Secrecy Analysis of Finite-Precision Compressive Cryptosystems

Secrecy Analysis of Finite-Precision Compressive Cryptosystems

Matteo Testa, Tiziano Bianchi, and Enrico Magli

***Abstract*—Compressed Sensing (CS) has recently emerged as an effective and efficient way to encrypt data. Under certain conditions, it has been shown to provide some secrecy notions. In theory, it could be considered to be a perfect match for constrained devices needing to acquire and protect the data with computationally cheap operations. However, theoretical results on the secrecy of compressive cryptosystems only hold under the assumption of infinite precision representation. With this work, we aim to close this gap and lay the theoretical foundations to support this practical framework. We provide theoretical upper bounds on the distinguishability of the measurements acquired through finite precision sensing matrices and experimentally validate them. Our main result is that the secrecy of a CS cryptosystem can be exponentially increased with a linear increase in the representation precision. This result confirms that CS can be an effective secrecy layer and provides tools to use it in practical settings.**

***Index Terms*—Compressed Sensing, Compressive cryptosystem, Quantization, Finite-precision, Secrecy, worst-case bounds.**

I. Introduction

A. Motivation

Compressed Sensing (CS) [1], [2] has been extensively studied over the last decade as an attractive way to perform dimensionality reduction. According to CS theory, signal acquisition and compression can be jointly performed by means of random projections allowing for sub-Nyquist acquisition rates [3], [4]. In more detail, a K-sparse signal, i.e. a signal with K non-zero entries, can be exactly recovered with overwhelming probability from its random linear measurements if some assumptions on the sensing matrix are satisfied [2], [4], [5]. The ability of performing low complexity and low energy consumption acquisition is one of the main reasons behind the rise of CS in the last years. Works such as [6]–[8] showed its advantages over traditional acquisition methods. Furthermore, in [9] the authors consider the specific application of CS to different Internet of Things (IoT) scenarios.

It has then become evident that, because of its structure, CS could also provide some notions of secrecy [10]. If the sensing matrix is not known, the original signal cannot be recovered and the CS framework acts as a private key cryptosystem. In this regard, the sensing matrix entries are the secret key, thus only shared among trusted parties, the original signal is the plaintext and the measurements are the ciphertext; the encryption is performed by means of CS acquisition while the decryption corresponds to CS recovery.

M. Testa, T. Bianchi and E. Magli are with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, 10129 ITALY e-mail: {name.surname}@polito.it

This means that all applications making use of CS can also provide some kind of privacy, with little or no added cost. This is extremely advantageous for the wide range of low complexity devices which may acquire sensitive data, e.g. in the internet of things scenario, and might not be able to cope with standard encryption schemes such as AES [11]. Moreover, these devices are oftentimes heavily constrained in terms of representation precision, see e.g. [12]. Furthermore, it has to be highlighted that any practical system in which the sensing matrix has to be stored, automatically implies that its entries are represented with finite precision. If the sensing matrix entries are sub-Gaussian, then also its finite-precision counterpart will follow a sub-Gaussian distribution. In turn, this implies that a sparse or compressible signal acquired with such sensing matrix and enough measurements can be recovered with high probability. Nevertheless, if we focus on the security aspects, it is not clear how fast the secrecy may decrease when the representation precision is reduced. This indeed motivates us to explore the secrecy of CS cryptosystems exploiting sensing matrices under finite-precision representation.

B. Our contribution

Starting from the preliminary experimental results on the secrecy of finite-precision sensing matrices presented in [13], with this work we formalize and extend the results from a information-theoretic perspective. This will allow us to address the gap existing in literature between the secrecy provided by CS cryptosystems exploiting sensing matrices under infinite precision representations and more practical scenarios requiring finite precision. To the best of our knowledge, this is the first work in literature dealing with finite precision representations lying on a larger-than-binary alphabet.

More specifically, we prove that a CS cryptosystem exploiting quantized and truncated Gaussian entries can achieve a secrecy which is a function of the available representation precision. Next, we show that the same result can also be extended to the case of sensing matrices with i.i.d. entries drawn from a discrete Gaussian distribution.

More in detail, we derive upper bounds for the secrecy (in terms of θ-distinguishability [14]) in the worst case scenario of cryptosystems employing 1) quantized and 2) discrete Gaussian sensing matrices. If equal energy signals are considered, then we show that there exists a regime condition for which the secrecy exponentially increase with the number of bits employed for the quantization of the sensing matrix entries. On the other hand, in case of signals having arbitrary energy, we

prove that the secrecy is not only related to the representation precision of the sensing matrix entries, but it also depends on the energy mismatch and derive an upper bound on this latter term.

Lastly, the validity of the obtained bounds is evaluated through extensive experimental simulations.

As a remark, it is important to consider that the results of this paper can also be exploited in works which rely on known secrecy properties of specific sensing matrices constructions. As an example, in [15], where the focus is put on practical and secure sensing matrix generation schemes, the results are based on the assumption of using sensing matrices for which the secrecy has already been proven.

C. Relation to prior work

The seminal paper of Rachlin and Baron [10] was the first study related to the secrecy capabilities of CS. Further studies such as [16] focused on the asymptotic secrecy properties of compressive encryption, showing that measurements of equal energy signals become indistinguishable as the size of the original signals tends to infinity. Non-asymptotic analysis of the distinguishability of measurements sampled from Gaussian i.i.d. sensing matrices was carried out in [14]. In this latter work, the authors show that normalizing to unit energy the signal leads to perfect secrecy under the assumption of one time sensing (OTS) acquisition, i.e. the sensing matrix is re-generated at each encryption. A similar analysis was also extended to the case of circulant sensing matrices in [17], where the authors characterize the increased information leakage of the measurements due to the structured nature of the sensing matrix. Further, a more comprehensive review on the secrecy properties of different classes of sensing matrices and signals was given in [18].

From all the above works, it has emerged that, because of the linearity of the sensing process, the measurements will always reveal *at least* the energy of the original signal. The best case, in which only the original signal energy is leaked in the non-asymptotic case, is that of sensing matrices made of real-valued Gaussian i.i.d. entries. In order to overcome this problem, proposed solutions consider either to normalize the signal to unit energy as in [14] or to obfuscate the energy as in [19]. In this latter work, the authors propose a method to obfuscate the energy of the original signal through scalar multiplication, avoiding the encryption and transmission burden related to the transmission of plaintext energy. Interestingly, the authors show that this method also allows trusted parties to perform basic signal processing operations in the encrypted domain, i.e. anomaly detection. Differently, in [20] the author considered the effect of the energy mismatch in case of compressive cryptosystems employing both Gaussian and Bernoulli sensing matrices. In more detail, they obtain Total Variation bounds on the measurements distinguishability and show that in the case of signals with unequal energy, the measurements are nearly distinguishable based on the system parameters.

Other works shifted the emphasis on the encryption models. In [21], similarly to standard private key cryptosystems, modes of operation for compressive encryption are introduced which, along with the use of Bernoulli sensing matrices, make the considered scenario suitable for practical implementations. On a similar line, in [22] different encryption models, including a model based on optical imaging, are discussed. In addition, the authors also show how a practical parallel compressed sensing scheme with random permutation can achieve asymptotical spherical secrecy. Lastly, in [23] the authors discuss how the secrecy notions of CS can be effectively used in applications such as multimedia, cloud computing and IoT.

Regarding practical systems, other works targeted a specific case of practical sensing matrix: the one made of Bernoulli distributed entries. As an example, in [16], [24], the authors also consider Bernoulli sensing matrices and prove their asymptotical spherical secrecy. However, higher dimensional finite alphabets were only considered in [13] from an experimental perspective.

D. Organization

The remainder of this paper is organized as follows. In Section II we provide some background and notation about compressed sensing, the finite-precision Gaussian distributions and the secrecy metrics we will adopt for the rest of the paper. The main results of this work are presented in Section III. The experimental results and their discussion are shown in Section IV.

II. Background

A. Compressed Sensing

Given a signal $\mathbf{x} \in \mathbb{R}^{n\times 1}$ being K-sparse, thus having at most K non-zero entries, i.e. $||\mathbf{x}||_0 \leq K$, and $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ with $m \ll n$ being the sensing matrix, then the CS acquisition process can be written as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where $\mathbf{y}$ is the measurements vector. As long as $m \geq 2K$, we have that $\mathbf{x}$ can be exactly recovered from $\mathbf{y}$ by solving the following minimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_0 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\mathbf{x}.$$

However, the above problem is NP-hard. To overcome this problem, it can be shown that if sensing matrix entries are i.i.d. drawn from a sub-Gaussian distribution, then we can recover the original signal with overwhelming probability by relaxing the ℓ_0-norm by the ℓ_1-norm as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{2}$$

for sufficiently large m. It is worth noting that different algorithms designed to solve the above problem are available in the literature. Among the most efficient ones, which follow a greedy approach, we can identify the Orthogonal Matching Pursuit [25] and CoSaMP [26], and thresholding algorithms such as [27]–[29]. Furthermore, if higher reconstruction accuracy is needed, though more computationally demanding, convex-based algorithms which solve (2) can be employed, see [30]–[33]. Lastly, let us highlight that oftentimes, natural

signals are *compressible* in some basis rather than exactly K-sparse in the original domain. The compressibility implies that, in some basis, the coefficients of the signal, when sorted, rapidly decrease to values close to zero; hence can be well approximated by a K-sparse signal. Even though in the case of compressible signals there is not guarantee to *exactly* recover the original signal, an accurate recovery is possible for fast decaying coefficients and enough measurements. Although compressible signals are of interest in many applications, for the sake of simplicity, in the following we will specifically consider the case of K-sparse signals. Nevertheless, in Sec. III we provide an intuitive explanation of the behavior of such compressible signals within the considered cryptosystems.

B. Finite-precision Gaussian distributions

As will become clearer in the following, in order to derive our results at first we need to consider the statistical distance between a quantized Gaussian distribution and a discrete Gaussian distribution. Both these distributions can be defined over a one-dimensional lattice $r\Lambda = \{rz \ : r \in \mathbb{R},\, z \in \mathbb{Z}\}$, however without loss of generality and to improve the tractation, unless differently specified from now on we consider Λ to be a lattice defined by $r = 1$ which corresponds to consider the integer set $\mathbb{Z}$. We can now define the two distributions we will consider throughout this work. A zero-mean quantized Gaussian distribution over a lattice Λ can be defined as

$$\mathcal{G}_{\Lambda,\sigma}(z) = \int_{z-1/2}^{z+1/2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt \ \text{ with } \ z \in \Lambda.$$

Conversely, a zero-mean discrete Gaussian distribution over a lattice Λ is defined as

$$\mathcal{D}_{\Lambda,\sigma}(z) = \frac{\rho_\sigma(z)}{\rho_\sigma(\Lambda)} \ \text{ with } \ z \in \Lambda,$$

where $\rho_\sigma(z) = \exp(-z^2/2\sigma^2)$ and $\rho_\sigma(\Lambda) = \sum_{z\in\Lambda} \exp(-z^2/2\sigma^2)$ is the normalization factor.

