Information-Theoretic Limits for Steganography in Multimedia

Hassan Y. El-Arsh, Amr Abdelaziz, *Member, IEEE,* Ahmed Elliethy, *Member, IEEE,* and Hussein A. Aly, *Senior Member, IEEE*

***Abstract*—Steganography in multimedia aims to embed secret data into an innocent multimedia cover object. The embedding introduces some distortion to the cover object and produces a corresponding stego-object. The embedding distortion is measured by a cost function that determines the probability of detection of the existence of secret embedded data. An accurate definition of the cost function and its relation to the maximum embedding rate is the keystone for the proper evaluation of a steganographic system. Additionally, the statistical distribution of multimedia sources follows the Gibbs distribution which is a complex statistical model that prohibits a thorough mathematical analysis. Previous multimedia steganographic approaches either assume a relaxed statistical distribution of multimedia sources or presume a proposition on the maximum embedding rate then try to prove the correctness of the proposition. Alternatively, this paper introduces an analytical procedure for calculating the maximum embedding rate within multimedia cover objects through a constrained optimization problem that governs the relationship between the maximum embedding rate and the probability of detection by any steganographic detector. In the optimization problem, we use the KL-divergence between the statistical distributions for the cover and the stego-objects to be our cost function as it upper limits the performance of the optimal steganographic detector. To solve the optimization problem, we establish an equivalence between the Gibbs and the correlated-multivariate-quantized-Gaussian distributions for mathematical thorough analysis. The solution to our optimization problem provides an analytical form for the maximum embedding rate in terms of the WrightOmega function. Moreover, we prove that the achieved maximum embedding rate comes in agreement with the well-known square root law (SRL) of steganography. We also establish the relationship between the achieved maximum embedding rate and the experimental results obtained from several embedding and detection steganographic techniques.**

I. Introduction

Steganography aims to embed data within innocent-looking cover objects such as images, audio, video, and even text [1], to produce a stego-object that looks similar to the original cover object but with hidden data embedded. The embedding process implies some form of distortion to the cover object and this distortion is utilized by a warden to detect if there is hidden data or not using different steganalysis techniques. Consequently, the optimal target for steganography is to hide the maximum amount of data subject to achieving a minimum probability of detection by a warden. In other words, steganography has two competing goals: undetectability and embedding rate.

Undetectability is concerned with determining how to alter the cover object to embed data without making a notable distortion to the cover. The distortion is measured by a cost function which is either modeled as the difference between the cover and stego-objects [2], or as the expected warden detector sensitivity of changing certain features within the cover object [3]. On the other hand, the embedding rate is defined as the ratio between the number of hidden data bits to the number of cover data bits with an acceptable undetectability margin [4, Ch.4]. Maximizing the embedding rate with an acceptable undetectability margin is fundamental to any steganographic algorithm. Multimedia objects are excellent choices for steganography due to their rich structural features and their ubiquity over industry and daily life [5], [6].

To evaluate the performance of a multimedia steganographic approach, the relation between the two contradicting goals (undetectability and embedding rate) is a vital measure. Specifically, an accurate definition of the cost function and its relation to the maximum embedding rate is the keystone for proper evaluation of the approach. Additionally, dealing with multimedia incurs additional complexity as the statistical distribution of multimedia sources generally follows the Gibbs distribution [7]–[9], which is a complex statistical model. The complexity of the Gibbs distribution prohibits a thorough mathematical analysis, especially when the definition of the cost function requires an accurate definition of the statistical model of the cover object as in [4, Ch.13] in which the Kullback–Leibler divergence (KL-divergence or the relative entropy), is utilized as a global measure of cover distortion.

Previous approaches of steganography concerned with the definition of the cost function are either not globally accurate for all multimedia types or accurate with limited scope for a certain type of multimedia for a certain class of embedding or detection techniques [10]. Additionally, due to the complexity of Gibbs distribution, previous approaches either use more relaxed mathematical distributions such as Gaussian distribution [10]–[17] in the statistical modeling or presume a preposition on the maximum embedding rate then try to show the correctness of the preposition [11], [12]. Thus the estimated performance may not reflect the real one due to the model relaxation.

H. Y. El-Arsh, Amr Abdelaziz, A. Elliethy, and H. A. Aly are with Dept. of Computer Engineering, Military Technical College, Cairo, Egypt. (e-mail: hassan.yakout@ymail.com, amrashry@mtc.edu.eg, a.s.elliethy@mtc.edu.eg, haly@ieee.org).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

The work in[1] [18] partially overcomes the aforementioned problems and proposes a constrained optimization problem that models the undetectability by the KL-divergence and the embedding rate by the mutual information between statistical distributions of the cover and stego-objects. The optimization problem is solved through a rigorous mathematical procedure to present an analytical form for the relation between the embedding rate and undetectability. However, [18] utilized a relaxed statistical model for cover and stego-objects (Multivariate-Quantized-Gaussian-Distribution) and does not provide an achievability proof that verifies the reliable extraction of the embedded message at the receiver side when the exact cover realization is not present at the receiver side, in which there exists a probability of decoding error.

In this paper, we present several contributions beyond our preliminary work in [18]. Specifically, we propose a constrained optimization problem that calculates the upper limit that any multimedia steganographic method can achieve reliably (with an achievability proof), for a prespecified acceptable level of detectability by an adversary. Unlike [18], in this paper, we model the cover object by the Gibbs statistical distribution by establishing an equivalence relation between the Gibbs and the correlated-multivariate-quantized-Gaussian distribution (CMQGD) for rigid mathematical analysis. The solution to our optimization problem provides an analytical form for the maximum embedding rate in terms of the WrightOmega function. Additionally, we prove that the achieved maximum embedding rate comes in agreement with the well-known square root law (SRL) of steganography [4, Ch.13]. Finally, we introduce comparisons between our theoretically achieved maximum embedding rate with other practically calculated rates provided by [10], [16], and [17]. The results demonstrate that our calculated upper bound is relatively very small compared to the referenced practical stenographic methods. The reason is that our theoretical upper bound is calculated against the optimal detector, which may not be achieved yet for the referenced steganographic methods.

The paper is organized as follows. Section II briefly discusses the previous publications related to our work. Section III describes the main definitions and assumptions with clear mathematical representation for the proposed constrained optimization problem formulation. Section IV provides an equivalence relation between the Gibbs distribution and CMQGD. The solution to the proposed constrained optimization problem is introduced in Section V and the achievability proof is provided in Section VI. Section VII discusses the relation between the SRL and our results. Practical interpretation of our results is presented in Section VIII followed by a discussion in Section IX. The final conclusion is introduced in Section X.

II. Previous Work

This section summarizes the most notable prior work related to the scope of this paper. For more details, readers can refer to [4], [10]–[17].

[1]A preliminary version of this paper with the same authors.

Some of the previous approaches employ the KL-divergence as a statistical cost function and introduce a theoretical limit (under different assumptions) for the maximum number of bits that can be embedded in a designated class of multimedia cover objects under a specified probability of detection by a warden. For instance, [11] utilized a special case of Continuous-Gaussian-distributed covers (*AWGN* channels), and introduce additional proofs of achievability and converse. Additionally, [12] generalizes the technique used in [11] for Multivariate-Continuous-Gaussian-distributed covers (*MIMO* channels). Although these approaches utilized an accurate statistical model for the proposed cover (*AWGN* channels) with a rigorously defined cost function (KL-divergence), their approaches follow a pre-assumed preposition on the maximum embedding rate then try to prove the correctness of the preposition.

