Randomized Serially Concatenated LDGM Codes for the Gaussian Wiretap Channel

Alireza Nooraiepour and Tolga M. Duman, *Fellow, IEEE*

***Abstract*— We study the application of a special class of low-density parity-check codes to the wiretap channel. We construct a randomized coding scheme based on serially concatenated low-density generator matrix codes and their duals extending the approach used for convolutional and turbo codes. Furthermore, we propose an efficient iterative decoder for this scheme utilizing a joint iterative message passing algorithm. We demonstrate via numerical examples that this approach outperforms other available practical coding alternatives for the Gaussian wiretap channel in terms of the resulting security gap.**

***Index Terms*— Low density parity check (LDPC) codes, low density generator matrix (LDGM) codes, Gaussian wiretap channel, security gap.**

I. Introduction

WYNER [1] introduced the wiretap channel for studying physical layer security. In this model, a transmitter (Alice) wishes to transmit its message to a legitimate receiver (Bob) through the main channel while preventing an eavesdropper (Eve) to extract any information about it in an information theoretic sense.

In addition to the classical information theoretic metrics for evaluating the security performance, researchers have also developed alternative metrics for designing practical channel codes for the wiretap channel. Specifically, for the case of additive white Gaussian noise (AWGN) channels a difference between the qualities of the main and the eavesdropper's channels is needed to achieve a positive secrecy capacity [7], which is dubbed as the *security gap*. The security gap is directly extracted from the bit error rate (BER) curves by setting two values: 1) the maximum desired BER at Bob (P_{main}^{max}), and 2) the minimum desired BER at Eve (P_{eve}^{min}) [3]. These values correspond to two different signal to noise ratios (SNRs) measured in dBs, difference of which equals to the security gap. We consider positive security gaps in this letter, however, channels with feedback can result in negative security gaps as well [5]. We also emphasize that, although the security gap enables us to numerically evaluate and compare the performances of practical coding schemes, it provides only a necessary condition for achieving information-theoretic secrecy.

The literature of physical layer security includes several practical coding approaches designed to reduce the security gap. Klinc *et al.* [3] utilize punctured low density parity check (LDPC) codes for this purpose, while Baldi *et al.* [4] consider a scrambling method along with non-systematic codes. Extending the latter idea, Baldi *et al.* [5] apply scrambling, concatenation, and hybrid automatic repeat-request to different codes to further reduce the security gap. Furthermore, concatenation of polar and LDPC codes is studied in [6]. And, more recently, Nooraiepour and Duman [7] construct randomized codes using convolutional and turbo codes for the wiretap channel.

Application of LDPC codes to the wiretap channel was first investigated in [2] where it was proved that when the main channel is noiseless and the eavesdropper's channel is a binary erasure channel (BEC), using dual of an LDPC code and its cosets can achieve the secrecy capacity. Specifically, [2] aims at designing schemes based on LDPC codes, which satisfy (information-theoretic) security over BECs. However, considering the Gaussian wiretap channel, there is no practical decoder when the LDPC codes are applied to the randomized scheme. One of the reasons is that the dual of an LDPC code does not necessarily have a sparse parity check matrix. With this motivation, we propose the use of a special class of LDPC codes, namely the systematic low density generator matrix (LDGM) codes in the randomized coset coding scheme. The structure of these codes makes them desirable for the present set-up and enables us to propose an efficient decoder. We note that the systematic LDGM codes suffer from high error floors. In order to combat this problem, we specifically consider the use of serially concatenated LDGM (SCLDGM) codes proposed in [9]. We further concatenate SCLDGM with a convolutional code to be able to obtain a practical iterative decoding scheme for the randomized encoding set-up. Extensive simulation results show that our proposed scheme outperforms the other existing alternatives in terms of security gaps.

The rest of the letter is organized as follows: the preliminaries are introduced in Section II. The structure of LDGM codes used in this letter, and the SCLDGM codes are presented in Section III. The proposal of randomized SCLDGM codes is described in Section IV along with a detailed overview of the corresponding iterative decoder. Numerical examples and comparisons with the existing state of the art coding schemes are provided in Section V, and finally, the letter is concluded in Section VI.

II. Preliminaries

A. Channel Model

We assume that both the main and eavesdropper's channels are AWGN expressed as $y = x + N$ where $x = (-1)^c$ is the binary phase-shift keying (BPSK) modulated version of the transmitted bit c. N represents Gaussian noise with zero mean and variance $N_0/2$ where different noise components are assumed to be independent and identically distributed (i.i.d.).

