Low-Complexity Multiclass Encryption by Compressed Sensing

Low-Complexity Multiclass Encryption by Compressed Sensing

Valerio Cambareri, *Student Member, IEEE*, Mauro Mangia, *Member, IEEE*, Fabio Pareschi, *Member, IEEE*, Riccardo Rovatti, *Fellow, IEEE*, Gianluca Setti, *Fellow, IEEE*

***Abstract*—The idea that compressed sensing may be used to encrypt information from unauthorised receivers has already been envisioned, but never explored in depth since its security may seem compromised by the linearity of its encoding process.**

In this paper we apply this simple encoding to define a general private-key encryption scheme in which a transmitter distributes the same encoded measurements to receivers of different classes, which are provided partially corrupted encoding matrices and are thus allowed to decode the acquired signal at provably different levels of recovery quality.

The security properties of this scheme are thoroughly analysed: firstly, the properties of our multiclass encryption are theoretically investigated by deriving performance bounds on the recovery quality attained by lower-class receivers with respect to high-class ones. Then we perform a statistical analysis of the measurements to show that, although not perfectly secure, compressed sensing grants some level of security that comes at almost-zero cost and thus may benefit resource-limited applications.

In addition to this we report some exemplary applications of multiclass encryption by compressed sensing of speech signals, electrocardiographic tracks and images, in which quality degradation is quantified as the impossibility of some feature extraction algorithms to obtain sensitive information from suitably degraded signal recoveries.

***Index Terms*—Compressed sensing, encryption, security, secure communications**

I. Introduction

WITH the rise of paradigms such as wireless sensor networks [1] where a large amount of data is locally acquired by sensor nodes and transmitted remotely for further processing, defending the privacy of digital data gathered and distributed by such networks is a relevant issue. This privacy requirement is normally met by means of encryption stages securing the transmission channel [2], implemented in the digital domain and preceded by analog-to-digital conversion of the signal. Due to their complexity, these cryptographic modules (*e.g.* those implementing the Advanced Encryption Standard (AES) [3]) may require a considerable amount of resources, especially in terms of power consumption.

V. Cambareri and R. Rovatti are with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Italy (e-mail: valerio.cambareri@unibo.it, riccardo.rovatti@unibo.it).

M. Mangia is with the Advanced Research Center on Electronic Systems (ARCES), University of Bologna, Italy (e-mail: mmangia@arces.unibo.it).

F. Pareschi and G. Setti are with the Engineering Department in Ferrara (ENDIF), University of Ferrara, Italy (e-mail: fabio.pareschi@unife.it, gianluca.setti@unife.it).

Compressed Sensing (CS) [4], [5] is a mature signal processing technique used in the development of novel data acquisition schemes. CS exploits the structure of certain signals to simultaneously perform data compression and acquisition at the physical interface between the analog and digital domain, thus allowing acquisition at sub-Nyquist rates [6], [7]. This efficient acquisition is commonly followed by a decoding algorithm that maps an undersampled set of CS-encoded measurements into a recovery of the original signal. Within this framework, many sensing architectures have been proposed for the acquisition of a variety of signals [8]–[10].

We investigate on the possibility of using CS with Bernoulli random encoding matrices [11] as a physical-layer method to embed security properties in the acquisition process. Although it is well known that CS cannot be regarded as perfectly secure [12] we will formalise its main weaknesses and strengths as an exploration of the trade-off between achievable security properties and resource requirements in low-complexity acquisition systems, for which an almost-zero cost encryption mechanism is an appealing option.

In more detail, we here devise an encryption strategy relying on the fact that any receiver attempting to decode the CS measurements must know the true encoding matrix used in the acquisition process to attain exact signal recovery. In partial or complete defect of this information, the recovered signal will be subject to a significant amount of recovery noise [13].

We exploit this decoder-side sensitivity to provide multiple recovery quality-based levels (*i.e. classes*) of access to the information carried in the signal. In fact, when the true encoding matrix is completely unknown the signal is fully encrypted, whereas if a receiver knows it up to some random perturbations the signal will still be recoverable, albeit with limited quality. We therefore aim to control the recovery performances of users (receivers) belonging to the same class by exploiting their ignorance of the true encoding matrix. Since these encoding matrices are generated from the available private keys at the corresponding decoders, high-class receivers are given a complete key and thus the true encoding matrix, lower-class receivers are given an incomplete key resulting in a partially corrupted encoding matrix. To ensure that this mismatch goes undetected by lower-class receivers, we only alter the sign of a randomly chosen subset of the entries of the true encoding matrix, which is itself assumed to be a realisation of a ± 1-valued Bernoulli random matrix.

This contribution is structured as follows: in Section II we briefly review the theoretical framework of CS, introduce the mathematical model of two-class and multiclass CS, and

perform upper and lower bound analyses on the recovery error norm suffered by lower-class receivers depending on the chosen amount of perturbation.

Section III addresses the robustness of CS with universal random encoding matrices [11] against straightforward statistical attacks. While CS is not perfectly secret in the Shannon sense [12] and in general suffers from continuity due to its linear nature, we prove that, asymptotically, nothing can be inferred about the encoded signal except for its power and formalise this fact in a relaxed secrecy condition. Moreover, we show how the convergence to this behaviour is sharp for finite signal dimensions. Hence, eavesdroppers are practically unable to extract but a very limited amount of information from the sole statistical analysis of CS measurements. Other non-statistical attacks of a more threatening nature will be treated separately in a future contribution.

