On Secure Communications Over Gaussian Wiretap Channels via Finite-Length Codes

Alireza Nooraiepour, Sina Rezaei Aghdam, and Tolga M. Duman

***Abstract*— Practical codes for the Gaussian wiretap channel are designed aiming at satisfying information-theoretic metrics to ensure security against a passive eavesdropper (Eve). Specifically, a design criterion is introduced for the coset coding scheme in order to satisfy a strong secrecy condition described with the mutual information between the secret message and Eve's observation. In addition, mutual information neural estimation (MINE) powered from deep learning tools is applied in order to directly compute the information-theoretic security constraint, and verify the proposed solutions. It is shown that finite-length coset codes can indeed ensure secure transmission from an information-theoretic perspective.**

***Index Terms*— Gaussian wiretap channel, information-theoretic secrecy, coset coding, mutual information neural estimation.**

I. Introduction

DESIGNING explicit secure codes which can be practically implemented in wireless networks is an important step towards achieving physical layer security. Several works have proposed finite-length codes for security by utilizing them with the coset coding method as a tool to confuse the eavesdropper [1], [2]. This approach, which also appears under the name *randomized encoding scheme* [1], [3], maps each message to a randomly selected codeword in a coset of a code. The authors in [3] propose randomized convolutional codes with efficient encoding, and provide a low complexity decoder exploiting the trellis structure of the convolutional codes. Based on this result, turbo codes and LDPC codes are utilized for physical layer security in [4], [5] to improve the system performance. Besides randomized coding, scrambling [6] has also been widely used in the literature to maximize the bit error probability (BEP) at the eavesdropper through propagating errors in the decoding process. Furthermore, application of punctured LDPC codes for the Gaussian wiretap channel is studied in [7]. All of these works rely on the BEP performance as the security constraint, and assume that a BEP close to $1/2$ satisfies the security criterion. In an attempt to characterize the system performance from an information theoretic point of view, the authors in [8], [9] provide a lower bound on the mutual information between the message and the eavesdropper's observation, and use it to find the minimum equivocation obtained by LDPC codes.

In this work, we aim at designing finite-length coded systems for the Gaussian wiretap channel based on the randomized coding scheme which provide security in an information-theoretic sense. Specifically, we use the mutual information between the secret message (M) and the eavesdropper's observation ($\mathbf{Y}$), i.e., $\mathbb{I}(M;\mathbf{Y})$, as the security metric, and refer to a system which satisfies $\mathbb{I}(M;\mathbf{Y}) \leq \kappa$ as κ-strongly-secure. One should note the difference with the "strong secrecy" condition [1], which is studied for asymptotic code lengths. We develop a theorem which provides a sufficient condition on the randomized linear codes to be κ-strongly-secure over AWGN channels. Furthermore, we present true characterizations of secrecy provided by several finite-length codes utilizing a recent tool developed in the deep learning literature called mutual information neural estimation (MINE). Numerical results demonstrate that finite-length randomized codes achieve points on boundaries of the equivocation region corresponding to the maximum equivocation which verifies that they can indeed provide information theoretic secrecy.

Organization of the letter is as follows. Section II presents the system model for the Gaussian wiretap channel. Section III develops a result to identify a criterion for the randomized codes to be κ-strongly-secure with a smaller value of κ. Detailed proof of this main result is provided in Section IV. Numerical examples are presented in Section V, and finally, the letter is concluded in Section VI.

II. System Model

We consider a wiretap channel where both the main and the eavesdropper channels are additive white Gaussian noise (AWGN) channels. The main channel is between Alice (Tx) and Bob (Rx), while the Eve's channel connects Alice and Eve. For an input $\mathbf{x}_i \in \{-1,+1\}^n$ of length n, the output of an AWGN channel is obtained by $\mathbf{y} = \mathbf{x}_i + \mathbf{z}$, where $\mathbf{z}$ is a length n Gaussian noise vector whose components are independent and identically distributed (i.i.d.) with zero mean and variance $N_0/2$. E_b/N_0 is referred to as signal-to-noise ratio (SNR).

In the absence of tools to compute information-theoretic metrics, alternatives based on BEP, e.g., security gap, have been widely used [3], [6] to assess security. In this work, however, we employ an information-theoretic metric. Specifically, we define κ-strongly-secure codes as those that satisfy

Manuscript received April 3, 2020; revised May 1, 2020; accepted May 2, 2020. Date of publication May 14, 2020; date of current version September 12, 2020. The associate editor coordinating the review of this letter and approving it for publication was C. Condo. *(Corresponding author: Alireza Nooraiepour.)*

Alireza Nooraiepour is with the WINLAB, Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854 USA (e-mail: alireza.nooraiepour@rutgers.edu).

Sina Rezaei Aghdam is with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden (e-mail: sinar@chalmers.se).