It is important to note that we are taking into account the physical limitations of a practical system which employs finite precision representations. Since we obtain samples from a quantized Gaussian distribution, these limitations translate into tails truncation. In more detail, given a fixed amount of bits N_b we consider to truncate the tails at $k\sigma$ where $\sigma = (2^{N_b-1} - 0.5)/k$ is a function of the number of available bits. We denote as $\Lambda^C \subset \Lambda$ the finite set of all the elements of Λ which fall inside the truncation interval $[-k\sigma, k\sigma]$. Moreover, without loss of generality we assume that $k\sigma = 0.5 + l, \ l \in \mathbb{Z}$ to provide a truncation which is consistent with the quantization intervals. We can now define the truncated quantized Gaussian distribution as

$$\overline{\mathcal{G}}_{\Lambda^C,\sigma}(z) = w\mathcal{G}_{\Lambda^C,\sigma}(z) \ \text{ with } \ z \in \Lambda^C,$$

where $w = (1 - g_T)^{-1}$ is a normalization factor and $g_T = 2\int_{k\sigma}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt$. In a similar fashion, we also define the truncated discrete Gaussian distribution as

$$\overline{\mathcal{D}}_{\Lambda^C,\sigma}(z) = \frac{\rho_\sigma(z)}{\rho_\sigma(\Lambda^C)} \ \text{ with } \ z \in \Lambda^C.$$

C. Security model

The private key cryptosystem we will consider throughout the paper is defined as follows. The signal $\mathbf{x}$ is the plaintext, the measurements $\mathbf{y}$ are ciphertext, and the sensing matrix $\mathbf{\Phi}$ is the secret key. The encryption $e_{\mathbf{\Phi}}(\mathbf{x})$ is performed through CS acquisition as defined in (1) and the decryption $d_{\mathbf{\Phi}}(\mathbf{y})$ corresponds to any CS recovery algorithm which can be used to solve the minimization problem in (2).

The model we consider is based on the one-time sensing (OTS) setting, namely the sensing matrix is re-generated at each encryption. This assumption makes the cryptosystem resistant to both known ciphertext attacks (KPA) and chosen ciphertext attacks (CPA). In the first attack, the attacker tries to break the system given the knowledge of a number of plaintext-ciphertext $(\mathbf{x}, \mathbf{y})$ pairs. This approach cannot succeed under OTS since to solve the mn linear system of equations defined by $\mathbf{\Phi}$ the attacker would require n pairs acquired through the same sensing matrix. Concerning the CPA, an attacker may choose the canonical basis vectors as $\mathbf{x}$ and obtain, at each encryption, a column of the secret key $\mathbf{\Phi}$. As before, this attack cannot succeed under OTS since it would require n ciphertexts performed with the same sensing matrix in order to obtain the secret key $\mathbf{\Phi}$.

D. Security metrics

Given the cryptosystem we defined above, different metrics can be used to characterize its security properties. From an information theoretic perspective, a cryptosystem is said to be perfectly secure if

$$P[\mathbf{y}|\mathbf{x}] = P[\mathbf{y}],$$

where $\mathbf{y}$ denotes the ciphertext and $\mathbf{x}$ denotes the plaintext. Namely, the posterior probability of the ciphertext given plaintext is independent of the plaintext. This implies that an attacker cannot be more successful than random guessing the plaintext. In spite of this very strong definition, another widely used definition is that of *computational secrecy*. In this case, a cryptosystem is said to be computationally secure if breaking the system corresponds to solve an NP-hard problem.

The information theoretic approach is stronger than the computational one since it characterizes the amount of information an attacker can have access to. If no sufficient information is available, then even with unbounded computational capabilities the attack cannot succeed. This, and the fact that we want to characterize the information leakage are the reasons behind our choice to proceed with an information theoretic approach.

From [14] is it known that, under the assumption of $\mathbf{\Phi}$ being made of i.i.d. Gaussian entries, and $\mathbf{y}$ obtained through eq. (1), the following holds

$$P[\mathbf{y}|\mathbf{x}] = P[\mathbf{y}|\varepsilon_{\mathbf{x}}],$$

where $\varepsilon_{\mathbf{x}}$ is the energy of $\mathbf{x}$. This means that only the energy of the original signal is leaked through the measurements, and that if $\mathbf{x}$ is normalized to have unit energy, then it is possible to achieve perfect secrecy.

However, the distributions we consider for the entries of $\mathbf{\Phi}$ are not Gaussian. This means that perfect secrecy, at least in non-asymptotic sense, cannot be achieved. Thus, it is important to characterize the information leakage. To achieve this goal, we employ another metric, introduced in [14], which is the θ-distinguishability. This metric, which is defined by means of a detection experiment, is inspired by the distinguishability definitions commonly used in cryptography. Given two signals $\mathbf{x}_1, \mathbf{x}_2$ we consider a simple detection test in which the attacker, by using a detector $D(\mathbf{y})$, has to guess whether $\mathbf{y}$ comes from $P[\mathbf{y}|\mathbf{x}_1]$ or $P[\mathbf{y}|\mathbf{x}_2]$.

Therefore, we will say that CS measurements are θ-indistinguishable if, for every possible detector $D(\mathbf{y})$, $P_d - P_f \leq \theta$, where P_d and P_f are the probability of detection and false alarm of the detector, respectively. It is evident that $\theta = 0$ corresponds to perfect secrecy, namely no detector can distinguish the two signals better than guessing at random. As shown in the following, evaluating θ boils down to the evaluation of a distance measure between the measurements distributions. Thus, let us recall two important distance measures which will be used in the following.

Definition II.1. *The Total Variation (TV) distance between two discrete probability distributions P and Q is defined as*

$$\delta_{\mathrm{TV}} = \frac{1}{2}||P - Q||_1.$$

Definition II.2. *The Kullback-Leibler (KL) divergence between two discrete probability distributions P and Q is defined as*

$$\delta_{\mathrm{KL}} = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

At this point, in order to evaluate the value of θ, we can rely on the following Lemma.

Lemma II.3. *(Lemma 4 in [14]) OTS measurements are at least $\delta_{\mathrm{TV}}(\mathbf{\Phi}\mathbf{x}_1, \mathbf{\Phi}\mathbf{x}_2)$-indistinguishable with respect to two signals $\mathbf{x}_1$ and $\mathbf{x}_2$, where $\delta_{\mathrm{TV}}(\cdot,\cdot)$ corresponds to the TV distance.*

Thanks to the above Lemma, in order to derive the distinguishability of two signals under the quantized OTS setting, it is sufficient to evaluate the TV distance between their probability distributions. Nonetheless, when it is not possible to explicitly derive the TV distance, it can be upper bounded with the Kullback-Leibler (KL) divergence through the Pinsker's inequality [34]. In this case we have that CS measurements are at least $\sqrt{0.5 \cdot \delta_{\mathrm{KL}}(p(\mathbf{y}|\mathbf{x}_1), p(\mathbf{y}|\mathbf{x}_2))}$-indistinguishable w.r.t. $\mathbf{x}_1, \mathbf{x}_2$, where $\delta_{\mathrm{KL}}(\cdot,\cdot)$ corresponds to the KL divergence. Throughout this paper we will use both TV distance and KL divergence in order to evaluate the θ-distinguishability.

III. Main results

A. Preliminaries

Before starting the derivation of our main results, it is important to highlight that throughout the paper we will consider the signal to be acquired having integer entries, i.e. $\mathbf{x} \in \mathbb{Z}^n$. This assumption, which can be made without loss of generality, is necessary since we are considering a finite precision cryptosystem.

We can now present some useful lemmas relating the probability distributions of truncated quantized and truncated discrete Gaussian distributions with that of a discrete Gaussian distribution.

Lemma III.1. *Let $X \sim \overline{\mathcal{G}}_{\Lambda^C,\sigma}$ and $Y \sim \mathcal{D}_{\Lambda,\sigma}$ and $z \in \Lambda^C$ then*

$$P[X = z] \leq (1 + \delta_{Q,\sigma})\mathcal{D}_{\Lambda,\sigma}(z),$$

where there is a regime of N_b, k for which $\delta_{Q,\sigma}$ approaches zero exponentially fast as the number of bits $N_b \to \infty$.

Proof. The proof is presented in the Appendix. □

From this first lemma it is possible to see that, as N_b increases, a truncated quantized Gaussian distribution can approach a discrete Gaussian distribution exponentially fast. Because of the relationship existing between the truncation factor k and N_b, the exponential behavior is valid under a specific regime which holds for

$$k^* = \{k : k^2 = \beta N_b\},$$

where β is a positive constant. As an example, the regime condition is satisfied for $k = \sqrt{N_b}$. Let us remark that, in the following, when referring to the exponential behavior of δ_Q, even if not specified, we will refer to the above regime condition.

In a similar fashion, it is possible to relate a truncated discrete Gaussian distribution with its non-truncated counterpart as described in the following Lemma.

Lemma III.2. *Let $X \sim \overline{\mathcal{D}}_{\Lambda^C,\sigma}$ and $Y \sim \mathcal{D}_{\Lambda,\sigma}$ and $z \in \Lambda^C$ then*

$$P[X = z] \leq (1 + \delta_{T,k})\mathcal{D}_{\Lambda,\sigma}(z)$$

where $\delta_{T,k}$ approaches zero exponentially fast as the truncation factor k increases.

Proof. The proof is presented in the Appendix. □

In this case we have that, because of the exponential relationship, the discrete and truncated discrete Gaussian distributions can be made arbitrarily close for a sufficiently large value of k.

We can now use these results in order to characterize the distribution resulting from a linear combination of two truncated and quantized Gaussian distributions.