In Section IV we propose example applications of multiclass CS to concealing sensitive information in images, electrocardiographic tracks and speech signals. The recovery performances are evaluated in a signal processing perspective to prove the efficacy of this strategy at integrating some security properties in the sensing process, with the additional degree of freedom of allowing multiple quality levels and eventually hiding selected signal features to certain user classes.

A. Relation to Prior Work

This contribution mainly improves on two separate lines of research: (i) statistical security analyses of the CS encoding and (ii) the effect of encoding matrix perturbations on signal recovery. Line (i) stems from the security analysis in [12]. Both [12] and [14] showed how brute-force attacks are computationally infeasible, so that some security properties could indeed be provided in relevant applications [15], [16]. We deepen the results in [12] by introducing an asymptotic notion of secrecy for signals having the same power, and by verifying it for CS. We assess its consequences for finite dimensions by means of hypothesis testing and develop a non-asymptotic analysis of the rate at which acquired signals having the same energy become indistinguishable when observing the probability distribution of their subgaussian CS measurements. This is obtained by adapting a recent result in probability theory [17]. On the other hand, line (ii) relates to studying the effect of our particular sparse random perturbation matrix on signal recovery performances, a field that is closely related to statistical sparse regression with corrupted or missing predictors [18], [19]. The authors of [13] quantify this effect in a general framework: we will adapt these results to our case for a formal analysis of the *worst-case* signal recovery performances of lower-class users, while developing some new arguments to find their *best-case* recovery performances – our aim being the distinction between different user classes.

II. Multiclass Compressed Sensing

A. Brief Review of Compressed Sensing

Compressed sensing [4], [5] is summarised by considering the following setting: let x be a vector in $\mathbb{R}^n$ and $y \in \mathbb{R}^m$ a vector of *measurements* obtained from x by applying a linear dimensionality-reducing transformation $y = Ax$, *i.e.*

$$y_j = \sum_{l=0}^{n-1} A_{j,l} x_l \,, \;\; j = 0, \ldots, m-1 \tag{1}$$

with $A_{m\times n}$ the *encoding matrix*. Under suitable assumptions, fundamental results [4], [20], [21] showed it is possible to recover x from y even if $m < n$.

The first of such assumptions is that x has a *k-sparse* representation, *i.e.* we assume that there exists a *sparsity basis* $D_{n\times n}$ such that $x = Ds$, with $s \in \mathbb{R}^n$ having a support of cardinality k. This cardinality is also indicated as $\|s\|_0 = k$ with $k \ll n$. Asserting that x is represented by $k < m < n$ non-zero coefficients in a suitable domain intuitively means that its intrinsic information content is smaller than the apparent dimensionality. In the following we will assume that D is an orthonormal basis (ONB).

A second assumption must be made on the structure of A. Many conditions have been formulated in the literature (*e.g.* the *restricted isometry property*, RIP [22]) to guarantee that the information in s is preserved through the mapping $y = ADs$. To the purpose of this paper it suffices to say that the most *universal* option (*i.e.* independently of D) is choosing A as typical realisations of a random matrix with independent and identically distributed (i.i.d.) entries from a subgaussian distribution, *e.g.* an i.i.d. Gaussian or Bernoulli random matrix [11]. We will let A be an $m \times n$ i.i.d. Bernoulli random matrix[1] unless otherwise noted.

When both these conditions hold, s can be recovered from $y = ADs$ as the sparsest vector solving the hard problem

$$s = \arg\min_{\xi\in\mathbb{R}^n} \|\xi\|_0 \text{ s. t. } y = AD\xi \tag{P_0}$$

Moreover, if the dimensionality of the measurements $m \sim k \log \frac{n}{k}$ is not too small w.r.t. that of x and its sparsity w.r.t. D, (P_0) can be relaxed to the convex ℓ_1-norm minimisation

$$\hat{s} = \arg\min_{\xi\in\mathbb{R}^n} \|\xi\|_1 \text{ s. t. } y = AD\xi \tag{P_1}$$

still yielding $\hat{s} = s$ (provided that A is carefully chosen [21]), with (P_1) being a linear programming problem solved with polynomial-time algorithms [20]. In the following, we will refer to this problem as the $\min \ell_1$ decoder (also known as *basis pursuit*, BP).

B. A Cryptographic Perspective

Standard CS may be interpreted as a private key cryptosystem where x is the *plaintext*, the measurement vector y is the *ciphertext* and the *encryption algorithm* is a linear transformation operated by the encoding matrix A defining the acquisition process. In the classic setting, Alice acquires a plaintext x by CS using A and sends to Bob the ciphertext y; Bob is able to successfully recover x from y if he is provided with A or equivalently the private key required to generate it.

[1]This notation is used both for the random matrix and its realisations, disambiguating with the term *instance* where needed.