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

Digital Object Identifier 10.1109/LCOMM.2020.2994884

$\mathbb{I}(M;\mathbf{Y}) \leq \kappa$, where M and $\mathbf{Y}$ denote random variables corresponding to the secret message and Eve's observation, respectively, and use this metric to evaluate the performance of short-length codes. The design parameter κ is a predefined (small) value.

III. Strongly-Secured Codes

In this section, we describe how κ-strongly-secure codes can be realized through the randomized encoding scheme, and briefly introduce MINE, a tool powered from deep learning for computing mutual information in a data-driven manner.

A. Criterion for κ-Strongly-Secure Codes

Coset coding approach (also referred as the randomized encoding scheme) can be described as a matrix multiplication. Let $\mathbf{s}$ and $\mathbf{v}$ denote the message and random bit vectors of length k and r, respectively. Then, a codeword of length n is generated by $\mathbf{c} = \begin{bmatrix}\mathbf{s} & \mathbf{v}\end{bmatrix}\begin{bmatrix}\mathbf{H}\\ \mathbf{G}\end{bmatrix}$, where $\mathbf{H}$ and $\mathbf{G}$ are $k \times n$ and $r \times n$ generator matrices whose rows are linearly independent. We note that $r \leq n-k$ where the equality corresponds to a scheme with full randomization. For a given message $\mathbf{s}$, the scheme picks a random codeword $\mathbf{c}$ from the coset corresponding to $\mathbf{s}$ via a randomly generated $\mathbf{v}$. We refer to $\mathbf{G}$ as the generator matrix of the small code which is equivalent to the coset corresponding to $\mathbf{s} = \mathbf{0}$. For an AWGN channel, the maximum a posteriori probability (MAP) decoder for the coset coding method boils down to [3], $\hat{i} = \underset{i}{\operatorname{argmax}} \sum_{j=1}^{N} e^{\frac{-\|\mathbf{y}-\mathbf{c}_{ji}\|^2}{N_0}}$, where N is the number of codewords in each coset and $\mathbf{c}_{ji}$ denotes the Binary Phase Shift Keying (BPSK) modulated version of the jth codeword in the ith coset.

Theorem 1: Consider a randomized scheme containing all the n-tuples with m distinct messages where $\mathbf{G}$ has distinct non-zero columns. Then, there exists an $L > 0$ such that for an observation vector $\mathbf{y}$ satisfying $||\mathbf{y}||_\infty \leq L$, we have

$$\mathbb{I}(M;\mathbf{y}) \leq B_1(\mathbf{y}) \triangleq \log_2(1-m\eta) - m\eta\log_2\left(\frac{1-m\eta}{m}\right), \tag{1}$$

where η is a value which depends on $\mathbf{y}$ satisfying $\eta \leq 1/m$, and $||\mathbf{y}||_\infty = \max_i \ y_i$ for $\mathbf{y} = [y_1, \ldots, y_n]$.

Proof: The proof is given in Section IV. □

From the perspective of posterior beliefs on the messages, i.e., $P(m_i|\mathbf{y})$'s, the above theorem states that there exists $\eta \leq 1/m$ such that $|P(m_i|\mathbf{y}) - \frac{1}{m}| \leq \eta$ for the proposed randomized codes, where m_i's, $i = 1, \ldots, m$, are the set of indices from which M is chosen. Intuitively, as messages are represented by different cosets, the proposed choice of $\mathbf{G}$ generates the so-called *symmetric* cosets defined in Section IV-A, which could potentially result in posterior beliefs in the vicinity of $\frac{1}{m}$. The following result proves that the class of codes introduced in the above theorem can be κ-strongly-secure for some $\kappa > 0$.

Theorem 2: If $\mathbb{I}(M;\mathbf{y}) \leq B_1(\mathbf{y})$ for any $||\mathbf{y}||_\infty \leq L$, then

$$\mathbb{I}(M;\mathbf{Y}) \leq \kappa \triangleq \sup_{||\mathbf{y}||_\infty \leq L} \alpha B_1(\mathbf{y}) + (1-\alpha)\mathbb{H}(M) \tag{2}$$

where $\alpha = \left(\Phi\left(\frac{L-1}{\sqrt{N_0/2}}\right) - \Phi\left(\frac{-L-1}{\sqrt{N_0/2}}\right)\right)^n$, with $\Phi(\cdot)$ denoting the cumulative distribution function of a standard Gaussian random variable.