Lemma III.3. *Let $\boldsymbol{\phi} = [\phi_1 \;\; \phi_2]^\intercal$ be made of two independent random variables distributed as $\overline{\mathcal{G}}_{\Lambda^C,\sigma}$. Assume $\sigma > \sqrt{2(x_1^2 + x_2^2)}\eta_\epsilon(\Lambda)$ holds for negligible ϵ, where $\eta_\epsilon(\Lambda)$ is the smoothing parameter of Λ as defined in [35]. Moreover, let us define $\sigma_1 = |x_1|\sigma$ and $\sigma_2 = |x_2|\sigma$. Then, $y = \boldsymbol{\phi}^\intercal\mathbf{x}$ where $\mathbf{x} = [x_1 \; x_2]^\intercal \in \mathbb{Z}^2$ is statistically close to $\mathcal{D}_{G\Lambda,\sqrt{\sigma_1^2+\sigma_2^2}}$, i.e.*

$$P[Y = y] \leq (1 + \delta_Q)^2(1 + \delta_D)\mathcal{D}_{G\Lambda,\sqrt{\sigma_1^2+\sigma_2^2}}(y),$$

where $G = \gcd(x_1, x_2)$ is the greatest common divisor between x_1 and x_2, and δ_Q, δ_D approach zero exponentially fast as N_b increases.

Proof. The proof is presented in the Appendix. □

This result allows us to establish a relationship between a linear combination of two quantized and truncated Gaussian random variables and a random variable distributed as a discrete Gaussian distribution. Interestingly, this bound is a function of the system parameters N_b and k and can be made arbitrarily tight.

With the following lemma we show that the above result can be extended to an arbitrary linear combination of truncated and quantized Gaussian distributions weighted by integer coefficients.

Lemma III.4. *Let $Y = \boldsymbol{\phi}^\intercal \mathbf{x}$ with $\mathbf{x} = [x_1 \dots x_K]^\intercal \in \mathbb{Z}^K$ and $\boldsymbol{\phi} = [\phi_1 \dots \phi_K]^\intercal$ be mutually independent random variables sampled from $\overline{\mathcal{G}}_{\Lambda^C,\sigma}$. Moreover, suppose that $\sigma > \sqrt{2}||\mathbf{x}||_2\eta_\epsilon(\Lambda)$ holds for negligible ϵ. Then Y is statistically close to $\mathcal{D}_{G\Lambda,||\mathbf{x}||\sigma}$, i.e.*

$$P[Y = y] \leq (1+\delta_Q)^K(1+\delta_D)^{K-1}\mathcal{D}_{G\Lambda,||\mathbf{x}||\sigma}(y),$$

where $G = \gcd(\mathbf{x})$ is the greatest common divisor among the elements of $\mathbf{x}$ and δ_Q, δ_D approach zero exponentially fast as $N_b \to \infty$.

Proof. The proof is presented in the Appendix. □

The above Lemma states that a linear combination of quantized Gaussian random variables is close to a Discrete Gaussian distribution lying on a different lattice whose structure depends on the weights of the linear combination. It becomes evident that, as we are interested in assessing the distinguishability of two linear combinations of quantized Gaussian distributions, if they do not share the same support, they can always be distinguished based on this information. For this reason, it is necessary to re-quantize these linear combinations with a suitable and *common* quantization scheme. The following Lemma provides a bound on the distribution of a re-quantized discrete Gaussian distribution in terms of a suitable discrete Gaussian distribution over a different lattice.

Lemma III.5. *Let us assume $P[Y = y] \leq (1+\delta)\mathcal{D}_{G\Lambda,\sigma}(y)$ for $|y| \leq k'\sigma$ and $P[Y = y] = 0$ for $|y| > k'\sigma$. Let $Z = Q_H(Y) = \lceil Y/H \rfloor$ be the same random variable re-quantized with a scalar quantizer over bins of size $H \geq G$, then*

$$P[Z = z] \leq (1+\delta)(1+\delta_R)\mathcal{D}_{H\Lambda,\sigma}(z),$$

where δ_R approaches zero exponentially fast as N_b increases.

Proof. The proof is presented in the Appendix. □

B. Equal-energy signals

We are now ready to state our first main result. It allows us to quantify the statistical distance between two linear combinations of equal-energy signals whose entries are weighted by means of quantized Gaussian sensing matrices. As previously discussed, non unit-energy signals cause an information leakage through the measurements which is directly proportional to their energy. For this reason, in the following theorem we consider signals lying on the surface of a unitary hyper-sphere; this allows us to explicitly consider only the secrecy loss due to the finite precision representation of the sensing matrix entries.

Theorem III.6. *Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{Z}^K$ be two equal-energy signals with $\|\mathbf{x}_1\|_0 = K_1$ and $\|\mathbf{x}_2\|_0 = K_2$. Then, define $Y_1 = \boldsymbol{\phi}^\intercal \mathbf{x}_1$ and $Y_2 = \boldsymbol{\phi}^\intercal \mathbf{x}_2$ with $\boldsymbol{\phi}$ be made of independent random variables sampled from $\overline{\mathcal{G}}_{\Lambda^C,\sigma}$. Moreover, let us assume that Y_1 and Y_2 are re-quantized as $Z_1 = Q_H(Y_1)$ and $Z_2 = Q_H(Y_2)$, where $H = \max(\gcd(\mathbf{x}_1), \gcd(\mathbf{x}_2))$. Then, the statistical distance between Z_1 and Z_2 can be upper bounded as*

$$\delta(Z_1, Z_2) \leq (K_1+K_2)\delta_Q + (K_1+K_2-2)\delta_D + 2\delta_R + \epsilon,$$

where $\delta_Q, \delta_D, \delta_R$ approach zero exponentially fast as $N_b \to \infty$ and $\epsilon = o(\delta_Q) + o(\delta_D) + o(\delta_R)$. Moreover, if $\gcd(\mathbf{x}_1) = \gcd(\mathbf{x}_2)$, then $\delta_R = 0$.

From Theorem III.6 it can be seen that as long as the Gaussian random variable is quantized with enough bits and it is truncated for small tails, then the statistical distance between two linear combinations of quantized Gaussian random variables can be made arbitrarily small. It can also be noted that the bound linearly increases with the sparsity of the two signals. However, the statistical distance should linearly decrease with K_1, K_2. In fact, since the entries of the sensing matrices are i.i.d., as the size of the linear combination of i.i.d. elements increases, the result will tend to a Gaussian distribution. In the case of the limit $K_1, K_2 \to \infty$, two signals will result in equally distributed measurements and thus their statistical distance will approach zero. Thus we can state that, while tight for N_b, k, the bound of Theorem III.6 is not very tight with respect to sparsity of the original signals because of the successive approximation employed in its derivation. This behavior is showed in detail in the experimental section.

At this point, we can employ the results of Theorem III.6 to derive the following corollary which provides a bound on the θ-distinguishability for measurements vectors of two equal-energy signals acquired by means of quantized Gaussian sensing matrices.

Corollary III.6.1. *Let $\boldsymbol{\Phi} \in \Lambda^{C^{m\times n}}$ be a quantized sensing matrix whose entries are distributed according to $\overline{\mathcal{G}}_{\Lambda^C,\sigma}(z)$. Also, let $\mathbf{x}_1, \mathbf{x}_2$ be two signals having the same energy $||\mathbf{x}||^2$ which are encrypted by means of the linear operator $\mathbf{y}_{\{1,2\}} = \boldsymbol{\Phi}\mathbf{x}_{\{1,2\}}$. Then, assuming that the measurements have been quantized as in Theorem III.6, the θ-distinguishability as defined in [14] is upper bounded as*

$$\begin{aligned}\theta_\Phi(\mathbf{x}_1, \mathbf{x}_2) \leq & m(K_1+K_2)\delta_Q + m(K_1+K_2-2)\delta_D \\ & + 2m\delta_R + \epsilon,\end{aligned}$$

where δ_Q, δ_D approach zero exponentially fast as $N_b \to \infty$ and $\epsilon = o(\delta_Q) + o(\delta_D) + o(\delta'_Q)$.

Proof. The proof follows by the use of Theorem III.6 and the fact the the elements of $\mathbf{y}_1, \mathbf{y}_2$ are i.i.d. □

In light of the above results we can say that the secrecy of a CS cryptosystem under finite precision representation exponentially increases with the number of employed bits N_b. This is an important result because it means that, in the worst case scenario: 1) Bernoulli sensing cryptosystems achieve the

highest possible distinguishability under finite precision representation and 2) the secrecy can be exponentially improved by using additional bits in the representation.

C. Arbitrary signals

Up to this point, we considered the signals to be unit-energy in order to isolate the security loss effect due to the finite precision representation. Nevertheless, if we assume that two signals $\mathbf{x}_1, \mathbf{x}_2$ are represented with finite precision, then the constraint of having unit-energy might not be satisfied. For this reason we start by evaluating the energy mismatch due to the quantization of two unit-energy signals. In fact, a practical scenario may involve signals which lie on the surface of the same $\mathbb{R}^n$ hypersphere and which are *then* quantized. The result is an energy mismatch which is characterized by the following Lemma.

Lemma III.7. *Let $\bar{\mathbf{x}}', \bar{\mathbf{x}}'' \in \mathbb{R}^n$ be two signals having the same energy. Then, assume that these signals are quantized with N_x bits into $\mathbf{x}'$ and $\mathbf{x}''$ respectively. Then, the energy mismatch on the quantized signals $q_e = \left| ||\mathbf{x}'||^2 - ||\mathbf{x}''||^2 \right| \leq t||\mathbf{x}'||^2$ with probability $1 - \delta(t, N_x)$, where there is a regime of t, N_x for which t and $\delta(t, N_x)$ approach zero exponentially fast in N_x.*

From the above Lemma it can be seen that the energy mismatch q_e can be made arbitrarily small with high probability for large enough N_x. More in detail, it tends to zero with probability approaching 1 exponentially fast in t, N_x under a regime condition which holds for

$$t^* = \{t : t \leq 2^{-\alpha N_x}, t^2 2^{2N_x} = \beta N_x\},$$

where α and β are positive constants. As an example, by letting $t = \sqrt{N_x} 2^{-N_x}$ the regime condition is satisfied.

The following Theorem generalizes Corollary III.6.1 and provides a characterization of the distinguishability of signals which have different energy in both a more general case and in the specific one of quantized unit-energy signals.