The work in [15] provides a systematic technique for constructing the distortion function based on the relation between steganographic Fisher information and KL-divergence. In [15], the additive distortion approach is utilized to model the distortion for digital images, then adaptively calculating the individual pixel costs that minimize the KL-divergence when embedding using the least-significant bit matching approach. Although [15] utilizes a simple cover model (zero-mean independent multivariate quantized Gaussian distribution), the achieved security outperforms the state-of-the-art algorithm HUGO [19]. The authors improved their work in [13] by replacing their previous model in [15] with the generalized multivariate Gaussian (referred to as MVGG) combining it with an improved variance estimator. These improvements enhanced the steganographic performance by allowing embedding changes with larger amplitudes in complex (highly textured) regions, thanks to a thicker-tail MVGG model. The MVGG provides comparable performance with respect to pentary coded HiLL [20] and S-UNIWARD [21] against maxSRMd2 [22] and SRM [23] feature-based steganalysis methods. Although these techniques provide enhanced practical outcomes and an accurate analytical formulation for the relation between the embedding rate and the cost function, their results are based on a relaxed statistical model for the cover.

In [16], the proposed approach follows the non-additive model assumption and utilizes the Gaussian Markov Random Field (GMRF) with four-elements cross neighborhood. The proposed GMRF with low-dimensional clique structures provides the ability to capture the interdependencies among spatially contiguous pixels. The cost function design approach is formulated as the minimization of KL-divergence between the original cover image and the stego one. With GMRF, the cover image is split into two disjoint sub-images, which are conditionally independent. Then, an alternating iterative optimization technique is applied to achieve an efficient embedding incorporated with minimizing the total KL-divergence. The paper provides experiments demonstrating that the performance of GMRF outperforms state-of-the-art steganography MiPOD [14] and HiLL techniques in terms of secure payload against SRM and maxSRMd2. This work is expanded in [17] with higher-dimension clique structure (eight-elements

TABLE I: List of commonly used symbols through this paper.

$\mathbf{A}$	The sender
$\mathbf{B}$	The receiver
$\mathbf{E}$	The eavesdropper
$\mathbb{V}$	Alphabet of cover and stego-elements
$\mathbb{X}$	Alphabet of the original message
$\mathbb{M}$	Alphabet of the coded message
$\mathbf{c}$	The cover object
$\mathbf{s}$	The stego-object
$\mathbf{x}$	The original message
$\mathbf{m}$	The coded message
P_D	The probability of steganalyzer correct detection at $\mathbf{E}$
P_e	The Total probability of steganalyzer error at $\mathbf{E}$
P_E	The average probability of of steganalyzer error at $\mathbf{E}$
P_B	The probability of decoding error at $\mathbf{B}$ side
P_c	The joint probability distribution of the cover object
P_s	The joint probability distribution of the stego-object
P_m	The joint probability distribution of the coded message
c_i	The i^{th} cover element
s_i	The i^{th} stego-element
P_{c_i}	The probability distribution of the i^{th} cover object
P_{s_i}	The probability distribution of the i^{th} stego-object

GMRF). Although [16], [17] provided improved results and an analytical formulation for the relation between the embedding rate and the cost function, the provided results are based on a relaxed statistical model for the cover.

Another approach in [10] focused on gray-scale images as cover objects with an initial assumption that the cover, the message, and the stego-object to be statistically modelled as a Multivariate-Quantized-Gaussian-Distribution. This approach provides a theoretical limit for three popular detection strategies for the likelihood ratio test: Bayes, Minimax, and Neyman-Pearson. This approach maximizes the detection error of these three detectors related to a specified payload. Although this approach provides an accurate analytical form of the relation between the embedding rate and the cost function, the scope of the defined cost function is limited (only three types of detectors) with a relaxed statistical model (Gaussian distribution).

The Square Root Law (SRL) [4, Ch.13] is regarded as the first successful work that mathematically models the complex statistical relation between the embedding rate and undetectability in a global framework for multimedia. The SRL states that the embedding rate for any steganographic technique is proportional to the square root of the number of cover elements. For example, assuming cover A contains x elements and cover B contains $4x$ elements. If we can embed y bits of a secret message within A with a probability of detection $\mathcal{P}$ by a certain steganalyzer, then for the same steganalyzer with the same probability of detection $\mathcal{P}$, we can only embed $2y$ bits. Also, for cover A, the embedding rate per each cover element will be $\frac{y}{x}$, but for cover B will be $\frac{y}{2x}$. The SRL is valid for any steganographic approach regardless of the embedding and the steganalysis methods. This approach produces only relative approximated order results for any defined cost function without calculating the scaling constant.

III. System Model and Problem Statement

A. Communication Model

Elements of a cover object are assumed to be the communication channel between two entities: sender $\mathbf{A}$ and its corresponding receiver $\mathbf{B}$. An eavesdropper $\mathbf{E}$ is fully monitoring this communication channel between $\mathbf{A}$ and $\mathbf{B}$. Fig.1 introduces the general communication model for steganography. Through this paper, we consider an *innocent* original cover $\mathbf{c} = [c_1, c_2, \dots, c_n]$ and a stego-object $\mathbf{s} = [s_1, s_2, \dots s_n]$ with probability distributions P_c and P_s, respectively, where n is the number of elements of the cover or the stego-object. Both $\mathbf{c}, \mathbf{s} \in \mathbb{V}^n$, where $\mathbb{V}$ is the set of allowed cover values. For example, when the cover is a gray-scale image, then $\mathbb{V} = \{0, 1, \dots, 255\}$ is the set of all available pixel values. It should be noted that $\mathbb{V}$ is not limited to pixels, but can be considered for DC-coefficients [24], video motion vectors [2], [25], ...etc. Similarly, we consider an original message $\mathbf{x} \in \mathbb{X}^k$, where k is the number of elements of the message and $\mathbb{X}$ is the set of allowed message values (alphabet). The stego-object $\mathbf{s}$ is obtained by embedding a coded message $\mathbf{m} \in \mathbb{M}^n$ via an addition process, i.e., $\mathbf{s} = \mathbf{c} + \mathbf{m}$. The coded message $\mathbf{m}$ is obtained from $\mathbf{x}$ using a certain codebook that maps $\mathbf{x} \mapsto \mathbf{m}$. The coded message $\mathbf{m}$ is statistically distributed as P_m.

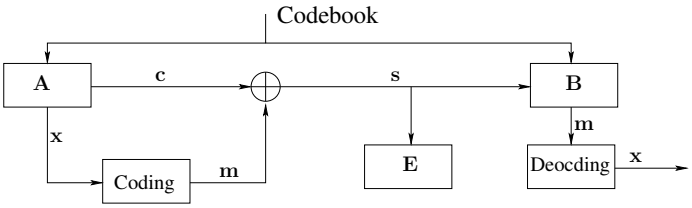

Fig. 1: General communication model for steganography.

B. Problem Statement

Consider the scenario in which $\mathbf{A}$ has access to a multimedia cover-object $\mathbf{c}$, that is drawn from some distribution P_c which is known by both $\mathbf{B}$ and $\mathbf{E}$. To conceal secret information within $\mathbf{c}$, $\mathbf{A}$ needs to transmit a slightly perturbed version of $\mathbf{c}$ called the stego-object $\mathbf{s}$ that only $\mathbf{B}$ can reveal the context of these perturbations. Meanwhile, $\mathbf{A}$ needs to keep his probability of being detected by $\mathbf{E}$ as minimum as possible.