Manuscript received October 8, 2017; revised November 21, 2017 and December 12, 2017; accepted December 18, 2017. Date of publication January 9, 2018; date of current version April 7, 2018. The associate editor coordinating the review of this paper and approving it for publication was M. Baldi. *(Corresponding author: Tolga M. Duman.)*

A. Nooraiepour is with the Electrical and Computer Engineering Department, Rutgers University, New Brunswick, NJ 08854 USA (e-mail: alireza.nooraiepour@rutgers.edu).

T. M. Duman is with the Department of Electrical and Electronics Engineering, Bilkent University, TR-06800 Ankara, Turkey (e-mail: duman@ee.bilkent.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2018.2789346

1558-2558 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

We have $E_b = 1/R$ with E_b denoting the energy per bit and R denoting the transmission rate. We define the SNR as E_b/N_0.

B. The Randomized Coding Scheme for Secrecy

In this method, each message gets mapped to a unique coset of a certain code. To transmit a k-bit message, 2^k cosets are used. Assuming that each coset consists of 2^r codewords of length n, in order to cover all the codewords, a linear code of length n and dimension at least $k + r$ (which we call the *big code*) is used. We refer to the coset corresponding to the all-zero message as the *small code* denoted by $\mathcal{C}$ with the generator matrix $\mathbf{G}$ of size $r \times n$.

A secret message denoted by *data bits* $\mathbf{s} = [s_1, s_2, ..., s_k]$ is mapped to the coset obtained by adding the n-tuple $s_1\mathbf{h}_1 + s_2\mathbf{h}_2 + ... + s_k\mathbf{h}_k$ to all the codewords in $\mathcal{C}$ where $\mathbf{h}_i$'s are linearly independent n-tuples outside $\mathcal{C}$. Assuming $\mathbf{h}_i$'s form a matrix $\mathbf{H}$ of size $k \times n$, the transmitted codeword through the channel is

$$\mathbf{c} = \begin{bmatrix}\mathbf{s} & \mathbf{v}\end{bmatrix}\begin{bmatrix}\mathbf{H}\\ \mathbf{G}\end{bmatrix}, \tag{1}$$

where $\mathbf{v} = [v_1, v_2, \ldots, v_r]$ denotes the random bit vector used for choosing a codeword in the corresponding coset uniformly randomly. In other words, the randomized encoding scheme needs two sets of generators: one for the random bits and the other for the data bits.

One can easily obtain a $\mathbf{G}$ matrix by choosing a linear code as the small code $\mathcal{C}$, however, finding the $\mathbf{h}_i$ vectors requires an exhaustive search, which may not be practical. As a remedy, in [7] and [8], it is proved that one can use generators of dual of the small code $\mathcal{C}^{\perp}$ as $\mathbf{h}_i$'s if $\mathcal{C}^{\perp}$ is not *pseudo-self-dual*.[1]

C. Decoding Rule

The maximum a posteriori probability (MAP) decoder for the randomized encoding scheme can be written explicitly. Specifically, for an AWGN channel and equiprobable cosets, the decoding rule is given by [7]

$$\hat{i} = \underset{i=1,2,...,M}{\operatorname{argmax}} \sum_{j=1}^{N} e^{\frac{-\|\mathbf{y}-\mathbf{c}_{ji}\|^2}{N_0}}, \tag{2}$$

where M and N denote the number of cosets and the number of codewords within each coset, respectively. We note that the optimal MAP decoder is different through the main and the eavesdropper's channels since they have different noise variances. A direct observation from (2) is that this decoder is prohibitively complex ruling out its implementation in practice when M and/or N is large.

III. LDGM and SCLDGM Codes

The main purpose of this letter is to apply LDGM codes to the randomized coset coding scheme, which has not been explored in the literature. Our main motivation is that using systematic LDGM codes enables us to obtain an efficient iterative decoder for the randomized coding scheme, which is not available for the case of standard LDPC codes in general. This is partly due to the fact that belief propagation (BP) algorithm is applicable to such codes and their duals (which are needed for the randomized coding scheme as illustrated in Section II-B) in an efficient manner owing to the fact that the parity check matrices for both are sparse.