Since many CS-based acquisition systems [8], [9] entail the use of i.i.d. Bernoulli random matrices generated by a pseudo-random number generator (PRNG) we define *encryption key* (or *shared secret*) the initial seed which is expanded by the PRNG to generate a sequence of encoding matrices. In the following we will assume that the period of this sequence is sufficiently long to guarantee that in a reasonable observation time no two encoding matrices will be the same, *i.e.* that each plaintext will be encoded with a different matrix. With this hypothesis, we let each instance of A be a generic, unique element of the aforementioned sequence.

C. Signal Models and Assumptions

This paper will analyse the security properties of CS starting from some statistical properties of the signal being encoded as in (1). Rather than relying on its *a priori* distribution, our analysis uses general moment assumptions that may correspond to many probability distributions on the signal domain. We will therefore adopt the following signal models:

(m_1) for finite n, we let $X = \{X_j\}_{j=0}^{n-1}$ be a real random vector (RV). Its realisations (finite-length plaintexts) $x = \left(x_0, \cdots, x_{n-1}\right) \in \mathbb{R}^n$ are assumed to have finite energy $e_x = \|x\|_2^2$. We will let each $x = Ds$ with D an ONB and s being k-sparse to comply with sparse signal recovery guarantees [21]. X is mapped to the measurements' RV $Y = \{Y_j\}_{j=0}^{m-1}$ (whose realisations are the ciphertexts y) as $Y = AX$, *i.e.* each realisation of (Y, A, X) is an instance of (1).

(m_2) for $n \to \infty$, we let $\mathbf{X} = \{X_j\}_{j=0}^{+\infty}$ be a real random process (RP). Its realisations (infinite-length plaintexts) x are assumed to have finite power $W_x = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} x_j^2$. We denote them as sequences $x = \{x^{(n)}\}_{n=0}^{+\infty}$ of finite-length plaintexts $x^{(n)} = \left(x_0, \cdots, x_{n-1}\right)$. $\mathbf{X}$ is mapped to either a RV Y of realisations (ciphertexts) y for finite m, or a RP $\mathbf{Y} = \{Y_j\}_{j=0}^{+\infty}$ of ciphertexts y for $m, n \to \infty, \frac{m}{n} \to q$. Both cases are comprised of random variables $Y_j = \frac{1}{\sqrt{n}} \sum_{l=0}^{n-1} A_{j,l} X_l$. The $\frac{1}{\sqrt{n}}$ scaling is not only theoretically needed for normalisation purposes, but also practically required in the design of finite quantiser ranges for CS-based acquisition front-ends.

When none of the above models is specified, a single realisation of (1) is considered as in the standard CS framework (Section II-A).

D. Multiclass Encryption by Compressed Sensing

Let us consider a scenario where multiple users receive the same measurements y, know the sparsity basis D, but are made different by the fact that some of them know the true A, whereas the others only know an approximate version of it. The resulting mismatch between A and its approximation used in the decoding process by the latter set of receivers will limit the quality of signal recovery as detailed below.

1) Two-Class Scheme: With this principle in mind a straightforward method to introduce perturbations is flipping the sign of a subset of the entries of the encoding matrix in a random pattern. More formally, let $A^{(0)}$ denote the *initial encoding matrix* and $C^{(0)}$ a subset of $c < m \cdot n$ index pairs chosen at random for each $A^{(0)}$. We construct the *true encoding matrix* $A^{(1)}$ by

$$A_{j,l}^{(1)} = \begin{cases} A_{j,l}^{(0)}, & (j,l) \notin C^{(0)} \\ -A_{j,l}^{(0)}, & (j,l) \in C^{(0)} \end{cases}$$

and use it to encode x as in (1). Although this alteration simply involves inverting c randomly chosen sign bits in a buffer of $m \cdot n$ pseudorandom symbols, we will use its linear model

$$A^{(1)} = A^{(0)} + \Delta A \tag{2}$$

where ΔA is a c-sparse random perturbation matrix of entries

$$\Delta A_{j,l} = \begin{cases} 0, & (j,l) \notin C^{(0)} \\ -2A_{j,l}^{(0)}, & (j,l) \in C^{(0)} \end{cases} \tag{3}$$

or equivalently

$$\Delta A_{j,l} = \begin{cases} 0, & (j,l) \notin C^{(0)} \\ 2A_{j,l}^{(1)}, & (j,l) \in C^{(0)} \end{cases} \tag{4}$$

with density $\eta = \frac{c}{mn}$, *i.e.* the ratio of non-zero entries w.r.t. the product of the dimensions of ΔA. By doing so, any receiver is still provided an encoding matrix differing from the true one by an instance of ΔA. This perturbation is *undetectable*, *i.e.* $A^{(1)}$ and $A^{(0)}$ are statistically indistinguishable since they are equal-probability realisations of the same i.i.d. Bernoulli random matrix ensemble [11] with all points in $\{-1, 1\}^{m\times n}$ having the same probability.

