Matching Criterion Between Source Statistics and Source Coding Rate

Chen Chen, Lin Wang, *Senior Member, IEEE*, and Zixiang Xiong, *Fellow, IEEE*

***Abstract*—Joint source–channel coding (JSCC) based on double protograph low-density parity-check (DP-LDPC) codes has been shown to be a good solution for wireless communications. However, the premature error floor region of the DP-LDPC codes usually does not meet the bit-error-rate requirements. In this letter, we provide a matching criterion between the source statistics and the source coding rate using a novel protograph extrinsic information transfer (PEXIT) algorithm with a new decoding threshold. The resulting matching criterion helps predict the error floor region performance and give directions for source coding rate designs.**

***Index Terms*—Matching criterion, error floor region, PEXIT.**

I. Introduction

THE traditional tandem coding structure that implements source and channel coding separately is known to be suboptimal for finite-block length coding in practical communications, because the residual redundancy left by the source coding cannot be exploited by the channel decoder [1]. Therefore, many researchers have studied joint source-channel coding (JSCC) schemes [2], [3]. Among the different source-channel coding pairs used in JSCC, low-density parity-check (LDPC) codes [4] as both source and channel codes are widely used due to their outstanding error correction performance.

Owing to the cast encoding structure and their high decoding speed, protograph LDPC (P-LDPC) codes [5] were introduced as good candidates for JSCC systems in [6]. The issue of source statistics, which affects the performance of double LDPC (D-LDPC) codes, was raised in [6], and in [7], the source statistics was taken into account in a novel JSCC framework based on DP-LDPC codes for transmitting radiography images. Further, using unequal error protection (UEP) in the DP-LDPC systems was shown to give outstanding performance in transmitting radiography images [8]. It was also observed in [9] that the source statistics plays a more important role than diversity orders in JSCC systems and that the main components of JSCC (e.g., the source statistics, source coding rate, channel coding rate, and channel conditions) need to be matched internally to achieve good performance. A scheme that can effectively improve the system performance by allocating coding rates according to the source statistics and channel conditions was proposed in [10]. However, the simulation-based scheme in [10] is ad hoc in nature.

Manuscript received January 7, 2015; revised June 16, 2015; accepted June 27, 2015. Date of publication July 9, 2015; date of current version September 4, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61271241, by the European Union FP7 through CoNHealth under Grant 294923, by the Shenzhen City Science and Technology Project under Grant JC201006030862A, and by the National Science Foundation under Grant 1216001. The associate editor coordinating the review of this paper and approving it for publication was E. Paolini.

C. Chen and L. Wang are with the Shenzhen Research Institute, Xiamen University, Shenzhen 518000, China, and also with the Department of Communication Engineering, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: wanglin@xmu.edu.cn).

Z. Xiong is with the Department of Electrical and Computer Engineering, College of Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zx@ece.tamu.edu).

Digital Object Identifier 10.1109/LCOMM.2015.2454505

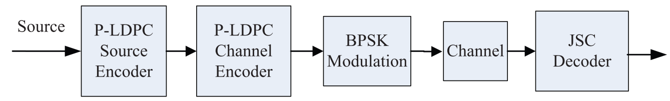

Fig. 1. System model of JSCC system with DP-LDPC codes.

In this work, we analytically study the criterion for matching the source statistics with source coding rate using an innovative PEXIT chart technique. EXIT chart [11] is an accurate and convenient tool for predicting the performance of LDPC codes. It was introduced into D-LDPC code systems in [1] and extended to P-LDPC codes in [12] as PEXIT chart. Most existing works focus on employing this tool to analyze the decoding threshold of the water-fall region and to design capacity-approaching LDPC codes for channel coding. In contrast, we extend PEXIT in [12] to error-floor performance analysis. Specifically, we look into the matching criterion between the source statistics and source coding rate to provide guidance for source coding rate designing and help reach an error floor that is below 10^{-6}. This is done by proposing a new PEXIT algorithm that predicts the convergence of the source code by using the source entropy (instead of the SNR) as the main variable in the algorithm.

Section II gives the background of JSCC system based on DP-LDPC codes. Section III details our proposed PEXIT analysis. Simulation results are shown and discussed in Section IV. Section V concludes the paper.