Theorem III.8. *Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{Z}^n$ be two signals with $\|\mathbf{x}_1\|_0 = K_1$ and $\|\mathbf{x}_2\|_0 = K_2$ and their energies be $||\mathbf{x}_1||^2$ and $||\mathbf{x}_2||^2$. Then, define $\mathbf{y}_1 = \mathbf{\Phi}\mathbf{x}_1$ and $\mathbf{y}_2 = \mathbf{\Phi}\mathbf{x}_2$ with $\mathbf{\Phi}$ having size $m \times n$ be made of independent random variables sampled from $\overline{\mathcal{G}}_{\Lambda^C,\sigma}$. Then, assuming that the measurements have been quantized as in Theorem III.6, the θ-distinguishability between $\mathbf{x}_1$ and $\mathbf{x}_2$ can be upper bounded as*

$$\begin{aligned}\theta_{\mathbf{\Phi}}(\mathbf{x}_1, \mathbf{x}_2) \leq & m(K_1 + K_2)\delta_Q + m(K_1 + K_2 - 2)\delta_D \\ & + 2m\delta_R + m\delta_M + \epsilon\end{aligned}$$

where exists a regime for N_b, k for which $\delta_D, \delta_Q, \delta_R$ approach zero exponentially fast in N_b and $\epsilon = o(\delta_Q) + o(\delta_D) + o(\delta_R)$. Moreover, we have that $\delta_M = \sqrt{\frac{1}{2}\log\frac{||\mathbf{x}_1||^2}{||\mathbf{x}_2||^2} + \epsilon_M + C\frac{||\mathbf{x}_2||^2 - ||\mathbf{x}_1||^2}{2||\mathbf{x}_2||^2}}$ where $\varepsilon_M \to 0$ exponentially fast for $N_b \to +\infty$ and $C \approx 1$. If we consider the specific case in which $\mathbf{x}_{\{1,2\}}$ is quantized as described in Lemma III.7, then $\delta_M < t$ with probability $1 - \delta(t, N_x)$, where there is a regime of t, N_x, N_b for which t and $\delta(t, N_x)$ approach zero exponentially fast as N_x and N_b increase.

We showed that under some reasonable assumptions the secrecy of a compressive cryptosystem can be made arbitrarily high. In more detail, the parameters of a practical implementation which directly affect the secrecy of compressive cryptosystem are the number of bits N_b and the Gaussian truncation parameter k which can be used to make the θ-distinguishability as small as desired when considering truncated and quantized Gaussian sensing matrix entries. In the following we show that these results can be extended to the case of sensing matrices made of truncated discrete Gaussian random variables.

Corollary III.8.1. *Let $\mathbf{\Phi} \in \Lambda^{C^{m\times n}}$ be a quantized sensing matrix whose entries are distributed according to $\overline{\mathcal{D}}_{\Lambda^C,\sigma}(z)$. Then, the results of Theorem III.6 and III.8 hold, respectively as*

$$\begin{aligned}\theta_{\mathbf{\Phi}}(\mathbf{x}_1, \mathbf{x}_2) \leq & m(K_1 + K_2)\delta_T + m(K_1 + K_2 - 2)\delta_D \\ & + 2m\delta_R + \epsilon,\end{aligned}$$

and

$$\begin{aligned}\theta_{\mathbf{\Phi}}(\mathbf{x}_1, \mathbf{x}_2) \leq & m(K_1 + K_2)\delta_T + m(K_1 + K_2 - 2)\delta_D \\ & + 2m\delta_R + m\delta_M + \epsilon,\end{aligned}$$

where δ_T approaches zero exponentially fast as the truncation factor k increases and $\epsilon = o(\delta_T) + o(\delta_D)$.

Proof. This proof easily follows by the application of Lemma III.2 to the proofs of Theorems III.6 and III.8. □

Lastly, it is important to recall that throughout the above derivations we considered the case of K-sparse signals. However, as hinted in Sec. II most of natural signals are *compressible* rather than exactly sparse. In this regard one may wonder how the secrecy of sparse signals encrypted by means of quantized Gaussian CS cryptosystems compares with that of compressible signals. Given that the recovery process of CS will recover the K-sparse approximation of a compressible signal, we can write the compressible signals as $\mathbf{c} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x}$ is a K-sparse signal and $\mathbf{n}$ is a noise term. Similarly, we can write the measurements of $\mathbf{c}$ as $\mathbf{y}_c = \mathbf{y}_x + \mathbf{y}_n$ where $\mathbf{y}_x$ are the measurements of the sparse approximation and $\mathbf{y}_n$ are those of the noise term. Given this processing chain, by invoking the data processing inequality, we can state that the mutual information between $\mathbf{y}_c$ and $\mathbf{x}$ is smaller with respect to that between $\mathbf{y}_x$ and $\mathbf{x}$. As a result, the noise term which accounts for the fact that the signal is not exactly K-sparse leads to a secrecy increase. In light of the above, our results also guarantee the secrecy of the K-sparse approximation when *compressible* signals are acquired.

IV. Experiments

In this section we validate the bounds we obtained in the previous section by providing some numerical results on the distinguishability of signals which are encrypted by means of CS using quantized sensing matrices.

A. Methods

Let us start with a simple detection experiment which we will use as a benchmark to evaluate the performance of a cryptosystem. Given two signals $\mathbf{x}_1, \mathbf{x}_2$ we consider a simple detection test in which the attacker, by using a detector $D(\mathbf{y})$, has to guess whether $\mathbf{y}$ comes from $P[\mathbf{y}|\mathbf{x}_1]$ or $P[\mathbf{y}|\mathbf{x}_2]$. According to the Neyman-Pearson (NP) lemma, and a given probability of false alarm $P_f = \alpha$, we have that the maximizer of the probability of detection P_d is given by letting $D(\mathbf{y}) = \mathbf{x}_1$ if

$$\gamma(\mathbf{y}) = \frac{P[\mathbf{y}|\mathbf{x}_1]}{P[\mathbf{y}|\mathbf{x}_2]} \geq \theta,$$

where $P[\gamma(\mathbf{y}) \geq \theta|\mathbf{x}_2] = \alpha$. Moreover, as we already highlighted before the entries of $\mathbf{y}$ are i.i.d.; this allows us to rewrite the NP-test as

$$\gamma_l(\mathbf{y}) = \sum_{i=1}^{m} (\log(P[\mathbf{y}_i|\mathbf{x}_1]) - \log(P[\mathbf{y}_i|\mathbf{x}_2])) \geq \theta_l.$$

In order to find $P[\mathbf{y}_i|\mathbf{x}]$, we can notice that $\mathbf{y}_i$ is a linear combination of n sensing matrix entries $\boldsymbol{\phi}$ and thus, its characteristic function can be written in product fashion as

$$\phi_{\mathbf{y}_i|\mathbf{x}}(t) = \prod_{k=1}^{n} \tilde{\phi}(\mathbf{x}_k t), \tag{3}$$

where $\tilde{\phi}(t)$ is the characteristic function of a truncated and quantized Gaussian distribution. According to [36], the characteristic function of a truncated Gaussian distribution whose realizations are quantized through area sampling can be written as

$$\tilde{\phi}(t) = \sum_{l=-\infty}^{+\infty} \phi_T(t + l\Psi) \mathrm{sinc}\left(\frac{q(t + l\Psi)}{2}\right),$$

where $\Psi = \frac{2\pi}{q}$, $\phi_T(t)$ is the characteristic function of a truncated Gaussian distribution and q is the width of the quantization bin. Lastly, we can write

$$\phi_T(t) = \phi_G(t) * \frac{\sin(2T_R t)}{2T_R},$$

where $2T_R$ is the truncation interval and $\phi_G(t)$ is the characteristic function of a Gaussian distribution. At this point, we can compute $p(\mathbf{y}_i|\mathbf{x})$ for a given $\mathbf{x}_1$ and $\mathbf{x}_2$ by using (3) and performing the inverse Fourier transform. Lastly, as done in [14], though hard to be obtained analytically the error probability for the detector described above can be upper bounded by the TV distance between $p(\mathbf{y}_i|\mathbf{x}_1)$ and $p(\mathbf{y}_i|\mathbf{x}_2)$ as

$$P_d - P_f \leq \delta_{\mathrm{TV}}(p(\mathbf{y}_i|\mathbf{x}_1) - p(\mathbf{y}_i|\mathbf{x}_2)).$$

B. Experimental results

The experiments we discuss in the following are performed by considering two different scenarios, namely worst-case and average-case. More in detail, we follow a twofold approach: at first, we validate the proposed upper bounds under the worst-case scenario for equal energy signals and next, we show the results for *approximately* equal energy signals under the average-case scenario.

1) Worst-case scenario: Let us recall that, since the bound obtained in Corollary III.6.1 is an *upper* bound that does not depend on the sensed signal, it holds in the *worst case* scenario, i.e., for every possible $\mathbf{x}_1$, $\mathbf{x}_2$ pair. In the first experiments we analyze the worst case scenario by evaluating the θ-distinguishability through its upper bound (TV distance) and compare the results with those of the bound in Corollary III.6.1. More in detail, we consider two *equal energy* signals of length 64 whose entries can be represented with $N_x = 3$ and sparsity $K_1 = 1$ and $K_2 = 64$: $\mathbf{x}_1$ has a single entry with value 8 and $\mathbf{x}_2$ has 64 entries with value 1. Besides, the measurements are re-quantized over N_b bits, so as to ensure that the distributions of the two sets of measurements are defined on the same support.

As depicted in Fig. 1, it can be seen that the general trend is an exponential decrease of the θ-distinguishability as the value of N_b increases. Furthermore, it can be noted that this trend is, in practice, limited by the employed truncation factor k; if k it is not large enough with respect to N_b the θ-distinguishability reaches a plateu due to the truncation error and cannot decrease. More in detail, as expected, as the number of bits employed for the quantization of the sensing matrix N_b is increased, a larger truncation factor k is required in order to reach the optimal θ-distinguishability for that specific configuration. This effect can also be appreciated in Fig. 2 where the distinguishability is shown for different N_b in function of the truncation factor k. It can be seen that small values of k limit the achievable distinguishability from below: the error term due to the truncation is higher than the one corresponding to the employed number of bits.

It has become evident that N_b and k act on the θ-distinguishability in a joint fashion. As discussed in Corollary III.6.1, there exists a regime for k, N_b such that the decrease of the θ-distinguishability is exponential. In Fig. 3 this behavior is depicted for $k = 2\sqrt{N_b}$. It is immediate to notice that, if the regime condition is satisfied, then the θ-distinguishability exponentially decreases in N_b.

Lastly, let us consider how the derived bound compares with the obtained results. It can be seen that, though being loose for small N_b, the bound tightens to the simulated TV distance as N_b increases. As a matter of fact, because of their derivation, the bounds do overestimate the θ-distinguishability for the cases in which 2^{N_b} is small. Nevertheless, as can be seen e.g. in Fig. 1, the bound is tight for reasonably small values of N_b.