A *first-class* user receiving $y = A^{(1)}x = (A^{(0)} + \Delta A)x$ and knowing $A^{(1)}$ is able to recover, in absence of other noise sources and with m sufficiently larger than k, the exact sparse solution $\hat{s} = s$ by solving (P_1) [20], [21]. A *second-class* user only knowing y and $A^{(0)}$ is instead subject to an equivalent signal- and perturbation-dependent noise term ε due to missing pieces of information on $A^{(1)}$, *i.e.*

$$y = A^{(1)}x = A^{(0)}x + \varepsilon \tag{5}$$

where $\varepsilon = \Delta A x$ is a pure disturbance since both ΔA and x are unknown to the second-class decoder.

In general, performing signal recovery in the erroneous assumption that $y = A^{(0)}x$, *i.e.* with a corrupted encoding matrix will lead to a noisy recovery of x. Nevertheless, upper bounds on the recovery error norm $\|\hat{x} - x\|_2$ (with $\hat{x} = D\hat{s}$, $\hat{s}$ an approximation of s) are well known for measurements affected by generic additive noise [21, Theorem 1.1, 1.2]. These bounds have been extended in [13] to a general perturbed encoding matrix model that encompasses (5). We adapt these results in Section II-E2 to obtain a worst-case analysis of the second class recovery error norm.

Moreover, to prove the difference between first- and second-class recovery performances, in Section II-E1 we develop a lower bound, *i.e.* a best-case analysis of the second-class recovery error norm. Both performance bounds show a clear dependence on the perturbation density η, which is suitably chosen to fix the desired quality range for each class.

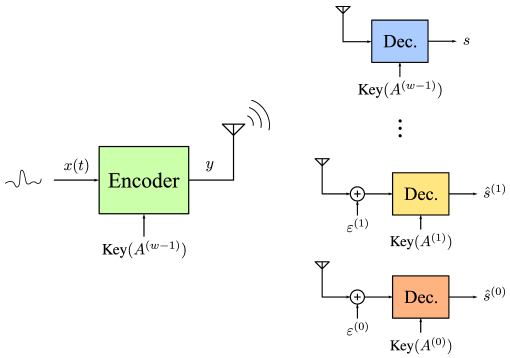


Fig. 1. A multiclass CS network: the encoder acquires an analog signal $x(t)$ at sub-Nyquist rate and transmits the measurement vector y. Low-quality decoders reconstruct a signal approximation with partial knowledge of the encoding, resulting in additive perturbation noise $\varepsilon^{(u)}$ and leading to an approximate solution $\hat{s}^{(u)}$ for the u-th user class.

2) Multiclass Scheme: The two-class scheme may be iterated to devise an arbitrary number w of user classes: sign-flipping is now applied on disjoint subsets of index pairs $C^{(u)}, u = 0, \ldots, w-2$ of $A^{(0)}$ so that

$$A_{j,l}^{(u+1)} = \begin{cases} A_{j,l}^{(u)}, & (j,l) \notin C^{(u)} \\ -A_{j,l}^{(u)}, & (j,l) \in C^{(u)} \end{cases}$$

If the plaintext x is encoded with $A^{(w-1)}$ then we may distinguish *high-class* users knowing the complete encoding $A^{(w-1)}$, *low-class* users knowing only $A^{(0)}$ and *mid-class* users knowing $A^{(u+1)}$ with $u = 0, \ldots, w-3$. This simple technique can be applied to provide multiple classes of access to the information in x by having different signal recovery performances at the decoder.

3) A System Perspective: The strategy described in this section provides a multiclass encryption architecture where the shared secret between the CS encoder and each receiver is distributed depending on the quality level granted to the latter.

In particular, the full encryption key of a w-class CS system is composed of w seeds, *i.e.* low-class users are provided the secret Key($A^{(0)}$), class-1 users are provided Key($A^{(1)}$) $=$ $\big($Key($C^{(0)}$), Key($A^{(0)}$)$\big)$ up to high-class users with Key($A^{(w-1)}$) $=$ $\big($Key($C^{(w-2)}$), $\cdots$, Key($C^{(0)}$), Key($A^{(0)}$)$\big)$. A sample network scenario is depicted in Fig. 1.

From the resources point of view, multiclass CS can be enabled with very small computational overhead. The encoding matrix generator is the same at both the encoder and high-class decoder side, whereas lower-class decoders may use the same generation scheme but are unable to rebuild the true encoding due to the missing private keys Key($C^{(u)}$).

The initial matrix $A^{(0)}$ is updated from a pseudorandom binary stream generated by expanding Key($A^{(0)}$) with a PRNG. The introduction of sign-flipping is a simple post-processing step carried out on the stream buffer by reusing the same PRNG architecture and expanding the corresponding Key($C^{(u)}$), thus having minimal computational cost.

Since the values generated by this PRNG are never exposed, cryptographically secure generators may be avoided, provided that the period with which the matrices are reused is kept sufficiently large – this requirement is crucial to avoid attacks that could exploit multiple plaintext-ciphertext pairs to fully or partially recover the encoding matrix.

E. Lower-Class Recovery Performance Bounds

In order to quantify the recovery quality performance gap between low- and high-class users receiving the same CS measurements from the network of Fig. 1, we now provide performance bounds on the recovery error in the simple two-class case, starting from the basic intuition that if the sparsity basis of x is not the canonical basis, then most plaintexts $x \notin \mathrm{Ker}\,(\Delta A)$, so the perturbation noise $\varepsilon = \Delta A x \neq 0$.