II. System Model

The general transmission system model of JSCC based on DP-LDPC codes is the same as [6], which is depicted in Fig. 1. At the transmitter, we consider a binary i.i.d. Bernouli (p) source with entropy

$$H = -p \log_2 p - (1 - p) \log_2 (1 - p), \qquad (1)$$

where $p < 1/2$. The source is first compressed by an unpuncture P-LDPC code, and then protected by another puncture P-LDPC code. The source-channel encoded source is then BPSK modulated before transmission over an AWGN channel. At the receiver, the source-channel decoder shown in Fig. 2 is employed to reconstruct the source.

In this work, three different coding rates of repeat-by-4-jagged-acccumulate (R4JA) code and one rate-1/2 accumulate-repeat-by-3-accumulate (AR3A) code are utilized for source-channel coding [13], respectively. The protographs of these two codes (with rate 1/2) are shown in Fig. 3. Blank circles denote the punctured variable nodes, which are not

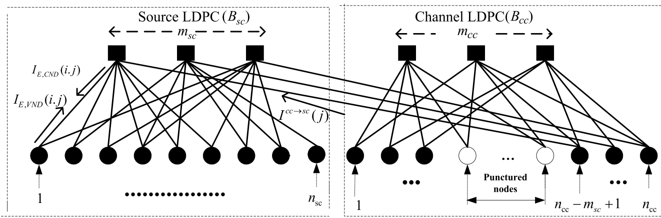

Fig. 2. Joint decoder of JSCC system with mutual information updating.

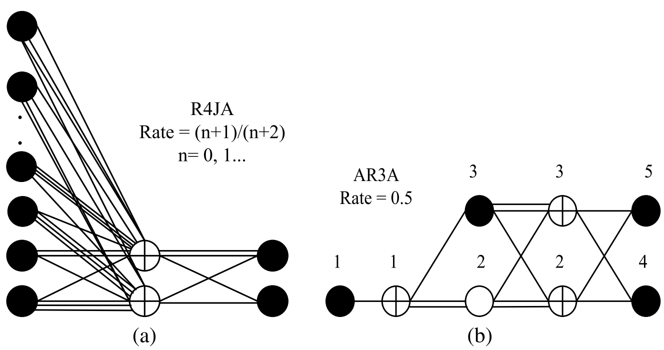

Fig. 3. The protographs of (a) an R4JA code and (b) an AR3A code.

transmitted through the channel. The transmitted variable nodes are shown with dark circles, and the check nodes are those circles with plus sign inside.

III. Proposed PEXIT Analysis

A. Background

In PEXIT chart, the space between the inner and outer coder curves is defined as the decoding tunnel. Only when the two PEXIT curves do not touch or cross each other except at the value of unity can the decoder converge successfully. Therefore, a critical state occurs when the outer coder curve is tangent to the inner coder curve. In our proposed PEXIT algorithm, both the inner and outer coder curves are related to the source coding rate, but the inner coder curve only depends on the source entropy. So, at a fixed source coding rate, adjusting source entropy can help make the inner coder curve tangent to the outer coder curve. At this point, the parameter p of the source is considered to be a critical value and denoted as p^{th} for the specific source coding rate, and the corresponding source entropy, denoted as H^{th}, represents the decoding threshold.

We point out that H^{th} in our proposed algorithm has new meaning. In conventional PEXIT analysis, the decoding threshold represents the smallest SNR above which an arbitrarily small BER can be achieved. In our proposed PEXIT algorithm, it represents the highest entropy below which the corresponding source can be reconstructed at the source decoder as the code length approaches infinity. In addition, the decoding threshold H^{th} is a function of the source coding rate.

Although the source is theoretically Bernouli (p), the actual p in each finite-length frame has a small degree of variations, which leads to the variation of the inner coder curve. Since any performance analysis should consider the worst case scenario in accordance with the required confidence level, an extra PEXIT curve representing the lower bound of inner coder curves is added to represent the worst case. The threshold derived from the worst case is considered to be the decoding threshold in a strict sense and denoted as H^{th_strict}. Consequently, when the given source entropy is less than H^{th_strict}, the source can still be successfully reconstructed by the source decoder even the worst case occurs. Therefore,

$$H < H^{th\_strict} \tag{2}$$

can be viewed as the matching criterion between the source entropy and source coding rate. Similarly, the decoding threshold in the general case is denoted as $H^{th\_general}$, which means a source with entropy higher than $H^{th\_general}$ cannot be reconstructed by the source decoder.