The fundamental problem here is to quantify, for a given P_c, the information-theoretic limits that govern how much information can be exchanged reliably between $\mathbf{A}$ and $\mathbf{B}$ while at the same time, maintaining a certain level of probability of their communication being detected by $\mathbf{E}$, namely P_D. Since $\mathbf{A}$ does not know $\mathbf{E}$'s detection strategy, the transmission strategy should be designed against the optimal detector that $\mathbf{E}$ can deploy. In addition, we assume no prior knowledge of the exact realization of the cover-object neither at $\mathbf{B}$ nor at $\mathbf{E}$. In other words, the cover-object realization is selected at random by $\mathbf{A}$ at the time of communication.

We assume that $\mathbf{E}$ will perform a binary hypothesis test for $\mathcal{H}_0$ and $\mathcal{H}_1$, where $\mathcal{H}_0$ means $\mathbf{A}$ is not embedding any secret message whereas $\mathcal{H}_1$ means $\mathbf{A}$ does. Thus, we have two types of errors:

- Type $\mathbf{I}$ error: Deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is true, also called false positive. We denote the probability of this type of errors by α.
- Type $\mathbf{II}$ error: Deciding $\mathcal{H}_0$ when $\mathcal{H}_1$ is true, also called false negative (miss-detection) and we denote the

probability of this type of errors by β.

Assuming equal priors for $\mathbf{E}$'s optimal hypothesis test, the total probability of error P_e will be

$$P_e = 1 - P_D = \alpha + \beta = 2P_E, \tag{1}$$

where P_E is the average probability of error for $\mathbf{E}$ under equal priors. According to [26], P_e can be calculated as

$$P_e = 1 - \mathcal{V}(P_s, P_c), \tag{2}$$

where $\mathcal{V}(P_s, P_c)$ is the total variation distance between P_s and P_c, defined in [26] as

$$\mathcal{V}(P_s, P_c) = \frac{1}{2} \left\| P_s - P_c \right\|_1, \tag{3}$$

where $\|.\|_1$ is the L_1 norm. As L_1 norm's analytic calculations are not wieldy, $\mathcal{V}(P_s, P_c)$ is related to the KL-divergence as [27, Ch.11]:

$$\mathcal{V}(P_s, P_c) \leq \sqrt{\frac{1}{2}\mathcal{D}(P_s \parallel P_c)}, \tag{4}$$

where the KL-divergence is given by:

$$\mathcal{D}(P_s \parallel P_c) = \sum_{j=1}^{n} P_{s_j} \ln \frac{P_{s_j}}{P_{c_j}}. \tag{5}$$

where P_{s_j} and P_{c_j} are the bins of probability mass function for cover and stego-objects, respectively. For $\mathbf{A}$ to guarantee a low detection probability at $\mathbf{E}$'s optimal detector, $\mathbf{A}$ needs to bound $\mathcal{V}(P_s, P_c)$ by some ϵ chosen according to the desired probability of $\mathbf{E}$'s detection. Consequently, $\mathbf{A}$ ensures that the sum of error probabilities at $\mathbf{E}$ is bounded as $\alpha + \beta = 1 - \mathcal{V}(P_s, P_c)$. Using (4), $\mathbf{A}$ can achieve this goal by designing a steganographic technique such that

$$\mathcal{D}(P_s \parallel P_c) \leq 2\epsilon^2. \tag{6}$$

On the other hand, $\mathbf{B}$ must be able to correctly decode the message $\mathbf{x}$ from $\mathbf{m}$ sent by $\mathbf{A}$. Thus the chosen codebook must be selected such that P_m maximizes the mutual information between both P_s and P_m, which is defined as

$$I(P_m; P_s) = H(P_m) - H(P_m \mid P_s), \tag{7}$$

where $H(P_m)$ is the entropy of P_m and $H(P_m \mid P_s)$ is the conditional entropy of P_m given P_s. Thus, our constrained optimization problem can be written as

$$\underset{P_m}{\operatorname{argmax}} \quad I(P_m; P_s) \qquad \text{s.t.} \qquad \mathcal{D}(P_s \parallel P_c) \leq 2\epsilon^2. \tag{8}$$

Note that, $I(P_m; P_s)$ in (7) can be written as [27, Ch.2]

$$I(P_m; P_s) = H(P_s) - H(P_s \mid P_m) = H(P_s) - H(P_c), \tag{9}$$

and because $H(P_c)$ is given, maximizing $I(P_m; P_s)$ is equivalent to maximizing $H(P_s)$. Thus, the optimization problem in (8) can be restated as

$$\boxed{\underset{P_m}{\operatorname{argmax}} \quad H(P_s) \qquad \text{s.t.} \qquad \mathcal{D}(P_s \parallel P_c) \leq 2\epsilon^2}. \tag{10}$$

IV. Multimedia stochastic model

In this section, a mathematical illustration for our basic assumptions about the general statistical distribution model for the multimedia (P_c) is introduced. Consequently, for a clear illustration of the multimedia stochastic model, we need first to define the Markov-Random-Field (MRF) and Gibbs distribution.

MRF is a multidimensional random process, which generalizes the single dimensional Markov random process [9]. Let $\ddot{X}$ be a coordination system in R^N and $\rho(i)$ is a function representing the neighbourhood for each element $i \in \ddot{X}$, such that $i \notin \rho(i)$ and $i \in \rho(j)$ *iff* $j \in \rho(i)$. For example, the neighbourhood may be defined as the immediate left, right, top and bottom neighbours of i. Let $\ddot{Y}$ be the neighbourhood system representing the set of neighbourhoods of all elements $i, j, \ldots \in \ddot{X}$.

A random field ξ over $\ddot{X}$ is a multidimensional random process where each element $i \in \ddot{X}$ is assigned a random variable ξ_i with $\mathfrak{f}_i$ be its associated realization for all $i \in \ddot{X}$. ξ is called a MRF if it achieves the following conditions:

- Positivity property: $P(\xi = \mathfrak{f}) > 0$, $\forall \mathfrak{f} \in \ddot{S}$, and
- Markovianity property: $P(\xi_i = \mathfrak{f}_i | \xi_j = \mathfrak{f}_j, \forall j \neq i) = P(\xi_i = \mathfrak{f}_i | \xi_j = \mathfrak{f}_j, \forall j \in \rho(i)), \forall i \in \ddot{X}, \forall \mathfrak{f} \in \ddot{S}$,

where P is the probability measure and $\ddot{S}$ is the state space for the MRF ξ.

According to [9], Hammersley-Clifford theorem states that ξ is a MRF over $\ddot{X}$ with respect to $\ddot{Y}$ if and only if its probability distribution follows the Gibbs distribution with respect to $\ddot{X}$ and $\ddot{Y}$. To explain the Gibbs distribution, the idea of *clique* must to be clarified. A clique ω is a correlated group of neighboured elements, where $\omega \subset \ddot{X}$ with respect to $\ddot{Y}$ (i.e.: ω consists of a single element i or multiple elements $i, j, \ldots$ which are neighbours). Also, $\omega \in \ddot{\Omega}$, where $\ddot{\Omega}$ denotes the set of all cliques.

According to [7]–[9], it is accurate to statistically model the multimedia objects by the Gibbs distribution, which implies the above described properties (positivity and markovianity), in which all elements for each multimedia object are organized in *cliques*. Gibbs distribution quantifies the probability $P(\xi = \mathfrak{f})$ through an energy function $U_{\mathfrak{f}}$ and a temperature constant $\hat{T}$ as

$$P(\xi = \mathfrak{f}) = \frac{1}{Z} e^{-U_{\mathfrak{f}}/\hat{T}}, \tag{11}$$

where

$$Z = \sum_{k=1}^{M} e^{-U_k/\hat{T}}, \tag{12}$$

is the normalization denominator that is called the *partition function* and M is the number of all allowed states of the system (number of elements within $\ddot{S}$). The energy function is modelled as

$$U_{\mathfrak{f}} = \sum_{\omega \in \ddot{\Omega}} V_{\mathfrak{f}}(\omega), \tag{13}$$

where $V_{\mathfrak{f}}(\omega)$ is the potential function for $\mathfrak{f}$ calculated only within the clique $\omega \in \ddot{\Omega}$.