1) Second-Class Recovery Error – Lower Bound: The following results aim at predicting the best-case recovery quality of any second-class decoder that assumes y was encoded by $A^{(0)}$, whereas $y = A^{(1)}x$ in absence of other noise sources and regardless of the sparsity of x.

Theorem 1 (Second-class recovery error lower bound)**.** *Let:*

1) $A^{(0)}, A^{(1)}$ *be* $m \times n$ *i.i.d. Bernoulli random matrices as in* (2) *and* ΔA *the sparse random perturbation matrix in* (3) *of density* $\eta \leq \frac{1}{2}$*;*
2) X *be as in* (m_1) *with finite* $\mathcal{E}_x = \mathbb{E}[\sum_{j=0}^{n-1} X_j^2]$, $\mathcal{F}_x = \mathbb{E}[(\sum_{j=0}^{n-1} X_j^2)^2]$ *and* $Y = A^{(1)}X$ *be the corresponding measurements' RV;*

For all $\theta \in (0,1)$ *and any instance* $y = A^{(1)}x$*, any* $\hat{x}$ *that satisfies* $y = A^{(0)}\hat{x}$ *is such that the recovery error norm*

$$\|\hat{x} - x\|_2^2 \geq \frac{4\eta m \mathcal{E}_x}{\sigma_{\max}(A^{(0)})^2}\,\theta \tag{6}$$

with probability

$$\zeta = \frac{1}{1+(1-\theta)^{-2}\left\{\left[1+\frac{1}{m}\left(\frac{3}{2\eta}-1\right)\right]\frac{\mathcal{F}_x}{\mathcal{E}_x^2}-1\right\}} \tag{7}$$

where $\sigma_{\max}(\cdot)$ *denotes the maximum singular value of its argument.*

Corollary 1 (Asymptotic case of Theorem 1)**.** *Let:*

1) $A^{(0)}, A^{(1)}, \Delta A, \eta$ *be as in Theorem 1 as* $m, n \to \infty, \frac{m}{n} \to q$*;*
2) $\mathbf{X}$ *be as in* (m_2)*,* α*-mixing [23, (27.25)], with finite* $\mathcal{W}_x = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\sum_{j=0}^{n-1} X_j^2]$ *and uniformly bounded* $\mathbb{E}[X_j^4] \leq m_x$ *for some* $m_x > 0$*. Denote with* $\mathbf{Y}$ *the corresponding measurements' RP of instances* y*;*

For all $\theta \in (0,1)$ *and* $y = \frac{1}{\sqrt{n}}A^{(1)}x$*, any* $\hat{x}$ *that satisfies* $y = \frac{1}{\sqrt{n}}A^{(0)}\hat{x}$ *is such that the recovery error power*

$$W_{\hat{x}-x} = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}(\hat{x}_j - x_j)^2 \geq \frac{4\eta q \mathcal{W}_x}{(1+\sqrt{q})^2}\,\theta \tag{8}$$

with probability 1.

The proof of these statements is given in Appendix A. Simply put, Theorem 1 and Corollary 1 state that a second-class decoder recovering $\hat{x}$ such that $y = A^{(0)}\hat{x}$ is subject to a recovery error whose norm, with high probability, exceeds a quantity depending on the density η of the perturbation ΔA, the undersampling rate $\frac{m}{n}$ and the average energy $\mathcal{E}_x$ or

power $\mathcal{W}_x$ respectively. In particular, the non-asymptotic case in (6) is a probabilistic lower bound: as a quantitative example, by assuming it holds with probability $\zeta = 0.98$ and that $\frac{\mathcal{F}_x}{\mathcal{E}_x^2} = 1.0001, n = 1024, m = 512, \sigma_{\max}(A^{(0)}) \approx \sqrt{m} + \sqrt{n}$ (see [24]) one could take an arbitrary $\theta = 0.1 \Rightarrow \eta = 0.1594$ to obtain $\|\hat{x} - x\|_2^2 \geq 0.0109$ w.r.t. RVs having average energy $\mathcal{E}_x = 1$. In other words, with probability 0.98 a perturbation of density $\eta = 0.1594$ will cause a minimum recovery error norm of 19.61 dB.

A stronger asymptotic result holding with probability 1 on $W_{\hat{x}-x}$ is then reported in Corollary 1 under mild assumptions on the RP $\mathbf{X}$, where θ can be arbitrarily close to 1 and only affecting the convergence rate to this lower bound. The bounds in (6) and (8) are adopted as reference best-cases in absence of other noise sources for the second-class decoder, which exhibits higher recovery error for most problem instances and reconstruction algorithms as detailed in Section IV.

2) Second-Class Recovery Error – Upper Bound: We now derive a second-class recovery error upper bound by applying the theory in [13] which extends the well-known recovery guarantees in [21] to a perturbed encoding matrix model identical to (2). While adaptations exist [25] none tackle the unstructured, i.i.d. sparse random perturbation of (3) which we will now model.

