth (PEG) [7] is one of the most effective algorithms to enlarge the girth of the codes. Here, we also make use of the PEG algorithm to establish the connections in our LDPC codes edge-by-edge.

III. Results and Discussions

We first use our proposed algorithm to optimize the degree distributions for rate-0.75 and rate-0.82 SF-LDPC codes. We called the SF-LDPC codes built with the new constraint as constrained SF-LDPC (CSF-LDPC) codes. Table I presents the highest thresholds achieved by rate-0.75 and rate-0.82 CSF-LDPC codes as well as other best-known high rate codes. The corresponding parameters used are also tabulated. The results indicate that the "pure" DE produces a slightly larger σ^* compared with other LDPC codes. However, "pure" DE also produces the proportion of degree-2 symbol nodes almost two



(a)



(b)

Fig. 3. Performance of three different LDPC codes — "DE8", "CDE8"and "CSF14". Code lengths are 2800 and the code rate is 0.82. (a) BER and BLER; (b) average number of iterations to decode a codeword ($\bar{I}$).

times of $(1-R_{\text{code}})$, i.e., $F_2 \approx 2(1-R_{\text{code}})$. We also observe that constrained DE and constrained SF-LDPC produce very similar σ^* and $< k_v >$.

Finally, we study rate-0.82 codes built on (i) constrained SF-degree distribution with a maximum symbol-node degree of 14 (denoted by "CSF14"); (ii) DE-optimized symbol-node degree distribution given by (denoted by "DE8")

$$\lambda_{\text{DE8}}(x) = 0.2343x + 0.3406x^2 + 0.2967x^6 + 0.1284x^7; \quad (13)$$

and (iii) constrained DE-optimized symbol-node degree distribution given by [3], [15] (denoted by "CDE8")

$$\lambda_{\text{CDE8}}(x) = 0.1021x + 0.5895x^2 + 0.1829x^6 + 0.1262x^7. \quad (14)$$

Details of the codes are listed in Table I. The lengths of the three types of codes are also set to be 2800, i.e., $n = 2800$. Figure 3 plots the bit/block error BER/BLER and the average number of iterations to decode a codeword $\bar{I}$ for the three codes. We can observe that the CSF-LDPC code "CSF14" slightly underperforms compared with "DE8" and "CDE8" at low SNR but it achieves the lowest error at higher SNR values among the three codes. Moreover, Fig. 3(b) shows that

TABLE I

COMPARISON OF THE THRESHOLD VALUE AND THE AVERAGE NUMBER OF CONNECTIONS BETWEEN CONSTRAINED SF-LDPC CODES AND OTHER BEST-KNOWN LDPC CODES. CODE RATE R_{code} EQUALS 0.75 AND 0.82. THE LETTERS IN THE CODE NAME DENOTE THE TYPE OF CODE, INCLUDING DE, CONSTRAINED DE (ABBREVIATED BY "CDE") , AND CONSTRAINED SF-LDPC (ABBREVIATED BY "CSF"). THE DIGITS IN THE CODE NAME DENOTE THE MAXIMUM SYMBOL NODE DEGREE OF THE CODE.

Code Rate R_{code}	Code Name	d_v	$F_r(2)$	σ^*	$< k_v >$	d_c	γ
0.75	DE14	14	0.446	0.664	4.526	20	/
	CDE12	12	0.250	0.663	4.000	16	/
	CSF12	12	0.278	0.647	3.974	17	2.38
	CSF20	20	0.280	0.651	4.005	17	2.80
	CSF28	28	0.294	0.653	4.054	17	2.90
0.82	DE8	8	0.405	0.585	3.459	21	/
	CDE8	8	0.180	0.580	3.459	21	/
	CSF10	10	0.176	0.577	3.468	21	3.80
	CSF14	14	0.191	0.579	3.465	21	3.95

TABLE II

COMPARISON OF AVERAGE CONVERGENCE TIMES FOR RATE-0.82 CODES AT HIGH SNR VALUES.

SNR/Code length	6.0 dB/2800	5.8 dB/2800	5.6 dB/2800
Code type	DE8/CDE8/CSF14	DE8/CDE8/CSF14	DE8/CDE8/CSF14
$< k_v >$	3.46/3.46/3.46	3.46/3.46/3.46	3.46/3.46/3.46
I	4.96/4.61/4.55	5.84/5.40/5.32	7.10/6.53/6.42
$t_c = I\times < k_v >$	17.16/15.95/15.74	20.21/18.68/18.41	24.57/22.59/22.21
Normalized t_c	1.09/1.01/1.00	1.10/1.01/1.00	1.11/1.02/1.00

"CSF14" can decode the codewords with the smallest number of iterations on average. We also list the "average convergence time" (t_c) of the three rate-0.82 codes at high SNRs in Table II. The results show that "CDE8" and "CSF14", compared with "DE8", require 10% less time (resources) on average to decode a codeword.

IV. CONCLUSIONS

In this paper, we have proposed a simple algorithm for designing high-rate LDPC codes. With the exception of degree-2 symbol nodes, all the other symbol nodes have their degrees following a power-law distribution. The analytical and simulation results have shown that the proposed constrained SF-LDPC (CSF-LDPC) code can accomplish very similar achievable error performance (threshold), produce lower bit/block error rate at the high SNR region and require a smaller number of iterations for convergence.

ACKNOWLEDGEMENT

The work described in this paper was partially supported by a grant from the RGC of the Hong Kong SAR, China (Project No. PolyU 521809) and by the National Natural Science Foundation of China (Grant No. 60972037).

REFERENCES

[1] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[2] L. Chen, J. Lu, and J. Lu, "Synchronization of the time-varying discrete biological networks," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, May 2007, pp. 2650–2653.

[3] M. Chiani and A. Ventura, "Design and performance evaluation of some high-rate irregular low-density parity-check codes," in *IEEE GLOBECOM*. Texas,USA, 2001, pp. 990–994.

[4] P. Erdos and A. Renyi, "On Random Graphs," *Publ. Math. Debrecen*, vol. 6, no. 290, 1959.

[5] R. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[6] B. Gou, H. Zheng, W. Wu, and X. Yu, "Probability distribution of blackouts in complex power networks," in *Proc. IEEE Int. Symp. Circ. Syst*. New Orleans, USA, 2007, pp. 69–72.

[7] X. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 386–398, 2005.

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

[9] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, 2001.

[10] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[11] C. Shannon and W. Weaver, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, no. 2, pp. 632–656, 1948.

[12] W. Tam, F. Lau, and C. Tse, "Modeling the telephone call network," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, May 2007, pp. 453–456.

[13] ——, "Complex-network modeling of a call network," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 56, no. 2, pp. 416–429, Feb. 2009.

[14] D. Watts and S. Strogatz, "Collective dynamics of small-world networks," *The Structure and Dynamics of Networks*, vol. 393, pp. 440–442, 2006.

[15] M. Yang and W. Li, "Design of efficiently encodable moderate-length high-rate irregular LDPC codes," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 564–571, 2004.

[16] X. Zheng, F. Lau, and C. Tse, "Short-length LDPC Codes with Power-law Distributed Variable-node Degrees," in *Proc. International Symposium on Nonlinear Theory and its Applications*. Budapest, Hungary, Sept. 2008.

[17] X. Zheng, F. Lau, C. Tse, Y. He, and S. Hau, "Application of Complex-Network Theories to the Design of Short-length LDPC Codes," *IET Communications*, vol. 3, no. 10, pp. 1569–1577, Oct. 2009.