The LDGM codes used in this letter belong to a family of quasi-cyclic LDPC (QC-LDPC) codes whose parity-check and generator matrices are formed by circulant blocks [10]. For a given code length N, the parity-check matrices of these codes consist of a row of sparse circulant blocks, i.e.,

$$\mathbf{H} = [\mathbf{H}_0|\mathbf{H}_1|\ldots|\mathbf{H}_{N_b-1}], \tag{3}$$

where each $\mathbf{H}_i$ is of size $(N/N_b) \times (N/N_b)$. A circulant matrix $\mathbf{H}_i$ is defined over the Galois Field of order $p, GF(p)$, i.e.,

$$\mathbf{H}_i = \begin{bmatrix} h_0 & h_1 & h_2 & \ldots & h_{n-1} \\ h_{n-1} & h_0 & h_1 & \ldots & h_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_1 & h_2 & h_3 & \ldots & h_0 \end{bmatrix}, \tag{4}$$

where $h_i \in GF(p)$, $i = 0, 1, \ldots, n-1$. In fact, $\mathbf{H}_i$ is described by one of its rows and the other rows are obtained by cyclical shifts. We denote the number of nonzero symbols in each row (or column) of $\mathbf{H}_i$ (weight of the vector) by $W[\mathbf{H}_i]$.

Provided that at least one of the circulant blocks is full rank, the code rate is $(N_b - 1)/N_b$. Furthermore, the corresponding generator matrices for such codes can be found easily if one of the $\mathbf{H}_i$'s is an identity matrix. Specifically, we assume that $\mathbf{H}_{N_b-1} = \mathbf{I}_{(N/N_b \times N/N_b)}$ is used, resulting in a generator matrix of the systematic code as

$$\mathbf{G} = \begin{bmatrix} & \mathbf{H}_0^T \\ & \mathbf{H}_1^T \\ \mathbf{I}_{(N(N_b-1)/N_b \times N(N_b-1)/N_b)} & \vdots \\ & \mathbf{H}_{N_b-2}^T \\ & \mathbf{H}_{N_b-1}^T \end{bmatrix}, \tag{5}$$

where $\mathbf{I}$ denotes the identity matrix with the specified dimensions and superscript ^{T} denotes the transpose operator.

For systematic LDGM codes, some of the variable nodes have degree one and do not help in the BP decoding. As a result, the LDGM codes are known to have high error floors. Gonzalez-Lopez *et al.* [9] propose a concatenated scheme (called SCLDGM) formed by an outer high-rate LDGM code followed by an inner low-rate one to improve the resulting performance.

IV. Randomized SCLDGM (RSCLDGM) Codes for Physical Layer Security

A. Encoding

As pointed out in Section II-B, a linear block code and its dual can be used to construct a randomized coding scheme. In this section, we propose concatenation of an SCLDGM code with a recursive systematic convolutional (RSC) code as the small code as shown in Fig. 1. Here $\mathbf{G}_i$ denotes the generator matrix of the code corresponding to the ith block. A subset of dual of this code can be obtained by replacing each individual block with its dual as illustrated in Fig. 2, where $\mathbf{H}_i$ denotes the generator matrix for the ith block.

We note that the dual of LDGM codes can be obtained easily since they are systematic. Furthermore, one can use the approach described in [7] in order to obtain the dual of an RSC code. In the proposed setup, $\mathbf{G}_i(\mathbf{H}_i)^T = \mathbf{0}$ for

[1] A linear code $\mathcal{C}(n, r)$ with generator matrix $\mathbf{G}$ is defined as *pseudo-self-dual* if $\mathbf{G}\mathbf{G}^T$ is rank-deficient [7].

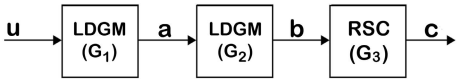

Fig. 1. The SCLDGM code concatenated with an RSC code.

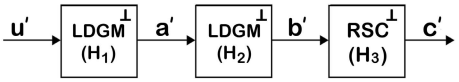

Fig. 2. The encoder for a subset of the dual of the code in Fig. 1.

$i = 1, 2, 3$, and the transmitted codeword through the channel is the modulo-2 sum $\mathbf{c} + \mathbf{c}'$ where $\mathbf{c}$ and $\mathbf{c}'$ are depicted in Figs. 1 and 2, respectively. We use one of the encoders in Figs. 1 and 2 for encoding the random bits and the other for the data bits to obtain the overall randomized encoding scheme. We emphasize that the RSC blocks are utilized to make the decoding process implementable and practical as will be clarified in next.