The framework of [13] analyses the recovery error upper bound when using $A^{(0)} = A^{(1)} - \Delta A$ as the encoding matrix in the reference *basis pursuit with denoising* (BPDN) problem

$$\hat{s} = \arg\min_{\xi \in \mathbb{R}^n} \|\xi\|_1 \text{ s. t. } \|y - A^{(0)} D\xi\|_2 \leq \gamma \qquad (P_2)$$

for a given noise parameter $\gamma \geq \|\varepsilon\|_2$. Let $\sigma^{(k)}_{\min/\max}(\cdot)$ denote the extreme singular values among all k-column submatrices of a matrix, and define the perturbation-related constants

$$\epsilon^{(k)}_{A^{(1)}} \geq \frac{\sigma^{(k)}_{\max}(-\Delta AD)}{\sigma^{(k)}_{\max}(A^{(1)}D)}; \ \epsilon_{A^{(1)}} \geq \frac{\sigma_{\max}(-\Delta AD)}{\sigma_{\max}(A^{(1)}D)} \geq \epsilon^{(k)}_{A^{(1)}} \qquad (9)$$

and the well-known *restricted isometry constant* (RIC, [22]) $\delta^{(k)} = \max\{\sigma^{(k)}_{\max}(A^{(1)}D)^2 - 1, 1 - \sigma^{(k)}_{\min}(A^{(1)}D)^2\}$. We estimate and plug these quantities in the upper bound of [13, Theorem 2]. For the sake of simplicity, we report it below in the case of plaintexts x having exactly k-sparse representations in absence of other noise sources.

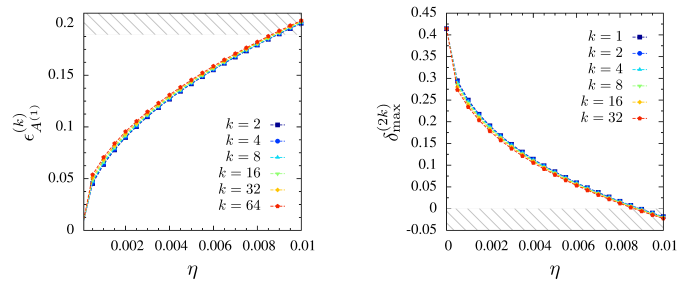
Proposition 1 (Second-class recovery error upper bound, adapted from [13])**.** *Let*

1) $A^{(0)}, A^{(1)}, \Delta A, \eta$ *be as in Theorem 1;*
2) X *be as in* (m_1)*, with* $x = Ds$*,* D *an ONB and* s *being* k*-sparse;*
3) $\epsilon^{(2k)}_{A^{(1)}} < 2^{\frac{1}{4}} - 1$ *and* $\delta^{(2k)} < \delta^{(2k)}_{\max} = \sqrt{2}(1+\epsilon^{(2k)}_{A^{(1)}})^{-2} - 1$*;*

For any instance $y = A^{(1)}x$*, a vector* $\hat{x} = D\hat{s}$ *with* $\hat{s}$ *the solution of* (P_2) *with noise parameter* $\gamma = \epsilon^{(k)}_{A^{(1)}} \sqrt{\frac{1+\delta^{(k)}}{1-\delta^{(k)}}} \|y\|_2$ *obeys [13]*

$$\|\hat{x} - x\|_2 \leq \overline{C}\gamma, \overline{C} = \frac{4\sqrt{1+\delta^{(2k)}}(1+\epsilon^{(2k)}_{A^{(1)}})}{1-(\sqrt{2}+1)\left[(1+\delta^{(2k)})(1+\epsilon^{(2k)}_{A^{(1)}})^2 - 1\right]} \qquad (10)$$

Such a guarantee depends on $\epsilon^{(k)}_{A^{(1)}}, \epsilon^{(2k)}_{A^{(1)}}$: theoretical results exist for estimating their value by bounding the maximum singular values in (9) since the entries of $A^{(1)}$ and ΔA are i.i.d. (in particular, [24] applies to $A^{(1)}$, [26] to ΔA). Yet, they would hold only when D is the identity and involve universal constants whose values would require numerical evaluation. For these reasons we choose to estimate the required quantities directly by Monte Carlo simulation. As an example, we calculate (9) for 10^4 instances of submatrices of $A^{(1)}$ and ΔA with $m = 512, k = 2, 4, \ldots, 64$ and $\eta \in [5 \cdot 10^{-4}, 10^{-2}]$. This allows us to find typical values of $\epsilon^{(k)}_{A^{(1)}}$ as reported in Fig. 2a. In the same setting $\epsilon^{(k)}_{A^{(1)}} < 2^{\frac{1}{4}} - 1$ only when $\eta \leq 8 \cdot 10^{-3}$. In Fig. 2b we report the corresponding range of allowed RIC $\delta^{(2k)} \leq \delta^{(2k)}_{\max}$ that comply with Proposition 1, *i.e.* the RIC constraints the encoding matrices must meet so that (10) holds.



(a) Empirical values of $\epsilon^{(k)}_{A^{(1)}}$ (b) Maximum allowed values of $\delta^{(2k)}$

Fig. 2. Empirical evaluation of the constants in Proposition 1 based on a large number of $A^{(1)}, \Delta A$ with $m = 512, \eta \in [5 \cdot 10^{-4}, 10^{-2}]$ and D a random ONB.