B. PEXIT Algorithm of Protograph LDPC Codes for the Source Decoder

Consider a source code with base matrix B_{sc} of dimension $m_{\text{sc}} \times n_{\text{sc}}$ and channel code with base matrix B_{cc} of dimension $m_{\text{cc}} \times n_{\text{cc}}$. The (i, j)-th element of B_{sc}, denoted as $b_{\text{sc}}^{(i,j)}$, represents the number of edges connecting the i-th variable node to the j-th check node in B_{sc}. To establish a correspondence relationship with source statistics, we introduce a "virtual" binary symmetric channel (BSC) with crossover probability p [1].

Five types of mutual information (MI) are defined as follows:

1) $I_{\text{E,VND}}(i, j)$: the extrinsic mutual information from the j-th variable node to the i-th check node in B_{sc}.
2) $I_{\text{A,VND}}(i, j)$: the extrinsic mutual information from the i-th check node to the j-th variable node in B_{sc}.
3) $I_{\text{E,CND}}(i, j)$: the extrinsic mutual information from the i-th check node to the j-th variable node in B_{sc}.
4) $I_{\text{A,CND}}(i, j)$: the extrinsic mutual information from the j-th variable node to the i-th check node in B_{sc}.
5) $I^{cc\to sc}(i)$: the extrinsic mutual information from the i-th variable node in B_{cc} to the connected check node in B_{sc}. As we only focus on the performance of error-floor, channel decoding is assumed to be error free, and $I^{cc\to sc}(i)$ is set to 1 for $i = 1, 2, \ldots, m_{\text{sc}}$.

1) Initiation: For a given probability p, a sequence of non-uniform memoryless information bits is generated and then the actual success probability of this sequence is calculated and denoted as $\hat{p}$. This step is repeated N times to obtain a set of $\{\hat{p}\}$'s. The pdf of the LLR information from the virtual channel can be calculated as

$$f(\psi) = \hat{p} \times \delta(\psi + L) + (1 - \hat{p}) \times \delta(\psi - L), \tag{3}$$

$$L = \ln\left((1 - \hat{p})/\hat{p}\right), \tag{4}$$

where L is the LLR information of source. Then the function $J_{\text{BSC}}(\cdot)$ with the prior probability from the source can be expressed as

$$J_{\text{BSC}}(\mu, \hat{p}) = (1 - \hat{p}) \times I\left(V; \chi^{(1-\hat{p})}\right) + \hat{p} \times I(V; \chi^{\hat{p}}), \tag{5}$$

where $\chi^{(1-\hat{p})} \sim N(\mu + L, 2\mu)$, $\chi^{(\hat{p})} \sim N(\mu - L, 2\mu)$ and $I(V; \chi)$ denotes the mutual information between the variable node of the source and χ.

2) Inner Coder—Variable Nodes to Check Nodes Updating: For a given $I_{\mathrm{A,VND}} \in [0, 1]$ and $\{\hat{p}\}$, the sequence $I_{\mathrm{E,VND}}$ is measured as below.

For $j = 1, \ldots, n_{\mathrm{sc}}$ and $i = 1, \ldots, m_{\mathrm{sc}}$, if $b_{\mathrm{sc}}^{(i,j)} \neq 0$,

$$I_{\mathrm{E,VND}}(i, j) = J_{\mathrm{BSC}} \left(\sum_{s \neq i} b_{\mathrm{sc}}^{(s,j)} J^{-1} \left(I_{\mathrm{A,VND}}(s, j) \right) + \left(b_{\mathrm{sc}}^{(i,j)} - 1 \right) J^{-1} \left(I_{\mathrm{A,VND}}(i, j) \right), \hat{p} \right). \quad (6)$$

If $b_{\mathrm{sc}}^{(i,j)} = 0$, $I_{\mathrm{E,VND}}(i, j) = 0$,

$$I_{\mathrm{E,VND}} = \sum_j \sum_i I_{\mathrm{E,VND}}(i, j) \times b_{\mathrm{sc}}^{(i,j)} / \sum_j \sum_i b_{\mathrm{sc}}^{(i,j)}. \quad (7)$$

3) $E[I_{E,VND}]$ and $min(I_{E,VND})$ Computation: The expected value of $I_{\mathrm{E,VND}}$ is calculated as $E[I_{\mathrm{E,VND}}]$, which can be considered as a typical value of all the frames. We compute the minimal value of $I_{\mathrm{E,VND}}$, denoted as $min(I_{\mathrm{E,VND}})$, and use it to represent the value of the worst case.