Gibbs distribution involves computing the partitioning function which is intractable for most models. Therefore, in this

paper, we propose an efficient approximation for the Gibbs distribution by modelling it with the *correlated-multivariate-quantized-Gaussian-distribution* (CMQGD) for more thorough mathematical analysis. One possible approximation to (11) is to model $V_f(\omega)$ in (13) as

$$V_{\mathfrak{f}}(\omega) = \frac{\hat{T}}{2}(\mathfrak{f}_\omega - \mu_\omega)\Sigma_\omega^{-1}(\mathfrak{f}_\omega - \mu_\omega)^T, \tag{14}$$

where

- $\mathfrak{f}_\omega$ is the realization of the random field ξ within the clique ω (i.e.: $\mathfrak{f}_i, \mathfrak{f}_j \in \mathfrak{f}_\omega \quad \forall i, j, \ldots \in \omega$ and $\mathfrak{f}_\omega \subset \mathfrak{f}$).
- μ_ω and Σ_ω are the mean vector for $\xi_\omega \subset \xi$ and its covariance matrix within the clique ω, respectively.

This approximation is valid for the following reasons:

- $V_{\mathfrak{f}}(\omega)$ can be mapped to any function depends only on the elements within the clique ω [9].
- The multiplication of $(\mathfrak{f}_\omega - \mu_\omega)$ and $(\mathfrak{f}_\omega - \mu_\omega)^T$, which is the distance from the mean calculated locally within the clique ω, controls the probability of certain realization (state) $\mathfrak{f}_\omega$ (and hence $P(\xi = \mathfrak{f})$) within the clique ω for certain Σ_ω. In other words, the local coherency condition of multimedia elements: $P(\xi = \mathfrak{f}) \propto -U_{\mathfrak{f}} \propto -V_{\mathfrak{f}}(\omega)$ is achieved. This allows lower energy states to be always have a higher probability than the higher energy ones, which is one of the main Gibbs properties [7].
- Z, which is the normalization denominator, is equivalent to $\sqrt{2\pi|\Sigma|}$ as $\sqrt{2\pi|\Sigma|} = \sum_{\omega \in \ddot{\Omega}} e^{-\sum \frac{\hat{T}}{2}(\mathfrak{f}_\omega - \mu_\omega)\Sigma_\omega^{-1}(\mathfrak{f}_\omega - \mu_\omega)^T}$.

It should be noted that other valid assumptions for the potential function $V_\omega(\mathfrak{f})$ can be used to approximate the Gibbs distribution with other distributions. The approximation in (14) is one of them.

V. Main Results

In this section, we discuss the solution and results of the optimization problem (10) in Sec. III in subsection V-B using the assumptions in subsection V-A.

A. The main assumptions

To solve the optimization problem (10), we have the following assumptions:

- From here on, each clique within the cover will be considered as a single cover element c_i with *unknown distribution* P_{c_i}.
- Having established the equivalence relation between Gibbs and Gaussian distributions as discussed in section IV, we model the whole cover object (*i.e.*: the joint distribution for the cover object P_c) as a CMQGD with mean $\vec{\mu_c}$ and covariance matrix Σ_c. Rather, we have no assumptions on the marginal distribution P_{c_i} of each cover element c_i.
- The cover statistical parameters ($\vec{\mu_c}$, Σ_c) are known to all $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{E}$.
- Although we have no assumptions for both P_s and P_m, according to [28], to minimize the KL-divergence for any Gaussian distributed cover object, the corresponding stego-object distribution must be Gaussian. Thus, the embedding operation will produce another CMQGD P_s with parameters ($\vec{\mu_s}$, Σ_s).
- The realization of codebook from P_m (the exact codewords used between $\mathbf{A}$ and $\mathbf{B}$) are shared only between $\mathbf{A}$ and $\mathbf{B}$ and not known to $\mathbf{E}$.

B. The Solution to the optimization problem in (10)

To solve (10), we need to determine $H(P_c)$, $H(P_s)$ and the KL-Divergence $\mathcal{D}(P_s \parallel P_c)$. Using lemma 3 in appendix (A), $H(P_c)$ and $H(P_s)$ are computed as demonstrated in (46) and (47), respectively. The KL-Divergence between two quantized distributions can be evaluated using (17) alongside with the following lemma.

Lemma 1. *According to [27, Ch.11], KL-Divergence between any two uniformly quantized distributions $F(\dot{X})$ and $G(\dot{X})$ is bounded as:*

$$\mathcal{D}(F(\dot{X}) \parallel G(\dot{X})) \leq \mathcal{D}(f(\dot{x}) \parallel g(\dot{x})) \tag{15}$$

Where f and g are the continuous versions of F and G, respectively, and $\dot{X}$ is the uniformly quantized version of the continuous random variable $\dot{x}$.

From lemma 1 and the definition of the KL-divergence for multivariate-continuous-Gaussian-distributions in [12], the KL-Divergence for CMQGD can be bounded as

$$\mathcal{D}(P_s \parallel P_c) \leq \frac{1}{2}\Bigg(\mathrm{tr}(\Sigma_c^{-1}\Sigma_s) + (\vec{\mu_c} - \vec{\mu_s})^T\Sigma_c^{-1}(\vec{\mu_c} - \vec{\mu_s}) + \ln\frac{|\Sigma_c|}{|\Sigma_s|} - n\Bigg), \tag{16}$$

where n is number of cover elements. For more conservative evaluation for our optimization problem in (10), we set $\mathcal{D}(P_s \parallel P_c)$ to its upper bound in the mathematical relations in the rest of the paper. Specifically,

$$\boxed{\mathcal{D}(P_s \parallel P_c) = \frac{1}{2}\Bigg(\mathrm{tr}(\Sigma_c^{-1}\Sigma_s) + (\vec{\mu_c} - \vec{\mu_s})^T\Sigma_c^{-1}(\vec{\mu_c} - \vec{\mu_s}) + \ln\frac{|\Sigma_c|}{|\Sigma_s|} - n\Bigg)}. \tag{17}$$

Having established the upper bound of of the KL-Divergence for CMQGD in (17), the solution of the optimization problem in (10) as provided in the following theorem.

Theorem 1. *With the assumptions stated in subsection V-A, the solution of* (10) *can be achieved by setting:*

1) *The distribution of P_m must be CMQGD (same as cover and stego-objects), with the parameters (mean $\vec{\mu_m}$ and variance Σ_m) as*
   a) $\vec{\mu_m} = 0$.
   b) $\Sigma_m = \Big(-W(-\frac{4\epsilon^2}{n} - 1 - i\pi) - 1\Big)\Sigma_c$.

   for optimal steganalysis detector with detection capabilities limited by ϵ (i.e.: $P_D \leq \sqrt{\frac{\mathcal{D}(P_s \| P_c)}{2}} \leq \epsilon$), where $W(.)$ is the WrightOmega function [29].

2) *The maximum achievable embedding rate* $\boxed{I(P_s;P_m) \text{ is } \frac{n}{2}\ln(-W(-\frac{4\epsilon^2}{n}-1-i\pi))}$ *for sufficiently large* n.