Such RIP-based analyses provide very strong sufficient conditions for signal recovery (see [27, Section X]) which in our case result in establishing a formal upper bound for a small range of η and when solving (P_2). As observed by the very authors of [13], typical recovery errors are substantially smaller than this upper bound. We will therefore rely on another less rigorous, yet practically effective least-squares approach using the same hypotheses of Theorem 1 to bound the average recovery quality performances in Section IV.

III. A Cryptanalysis of Compressed Sensing

Consider a generic CS scheme as in Section II-A with $y = Ax$ linearly encoding a plaintext x into a ciphertext y. We now investigate the security properties and limits of such random linear measurements by letting x, y be realisations of either RVs (m_1) or RPs (m_2) with their respective *a priori* distributions as in the classic Shannon framework [28].

A. Security Limits

The encoding performed by CS is a linear mapping, and as such it cannot completely hide the information contained in a plaintext x. This has two main consequences: firstly, linearity propagates scaling. Hence, it is simple to distinguish a plaintext x' from another x'' if one knows that $x'' = \alpha x'$ for some scalar α. For the particular choice $\alpha = 0$ this leads to a known argument [12, Lemma 1] against the fundamental requirement for perfect secrecy that the conditional PDF $f_{Y|X}(y|x) = f_Y(y)$ (*e.g.* in model (m_1)). In the following, we will prove that a scaling factor is actually all that can be

inferred from the statistical analysis of CS-encoded ciphertexts.

Secondly, linearity implies continuity. Hence, whenever x' and x'' are close to each other for a fixed A, the corresponding y' and y'' will also be close to each other. This fact goes against the analog version of the *diffusion* (or *avalanche effect*) requirement for digital-to-digital ciphers [29]. If the encoding process did not entail a dimensionality reduction, this fact could be exploited every time a plaintext-ciphertext pair x', y' is known. If a new ciphertext y'' is available that is close to y', then it is immediately known that the corresponding plaintext x'' must be close to x' thus yielding a good starting point for *e.g.* a brute-force attack.

The fact that $m < n$ slightly complicates this setting since the counterimages of y'' through A belong to a whole subspace in which points arbitrarily far from x' exist in principle. Yet, encoding matrices A are chosen by design so that the probability of their null space aligning with x' and x'' (that are k-sparse w.r.t. a certain D) is overwhelmingly small [22]. Hence, even if with some relaxation from the quantitative point of view, neighbouring ciphertexts strongly hint at neighbouring plaintexts. As an objection to this seemingly unavoidable issue note that the previous argument only holds when the encoding matrix remains the same for both plaintexts, while by our assumption (Section II-B) on the very large period of the generated sequence of encoding matrices two neighbouring plaintexts x', x'' will most likely be mapped by different encoding matrices to non-neighbouring ciphertexts y', y''.

B. Achievable Security Properties

1) Asymptotic Security: While perfect secrecy is unachievable, we may introduce the notion of *asymptotic spherical secrecy* and show that CS with universal random encoding matrices has this property, *i.e.* no information can be inferred on a plaintext x in model (m_2) from the statistical properties of all its possible ciphertexts but its power. The implication of this property is the basic guarantee that a malicious eavesdropper intercepting the measurement vector will not be able to extract any information on the plaintext except for its power.

Definition 1 (Asymptotic spherical secrecy)**.** *Let $\mathbf{X}$ be a RP whose plaintexts have finite power $0 < W_x < \infty$, $\mathbf{Y}$ be a RP modelling the corresponding ciphertexts. A cryptosystem has asymptotic spherical secrecy if for any of its plaintexts x and ciphertexts y we have*

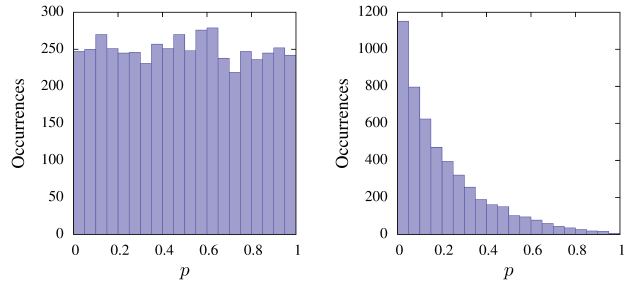
$$f_{\mathbf{Y}|\mathbf{X}}(y|x) \underset{\mathcal{D}}{\longrightarrow} f_{\mathbf{Y}|W_x}(y) \tag{11}$$

where $\underset{\mathcal{D}}{\rightarrow}$ denotes convergence in distribution as $m, n \to \infty$, $f_{\mathbf{Y}|W_x}$ denotes conditioning over plaintexts x with the same power W_x.

From an eavesdropper's point of view, asymptotic spherical secrecy means that given any ciphertext y we have

$$f_{\mathbf{X}|\mathbf{Y}}(x|y) \simeq \frac{f_{\mathbf{Y}|W_x}(y)}{f_{\mathbf{Y}}(y)} f_{\mathbf{X}}(x)$$

implying that any two different plaintexts with an identical, prior and equal power W_x remain approximately indistinguishable from their ciphertexts. In the asymptotic setting, the following proposition holds.



(a) $e_{x'} = e_{x''} = 1$; uniformity test p-value $= 0.4775$ implies uniformity at 5% significance.