Proof: As $\mathbb{I}(M;\mathbf{Y}) = \mathbb{H}(M) - \mathbb{H}(M|\mathbf{Y})$ with $\mathbb{H}(M|\mathbf{Y}) = \int_{\mathcal{Y}} \mathbb{H}(M|\mathbf{y})p(\mathbf{y})d\mathbf{y}$, assuming $L > 0$, we have

$$\mathbb{I}(M;\mathbf{Y}) \leq \mathbb{H}(M) - \int_{||\mathbf{y}||_\infty \leq L} \mathbb{H}(M|\mathbf{y})p(\mathbf{y})d\mathbf{y} \tag{3}$$

where $p(\mathbf{y})$ denotes the density of $\mathbf{y}$ given by

$$\sum_{\mathbf{c}_{ji}} p(\mathbf{y}|\mathbf{c}_{ji})P(\mathbf{c}_{ji}) = \frac{1}{2^{k+r}}\sum_{i=1}^{2^k}\sum_{j=1}^{2^r}\frac{1}{(\pi N_0)^{n/2}}e^{\frac{-||\mathbf{y}-\mathbf{c}_{ji}||^2}{N_0}}. \tag{4}$$

For an L which satisfies the conditions in Theorem 1, $\mathbb{H}(M|\mathbf{Y})$ can be bounded from below by

$$\inf_{||\mathbf{y}||_\infty \leq L} B_2(\mathbf{y}) \int_{||\mathbf{y}||_\infty \leq L} \prod_{k=1}^{n} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-(y_k - c_{ji}(k))^2}{N_0}} d\mathbf{y}, \tag{5}$$

where $\mathbb{H}(M|\mathbf{y}) > B_2(\mathbf{y}) = \mathbb{H}(M) - B_1(\mathbf{y})$ for $||\mathbf{y}||_\infty \leq L$, and $c_{ji}(k) \in \{-1, +1\}$ denotes the kth element of $\mathbf{c}_{ji}$. As the integral in (5) equals α, (3) results in

$$\begin{aligned}\mathbb{I}(M;\mathbf{Y}) &\leq \mathbb{H}(M) - \inf_{||\mathbf{y}||_\infty \leq L} B_2(\mathbf{y})\alpha \\ &= \sup_{||\mathbf{y}||_\infty \leq L} (1-\alpha)\mathbb{H}(M) + \alpha B_1(\mathbf{y}).\end{aligned} \tag{6}$$

□

We note that the tightness of the bound is controlled by the parameter η. Later on, we utilize the generalized Jensen's inequality [10] in order to obtain an upper bound on η.

With the above result, the proposed design criterion for a code to be κ-strongly-secure is two-fold: 1) all the possible n-tuples should be used in the randomized scheme, 2) the $\mathbf{G}$ matrix must have distinct non-zero columns.

B. MINE: Mutual Information Neural Estimation

To address the required mutual information calculation, we now utilize a recently proposed estimator in [11] which relies on Kullback Leibler (KL) divergence, i.e.,

$$\begin{aligned}\mathbb{I}(X;\mathbf{Y}) &= \mathbb{D}_{\mathrm{KL}}(p(x,y)||p(x)p(y)) \\ &\geq \mathbb{E}_{p(x,y)}[T] - \log(\mathbb{E}_{p(x)p(y)}[e^T])\end{aligned} \tag{7}$$

where $T : \Omega \to \mathbb{R}$ is a mapping from the sample space of the joint and marginal distributions (i.e., Ω) to $\mathbb{R}$. This representation relies on the choice of the function T in order to provide a tight lower bound on the mutual information. Specifically, the bound converges to the true mutual information for the optimal functions T^*. Therefore, utilizing a deep neural network to parameterize the set of functions T as $\{T_\theta\}_{\theta\in\Theta}$, MINE is defined as

$$\sup_{\theta\in\Theta} \ \mathbb{E}_{p(x,y)}[T_\theta] - \log(\mathbb{E}_{p(x)p(y)}[e^{T_\theta}]), \tag{8}$$

where the expectations are estimated using empirical samples. This can effectively be solved via stochastic gradient descent algorithms using mini batches of two datasets corresponding to the joint ($p(x,y)$) and marginal ($p(x)p(y)$) distributions.

IV. Proof of Theorem 1

Lemma 1: If the posterior probabilities satisfy $|P(m_i|\mathbf{y}) - \frac{1}{m}| \leq \eta$ *for* $i = 1, \ldots, m$, *for some* $\eta \leq 1/m$, *then*

$$\mathbb{I}(M;\mathbf{y}) \leq B_1 = \log_2(1-m\eta) - m\eta \log_2\Big(\frac{1-m\eta}{m}\Big). \quad (9)$$