Proof:
The Lagrangian $\mathcal{L}$ of (10) will be

$$\mathcal{L}(P_s,\lambda)=\mathcal{L}(\vec{\mu_s},\Sigma_s,\lambda)=H(P_s)-\lambda(\mathcal{D}(P_s\parallel P_c)-2\epsilon^2), \tag{18}$$

where λ is the Lagrange multiplier. To find the optimized parameters for (18), we set $\frac{\partial}{\partial\vec{\mu_s}}\mathcal{L}=0$ and $\frac{\partial}{\partial\Sigma_s}\mathcal{L}=0$. With the definitions of $H(P_s)$ and $\mathcal{D}(P_s\parallel P_c)$ in (47) and (17), respectively, we have

$$\frac{\partial}{\partial\vec{\mu_s}}H(P_s)=0, \tag{19}$$

$$\frac{\partial}{\partial\Sigma_s}H(P_s)\approx\frac{1}{2}(\Sigma_s^{-1})^T, \tag{20}$$

$$\frac{\partial}{\partial\vec{\mu_s}}\mathcal{D}(P_s\parallel P_c)=\frac{-1}{2}\Big(\Sigma_c^{-1}(\vec{\mu_c}-\vec{\mu_s})+(\Sigma_c^{-1})^T(\vec{\mu_c}-\vec{\mu_s})\Big), \tag{21}$$

$$\frac{\partial}{\partial\Sigma_s}\mathcal{D}(P_s\parallel P_c)=\frac{1}{2}\big(\Sigma_c^{-1}-\Sigma_s^{-1}\big). \tag{22}$$

As Σ_c and Σ_s are symmetric matrices, (21) can be rewritten as

$$\frac{\partial}{\partial\vec{\mu_s}}\mathcal{D}(P_s\parallel P_c)=-\Sigma_c^{-1}(\vec{\mu_c}-\vec{\mu_s}). \tag{23}$$

From (19) and (23), as $\frac{\partial}{\partial\vec{\mu_s}}\mathcal{L}=0$, we have

$$\vec{\mu_s}=\vec{\mu_c}. \tag{24}$$

From (20) and (22), as $\frac{\partial}{\partial\Sigma_s}\mathcal{L}=0$, we have

$$\begin{aligned}(\lambda+1)\Sigma_s^{-1}&\approx\lambda\Sigma_c^{-1},\\ \Sigma_s&\approx\frac{\lambda+1}{\lambda}\Sigma_c,\\ \Sigma_s&\approx a\Sigma_c.\end{aligned} \tag{25}$$

Through the paper, we name the factor $a=\frac{\lambda+1}{\lambda}$ as the *embedding factor*. From (25), it is clear that Σ_s has the exact eigenvectors as Σ_c and the eigenvalues of Σ_s are a scaled version of the eigenvalues of Σ_c with the embedding factor a. Let $\psi_c(\vec{t})$ and $\psi_s(\vec{t})$ be the characteristic functions for both the cover and the stego-objects where

$$\begin{aligned}\psi_c(\vec{t})&=e^{i\vec{\mu_c}^T\vec{t}+\frac{1}{2}\vec{t}^T\Sigma_c\vec{t}},\\ \psi_s(\vec{t})&=e^{i\vec{\mu_s}^T\vec{t}+\frac{1}{2}\vec{t}^T\Sigma_s\vec{t}}.\end{aligned}$$

As $\mathbf{s}=\mathbf{c}+\mathbf{m}$, this implies $P_s=P_c\oplus P_m$, where $\oplus$ represents the convolution operation. Thus, the characteristic function for the distribution of codebook of the message ($\psi_m(\vec{t})$) will be

$$\psi_m(\vec{t})=\frac{\psi_s(\vec{t})}{\psi_c(\vec{t})}=e^{i(\vec{\mu_s}^T-\vec{\mu_c}^T)\vec{t}+\frac{1}{2}\vec{t}^T(\Sigma_s-\Sigma_c)\vec{t}}.$$

Then from (24), we have

$$\psi_m(\vec{t})=e^{\frac{1}{2}\vec{t}^T(\Sigma_s-\Sigma_c)\vec{t}}=e^{\frac{1}{2}\vec{t}^T(\Sigma_m)\vec{t}}. \tag{26}$$

Thus

$$\boxed{\mu_m=0}. \tag{27}$$

This means that, the codebook of the message must be modelled as CMQGD with zero mean and variance

$$\begin{aligned}\Sigma_m\approx&\Sigma_s-\Sigma_c\\ =&\boxed{(a-1)\Sigma_c}.\end{aligned} \tag{28}$$

The embedding factor a can be calculated as follows. Starting from (25), we have

$$\mathrm{tr}(\Sigma_c^{-1}\Sigma_s)\approx na, \tag{29}$$

and

$$\ln\frac{|\Sigma_c|}{|\Sigma_s|}\approx-n\ln(a). \tag{30}$$

Substituting from (24), (29) and (30) into (17), we get

$$\begin{aligned}an-n\ln(a)&\approx 2\mathcal{D}(P_s\parallel P_c)+n,\\ a-\ln(a)=&\approx\frac{2}{n}\mathcal{D}(P_s\parallel P_c)+1.\end{aligned} \tag{31}$$

The solution of (31) can be obtained from [29] as

$$a\approx-W(-\frac{2}{n}\mathcal{D}(P_s\parallel P_c)-1-i\pi). \tag{32}$$

From (32) and according to [29], a is monotonically increasing with $\mathcal{D}(P_s\parallel P_c)$. From (25) and (47), a is monotonically increasing with $H(P_s)$. Thus, using the constraint in (10) by substituting $\mathcal{D}(P_s\parallel P_c)$ with $2\epsilon^2$ in (18) we have

$$\boxed{a^*=-W(-\frac{4\epsilon^2}{n}-1-i\pi)}, \tag{33}$$

where $a^*\in\mathbb{R}$ is the embedding factor[2] associated with the design parameter ϵ. Thus, (27), (28) and (33) complete the proof of the first part of Theorem 1. ■

From (9), we have:

$$I(P_s;P_m)=H(P_s)-H(P_c).$$

Then from (46) and (47):

$$I(P_s;P_m)\approx\frac{1}{2}\ln(2\pi e|\Sigma_s|)-\frac{1}{2}\ln(2\pi e|\Sigma_c|).$$

Then, from (28) and (33):

$$\begin{aligned}I(P_s;P_m)&\approx\frac{n}{2}\ln(a^*)+\frac{1}{2}\ln(2\pi e|\Sigma_c|)-\frac{1}{2}\ln(2\pi e|\Sigma_c|)\\ &=\frac{n}{2}\ln(a^*).\end{aligned} \tag{34}$$

Thus

$$I(P_s;P_m)\approx\frac{n}{2}\ln\Big(-W(-\frac{4\,\epsilon^2}{n}-1-i\pi)\Big). \tag{35}$$

Equation (35)[3] completes the proof of the second part of Theorem 1. ■

Please note that, despite $\mathcal{D}(P_s\parallel P_c)\neq\mathcal{D}(P_c\parallel P_s)$, the results obtained from Theorem 1 are also valid for $\mathcal{D}(P_c\parallel P_s)$. The proof of this claim in provided in the following lemma.

[2]We proved in Lemma 4 that $a^*\in\mathbb{R}$.

[3]It should be noted that for continuously distributed covers such as [11] and [12], the approximations in (15), (46), and (47) will be changed to equality. Consequently, for such cases, (35) will be valid with equality.