2) Average-case scenario: Up to this point, we performed a numerical validation of the bounds we obtained in the worst case under the assumption of equal energy signals. For the following experiment, we consider the *average case* scenario, namely we numerically evaluate the θ-distinguishability of randomly drawn signals. Since we do not make any equal energy assumption, the values of θ depicted in Fig. 4 also include the effect of the energy mismatch. Furthermore, because of unequal energy signals can be immediately distinguished based on their energy, we consider the specific case of *approximately* equal energy signals, as discussed in Lemma III.7.

The signals we consider are represented with $N_x = 6$ bits and evaluated at different sparsity levels $K = K_1 = K_2$. The

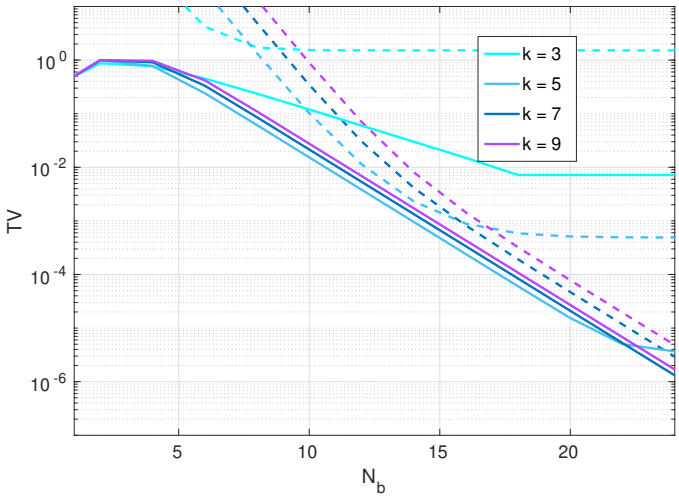

Fig. 1: TV distance simulated (solid line) and from Corollary III.6.1 (dashed line) as a function of N_b for different values of K.

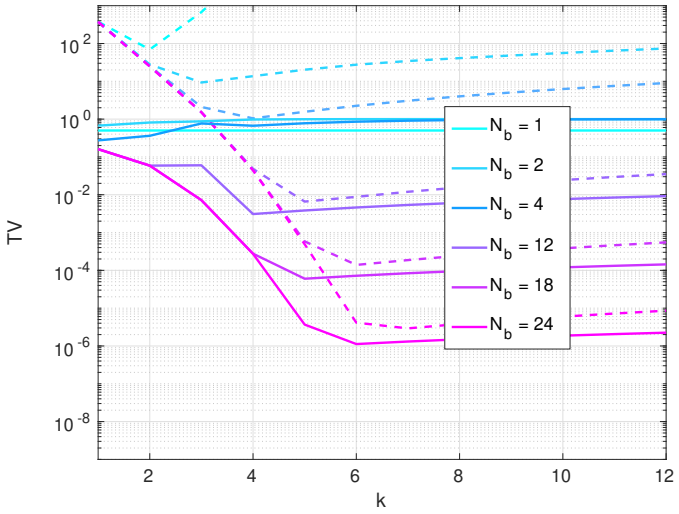

Fig. 2: TV distance simulated (solid line) and from Corollary III.6.1 (dashed line) as a function of K for different values of N_b.

values of N_b and k are chosen to be in the regime condition, namely $k = 2\sqrt{N_b}$. The measurements are then re-quantized to N_R bits accordingly to different policies, namely no re-quantization, $N_R = N_b$ and fixed $N_R = 10$. The TV distance is numerically evaluated over 200 pairs of randomly generated signals and the 95 percentile of the values is computed. The results are depicted in Fig. 4.

It is important to notice that we consider two different values of K in order to better appreciate the effects of the re-quantization. As will be discussed more in detail in Sec. V, higher values of K decrease the distinguishability as the distribution of the measurements coming from two different signals tend to the same distribution. However, this desired effect makes harder to isolate the effects of the re-quantization. For this reason, along with a large value of $K = 100$ we also consider a small value, namely $K = 4$.

Let us focus on the behavior of the distinguishability as a function of N_b for $K = 4$ (solid lines in Fig. 4). If the measurements are not re-quantized after sensing, it can be

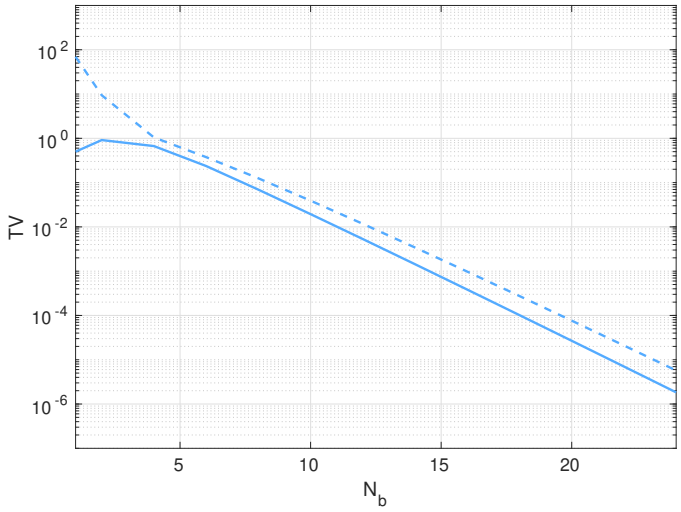

Fig. 3: TV distance simulated (solid line) and from Corollary III.6.1 (dashed line) as a function of N_b under $k = 2\sqrt{N_b}$.

immediately noticed that the distinguishability is extremely high. As discussed in Sec. III, two encrypted signals can be immediately distinguished if there is a mismatch in their support. This issue can be easily addressed by re-quantizing the measurements with a suitable number of bits $N_R \leq N_b$. In the same figure, it can also be seen that if the re-quantization is performed accordingly to $N_R = N_b$, then the effect due to the support mismatch is removed and the distinguishability follows a decreasing behavior for increasing N_b.

At this point it also important to highlight that a minimum number of bits to represent the measurements might be needed as a requirement at decryption side in order to achieve a lower distortion on the decrypted signal, see e.g. [37]. For this reason, we also consider a fixed number of bits for the re-quantization of the measurements, i.e. $N_R = 8$. It can be seen that, when the number of bits over which the measurements are represented, i.e. N_b+N_x, is smaller than N_R the distinguishability is high; a support mismatch will lead to TV distance close to 1. Conversely, when the measurement are re-quantized over a smaller number of bits ($N_b + N_x > N_R$), the distinguishability decreases with N_b.

It is worth noting that, even though this experiment is performed under the regime condition, the distinguishability does not exponentially decrease as expected, rather it reaches a plateau. This is due to the fact that we are considering *approximately* equal energy signals and thus the contribution of the energy mismatch between signals limits the exponential decrease.

In order to appreciate how the energy mismatch can drastically reduce the secrecy of a cryptosystem, in Fig. 5 we plot the θ-distinguishability as a function of N_x for a fixed $N_b = 12$, under the regime condition and with $N_R = N_b$. It can be immediately seen that the distinguishability decreases in N_x. As shown in Lemma III.7, two equal energy signals which are then re-quantized over N_x bits will exhibit an energy mismatch which depends on N_x. In turn, a larger energy mismatch will lead to a larger distinguishability (Th. III.8). This is indeed the behavior depicted in Fig. 5 where it can be

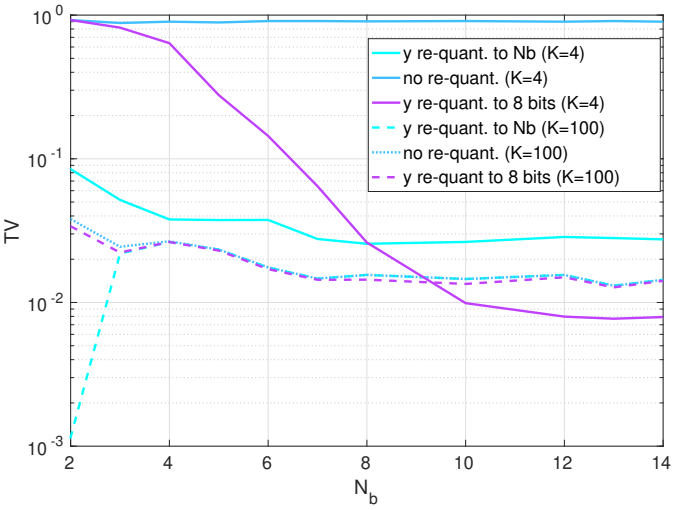

Fig. 4: θ-distinguishability simulated for approximately equal energy signals under the average case scenario as a function of N_b for different values of K and re-quantization of the measurements. $N_x = 6$. $K = 4$ solid line, $K = 100$ dashed line.

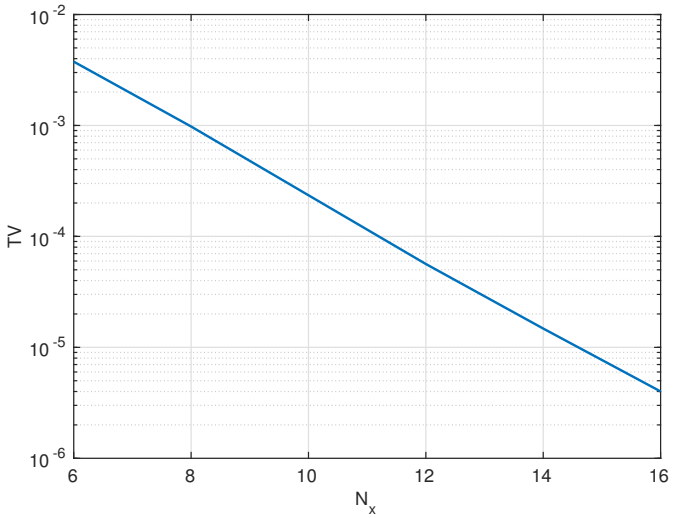

Fig. 5: θ-distinguishability simulated for approximately equal energy signals under the average case scenario as a function of N_x for $N_b = 12$.

further noticed that the energy mismatch is the highest source of distinguishability, being orders of magnitude larger than quantization and tail truncation effects.

V. Discussion

In the previous sections, we obtained and experimentally validated the bounds under the worst case scenario, namely for a pair of signals that result in very different distributions of the measurements, even when the above distributions have been constrained on the same support by using re-quantization. Moreover, we also showed, from an experimental point of view, the distinguishability of the measurements of quantized Gaussian matrices in the average case scenario, i.e., for randomly drawn pairs of signals. The most important result is that, by just employing few more bits than the single one needed to represent the entries of Bernoulli sensing matrices, it is possible to exponentially increase the secrecy of a CS-based cryptosystem.