Proof: Assume the conditions are satisfied, i.e., $|P(m_i|\mathbf{y}) - \frac{1}{m}| \leq \eta$ for some $\eta \leq 1/m$, since the function $-x\log x$ is strictly increasing for $x \leq 1/e$, we can write

$$\sum_{i=1}^{m} -P(m_i|\mathbf{y}) \log_2 P(m_i|\mathbf{y}) \geq \sum_{i=1}^{m} -\alpha \log_2(\alpha).$$

where $\alpha = 1/m - \eta$. Note that $\alpha \leq 1/e$ always holds assuming $m > e$. Using $H(M) = -\sum_{i=1}^{m} 1/m \log_2(1/m)$, with some rearrangement, we obtain $\mathbb{H}(M) - \mathbb{H}(M|\mathbf{y}) \leq B_1$. □

We point out that if there exists an $\eta \leq 1/m$ such that $|P(m_i|\mathbf{y}) - \frac{1}{m}| \leq \eta,\ i = 1, \ldots, m$ for the setting introduced in Theorem 1, the above lemma can be utilized to prove the result. We prove that such an η exists in two steps. First, we show that the codes described in Theorem 1 are *symmetric* randomized codes defined in Section IV-A. Secondly, in Section IV-B, we prove that the desired η exists for such codes.

A. Symmetric Randomized Codes

Consider a randomized scheme with m cosets each containing N codewords. Assume that $\mathbf{c}_{ji}$ is transmitted. The likelihood of the received vector $\mathbf{y}$ is given by $p(\mathbf{y}|\mathbf{c}_{ji}) = \frac{1}{\sqrt{\pi N_0}} e^{-a_{ji}}$ where $a_{ji} = \frac{\|\mathbf{y}-\mathbf{c}_{ji}\|^2}{N_0}$.

Definition 1: A randomized code is called symmetric if the sample mean and variance of a_{ji}*'s are constant for all the cosets, i.e.,* $\overline{a_i} = \frac{1}{N}\sum_{j=1}^{N} a_{ji} = \overline{a},\ \sigma_i^2 = \overline{a_i^2} - \overline{a_i}^2 = \sigma^2\ \forall i$.

Lemma 2: For a binary linear code $\mathcal{C}(n,k)$, *assume that the* $2^k \times n$ *matrix* $\mathbf{C}$ *includes all the* 2^k *codewords as its rows. Denoting the corresponding* j*th column by* $\mathbf{c}_j$, $\sum_{i=1}^{2^k} \mathbf{c}_j(i) = 2^{k-1}$ *if* $\mathbf{c}_j \neq \mathbf{0}$ *where* $\mathbf{c}_j(i)$ *is the* i*th element of* $\mathbf{c}_j$.

The proof is straightforward and omitted for brevity.

Lemma 3: Considering $\mathcal{C}(n,k)$ *and* $\mathbf{C}$ *in Lemma 2,* $\sum_{i=1}^{2^k} \mathbf{c}_j(i)\mathbf{c}_l(i) = 2^{k-2}$ *for* $j \neq l$ *if* $\mathbf{c}_j \neq \mathbf{c}_l \neq \mathbf{0}$.

Proof:

Consider a matrix $\mathbf{A}_k$ of size $2^k \times k$ whose rows are binary representations of integers from 0 to $2^k - 1$. For non-zero $\mathbf{c}_j \neq \mathbf{c}_l$, we can write $\mathbf{c}_j = \mathbf{A}_k\mathbf{v}$ and $\mathbf{c}_l = \mathbf{A}_k\mathbf{w}$ where $\mathbf{v} \neq \mathbf{w}$.

We will prove the result by induction. The base case $k = 2$ is easily verified. Assuming that the statement is true for k, we will prove it for $k+1$ case. The non-zero $\mathbf{v}$ and $\mathbf{w}$ of length $k+1$ differ in at least one element which, without loss of generality (wlog), is assumed to be the last one. Then, $\overline{\mathbf{v}}$ and $\overline{\mathbf{w}}$ denote the first k elements of $\mathbf{v}$ and $\mathbf{w}$, respectively. Two cases can happen: 1) $\overline{\mathbf{v}} \neq \overline{\mathbf{w}}$, or 2) $\overline{\mathbf{v}} = \overline{\mathbf{w}}$. For the sake of brevity, we only present the proof for the first case. For the first 2^k rows of $\mathbf{A}_{k+1} = \begin{bmatrix} \mathbf{A}_k & \mathbf{0}_k \\ \mathbf{A}_k & \mathbf{1}_k \end{bmatrix}$ (where $\mathbf{0}_k$ and $\mathbf{1}_k$ are all-zero and all-one column vectors of size k), element-wise multiplication of $\mathbf{A}_k\overline{\mathbf{v}}$ and $\mathbf{A}_k\overline{\mathbf{w}}$ has 2^{k-2} ones. For the last 2^k rows, (assuming wlog that the last element of $\mathbf{v}$ is 0 and that of $\mathbf{w}$ is 1), are four cases:

$$\begin{aligned} &\mathbf{a}_k^j\overline{\mathbf{v}} = 0 \Rightarrow \mathbf{a}_{k+1}^j\mathbf{v} = 0, \quad \mathbf{a}_k^j\overline{\mathbf{v}} = 1 \Rightarrow \mathbf{a}_{k+1}^j\mathbf{v} = 1, \\ &\mathbf{a}_k^j\overline{\mathbf{w}} = 0 \Rightarrow \mathbf{a}_{k+1}^j\mathbf{w} = 1, \ \mathbf{a}_k^j\overline{\mathbf{w}} = 1 \Rightarrow \mathbf{a}_{k+1}^j\mathbf{w} = 0, \end{aligned} \quad (10)$$

where $\mathbf{a}_k^j$ denotes the jth row of $\mathbf{A}_k$, $j = 2^k + 1, 2^k + 2, \ldots, 2^{k+1}$. It is known from induction hypothesis that 2^{k-2} entries of $\mathbf{A}_k\overline{\mathbf{v}}$ and $\mathbf{A}_k\overline{\mathbf{w}}$ are simultaneously 1. Furthermore, Lemma 2 states that each of these has 2^{k-1} ones. Therefore, 2^{k-2} entries of $\mathbf{A}_k\overline{\mathbf{v}}$ are 1 while the corresponding elements of $\mathbf{A}_k\overline{\mathbf{w}}$ are 0. Based on the above 4 cases, these are precisely the entries where $\mathbf{A}_{k+1}\mathbf{v}$ and $\mathbf{A}_{k+1}\mathbf{w}$ are simultaneously 1. This means that we have 2^{k-2} entries of the element-wise product that are 1 from the last 2^k rows of $\mathbf{A}_{k+1}$. Altogether, there are $2^{k-2} + 2^{k-2} = 2^{k-1}$ 1's in the element-wise multiplication of $\mathbf{A}_{k+1}\mathbf{v}$ and $\mathbf{A}_{k+1}\mathbf{w}$, as desired. □

We note that Lemmas (2) and (3) hold for both a linear code and its cosets. Equipped with these two lemmas, we are ready to prove that the condition in Theorem 1 offers symmetric randomized codes.

Theorem 3: A randomized code is symmetric if the generator matrix of its small code, i.e., $\mathbf{G}$*, has distinct non-zero columns.*

Proof: We begin with computing $\overline{a_i} = \frac{1}{N}\sum_{j=1}^{N} a_{ji}$,

$$\overline{a_i} = \frac{1}{N_0 N} \sum_{l=1}^{n} N_{l,1}(y_l + 1)^2 + N_{l,0}(y_l - 1)^2 \quad (11)$$

where $N_{l,1}$ and $N_{l,0}$ denotes the number of 1's and 0's, respectively, in the lth column of the ith coset. Since $\mathbf{G}$ has non-zero columns, all the columns of the ith coset are non-zero as well. Then, using Lemma 2 we get $N_{l,1} = N_{l,0} = N/2$, hence $\overline{a_i}$ does not depend on the coset index i. Next, we have

$$\begin{aligned} \overline{a_i^2} &= \frac{1}{NN_0^2} \sum_{j=1}^{N} a_{ji}^2 = \frac{1}{NN_0^2} \sum_{j=1}^{N} \Big(\sum_{l=1}^{n} (y_l - (c_{ji})_l)^2 \Big)^2 \\ &= \frac{1}{NN_0^2} \sum_{j=1}^{N} \Big(\sum_{l=1}^{n} y_l^2 - 2\sum_{l=1}^{n} y_l(\mathbf{c}_{ji})_l + \sum_{l=1}^{n} (\mathbf{c}_{ji})^2 \Big)^2, \end{aligned} \quad (12)$$

where $(\mathbf{c}_{ji})_l$ denotes the lth element of $\mathbf{c}_{ji}$ which is either $+1$ or -1. The terms that involve the index i are:

$$\sum_{j=1}^{N}\sum_{l=1}^{n} y_l(\mathbf{c}_{ji})_l = \sum_{l=1}^{n} (N_{l,1} - N_{l,0}) y_l = 0 \quad (13)$$

$$\sum_{j=1}^{N} \Big(\sum_{l=1}^{n} y_l(\mathbf{c}_{ji})_l \Big)^2 = \sum_{j=1}^{N}\sum_{l=1}^{n} (y_l(\mathbf{c}_{ji})_l)^2 + 2\sum_{j=1}^{N}\sum_{f=1}^{n}\sum_{d=1}^{f-1} y_f(\mathbf{c}_{ji})_f y_d(\mathbf{c}_{ji})_d. \quad (14)(15)$$

As $\mathbf{G}$ has distinct columns, each coset also has distinct columns. By utilizing Lemma 3, one can show that the result of $\sum_{j=1}^{N} (\mathbf{c}_{ji})_f (\mathbf{c}_{ji})_d$ for all $f, d,\ f > d$ is independent of i. Therefore, similar to the sample mean, $\overline{a_i^2}$ also does not vary for different cosets, hence we have a constant sample variance for all the cosets, i.e., $\sigma_i^2 = \sigma^2, \forall i$. □

We note that the usefulness of unique non-zero columns for the generator matrix of the coset coding scheme has already been discussed for the binary erasure channel (BEC) in [12].