Lemma 2. *$\vec{\mu_m}$ and Σ_m obtained from Theorem 1 are also valid for $\mathcal{D}(P_c \parallel P_s)$.*

Proof: Equation (17) can be modified as:

$$\mathcal{D}(P_c \parallel P_s) = \frac{1}{2}\bigg(\mathrm{tr}(\Sigma_s^{-1}\Sigma_c) + (\vec{\mu_s} - \vec{\mu_c})^T\Sigma_s^{-1}(\vec{\mu_s} - \vec{\mu_c}) + \ln\frac{|\Sigma_s|}{|\Sigma_c|} - n\bigg). \tag{36}$$

Consequently, (22) and (23) will be modified as:

$$\frac{\partial}{\partial\Sigma_s}\mathcal{D}(P_c \parallel P_s) = \frac{1}{2}\big(-(\Sigma_s^{-1})^T\Sigma_c(\Sigma_s^{-1})^T + (\Sigma_s^{-1})^T\big), \tag{37}$$

$$\frac{\partial}{\partial\vec{\mu_s}}\mathcal{D}(P_c \parallel P_s) = \Sigma_s^{-1}(\vec{\mu_s} - \vec{\mu_c}). \tag{38}$$

Following the same steps of the proof of Theorem 1 using (19), (20), (37), and (38), the results presented at (27), (28), (32), and (33) will be the same as Theorem 1. ■

VI. Achievability

In this section, we prove that the maximum embedding rate $I(P_s; P_m)$ in (35) is achievable with a low probability of decoding error P_B at $\mathbf{B}$. We couldn't include the achievablity constraint in the optimization problem in (10) and (18) as P_s and P_m will not be obtained till solving the optimization problem.

We utilize the *standard random coding argument* [27, Ch.7] as a base for our proof. Although the minimum distance decoder (maximum likelihood estimator) is the optimal detector [27, Ch.7], there is technical difficulties in using this detector to calculate P_B as we do not have the marginal distribution for each cover element; only the joint distribution for the whole cover. Thus, we utilize the *jointly typical decoder* [27, Ch.7] at $\mathbf{B}$ instead.

In our achievability proof, we utilize the same procedures for the Channel Capacity Theorem (Theorem (9.1.1) in [27]) for calculating P_B. Although this theorem mandates the constraint for the average transmission power, in our case we do not need this power constraint as Σ_m is very small due to the constraint of low probability of detection at $\mathbf{E}$ in (10). Without loss of generality, assuming transmitting the i^{th} codeword, we have two types of errors

- The transmitted codeword and the received one are not jointly typical. We denote this error by $\hat{E}_i$.
- The received codeword is jointly typical with another wrong non-intended codeword. This error is denoted by $\sum_{j=1,j\neq i}^{\mathcal{K}} E_j$, where $\mathcal{K}$ is the number of all codewords.

Thus, equation (9.26) in [27] will be modified as

$$P_B \leq P(\hat{E}_i) + \sum_{j=1,j\neq i}^{\mathcal{K}} P(E_j), \tag{39}$$

where $P(x)$ is the probability of an event x. Continuing the same procedures for the proof of Theorem (9.1.1) in [27]:

$$P_B \leq 2\delta, \tag{40}$$

where $\delta > 0$ [27, Ch.3] is an arbitrary small number. Thus, (40) prove that when n is sufficiently large and for an actual embedding rate $R \leq I(P_s; P_m) - 2\epsilon$, then P_B is sufficiently small. This proves the existence of a codebook ($\mathcal{K}$,n,ϵ) that achieves an embedding rate $R \leq I(P_s; P_m) - 2\epsilon$ at $\mathbf{B}$ with low P_B.

It should be noted that although related approaches such as [11] and [12] utilize a converse proof, in our case we have dropped the converse proof as we have got the maximum embedding rate as a solution for the optimization problem in (10), which guarantees that there is no other codebook generated by any other distribution can achieve higher rate than (35).

VII. Relation to the square root law of steganography

In this section, we establish the relation between the obtained expressions in (33) and (35) in terms of the WrightOmega function and the well known previously obtained results of the SRL in [4], [11], [12]. To do so, we state the following facts about the WrightOmega function:

- As the WrightOmega function in the form $-W(-\gamma - 1 - i\pi)$ is monotonically increasing with γ; where γ is a positive constant; $-W(-\gamma-1-i\pi)$ can be approximated by polynomial regression in the form
$$a^* = -W(-\gamma - 1 - i\pi) = 1 + \theta_1\gamma + \theta_2\gamma^2 + \theta_3\gamma^3 + \ldots, \tag{41}$$
where θ_i is an arbitrary constant.
- As $\gamma = \frac{4\epsilon^2}{n}$, then for large n we can ignore the higher order terms in (41) and use the following approximation[4]
$$a^* \approx 1 + \theta_1\frac{4\epsilon^2}{n}. \tag{42}$$
- From Lemma 5, as $a^* \geq 1$, then $\sqrt{a^*} \geq 1$. Also, as $(1+x)^{\frac{1}{2}} \leq 1 + x^{\frac{1}{2}}$ and $\ln(1+x) \leq x$, then the following approximation of (34) is valid
$$\begin{aligned} I(P_s; P_m) &\approx n\ln(\sqrt{a^*}) \\ &\approx n\ln\Big(\sqrt{1 + \frac{4\epsilon^2}{n}}\Big) \\ &\leq n\ln\big(1 + \frac{2\epsilon}{\sqrt{n}}\big) \\ &\leq n\frac{2\epsilon}{\sqrt{n}}, \\ \boxed{I(P_s; P_m) \leq 2\epsilon\sqrt{n}}&. \end{aligned} \tag{43}$$

Thus, (43) establishes the relation between (35) and the SRL. Hence, (43) proves that our results come in agreement with previously established SRL in [4] in the context of image steganography and [11], [12] in the context of covert communication. Using the justifications in section IV for approximating the Gibbs distribution with CMQGD and the proved relation to the SRL in (43), equation (35) can be considered as a rigorous analytic form for the SRL. ■

To illustrate the above results, and to prove the relation between P_E and both (35) and (43), we plot in Fig.2 the maximum embedding rate $I(P_s; P_m)$ against the number of

[4]It should be noted that θ_1 must be a positive constant according to (50).

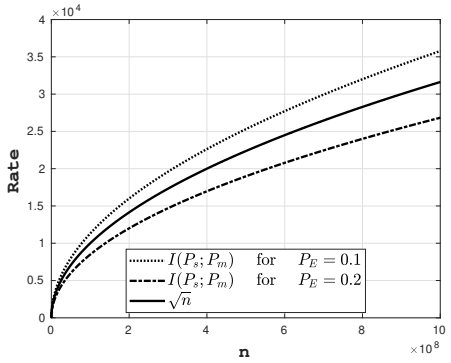

Fig. 2: Maximum achievable embedding rate $I(P_s; P_m)$ compared to the number of cover elements (n) for $P_E = 0.1$ and $P_E = 0.2$. Note that, we use the lower bound for a^* in (52) to compute $I(P_s; P_m)$ in (34).

the cover elements n using (34) and (52). In the figure, we use the lower bound for a^* in (52), i.e., we set $a^* = -W(-\frac{4\,(1-2P_E)^2}{n} - 1 - i\pi)$ to compute $I(P_s; P_m)$ in (34). We provided the plots for $P_E = 0.1$ and $P_E = 0.2$ and we used the implementation of the wrightOmega function in [30] to generate the plots. From the figure, we can conclude that (41), (42), and (43) are not over-approximated form of (35) as we plot the actual analytic form expression in (35) against $\sqrt{n}$.

VIII. Practical interpretation

In this section, we compare the practical experimental results for the stenographic methods in [10], [16], and [17] with our theoretically calculated limits. Specifically, we get the published results of each steganographic method that present the payload (bits per pixel) of the method against different P_E, where P_E is obtained using different steganalysis methods. Then, we plot these published results against our theoretical limit of the payload calculated using equation (35). In our comparison, we used the BOSSbase 1.01 dataset [31] where the size of each image is 512×512 pixels. We set $n = 512 \times 512 = 2^{18}$ in (35) to calculate the maximum achievable embedding rate that occurs in the case of independent cover elements, i.e. when each clique represents only a single pixel. We give more clarification for the relation between the achievable embedding rate and the clique size in Section IX.