As previously discussed, the obtained bounds are conservative as they hold in the worst case scenario. However, in practice the distinguishability can be on average much smaller than in the worst case scenario. Let us consider the probability p of K integer numbers chosen in $[1, x_{max}]$ being coprime, namely that their greatest common divisor is 1. Then, considering two signals having sparsity $K_1 = K_2 = K$ and nonzero values bounded by x_{max}, with probability p the support of the measurements of the two signals will be the same even before re-quantization. Conversely, with probability $(1-p)$ we are in the worst case and the measurements, before re-quantization, do not share the same support. It is a well-known result in number theory that for large x_{max} we have $p \approx 1/\zeta(K)$, where $\zeta(K)$ is the Riemann Zeta function [38]. Interestingly, as K increases p approaches 1 and thus, for large enough K, the probability of being in the good case is high.

Regarding the quantization of the measurements, we showed that this is a required operation in order avoid information leakage due to a support mismatch of the measurements. Nevertheless, it is interesting to note that this operation, even in the case of matching supports would still increase the secrecy of the whole cryptosystem. As a matter of fact, because of the data processing inequality, the mutual information between the measurements and the original signal decreases if additional operations are performed on the measurements as for the quantization. In turn, a decreased mutual information means that the secrecy of the system is increased. Thus, for these reasons, even though there is no support mismatch, re-quantization of the measurements is always advisable. This latter result, which we also covered in the experimental section, indicates that as we re-quantize the measurements with the least possible number of bits we have the highest possible secrecy. According to the results in Sec. III, we need to quantize the measurements with at most the number of bits employed for the quantization of the sensing matrix in order to avoid a complete distinguishability of the measurements based on their support. However, if less than N_b bits are employed during the re-quantization, the secrecy is increased.

While correct, this statements does not consider the functionality of the cryptosystem as a whole. In fact, being able to correctly recover the plaintext from the ciphertext is essential. Indeed, when the measurements are quantized with just few bits, the distortion of the recovered signals is increased, see e.g. [37], [39]. This raises an important trade-off between secrecy and recovery distortion which needs to be addressed during the design of a compressive cryptosystem.

VI. Conclusions

In this paper, we derived upper bounds on the distinguishability of a compressive cryptosystem based on quantized Gaussian random matrices. More in detail, the obtained bounds, which hold in the worst case scenario, have also been experimentally validated. The most important achievement is that, as the number of quantization bits employed for the sensing matrix entries is increased, the secrecy of the system

exponentially increases. This strong result demonstrates how a practical compressive cryptosystem can achieve very high secrecy when finite precision is taken into account. Moreover, we also analyzed the quantization of the measurements and showed that this operation is necessary in order to avoid information leakage due to possible support mismatch of the measurements.

Even though CS is not directly comparable with standard cryptographic systems, with this paper we showed that it can be used in practical systems to provide the required secrecy. In fact, the provided bounds can help the design of a practical compressive cryptosystem whose system parameters such as N_b and N_x can be carefully selected in order to achieve the required secrecy level.

ACKNOWLEDGMENT

This work results from the research cooperation with the Sony Technology Center Stuttgart (Sony EuTEC). We would especially like to thank Lev Markhasin and Oliver Erdler from Sony EuTEC for their valuable feedback.

APPENDIX

1) *Proof of Lemma III.1:* We want to upper bound the distance between $\overline{\mathcal{G}}_{\Lambda,\sigma}(z)$ and $\mathcal{D}_{\Lambda,\sigma}(z)$. By the triangle inequality we have

$$\begin{aligned}&|\overline{\mathcal{G}}_{\Lambda,\sigma}(z)-\mathcal{D}_{\Lambda,\sigma}(z)|\\&\quad\le|\overline{\mathcal{G}}_{\Lambda,\sigma}(z)-\mathcal{G}^a_{\Lambda,\sigma}(z)|+|\mathcal{G}^a_{\Lambda,\sigma}(z)-\mathcal{D}_{\Lambda,\sigma}(z)|,\end{aligned}\quad(4)$$

where $\mathcal{G}^a_{\Lambda,\sigma}(z)$ is the midpoint approximation of $\overline{\mathcal{G}}_{\Lambda,\sigma}(z)=w\int_{z-1/2}^{z+1/2}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}dt$. By using the midpoint integral approximation error [40], the first term in (4) can be upper bounded with $\frac{\max_{t\in[z-1/2,z+1/2]}|f''(t)|}{24}$ where $f''(t)=-\frac{t^2-\sigma^2}{\sigma^5\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}$ is the second derivative of the Gaussian function. Moreover, given that by assumption $|z|<k\sigma$, (4) can be upper bounded by $e^{-\frac{z^2}{2\sigma^2}}\left(w\frac{\sigma^2(k^2-1)+1/4+k\sigma}{24\sigma^5\sqrt{2\pi}}e^{\frac{2k\sigma-1/2}{2\sigma^2}}\left|\frac{w}{\sqrt{2\pi}\sigma}-\frac{1}{\rho_\sigma(\Lambda)}\right|\right)$.

Then, rearranging the terms we have that

$$\overline{\mathcal{G}}_{\Lambda^C,\sigma}(z)\le\mathcal{D}_{\Lambda,\sigma}(z)(1+\delta_{Q,\sigma}),$$

where $\delta_{Q,\sigma}=w\frac{\rho_\sigma(\Lambda)[\sigma^2(k^2-1)+1/4+k\sigma]}{24\sigma^5\sqrt{2\pi}}e^{\frac{2k\sigma-1/2}{2\sigma^2}}+\left|\frac{\rho_\sigma(\Lambda)}{\sqrt{2\pi}\sigma}-1+\frac{g_T}{(1-g_T)}\frac{\rho_\sigma(\Lambda)}{\sqrt{2\pi}\sigma}\right|$.

Now, in order to analyze the asymptotic behavior of $\delta_{Q,\sigma}$, let us start by finding the leading asymptotic term of $\rho_\sigma(\Lambda)$ as $\sigma\to\infty$. We can proceed to this derivation by equivalently employing the Poisson summation formula or the Jacobi imaginary transformation [41], we choose the second one. We have that $\lim_{\sigma\to\infty}\rho_\sigma(\Lambda)=\lim_{\sigma\to\infty}\sum_{j=-\infty}^{+\infty}e^{-\frac{j^2}{2\sigma^2}}$ where the infinite summation term is a Jacobi $\theta_3(q',\tau')$ function of parameters $q'=0$ and $\tau'=\frac{i}{2\pi\sigma^2}$. According to the Jacobi imaginary transformation we can write $\theta_3(0,\frac{i}{2\pi\sigma^2})=\sqrt{2\pi\sigma^2}\theta_3(0,2\pi i\sigma^2)$ and taking the first two terms of the series, for $\sigma\to\infty$ we have that the leading asymptotic term is given by $\rho_\sigma(\Lambda)\sim\sigma\sqrt{2\pi}(1+2e^{-2\pi^2\sigma^2})$. Then, considering this asymptotic behavior for $\sigma\to+\infty$ and k constant, we have that $\delta_{Q,\sigma}$ tends to $g_T/(1-g_T)+k^2\cdot O(1/\sigma^2)$. Since $g_T\le2e^{-\frac{k^2}{2}}$, if we set $k=\sqrt{N_b}$ and consider $\sigma=2^{N_b-1}/k$, we see that there exists a regime of k, N_b in which $\delta_{Q,\sigma}$ approaches zero exponentially fast in N_b.

2) *Proof of Lemma III.2:* Since we are considering a truncated distribution, we need to take into account a multiplicative factor $1+\delta'_T$ which re-normalizes the distribution to make it consistent. More in detail, we have that $\delta'_T=\frac{\epsilon_T}{1-\epsilon_T}$, where ϵ_T corresponds to probability of the truncated tails. We have that $\overline{\mathcal{D}}_{\Lambda^C,\sigma}(z)=(1+\delta'_T)\mathcal{D}_{\Lambda,\sigma}(\phi)$. From Lemma 4.4 in [42], the tail probability can be written as $\epsilon_T\le2e^{-\frac{k^2}{2}}$. This means that $\delta'_T\le\delta_T=\frac{2e^{-\frac{k^2}{2}}}{1-2e^{-\frac{k^2}{2}}}$ which approaches zero exponentially fast as k increases.

3) *Proof of Lemma III.3:* We are interested in the sum of two independent r.v. distributed as $\overline{\mathcal{G}}_{\Lambda,\sigma}(\phi)$ whose probability can be written as

$$\begin{aligned}&P[Y=y]\\&\quad=\sum_{\phi\in\Lambda^C}\overline{\mathcal{G}}_{x_1\Lambda^C,\sigma_1}(\phi)\overline{\mathcal{G}}_{x_2\Lambda^C,\sigma_2}(y-\phi)\\&\quad\le(1+\delta_{Q,\sigma_1})(1+\delta_{Q,\sigma_2})\sum_{\phi\in\Lambda}\mathcal{D}_{x_1\Lambda,\sigma_1}(\phi)\mathcal{D}_{x_2\Lambda,\sigma_2}(y-\phi),\end{aligned}$$

where the inequality comes from Lemma III.1.

Considering the asymptotic behavior of $\delta_{Q,\sigma}$ as in Lemma III.1, we have that $\delta_{Q,\sigma_1},\delta_{Q,\sigma_2}\le\delta_{Q,\sigma}$, since $\delta_{Q,\sigma*}$ achieves its maximum when $\sigma*=||\mathbf{x}||\sigma$ is at its smallest value and, since $\mathbf{x}\in\mathbb{Z}^n$, this value is achieved for $||\mathbf{x}||=1$. Moreover, as discusses in Lemma III.1 there exists a regime of N_b,k for which it approaches zero exponentially fast in N_b.

Moreover, combining the results of Lemma 4.12 in [43] and Lemma 2.7 in [44], we have that the probability of the linear combination of two discrete Gaussian random variables is upper bounded by $w\mathcal{D}_{\gcd(x_1,x_2)\Lambda,\sqrt{\sigma_1^2+\sigma_2^2}}(\phi)$ for some scalar $w\le\frac{1}{1-\epsilon}$ independent of ϕ. Then, according to Lemma A.2 in [43] we can write that $w\le(1+\delta_D)$ with $\delta_D\le\frac{\epsilon}{1-\epsilon}\le2\epsilon$. Since in Lemma 4.12 in [43] it is required $\sigma>\eta_\epsilon(\Lambda)$ to hold, where $\eta_\epsilon(\Lambda)$ is the smoothing parameter, we fix $\eta_\epsilon(\Lambda)=\frac{\sigma}{2}$ and have that $\rho_{\frac{\sigma}{2}}(\Lambda*\setminus\{0\})=\epsilon$. At this point, we have that $\delta_D\le2\epsilon\le4\sum_{k=1}^{+\infty}e^{-\frac{k^2\sigma^2||\mathbf{x}_{\{1,2\}}||^2}{8}}\le8\frac{e^{-\frac{\sigma^2}{8}}}{1-e^{-\frac{\sigma^2}{8}}}$ since this bound is maximized for $||\mathbf{x}_{\{1,2\}}||=1$. As it can be seen, this bound approaches zero exponentially fast as N_b increases.