B. Existence of $\eta \leq 1/m$

We are now ready to show that $\eta \leq 1/m$ exists for the setting presented in Theorem 1. We start by

$$P(m_i|\mathbf{y}) = \frac{p(\mathbf{y}|m_i)P(m_i)}{\sum_{j=1}^{m} p(\mathbf{y}|m_j)P(m_j)}, \tag{16}$$

which denotes the belief on message m_i given an observation $\mathbf{y}$. For equiprobable messages ($P(m_i) = \frac{1}{m}, i = 1, \ldots, m$),

$$P(m_i|\mathbf{y}) = \frac{\sum_{j=1}^{N} p(\mathbf{y}|\mathbf{c}_{ji})}{\sum_{i=1}^{m}\sum_{j=1}^{N} p(\mathbf{y}|\mathbf{c}_{ji})}, \tag{17}$$

where $p(\mathbf{y}|\mathbf{c}_{ji})$ is given in Section IV-A.

Lemma 4: (Taken from [10]) Let $\{x_i\}_{i=1}^{N}$ be N random samples drawn from X, a one-dimensional random variable such that $P(X \in (a,b)) = 1$ for $-\infty \leq a \leq b \leq \infty$. Let

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2,$$
$$g(x,\overline{x}) \triangleq \frac{\phi(x) - \phi(\overline{x})}{(x-\overline{x})^2} - \frac{\phi'(\overline{x})}{x-\overline{x}}, \tag{18}$$

where $\phi(x)$ is a twice differentiable function on (a,b). Then,

$$\sigma_x^2 \inf_{x\in[a,b]}\ g(x,\overline{x}) + \phi(\overline{x}) \leq \frac{1}{N}\sum_{i=1}^{N}\phi(x_i) \leq \sigma_x^2 \sup_{x\in[a,b]}\ g(x,\overline{x}) + \phi(\overline{x}), \tag{19}$$

where $a = \min\{x_1, \ldots, x_N\}$, $b = \max\{x_1, \ldots, x_N\}$, and g is monotonically decreasing in x if $\phi'(x)$ is concave. Also,

$$\inf\ g(x,\overline{x}) \geq \inf\ \phi''(x)/2,\ \sup\ g(x,\overline{x}) \leq \sup\ \phi''(x)/2. \tag{20}$$

Lemma 4 enables us to bound the posterior probabilities $P(m_i|\mathbf{y})$ in (17) in terms of the mean and variance of a_{ji} introduced in Section IV-A using $\phi(x) = e^{-x}$.

Lemma 5: For a symmetric randomized code, the posterior probabilities can be bounded as $|P(m_i|\mathbf{y}) - 1/m| \leq \eta = \frac{(h-l)\sigma^2}{m(l\sigma^2+e^{-\overline{a}})}$ for $i = 1, \ldots, m$ and some $1/2 \geq h \geq l \geq 0$. Furthermore, there exists an $L > 0$ such that $\eta \leq 1/m$ for $||\mathbf{y}||_\infty \leq L$ given that all the n-tuples are present in the code.

Proof: Utilizing Lemma 4 for the ith coset, one can write

$$l_i\sigma_i^2 + e^{-\overline{a}_i} \leq \frac{1}{N}\sum_{j=1}^{N} e^{-a_{ji}} \leq h_i\sigma_i^2 + e^{-\overline{a}_i}. \tag{21}$$

where l_i and h_i are the infimum and supremum of the g function introduced in Lemma 4 for $\phi(x) = e^{-x}$, respectively. The lower bound on l_i and the upper bound on h_i, presented in (20), equals to 0 and $1/2$, respectively. Utilizing the upper and lower limits in (21) for the sum in the nominator and denominator of (17), respectively, results in an upper bound for $P(m_i|\mathbf{y})$. The reverse order can be used for obtaining a lower bound. Therefore, one can write

$$\frac{l_i\sigma_i^2 + e^{-\overline{a}_i}}{m(h_i\sigma_i^2 + e^{-\overline{a}_i})} \leq P(m_i|\mathbf{y}) \leq \frac{h_i\sigma_i^2 + e^{-\overline{a}_i}}{m(l_i\sigma_i^2 + e^{-\overline{a}_i})}. \tag{22}$$