Figures 3 to 6 demonstrate the experimental results obtained from [16] plotted against the theoretical upper limit for the payload in (34) and (52). We set a^* to its lower bound in (52) to compute the theoretical upper limit for the payload against P_E and we plot the payload in log-scale. We also demonstrate in the supplementary material figures S-1 to S-4 and figures S-5 to S-11 for the methods in [17] and [10], respectively.

It can be concluded from the figures that our theoretically calculated upper limit is relatively very small compared to the referenced practical stenographic methods. The reason for this is that there exist other steganalysis methods that can be more optimum than the methods utilized by referenced stenographic methods. In other words, the steganalysis methods used in [10], [16], [17] can be regarded as non-optimal compared to the KL divergence $\mathcal{D}(P_s \parallel P_c)$ that is used in our proof

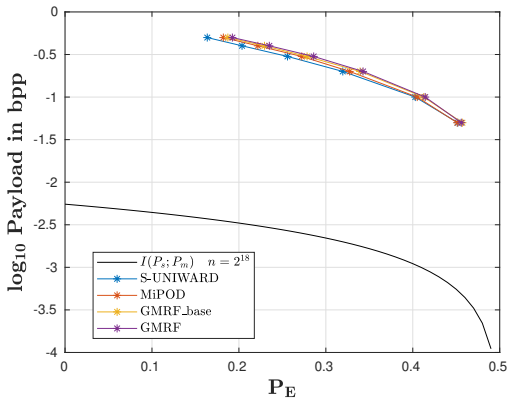

Fig. 3: Results from [16] comparing steganographic methods: S-UNIWARD, MIPOD, GMRF_BASE and GMRF with steganalyzer utilizing SRM feature.

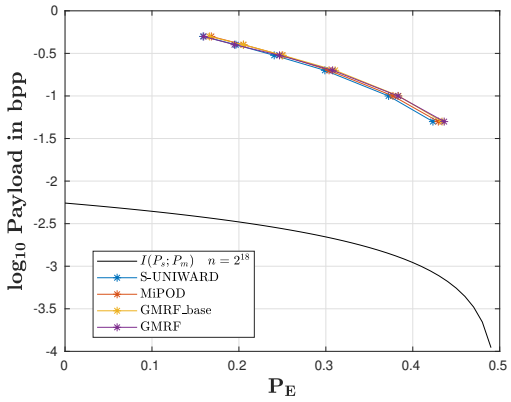

Fig. 4: Results from [16] comparing steganographic methods: S-UNIWARD, MIPOD, GMRF_BASE and GMRF with steganalyzer utilizing maxSRMd2 feature.

and can be regarded as the upper bound for any steganalysis detector. Consequently, using non-optimal steganographic detectors may be misleading as it can achieve a lower probability of detection error (i.e. P_E) when using higher embedding rates than our theoretically calculated limit. Likewise, using non-optimal embedding techniques may also be misleading as it can achieve lower embedding rates than the theoretically calculated limit with the same detection probability.

IX. Discussion

As $I(P_s; P_m) \propto \sqrt{n}$ as shown in (43), the upper limit for embedding rate occurs when each clique contains a single element, i.e. when n is large. This occurs in cases such as a highly textured noisy image or a motion field of a sand storm video. Meanwhile, the lowest embedding rate occurs when all the elements of the cover are fully correlated (*i.e.*: the whole cover can be considered as a single clique where $n = 1$), such as an image of a clear sky or a motion field of a video of the stationary scene with global camera motion. These cases imply relatively a small number of cliques. In other words, the upper limit occurs in the situation when all elements of the cover are independent, whereas the lowest one occurs in the case of the cover contains only a single clique.

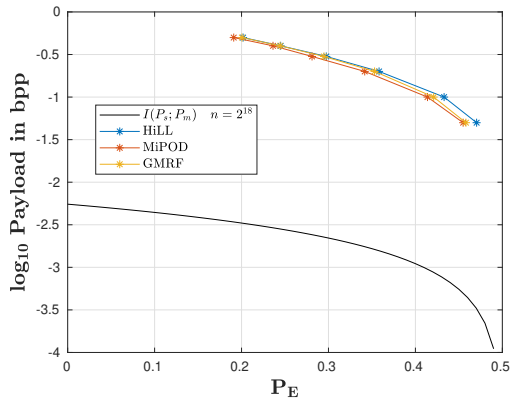

Fig. 5: Results from [16] comparing steganographic methods: MiPOD, HILL and GMRF enhanced by low-pass-filtered-cost method with steganalyzer utilizing SRM feature.

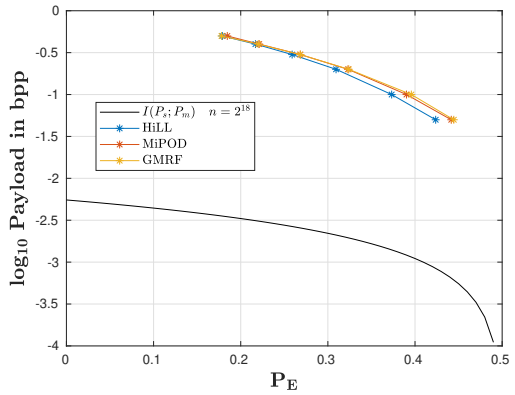

Fig. 6: Results from [16] comparing steganographic methods: MiPOD, HILL and GMRF enhanced by low-pass-filtered-cost method with steganalyzer utilizing maxSRMd2 feature.

Another important point is that using the information-theoretic upper bound for $I(P_s;P_m)$ in (35) is not a practical constructive approach due to the *random coding argument* assumption. In other words, it does not introduce concrete guidelines for how to design steganographic methods or codebooks to achieve this theoretic upper bound. Our work introduces a proof of existence in analytical form, implying that work is still needed to construct steganographic methods to be more closed to the calculated bounds.

X. Conclusion

Through this paper, we have calculated the information-theoretic upper bound for a predefined level of security for data embedding within a generalized case for multimedia covers. Our mathematical analysis has several advantages over previous ones. First, we introduced a mathematical justification for using CMQGD as a replacement for the accurate and mathematically intractable Gibbs distribution. Second, we provided a rigorous analytic form for the SRL, not an order result. Third, the provided analytic form is calculated not only in the case when the cover is known for both sender $\mathbf{A}$ and receiver $\mathbf{B}$, but also in the real-world case when $\mathbf{B}$ doesn't know the exact cover, only knows the cover distribution, thanks to the achievability proof in section VI. Forth, our calculated limit is applied to all types of stegnanalytic detectors (statistical, deep-learning, feature-based, ... *etc*) for any type of multimedia. Five, we have introduced the model parameters for the optimal message's codebook in an analytic form. It should be noted that for continuously distributed covers, our mathematical solutions will be exact closed-form solutions.