B. Decoding

In this section, we devise an iterative decoder for the randomized SCLDGM codes, which jointly decodes u_i's and u_i''s [2] given the received noisy vector $\mathbf{y}$. We note that given c_j and c_{j+1} (assuming RSC codes are of rate $1/2$) one can obtain the conditional probabilities[3] $P(b_j = 0|\mathbf{y})$ and $P(b_j = 1|\mathbf{y})$ using the BCJR algorithm and the trellis of the RSC code. Similarly, $P(b_j' = 0|\mathbf{y})$ and $P(b_j' = 1|\mathbf{y})$ are obtained using c_j' and c_{j+1}' and the trellis of RSC$^{\perp}$ code. Therefore, it is possible to obtain a vector $\mathcal{P} = [P_{00}, P_{01}, P_{10}, P_{11}]$ where each $P_{xz} = P(b_j = x, b_j' = z|\mathbf{y})$ can be computed using the BCJR algorithm on the *big trellis* obtained by combining the trellises corresponding to the RSC code and its dual RSC$^{\perp}$. Note that $\mathbf{y}$ is a noisy version of the transmitted codeword (i.e., $\mathbf{c} + \mathbf{c}'$).

Next we utilize the vector $\mathcal{P}$ in the joint BP decoding method introduced in [11] to obtain the extrinsic information on a_j's and a_j''s in Figs. 1 and 2. The decoder is composed of two LDPC graphs connected through *state nodes* as shown in Fig. 3. The state nodes receive information on the pair of bits $(P_{b_j b_j'})$ from $\mathcal{P}$ and exchange information with the two individual LDPC decoders. In the figure, m_{ab}^k denotes the message passed from node a to node b for the k'th Tanner graph (where Tanner graph corresponding to $\mathbf{H_2}$ is considered to be the first one) assuming $a, b \in \{v, c, s\}$ where v, c and s denote a variable node, a check node and a state node, respectively. m_{vc}^k, m_{cv}^k, m_{vs}^k are calculated using conventional BP. The only remaining messages m_{sv}^k's, which are passed from the state nodes to the variable nodes, can be computed as

$$\begin{cases} m_{sv}^1 = \log \dfrac{P_{00} e^{m_{vs}^2} + P_{01}}{P_{10} e^{m_{vs}^2} + P_{11}}, \\ m_{sv}^2 = \log \dfrac{P_{00} e^{m_{vs}^1} + P_{10}}{P_{01} e^{m_{vs}^1} + P_{11}}. \end{cases} \tag{6}$$

These equations basically compute the log-likelihood ratio (LLR) of the form $\log\left(\frac{P(b_j=0|\mathbf{y})}{P(b_j=1|\mathbf{y})}\right)$ for a specific variable node b_j in the Tanner graph of an LDPC code given the vector $\mathcal{P}$ and LLR of the corresponding variable node b_j'.

[2]Throughout this section we denote the i'th bit of a sample vector $\mathbf{v}$ with v_i where $\mathbf{v}$ is one of the vectors introduced in Figs. 1 and 2.

[3]We denote probability with P and probability density with p in this section.

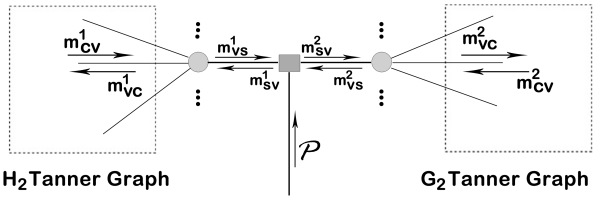

Fig. 3. Joint BP decoder on the factor graphs corresponding to $\mathbf{G_2}$ and $\mathbf{H_2}$.

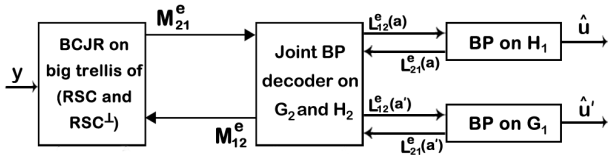

Fig. 4. Schematic of the overall decoder for the RSCLDGM scheme.