4) Outer Coder— Check Nodes to Variable Nodes Updating: For $j = 1, \ldots, n_{\mathrm{sc}}$ and $i = 1, \ldots, m_{\mathrm{sc}}$, if $b_{\mathrm{sc}}^{(i,j)} \neq 0$,

$$I_{\mathrm{E,CND}}(i, j) = 1 - J \left(\sum_{s \neq i} b_{\mathrm{sc}}^{(i,s)} J^{-1} \left(1 - I_{\mathrm{A,VND}}(i, s) \right) + \left(b_{\mathrm{sc}}^{(i,j)} - 1 \right) J^{-1} \left(1 - I_{\mathrm{A,CND}}(i, j) \right) + J^{-1} \left(1 - I^{cc \rightarrow sc}(i) \right) \right). \quad (8)$$

If $b_{\mathrm{sc}}^{(i,j)} = 0$, $I_{\mathrm{E,CND}}(i, j) = 0$,

$$I_{\mathrm{E,CND}} = \sum_j \sum_i I_{\mathrm{E,CND}}(i, j) \times b_{\mathrm{sc}}^{(i,j)} / \sum_j \sum_i b_{\mathrm{sc}}^{(i,j)}. \quad (9)$$

5) PEXIT Curves Description: Repeat Step 2 to Step 3 for different values of $I_{\mathrm{A,VND}} \in [0, 1]$. Then two inner coder curves, namely, the upper bound for the expected PEXIT curve and the lower bound for the worst case curve, can be obtained. Similarly, repeat Step 4 for different values of $I_{\mathrm{A,CND}} \in [0, 1]$ to obtain the corresponding outer coder curves.

6) Decoding Thresholds Estimation: Continue adjusting p to vary the inner coder curves until the lower bound of inner coder curve becomes tangent to the outer coder curve. The success probability in this case is denoted as p^{th_strict} and the corresponding source entropy is the decoding threshold of the worst case, denoted as H^{th_strict}. Moreover, $p^{th\_general}$ and $H^{th\_general}$ can be obtained when the outer coder curves are tangent lines of the upper bounds.

IV. Simulation Results and Discussion

First, the decoding thresholds of three different coding rates are given by means of the proposed PEXIT algorithm. Then, experiments with source entropies under three different statuses