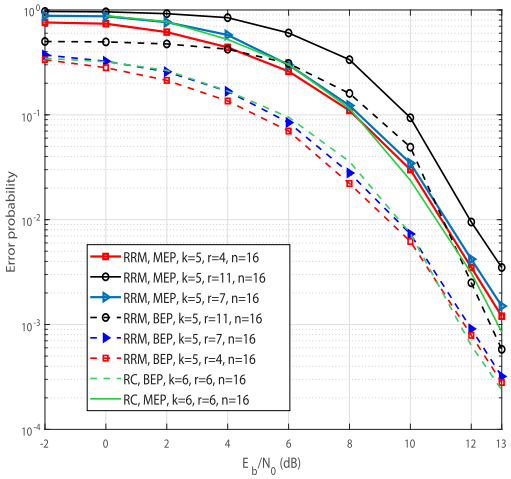

Fig. 1. MEP/BEP for the small-length randomized codes.

When $\overline{a}_i = \overline{a}$ and $\sigma_i^2 = \sigma^2\ \forall i$, the above can be written as $|P(m_i|\mathbf{y}) - \frac{1}{m}| \leq \eta$, for $i = 1, \ldots, m$, where $l = \min_i l_i$, $h = \max_i h_i$, and $\eta = \frac{(h-l)\sigma^2}{m(l\sigma^2+e^{-\overline{a}})}$.

Next, we show that there exists an L for which $\eta \leq 1/m$, or equivalently, $\zeta \triangleq e^{-\overline{a}}/\sigma^2 - h + 2l > 0$. We consider the extreme case where $|y_i| = L$ for $i = 1, \ldots, n$. To compute ζ, we need to obtain $\overline{a}$, σ^2, h and l. The first two can be computed through Theorem 3 as $\overline{a} = n(L^2+1)/N_0$ and $\sigma^2 = 4nL^2/N_0^2$. To compute h and l, we note that the g function in Lemma 4 is monotonically decreasing. Therefore, these value can be obtained through the minimum and maximum values of a_{ji}'s computed for all the codewords. As all the n-tuples are present in the code, one can verify that these extremes of a_{ji}'s are $n(L+1)^2/N_0$ and $n(L-1)^2/N_0$, which enables us to compute h and l. Then, plugging these values in ζ and simplifying $\zeta > 0$ results in $e^{2nt} - 2e^{-2nt} \leq 6nt + n$ for $t = L/N_0$, which imposes an upper bound on L in terms of n and N_0. It can also be shown that as $t \to 0$, η converges to 0 at a polynomial rate. As an example, we have computed this bound for a fixed N_0 and two code lengths $n = 16$ and $n = 32$. The results are $L \leq 0.11N_0$ and $L \leq 0.06N_0$, respectively. L is chosen accordingly to ensure $\eta \leq 1/m$. □

V. Numerical Results

Here we investigate the performance of randomized Reed-Muller (RRM) and randomized convolutional (RC) codes with $n = 16$, $k = 5$ and $r \in \{4, 6, 7, 11\}$. For the RRM case, the $(k \times n)$ matrix $\mathbf{H}$ is chosen to be the generator matrix of the Reed-Muller (RM) code with $k = 5$ and $n = 16$. Then, the $(r \times n)$ matrix $\mathbf{G}$ is obtained as the generator matrix of the dual code [3] where $r = 11$ and $n = 16$. For $r < 11$ cases, the first r rows of $\mathbf{G}$ are chosen as the generator matrix of the random bits. We note that the MAP decoder given in Section IV can be utilized for any $r \leq n - k$. For the RC codes, a convolutional code with generator [5 7] and its dual [7 5] (obtained in [3]) is used for encoding.

Fig. 1 illustrates message error probability (MEP) and BEP of RRM codes with the MAP decoder. It is shown that both BEP and MEP increase as r increases which emphasizes the effect of randomization. Also, performance of an RC code with

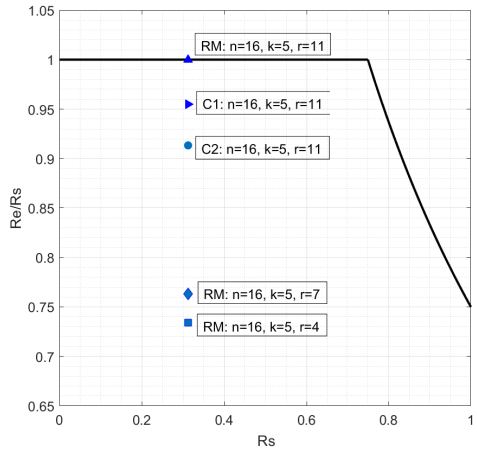

Fig. 2. Equivocation region for a wiretap channel with $\text{SNR}_E = -2\,\text{dB}$ and $\text{SNR}_B = 13\,\text{dB}$. $R_s = k/n$ and $R_e = \mathbb{H}(M|\mathbf{Y})/n$ denote data and equivocation rate, respectively.