4) *Proof of Lemma III.4:* We prove this Lemma by induction. In case of $K=1$, from the Lemma III.1 we have that $P[Y''=y]\le(1+\delta_Q)\mathcal{D}_{x_1\Lambda,||x_1||\sigma}(y)$. By inducing on K, let us suppose that the Lemma holds for $K-1$. Let us consider $Y'=\boldsymbol{\phi}'^{\mathsf{T}}\mathbf{x}'$ where the last element of $\boldsymbol{\phi}',\mathbf{x}'$ has been pruned. Applying the inductive hypothesis, we have that

$$P[Y'=y]\le(1+\delta_Q)^{K-1}(1+\delta_D)^{K-2}\mathcal{D}_{\gcd(\mathbf{x}')\Lambda,\|\mathbf{x}'\|\sigma}(y),$$

where $\gcd(\mathbf{x}')$ is the greatest common divisor among the entries in $\mathbf{x}'$.

If we now apply the Lemma, it can be shown that $Y=Y''+Y'$ is statistically close to $\mathcal{D}_{\gcd(\mathbf{x})\Lambda,\|\mathbf{x}\|\sigma}(y)$, and in particular

$$P[Y=y]\le(1+\delta_Q)^K(1+\delta_D)^{K-1}\mathcal{D}_{\gcd(\mathbf{x})\Lambda,\|\mathbf{x}\|\sigma}(y),$$

where $\gcd(\mathbf{x})$ is the greatest common divisor among the entries in $\mathbf{x}$.

5) *Proof of Lemma III.5*: Let us assume for simplicity $G = 1$. We have that the distribution of $Z = Q_H(Y)$ can be written as

$$\begin{aligned}
P[Z=z] &= \sum_{y=zH-H/2}^{zH+H/2-1} P[Y=y] \\
&\leq (1+\delta) \sum_{y=zH-H/2}^{zH+H/2-1} \mathcal{D}_{\Lambda,\sigma}(y) \\
&= \frac{(1+\delta)}{\rho_\sigma(\Lambda)} \sum_{y=zH-H/2}^{zH+H/2-1} e^{-\frac{y^2}{2\sigma^2}} \\
&\leq \frac{H(1+\delta)}{\rho_\sigma(\Lambda)} e^{-\frac{(|z|H-H/2)^2}{2\sigma^2}} \\
&\leq \frac{H(1+\delta)}{\rho_\sigma(\Lambda)} e^{-\frac{(zH)^2}{2\sigma^2}} e^{\frac{|z|H^2}{2\sigma^2}} \\
&\leq \frac{H(1+\delta)}{\rho_\sigma(\Lambda)} e^{-\frac{(zH)^2}{2\sigma^2}} e^{\frac{k'H^2}{2\sigma}} \\
&\leq \frac{H(1+\delta)(1+\delta'_R)}{\rho_\sigma(\Lambda)} e^{-\frac{(zH)^2}{2\sigma^2}}
\end{aligned}$$

where $\delta'_R = \frac{k'H^2}{2\sigma} + o\left(\frac{k'H^2}{2\sigma}\right)$ approaches zero exponentially fast as N_b increases. Concerning the scaling factor $\rho_\sigma(\Lambda)$, we have that $\sigma\sqrt{2\pi} - 1 \leq \rho_\sigma(\Lambda) \leq \sigma\sqrt{2\pi} + 1$. Hence, we can lower bound the scaling factor as

$$\begin{aligned}
\rho_\sigma(\Lambda) &\geq H \cdot \frac{\sigma}{H}\sqrt{2\pi} - 1 \\
&\geq H\left(\rho_{\sigma/H}(\Lambda) - 1\right) - 1 \\
&= H\rho_\sigma(H\Lambda) - H - 1 \\
&= H\rho_\sigma(H\Lambda)\left(1 - \frac{H+1}{H\rho_\sigma(H\Lambda)}\right) \\
&\geq H\rho_\sigma(H\Lambda)\left(1 - \frac{H+1}{\sigma\sqrt{2\pi}+1}\right) \\
&\geq H\rho_\sigma(H\Lambda)\left(1 - \delta'_R\right)
\end{aligned}$$

Putting all things together, we have

$$\begin{aligned}
P[Z=z] &\leq \frac{(1+\delta)(1+\delta'_R)e^{-\frac{(zH)^2}{2\sigma^2}}}{(1-\delta'_R)\rho_\sigma(H\Lambda)} \\
&= (1+\delta)(1+\delta_R)\mathcal{D}_{H\Lambda,\sigma}(z)
\end{aligned}$$

where $\delta_R = \frac{2\delta'_R}{1-\delta'_R}$.

6) *Proof of Theorem III.6:* It is easy to verify that $P[Y_{1,2} = y] = 0$ for $|y| > k||\mathbf{x}_{1,2}||_1\sigma$. Hence, combining Lemma III.4 and III.5 we can write

$$P[Z_{1,2}=z] \leq (1+\delta_Q)^{K_{1,2}}(1+\delta_D)^{K_{1,2}-1}(1+\delta_R)\mathcal{D}_{H\Lambda,||\mathbf{x}||\sigma}(z).$$

By applying the TV distance definition it is immediate to obtain

$$\begin{aligned}
\delta(Z_1,Z_2) &\leq \delta(Z_1,\mathcal{D}_{H\Lambda,||\mathbf{x}||\sigma}) + \delta(Z_2,\mathcal{D}_{H\Lambda,||\mathbf{x}||\sigma}) \\
&\leq (K_1+K_2)\delta_Q + o(\delta_Q) + (K_1+K_2-2)\delta_D + \\
&+ o(\delta_D) + 2\delta_R + o(\delta_R),
\end{aligned}$$

where the the terms approach zero exponentially fast for $N_b \to +\infty$ as described in the proof of Lemma III.3.

7) *Proof of Lemma III.7:* Let us define $\bar{\mathbf{x}}' \in \mathbb{R}^n$ to be a signal which is quantized with N_x bits to $\mathbf{x}' \in \mathbb{Z}^n$, moreover let us assume that $|\mathbf{x}'_i| \leq x_{\max}2^{N_x}$ and $||\mathbf{x}'||^2 \geq nx_{\min}^2 2^{2N_x}$, where $0 < x_{\min} \leq x_{\max} \leq 1$. Under the high rate assumption, the quantization error can be considered to be distributed as $U(-\frac{1}{2},\frac{1}{2})$. Accordingly, we have that $\mathbf{x}'_i = \bar{\mathbf{x}}'_i + \epsilon'$, where $\epsilon' \sim U(-1/2,1/2)$. Thus, we can write $S_n = ||\mathbf{x}'||^2 - ||\mathbf{x}''||^2 = \sum_{i=1}^n D_i$ with $D_i = 2\bar{\mathbf{x}}'_i\epsilon'_i - 2\bar{\mathbf{x}}''_i\epsilon''_i + \epsilon_i^{2\prime} - \epsilon_i^{2\prime\prime}$, and $b = \arg\max_i D_i = 2x_{\max}2^{N_x} + 1/4$, $a = \arg\min_i D_i = -2x_{\max}2^{N_x} - 1/4$. Lastly, since it is easy to show that $E[S_n] = 0$, by applying the Hoeffding's concentration inequality, we have that

$$\begin{aligned}
P[|||\mathbf{x}'||^2 - ||\mathbf{x}''||^2| < t||\mathbf{x}'||^2] &> 1 - 2e^{\frac{-2t^2||\mathbf{x}'||^4}{n(4x_{\max}2^{N_x}+\frac{1}{2})^2}} \\
&\geq 1 - 2e^{\frac{-2nt^2x_{\min}^4 2^{2N_x}}{25x_{\max}^2}} \\
&= 1 - \delta(t,N_x),
\end{aligned}$$

where there exists a regime for which both t and $\delta(t,N_x)$ approaches zero exponentially fast in N_x. As an example, this condition is satisfied for $t = \sqrt{N_x}2^{-N_x}$.

8) *Proof of Theorem III.8:* Since the entries of $\mathbf{y}_1, \mathbf{y}_2$ are i.i.d we start by considering a single entry for each of the two re-quantized measurements vectors, namely Z_1 and Z_2. By definition of TV and from Lemma III.4 and III.5 we have that

$$\begin{aligned}
\delta(Z_1,Z_2) &\leq \delta(Z_1,\mathcal{D}_{H\Lambda,||\mathbf{x}_1||\sigma}) + \delta(\mathcal{D}_{H\Lambda,||\mathbf{x}_1||\sigma},\mathcal{D}_{H\Lambda,||\mathbf{x}_2||\sigma}) \\
&+ \delta(\mathcal{D}_{H\Lambda,||\mathbf{x}_2||\sigma},Z_2) \\
&\leq (K_1+K_2)\delta_Q + (K_1+K_2-2)\delta_D + 2\delta_R + \epsilon \\
&+ \delta(\mathcal{D}_{H\Lambda,||\mathbf{x}_1||\sigma},\mathcal{D}_{H\Lambda,||\mathbf{x}_2||\sigma})
\end{aligned}$$

where the last term can be upper bounded by the KL divergence through the Pinsker's inequality as

$$\begin{aligned}
&\leq \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\mathcal{D}_{H\Lambda,||\mathbf{x}_1||\sigma} \,||\, \mathcal{D}_{H\Lambda,||\mathbf{x}_2||\sigma})} \\
&= \sqrt{\frac{1}{2}\left[\sum_{y\in H\Lambda} \frac{e^{-\frac{y^2}{2||\mathbf{x}_2||^2\sigma^2}}}{\rho_{||\mathbf{x}_2||\sigma}(H\Lambda)} \log\left(\frac{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)e^{-\frac{y^2}{2||\mathbf{x}_2||^2\sigma^2}}}{\rho_{||\mathbf{x}_2||\sigma}(H\Lambda)e^{-\frac{y^2}{2||\mathbf{x}_1||^2\sigma^2}}}\right)\right]} \\
&= \sqrt{\frac{1}{2}\left[\log\left(\frac{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)}{\rho_{||\mathbf{x}_2||\sigma}(H\Lambda)}\right) + \frac{||\mathbf{x}_2||^2 - ||\mathbf{x}_1||^2}{2\sigma^2||\mathbf{x}_1||^2||\mathbf{x}_2||^2}\sum_{y\in H\Lambda} y^2 \frac{e^{-\frac{y^2}{2||\mathbf{x}_1||^2\sigma^2}}}{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)}\right]} \\
&= \sqrt{\frac{1}{2}\left[\log\left(\frac{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)}{\rho_{||\mathbf{x}_2||\sigma}(H\Lambda)}\right) + \frac{||\mathbf{x}_2||^2 - ||\mathbf{x}_1||^2}{2||\mathbf{x}_2||^2}\frac{\tilde{\sigma}^2}{\sigma^2}\right]}
\end{aligned}$$