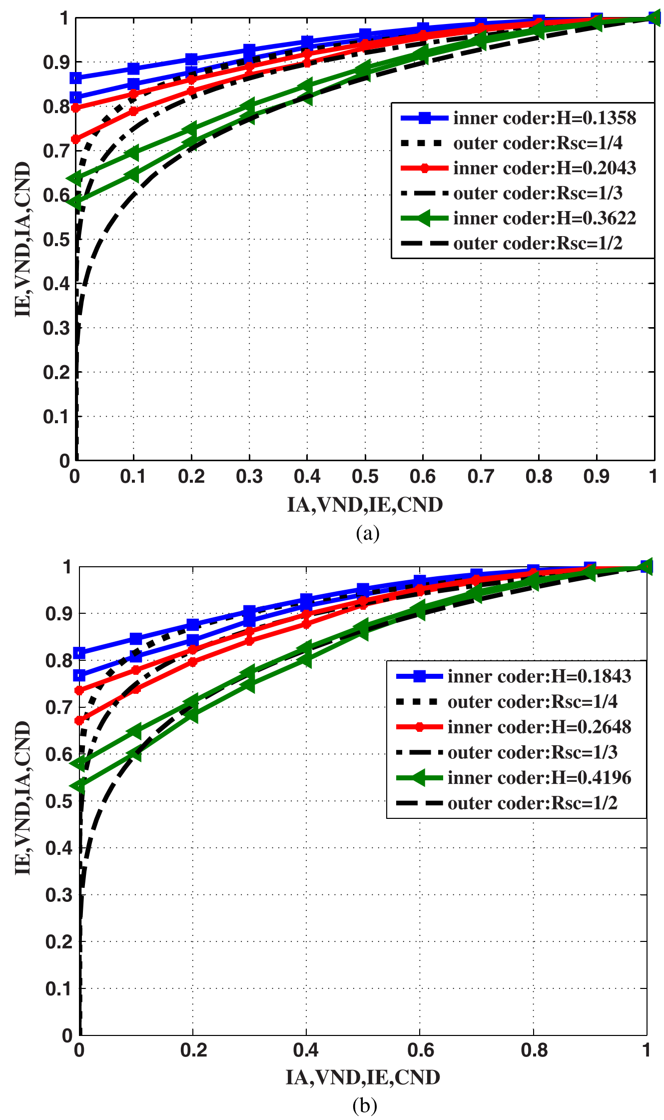

Fig. 4. PEXIT curves for inner coders are shown by solid lines, including upper-bound curves and lower-bound curves, outer coder PEXIT curves are represented by dotted lines.

(compared with decoding thresholds derived from PEXIT analysis) are performed at corresponding source coding rates. In all experiments, an AWGN channel is assumed and the frame length fixed at 3200 bits. To ensure accuracy, N is set to 10000. Since the mutual information values for protograph LDPC codes are not one-dimensional, we calculate the weighted averages of the mutual information values as (7) and (9) to obtain the one-dimensional EXIT charts. Finally, we plot the PEXIT curves in Fig. 4 (with axes switched for the outer decoder curves).

In the first experiment, using our proposed PEXIT algorithm, we obtain the decoding thresholds of R4JA codes with three different coding rates in Table I. In the second experiment, we simulate the DP-LDPC system with sources generated by different p's, which result in different entropies. For every coding rate, four different source entropies are considered. Compare with the thresholds in Table I, the sources can be classified into three types as shown in Table II. Sources with the same entropy, e.g., "Src3" and "Src6", can be in different classes.

TABLE I
DECODING THRESHOLDS FOR ERROR-FLOOR REGION WITH THREE DIFFERENT SOURCE CODING RATES

R_{sc}	p^{th_strict}	$p^{th\_general}$	H^{th_strict}	$H^{th\_general}$
1/4	0.019	0.028	0.1358	0.1843
1/3	0.032	0.045	0.2043	0.2648
1/2	0.069	0.085	0.3622	0.4196

TABLE II
PARAMETERS OF SOURCES WITH DIFFERENT ENTROPIES

R_{sc}	Source	p	H	Status
1/4	Src1	0.01	0.0808	$H < H^{th\_strict}$
	Src2	0.015	0.1123	$H < H^{th\_strict}$
	Src3	0.02	0.1414	$H^{th\_strict} < H < H^{th\_general}$
	Src4	0.03	0.1944	$H > H^{th\_general}$
1/3	Src5	0.01	0.0808	$H < H^{th\_strict}$
	Src6	0.02	0.1414	$H < H^{th\_strict}$
	Src7	0.04	0.2423	$H^{th\_strict} < H < H^{th\_general}$
	Src8	0.05	0.2864	$H > H^{th\_general}$
1/2	Src9	0.04	0.2423	$H < H^{th\_strict}$
	Src10	0.06	0.3274	$H < H^{th\_strict}$
	Src7	0.07	0.3659	$H^{th\_strict} < H < H^{th\_general}$
	Src8	0.09	0.4365	$H > H^{th\_general}$

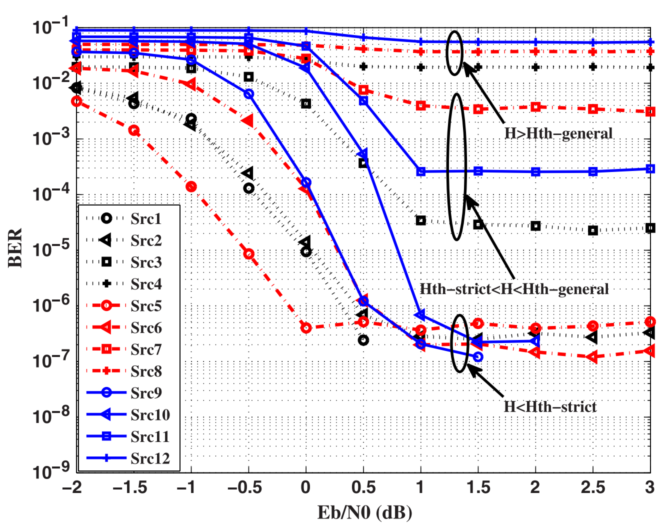

Fig. 5. Simulated BER results of sources under three different statuses.