The joint BP decoder based on $\mathbf{H_1}$ and $\mathbf{G_1}$ provides us with the extrinsic information on a_j's and a_j''s, which can be used to obtain the LLR's of u_i's and u_i''s, respectively. Fig. 4 illustrates the overall decoding algorithm where extrinsic information terms exchanged between the component decoders are also shown. For example, $\mathbf{M}_{21}^e$ is the vector $\mathcal{P}$ introduced earlier, and $L^e(a)$ denotes the extrinsic information on the bit a computed using the well-known BP expressions. We take $\mathbf{M}_{12}^e$ to be of the form $[P_{00}, P_{01}, P_{10}, P_{11}]$ with $P_{xz} = P(b_j = x|\mathbf{y})P(b_j' = z|\mathbf{y})$. We note that each iteration begins with passing the vector $\mathcal{P} = \mathbf{M}_{21}^e$ to the joint BP decoder. Then individual BP is ran for $\mathbf{H_1}$ and $\mathbf{G_1}$, which in part produces extrinsic information for the joint decoder ($L_{21}^e(a)$ and $L_{21}^e(a')$). Finally, an iteration is completed by sending $\mathbf{M}_{12}^e$ to the BCJR decoder. In practice, one can improve the decoding performance by increasing the number of iterations at the cost of increased computational burden. We also note that one can alternatively use probability densities of the form $p(\mathbf{y}|b_j = x, b_j' = z)$ in (6) as components of the vector $\mathcal{P}$, to initialize the iterative decoder.

Let us now explain the motivation behind including the RSC code blocks in the encoders in Figs. 1 and 2. Since for the proposed randomized coding scheme, modulo-2 sum of the two codewords is transmitted through the channel, we have $P_{00} = P_{11}$ and $P_{01} = P_{10}$. Incorporating these values in (6) results in $m_{sv}^1 = m_{sv}^2 = 0$ since $m_{vs} = m_{vs} = 0$ at the initialization stage. Therefore, the joint BP decoder fails to start for the randomized coding scheme. However, as illustrated in Fig. 4, the trellis based decoding of the RSC code and its dual RSC$^{\perp}$ enables us to initialize the joint BP decoder in an efficient manner, and to kick start the iterative decoder. We note that the joint BP decoder in [11] works since the authors assume that the sum of the BPSK versions of the codewords is transmitted.

V. Numerical Examples

In this section, we provide numerical examples on the performance of the randomized SCLDGM (RSCLDGM) codes to study their security gaps. We denote the codeword length of the big code, number of data bits and number of random bits by n, k and r, respectively. The utilized RSC code is of rate $1/2$ with the generator polynomial [1 5/7] (in octal form) whose dual is either [1 7/5] or [5/7 1] (see [7]); the former is used for the RSCLDGM scheme and the

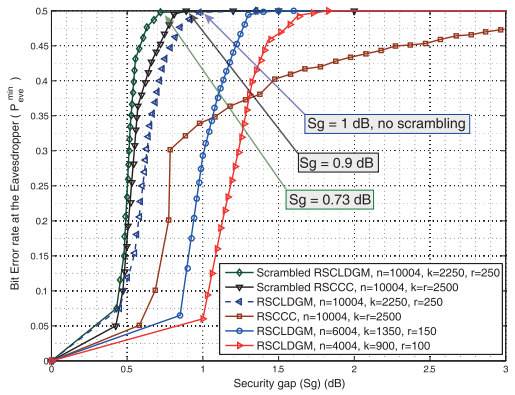

Fig. 5. P_{eve}^{min} versus the security gap (for $P_{main}^{max} = 10^{-5}$) for the RSCLDGM scheme of different lengths along with the RSCCC scheme proposed in [7]. The corresponding scrambled versions of these schemes are also included for the sake of comparison.

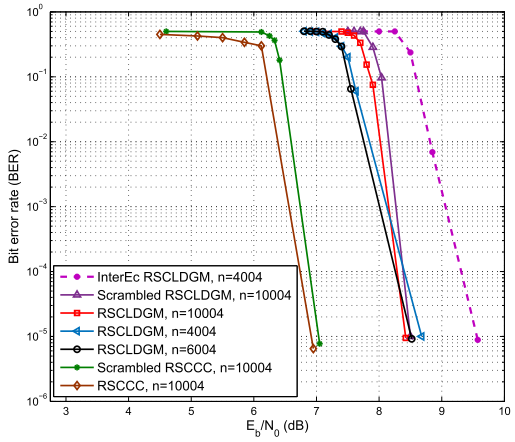

Fig. 6. BER performance of the codes used in Fig. 5.

latter for the randomized serially concatenated convolutional codes (RSCCCs) proposed in [7] and [12]. The number of iterations for exchanging the extrinsic information among different components of the decoder is set to 10. We randomly generate the parity check matrices for the utilized LDGM codes based on the structure given in Section III. Specifically, for length $n = 10004$, the parity check matrix of the outer LDGM is generated by using $N_b = 10$ and $W[\mathbf{H}_i] = \{4, 4, 3, 3, 3, 3, 3, 3, 3, 1\}$ for $i = 1, 2, \ldots, 10$ where each $\mathbf{H}_i$ is of size 250. Furthermore, for the parity check matrix of the inner LDGM code $N_b = 2$, $W[\mathbf{H}_1] = 5$, $W[\mathbf{H}_2] = 1$ and $\mathbf{H}_i$'s are of size 2500. We also have used two S-random interleavers between each of encoding stages in Figs. 1 and 2.