,

where $\tilde{\sigma}^2||\mathbf{x}_1||^2 = \sum_{y\in H\Lambda} y^2 \frac{e^{-\frac{y^2}{2||\mathbf{x}_1||^2\sigma^2}}}{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)}$ is the variance of $\mathcal{D}_{H\Lambda,||\mathbf{x}_1||\sigma}$. We now analyze the asymptotic behavior of the above bound. Let us start with the term $\log\left(\frac{\rho_{||\mathbf{x}_1||\sigma}(H\Lambda)}{\rho_{||\mathbf{x}_2||\sigma}(H\Lambda)}\right)$,

which according to the Proof of Lemma III.1 as $N_b \to +\infty$ tends to $\log\left(\frac{||\mathbf{x}_1||(1+\epsilon_1)}{||\mathbf{x}_2||(1+\epsilon_2)}\right) = \log\left(\frac{||\mathbf{x}_1||}{||\mathbf{x}_2||}\right) + \epsilon_M$, where $\epsilon_M = \log(1+\epsilon_1) - \log(1+\epsilon_2)$ and $\epsilon_{\{1,2\}} = 2e^{-2\pi||\mathbf{x}_{\{1,2\}}||^2\sigma^2} \leq 2e^{-2\pi\epsilon_{\max}\sigma^2}$, which approaches zero exponentially fast in N_b. Moreover, we have $\log(1+\epsilon_{\{1,2\}}) \leq \epsilon_{\{1,2\}}$ and thus $\epsilon_M \leq 2e^{-2\pi\epsilon_{\max}\sigma^2}$ which approaches zero exponentially fast in N_b.

Next, if we now focus on the log energy ratio, similarly to the steps above, we have that $\log\frac{||\mathbf{x}_1||^2}{||\mathbf{x}_2||^2} = \log\left(1+\frac{||\mathbf{x}_1||^2-||\mathbf{x}_2||^2}{||\mathbf{x}_2||^2}\right) = \log(1+\delta'_M) \leq \delta'_M$. Moreover, when Lemma III.7 holds we have that δ'_M becomes negligible for large values of N_x. For the same reason, the term $\delta''_M = \frac{||\mathbf{x}_2||^2-||\mathbf{x}_1||^2}{2||\mathbf{x}_2||^2}\frac{\tilde{\sigma}^2}{\sigma^2}$ becomes negligible for large values of N_x.

Lastly, since the entries in $\mathbf{y}_1, \mathbf{y}_2$ are i.i.d and the θ-distinguishability is upper bounded by their statistical distance, we have that

$$\theta_{\mathbf{\Phi}}(\mathbf{x}_1, \mathbf{x}_2) \leq m\left((K_1+K_2)\delta_Q + (K_1+K_2-2)\delta_D + \delta_R + \delta_M\right) + \epsilon,$$

where $\delta_M = \sqrt{\frac{1}{2}\left(\delta'_M + \delta''_M + \epsilon_M\right)}$ accounts for the energy mismatch and we have that $\delta_M < t$ with probability $1 - \delta(t, N_x)$, where, thanks to Lemma III.7 and $\epsilon_M \leq 2e^{-2\pi\epsilon_{\max}\sigma^2}$, there is a regime of t, N_x, N_b for which t and $\delta(t, N_x)$ approach zero exponentially fast as N_x and N_b increase.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[7] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[8] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Quantifying the feasibility of compressive sensing in portable electroencephalography systems," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 319–328.

[9] S. Li, L. Da Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, 2013.

[10] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 813–817.

[11] J. Schaad and R. Housley, "Advanced encryption standard (aes) key wrap algorithm," 2002.

[12] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, 2011.

[13] M. Testa, T. Bianchi, and E. Magli, "On the secrecy of compressive cryptosystems under finite-precision representation of sensing matrices," in *Circuits and Systems (ISCAS), 2018 IEEE International Symposium on*. IEEE, 2018, pp. 1–4.

[14] T. Bianchi, V. Bioglio, and E. Magli, "Analysis of one-time random projections for privacy preserving compressed sensing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 313–327, 2016.

[15] R. A. Djeujo and C. Ruland, "Secure matrix generation for compressive sensing embedded cryptography," in *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Oct 2016, pp. 1–8.

[16] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-complexity multiclass encryption by compressed sensing," *IEEE transactions on signal processing*, vol. 63, no. 9, pp. 2183–2195, 2015.

[17] T. Bianchi and E. Magli, "Analysis of the security of compressed sensing with circulant matrices," in *Information Forensics and Security (WIFS), 2014 IEEE International Workshop on*. IEEE, 2014, pp. 173–178.

[18] M. Testa, D. Valsesia, T. Bianchi, and E. Magli, *Compressed Sensing for Privacy-Preserving Data Processing*. Springer, 2018.

[19] M. Testa, T. Bianchi, and E. Magli, "Energy obfuscation for compressive encryption and processing," in *Information Forensics and Security (WIFS), 2017 IEEE International Workshop on*. IEEE, 2017.

[20] N. Y. Yu, "Indistinguishability and energy sensitivity of gaussian and bernoulli compressed encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1722–1735, 2018.

[21] R. Fay, "Introducing the counter mode of operation to compressed sensing based encryption," *Information Processing Letters*, vol. 116, no. 4, pp. 279–283, 2016.

[22] Y. Zhang, J. Zhou, F. Chen, L. Y. Zhang, K.-W. Wong, X. He, and D. Xiao, "Embedding cryptographic features in compressive sensing," *Neurocomputing*, vol. 205, pp. 472–480, 2016.

[23] Y. Zhang, *Secure compressive sensing in multimedia data, cloud computing and IoT*. Springer, 2018.

[24] V. Cambareri, J. Haboba, F. Pareschi, H. R. Rovatti, G. Setti, and K. w. Wong, "A two-class information concealing system based on compressed sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, May 2013, pp. 1356–1359.

[25] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[26] D. Needell and J. A. Tropp, "Cosamp: iterative signal recovery from incomplete and inaccurate samples," *Communications of the ACM*, vol. 53, no. 12, pp. 93–100, 2010.

[27] E. Huebner and R. Tichatschke, "Relaxed proximal point algorithms for variational inequalities with multi-valued operators," *Optimization Methods & Software*, vol. 23, no. 6, pp. 847–877, 2008.

[28] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[29] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[30] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[31] E. van den Berg and M. P. Friedlander, "Spgl1: A solver for large-scale sparse reconstruction," 2007.

[32] E. J. Candès and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming, oct. 2005."

[33] E. Van Den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[34] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[35] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on gaussian measures," *SIAM Journal on Computing*, vol. 37, no. 1, pp. 267–302, 2007.

[36] B. Widrow, I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *IEEE Transactions on instrumentation and measurement*, vol. 45, no. 2, pp. 353–361, 1996.

[37] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, 2010.

[38] J. Nymann, "On the probability that k positive integers are relatively prime," *Journal of Number Theory*, vol. 4, no. 5, pp. 469 – 473, 1972. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022314X72900388
[39] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*. IEEE, 2008, pp. 16–21.
[40] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*. Courier Corporation, 2007.
[41] J. Fenton and R. Gardiner-Garden, "Rapidly-convergent methods for evaluating elliptic integrals and theta and elliptic functions," *The ANZIAM Journal*, vol. 24, no. 1, pp. 47–58, 1982.
[42] V. Lyubashevsky, "Lattice signatures without trapdoors," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2012, pp. 738–755.
[43] D. Boneh and D. M. Freeman, "Linearly homomorphic signatures over binary fields and new tools for lattice-based signatures." in *Public Key Cryptography*, vol. 6571. Springer, 2011, pp. 1–16.
[44] C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM, 2008, pp. 197–206.

Matteo Testa (S'14-M'17) received the B.Sc. degree and the M.Sc. degree in telecommunications engineering from the Politecnico di Torino, Turin, Italy, in 2011 and 2012, respectively, and the Ph.D. degree in electronic and communications engineering from the Electronics Department, Politecnico di Torino, in 2016, under the supervision of Prof. E. Magli. He currently holds a post-doctoral position with the IPL lab, Politecnico di Torino, led by Prof. E. Magli in collaboration with SONY EuTEC. His research interests include deep learning methods for security applications, random projections and bayesian inference.

Tiziano Bianchi Tiziano Bianchi (S03M05) received the M.Sc. degree (Laurea) in electronic engineering and the Ph.D. degree in information and telecommunication engineering from the University of Florence, Italy, in 2001 and 2005, respectively. Since 2012, he has been with the Department of Electronics and Telecommunications, Politecnico di Torino, as an Assistant Professor. From 2005 to 2012, he was with the Department of Electronics and Telecommunications, University of Florence, as a Research Assistant. He has authored over 100 papers on international journals and conference proceedings. His research interests include multimedia security technologies, signal processing in the encrypted domain, and security aspects of compressed sensing. He is an Associate Editor of the Journal of Visual Communication and Image Representation.

Enrico Magli Enrico Magli (S'97-M'01-SM'07-F'17) received the M.Sc. and Ph.D. degrees from the Politecnico di Torino, Turin, Italy, in 1997 and 2001, respectively. He is currently a Full Professor with the Politecnico di Torino. His research interests include compressive sensing, image and video processing, and deep learning. He is currently an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the EURASIP Journal on Image and Video Processing, and a former Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA. He was an IEEE Distinguished Lecturer from 2015 to 2016. He was the recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, the IEEE ICIP 2015 Best Student Paper Award (as senior author), and the 2010 and 2014 Best Associate Editor Award for the IEEE TRANSACTIONS 1001 ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.