Fig. 5 plots simulation results of DP-LDPC system with sources specified in Table II. It is seen that the source class information can be utilized to predict the error floor performance. In Fig. 5, the error-floor performance can also be classified into three categories, which are aligned with the three source classes. In particular, an error-floor BER below 10^{-6} can be reached for sources that satisfy the matching criterion in (2). Finally, under this condition, further reducing the source entropy or compression ratio of source coding does not lower the error floor. Thus, (2) reveals that the source entropy and source coding rate are well matched, in accordance with our PEXIT analysis. We thus conclude that 1) Under specific source coding rate, a decoding threshold for the source decoder exists and it can be employed to analyze the error-floor region performance and derive the matching criterion in (2), and 2) the matching criterion in (2) provides guidance for designing proper source coding rates according to the source statistics. By taking the source entropy into account, priorities should be given to those coding rates which satisfy (2).

V. CONCLUSION

In this letter, we have derived a matching criterion between the source entropy and source coding rate by using a new PEXIT algorithm, which describes the asymptotic decoding trajectory of the source decoder. The resulting matching criterion can provide theoretical guidance for designing proper source coding rates to meet the quality demand of JSCC-based wireless communication systems and reach an error floor BER that is below 10^{-6}.

REFERENCES

[1] M. Fresia, F. Perez-Cruz, H. V. Poor, and S. Verdu, “Joint source and channel coding,” *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 104–113, Nov. 2010.

[2] L. Pu, Z. W. Wu, A. Bilgin, M. W. Marcellin, and B. Vasic, “LDPC-based iterative joint source-channel decoding for JPEG2000,” *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 577–581, Feb. 2007.

[3] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, “Joint source and channel turbo decoding of entropy-coded sources,” *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1680–1696, Sep. 2001.

[4] R. G. Gallager, “Low-density parity-check codes,” *IRE Trans. Inf. Theory*, vol. 8, pp. 21–28, Jan. 1962.

[5] J. Thorpe, “Low-density parity-check (LDPC) codes constructed from protographs,” Nat. Aeronaut. Space Admin. (NASA), Washington, DC, USA, IPN Progress Rep., pp. 42–154, Aug. 2003.

[6] J. G. He, L. Wang, and P. P. Chen, “A joint source and channel coding scheme base on simple protograph structured codes,” in *Proc. IEEE ISCIT*, Oct. 2012, pp. 65–69.

[7] H. H. Wu, J. G. He, L. L. Xu, and L. Wang, “Joint source-channel coding based on P-LDPC codes for radiograph images transmission,” in *Proc. IEEE Conf. IUCC*, Jun. 2012, pp. 2035–2039.

[8] L. L. Xu, H. Wu, J. G. He, and L. Wang, “Unequal error protection for radiography image transmission using protograph double LDPC codes,” in *Proc. WTS*, Apr. 2013, pp. 1–5.

[9] H. H. Wu, L. Wang, S. H. Hong, and J. G. He, “Performance of joint source-channel coding based on protograph LDPC codes over Rayleigh fading channels,” *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 652–655, Apr. 2014.

[10] C. Chen, L. Wang, and Z. H. Jiang, “Adaptive rate allocation scheme for joint source-channel coding based on double protograph LDPC codes,” in *Proc. Conf. WPMC*, Sep. 2014, pp. 158–162.

[11] S. ten Brink, G. Kramer, and A. Ashikhmin, “Design of low-density parity-check codes for modulation and detection,” *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[12] G. Liva and M. Chiani, “Protograph LDPC codes design based on EXIT analysis,” in *Proc. IEEE GLOBECOM*, 2007, pp. 3250–3254.

[13] D. Divsalar, S. Dolinar, C. R. Jones, and K. Andrews, “Capacity-approaching protograph codes,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 876–888, Sep. 2001.