Appendix

A. The Entropy of Multivariate Quantized Gaussian Distribution

Lemma 3. *Theorem (8.3.1) in [27] defines the entropy of quantized distribution as*

$$H(P) \approx h(p) + b, \tag{44}$$

where p is any continuous distribution, P is a quantized version of p with b quantization bits, $H(P)$ is the entropy of the quantized distribution and $h(p)$ is the differential entropy of the continuous distribution. Thus, as the entropy of CMQGD p_g is defined in [12] as

$$h(p_g) = \frac{1}{2}\ln(2\pi e|\Sigma_g|), \tag{45}$$

the entropy of both P_c and P_s can be defined w.r.t their continuous versions P_c, P_s as

$$H(P_c) \approx \frac{1}{2}\ln(2\pi e|\Sigma_c|) + b, \tag{46}$$

$$H(P_s) \approx \frac{1}{2}\ln(2\pi e|\Sigma_s|) + b. \tag{47}$$

B. Additional properties about the embedding factor a^*

Lemma 4. *a^* cannot be a complex number:*
It should be noted that $a^* \in \mathbb{R}$ despite $W(.) \in \mathbb{C}$. From [29], the relation between the Lambert W Function $\mathcal{W}(.)$ and the WrightOmega function $W(.)$ can be modelled as

$$W(x) = \mathcal{W}_{\left\lceil \frac{\mathrm{Im}(x)-\pi}{2\pi} \right\rceil}(e^x). \tag{48}$$

If we set $x = -\frac{4\epsilon^2}{n} - 1 - i\pi$ as in (33)*, then, $\mathrm{Im}(x) = \pi$ and therefore* (48) *can be re-written as*

$$W(x) = \mathcal{W}_{-1}(e^x), \tag{49}$$

which is a special case of Lambert W Function where $\mathcal{W}_{-1}(.) \in \mathbb{R}$ [32]. Thus, $a* \in \mathbb{R}$. The same applies for a in (32)*.*

Lemma 5. *a^* cannot be less than 1:*
As Σ_s, Σ_c and Σ_m are positive semi-definite matrices, then from (28) *a and a^* must be greater than 1, with equality when no embedding occurs, as explained in* (25) *and* (28)*. Thus:*

$$a \geq 1. \tag{50}$$

Lemma 6. *Relation between P_E and a^*:*
Using (1) *through* (4) *and* (6)*, we have:*

$$2\left(1 - 2P_E\right)^2 \leq \mathcal{D}(P_s \parallel P_c) \leq 2\epsilon^2, \tag{51}$$

Then, as the WrightOmega function in the form $-W(-\gamma - 1 - i\pi)$ *is monotonically increasing with* γ *[29], where* γ *is a positive constant, we can prove that*

$$-W(-\frac{4\,(1-2P_E)^2}{n}-1-i\pi) \leq a \leq -W(-\frac{4\epsilon^2}{n}-1-i\pi)$$
$$-W(-\frac{4\,(1-2P_E)^2}{n}-1-i\pi) \leq a \leq a^*, \tag{52}$$

by utilizing equations (32)*,* (33) *and* (51)*.*

REFERENCES

[1] P. Joseph and S. Vishnukumar, "A Study On Steganographic Techniques," in *Global Conference on Communication Technologies (GCCT)*, 2015, pp. 206–210.

[2] H. A. Aly, "Data Hiding in Motion Vectors of Compressed Video Based on Their Associated Prediction Error," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 14–18, March 2011.

[3] A. Westfeld, "F5—A Steganographic Algorithm," in *Information Hiding*, I. S. Moskowitz, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 289–302.

[4] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, Dec 2009.

[5] J. M. Jenifer, S. R. Ratna, J. S. Loret, and D. M. Gethsy, "A survey on different video steganography techniques," in *International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 627–632.

[6] Y. Tew and K. Wong, "An Overview of Information Hiding in H.264/AVC Compressed Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 305–319, Feb 2014.

[7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, Nov 1984.

[8] E. Dubois and J. Konrad, "Motion Estimation And Motion-Compensated Filtering Of Video Signals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 1993, pp. 95–98 vol.1.

[9] J. Konrad, "CHAPTER 3 - Motion Detection and Estimation," in *The Essential Guide to Video Processing*, A. Bovik, Ed. Boston: Academic Press, 2009, pp. 31 – 67.

[10] M. Sharifzadeh, M. Aloraini, and D. Schonfeld, "Adaptive Batch Size Image Merging Steganography and Quantized Gaussian Image Steganography," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 867–879, 2020.

[11] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of Reliable Communication with Low Probability of Detection on AWGN Channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, Sep 2013.

[12] A. Abdelaziz and C. E. Koksal, "Fundamental limits of covert communication over MIMO AWGN channel," in *IEEE Conference on Communications and Network Security (CNS)*, Oct 2017, pp. 1–9.

[13] V. Sedighi, J. Fridrich, and R. Cogranne, "Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model," in *Media Watermarking, Security, and Forensics 2015*, A. M. Alattar, N. D. Memon, and C. D. Heitzenrater, Eds., vol. 9409, International Society for Optics and Photonics. SPIE, 2015, pp. 144 – 156.

[14] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-Adaptive Steganography by Minimizing Statistical Detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[15] J. Fridrich and J. Kodovský, "Multivariate Gaussian model for designing additive distortion for steganography," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 2949–2953.

[16] W. Su, J. Ni, X. Hu, and J. Fridrich, "Image Steganography with Symmetric Embedding using Gaussian Markov Random Field Model," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2020.

[17] Y. Tong, J. Ni, and W. Su, "Image Steganography Using an Eight-Element Neighborhood Gaussian Markov Random Field Model," in *Digital Forensics and Watermarking*, H. Wang, X. Zhao, Y. Shi, H. J. Kim, and A. Piva, Eds. Cham: Springer International Publishing, 2020, pp. 247–255.

[18] H. Y. El-Arsh, A. Abdelaziz, A. Elliethy, and H. A. Aly, "Fundamental Limits Of Steganographic Capacity For Multivariate-Quantized-Gaussian-Distributed Multimedia," in *IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 1246–1250.

[19] T. Pevný, T. Filler, and P. Bas, "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography," in *Information Hiding*, R. Böhme, P. W. L. Fong, and R. Safavi-Naini, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 161–177.

[20] V. Holub, J. Fridrich, and T. Denemark, "Universal Distortion Function For Steganography In An Arbitrary Domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, p. 1, Jan 2014.

[21] B. Li, M. Wang, J. Huang, and X. Li, "A New Cost Function For Spatial Image Steganography," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4206–4210.

[22] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for Steganalysis of digital images," in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 48–53.

[23] J. Kodovsky and J. Fridrich, "Steganalysis of JPEG Images Using Rich Models," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 8303, pp. 7–, 02 2012.

[24] J. Jia, Z. Xiang, L. Wang, and Y. Xu, "An Adaptive JPEG Double Compression Steganographic Scheme Based on Irregular DCT Coefficients Distribution," *IEEE Access*, vol. 7, pp. 119 506–119 518, 2019.

[25] D. Xu, "Commutative Encryption and Data Hiding in HEVC Video Compression," *IEEE Access*, vol. 7, pp. 66 028–66 041, 2019.

[26] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

[27] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.

[28] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, "Gaussian Signalling for Covert Communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3542–3553, 2019.

[29] R. M. Corless and D. J. Jeffrey, "The Wright Omega Function," in *Proceedings of the Joint International Conferences on Artificial Intelligence, Automated Reasoning, and Symbolic Computation*, ser. AISC '02/Calculemus '02. Berlin, Heidelberg: Springer-Verlag, 2002, p. 76–89.

[30] A. Horchler. Complex double-precision evaluation of the Wright omega function, a solution of W+LOG(W) = Z. [Online]. Available: https://github.com/horchler/wrightOmegaq

[31] P. Bas, T. Filler, and T. Pevný, "Break Our Steganographic System: The Ins and Outs of Organizing BOSS," in *Information Hiding*, T. Filler, T. Pevný, S. Craver, and A. Ker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 59–70.

[32] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the LambertW function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, Dec 1996.