Fig. 5 demonstrates the resulting security gaps for the RSCLDGM code for three different code lengths. The results show that a security gap of about 1 dB can be obtained using the newly proposed scheme for a code of length 10^4, $P_{eve}^{min} = 0.5$ and $P_{main}^{max} = 10^{-5}$. We emphasize that this result is obtained without using any scrambler at the transmitter (used to amplify the errors at the eavesdropper [4] and [7]). Furthermore, for similar code lengths and code rates, using scrambling along with the RSCLDGM outperforms the scrambled RSCCCs (which to the best of our knowledge is the best existing result in the literature) by about 0.2 dB. Note that RSCCCs are a class of randomized turbo codes proposed in [12] and [7] where serially concatenated convolutional codes are used to encode data and random bits.

Fig. 6 illustrates the BER performances for the codes utilized in Fig. 5. Specifically, the results show the corresponding working SNRs for Bob and Eve for different scenarios. A stronger error amplification is introduced in [13] as inter frame error contamination (InterEC) where a single bit error can affect many different frames. We have also investigated the effect of InterEC idea along with the RSCLDGM codes. We depict the resulting error rate for RSCLDGM with a block length of 4004 in Fig. 6. We observe that InterEC provides a BER of 0.5 for a wider range of SNRs and also deteriorates the performance at the legitimate receiver such that the resulting security gap does not get improved for our set-up.

VI. Conclusions

We apply SCLDGM codes to the randomized encoding scheme as a solution for physical layer security. Our results demonstrate that a special class of LDPC codes (namely, SCLDGM codes) results in the lowest security gaps (as small as 1 dB) reported in the literature without introducing a scrambler. Therefore, this specific class may be particularly attractive for achieving secrecy with information-theoretic metrics as well, however, this requires further investigations.

References

[1] A. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
[2] A. Thangaraj *et al.*, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.
[3] D. Klinc *et al.*, "LDPC codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 532–540, Sep. 2011.
[4] M. Baldi *et al.*, "Non-systematic codes for physical layer security," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Dublin, Ireland, Aug. 2010, pp. 1–5.
[5] M. Baldi *et al.*, "Coding with scrambling, concatenation, and HARQ for the AWGN wire-tap channel: A security gap analysis," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 883–894, Jun. 2012.
[6] Y. Zhang *et al.*, "Polar-LDPC concatenated coding for the AWGN wiretap channel," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1683–1686, Oct. 2014.
[7] A. Nooraiepour and T. M. Duman, "Randomized convolutional codes for the wiretap channel," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3442–3452, Aug. 2017.
[8] A. Nooraiepour, "Randomized convolutional and concatenated codes for the wiretap channel," M.S. thesis, Dept. Elect. Eng., Bilkent Univ., Ankara, Turkey, 2016.
[9] M. Gonzalez-Lopez *et al.*, "Serially-concatenated low-density generator matrix (SCLDGM) codes for transmission over AWGN and Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2753–2758, Aug. 2007.
[10] M. Baldi *et al.*, "On a family of circulant matrices for quasi-cyclic low-density generator matrix codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6052–6067, Sep. 2011.
[11] A. Roumy and D. Declercq, "Characterization and optimization of LDPC codes for the 2-user Gaussian multiple access channel," *EURASIP J. Wireless Commun. Netw.*, vol. 2007, May 2007, Art. no. 074890.
[12] A. Nooraiepour and T. M. Duman, "Randomized turbo codes for the wiretap channel," in *Proc. IEEE GLOBECOM*, Singapore, Dec. 2017.
[13] M. H. Taieb and J.-Y. Chouinard, "Reliable and secure communications over Gaussian wiretap channel using HARQ LDPC codes and error contamination," in *Proc. IEEE CNS*, Florence, Italy, Sep. 2015, pp. 158–163.