similar code parameters is presented where the trellis-based approach proposed in [3] is utilized for decoding. In order to characterize the system performance in the equivocation region, we utilize MINE to compute $\mathbb{I}(M, \mathbf{Y})$. We note that due to the limitations of the current state of MINE, we mainly consider small-length codes. We have used a fully connected feed-forward neural network in MINE with 4 hidden layers, each with 400 neurons and using rectified linear unit (ReLU) as the activation function. The input layer has 21 neurons, and 4 million samples from the distributions $P(M)p(\mathbf{Y})$ and $p(M, \mathbf{Y})$ are generated as the dataset, and Adam optimizer [11] with a learning rate of 10^{-4} is used for training [13].

Fig. 2 demonstrates the equivocation region [9] based on the secrecy capacity of binary input AWGN channel [1] when $\text{SNR}_E = -2\,\text{dB}$ and $\text{SNR}_B = 13\,\text{dB}$ and the achieved points at Eve with the explicit randomized codes. It is shown that as r gets large, the achieved points by RRM codes get closer to the boundary. Specifically, the case $r = 11$ (full randomization) achieves a point very close (as close as 2×10^{-3}) to the boundary corresponding to $\mathbb{I}(M; \mathbf{Y}) \leq 0.01$, while Bob can achieve a BEP less than 7×10^{-4} at $\text{SNR}_B = 13\,\text{dB}$. One can also verify that BEP is ≈ 0.5 at Eve for this case. This scheme satisfies the condition in Theorem 1, and the bound in Theorem 2 can be obtained as 1.73, 1.55 and 1.32 for SNRs -4, -5 and -6 dB, respectively. The estimated values from MINE for these SNRs (with the same order) are obtained as 8.4×10^{-3}, 7.8×10^{-3} and 6.9×10^{-3}. One can verify that the absolute value of the difference between the bound and the computed mutual information decreases in lower SNRs. We note that even though the bound in Theorem 2 is not tight, the conditions based on which the bound is obtained, are verified to be useful for designing secure randomized codes. To see this more clearly, we have also included achievable points from two other randomized codes, denoted by $C1$ and $C2$ with $n = 16,\ r = 11,\ k = 5$, in Fig. 2 which are not symmetric. In fact, the generator matrices for $C1$ and $C2$ have 2 and 3 repeated columns, respectively. Although these codes have the same parameters as the RRM code, one can see that they have an inferior performance in terms of secrecy.

VI. Conclusion

We have considered finite-length codes for the Gaussian wiretap channel based on randomized (coset) coding in order to provide security from an information-theoretic perspective. Specifically, we have shown that for a Gaussian wiretap channel it is desirable to utilize all the n-tuples in the coset coding scheme and pick the generator matrix of the small code with distinct non-zero columns.

References

[1] S. Rezaei Aghdam, A. Nooraiepour, and T. M. Duman, "An overview of physical layer security with finite-alphabet signaling," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1829–1850, 2nd Quart., 2019.

[2] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[3] 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.

[4] A. Nooraiepour and T. M. Duman, "Randomized turbo codes for the wiretap channel," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[5] A. Nooraiepour and T. M. Duman, "Randomized serially concatenated LDGM codes for the Gaussian wiretap channel," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 680–683, Apr. 2018.

[6] M. Baldi, M. Bianchi, and F. Chiaraluce, "Non-systematic codes for physical layer security," in *Proc. IEEE Inf. Theory Workshop*, Aug. 2010, pp. 1–5.

[7] D. Klinc, J. Ha, S. W. McLaughlin, J. Barros, and B.-J. Kwak, "LDPC codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 532–540, Sep. 2011.

[8] M. Baldi, G. Ricciutelli, N. Maturo, and F. Chiaraluce, "Performance assessment and design of finite length LDPC codes for the Gaussian wiretap channel," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 435–440.

[9] C. W. Wong, T. F. Wong, and J. M. Shea, "LDPC code design for the BPSK-constrained Gaussian wiretap channel," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Dec. 2011, pp. 898–902.

[10] J. G. Liao and A. Berg, "Sharpening Jensen's inequality," *Amer. Statistician*, vol. 73, no. 3, pp. 278–281, Jul. 2019, doi: 10.1080/00031305.2017.1419145.

[11] M. Ishmael Belghazi *et al.*, "MINE: Mutual information neural estimation," 2018, *arXiv:1801.04062*. [Online]. Available: http://arxiv.org/abs/1801.04062

[12] W. K. Harrison and M. R. Bloch, "On dual relationships of secrecy codes," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 366–372.

[13] [Online]. Available: https//:Github.com/sungyubkim/MINE-Mutual-Information-Neural-Estimation-/blob/mas ter/MINE.ipynb