(b) $e_{x'} = 1$, $e_{x''} = 1.01$; uniformity test p-value $\simeq 0$ implies non-uniformity.

Fig. 3. Outcome of second-level statistical tests to distinguish between two orthogonal plaintexts x', x''. In (a) x', x'' have $e_{x'} = e_{x''}$, spherical secrecy applies and the uniform distribution of p-values shows that the corresponding ciphertexts are statistically indistinguishable. In (b) x', x'' have $e_{x'} \neq e_{x''}$, spherical secrecy does not apply and the distribution of p-values shows that the corresponding ciphertexts are distinguishable.

Proposition 2 (Asymptotic spherical secrecy of random measurements)**.** *Let $\mathbf{X}$ be a RP with bounded-value plaintexts of finite power W_x, Y_j any variable of the RP $\mathbf{Y}$ as in (m_2). For $n \to \infty$ we have*

$$f_{Y_j|\mathbf{X}}(y_j) \underset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, W_x) \tag{12}$$

Thus, universal encoding matrices provide independent, asymptotically spherical-secret measurements as in (11).

The proof of this statement is given in Appendix B. Since the rows of A are independent, the measurements conditioned only on W_x are also independent and Proposition 2 asserts that, although not secure in the Shannon sense, CS with suitable encoding matrices is able to conceal the plaintext up to the point of guaranteeing its security for $n \to \infty$.

As a more empirical illustration of spherical secrecy for finite n, we consider an attack aiming at distinguishing two orthogonal plaintexts x' and x'' from their encryption (clearly, finite energy must be assumed as in (m_1)). The attacker has access to a large number χ of ciphertexts collected in a set $\mathcal{Y}'$ obtained by applying different, randomly chosen encoding matrices to a certain x' as in (1). Then, the attacker collects another set $\mathcal{Y}''$ of χ ciphertexts, all of them corresponding either to x' or to x'', and must tell which is the true plaintext between the two. This reduces the attack to an application of statistical hypothesis testing, the null assumption being that the distribution underlying the samples in $\mathcal{Y}''$ is the same as that underlying the samples in $\mathcal{Y}'$. For maximum reliability we adopt a two-level testing approach: we repeat the above experiment for many instances of random orthogonal plaintexts x' and x'', performing a two-way Kolmogorov-Smirnov (KS) test to compare the empirical distributions obtained from $\mathcal{Y}'$ and $\mathcal{Y}''$ produced by such orthogonal plaintexts.

Each of the above tests yields a p-value quantifying the probability that two data sets coming from the same distribution exhibit larger differences w.r.t. those at hand. Given

their meaning, individual p-values could be compared against a desired significance level to give a first assessment whether the null hypothesis (*i.e.* equality in distribution) can be rejected. Yet, since it is known that p-values of independent tests on distributions for which the null assumption is true must be uniformly distributed in $[0,1]$ we collect P of them and feed this second-level set of samples into a one-way KS test to assess uniformity at the standard significance level 5%.

This testing procedure is done for $n = 256$ in the cases $e_{x'} = e_{x''} = 1$ (same energy plaintexts) and $e_{x'} = 1, e_{x''} = 1.01$, *i.e.* with a 1% difference in energy between the two plaintexts. The resulting p-values for $P = 5000$ are computed by matching pairs of sets containing $\chi = 5 \cdot 10^5$ ciphertexts, yielding the p-value histograms depicted in Figure 3. We report the histograms of the p-values in the two cases along with the p-value of the second-level assessment, *i.e.* the probability that samples from a uniform distribution exhibit a deviation from a flat histogram larger than the observed one. When the two plaintexts have the same energy, all evidence concurs to say that the ciphertext distributions are statistically indistinguishable. In the second case, even a small difference in energy causes statistically detectable deviations and leads to a correct inference of the true plaintext between the two.

2) Non-Asymptotic Security: We have observed how asymptotic spherical secrecy has finite n effects (for additional evidence by numerical computation of the Kullback-Leibler divergence see [15]). From a more formal point of view, we may evaluate the convergence rate of (12) for finite n to obtain some further guarantee that an eavesdropper intercepting the measurements will observe samples of an approximately Gaussian RV bearing very little information in addition to the energy of the plaintext. We hereby consider X a RV as in (m_1), for which a plaintext x of energy e_x lies on the sphere $S_{e_x}^{n-1}$ of $\mathbb{R}^n$ (with radius $\sqrt{e_x}$).

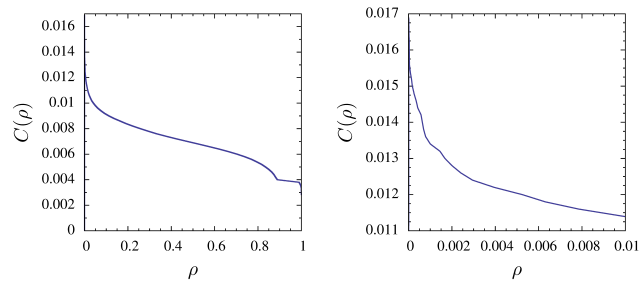
The most general convergence rate for sums of i.i.d. random variables is given by the well-known Berry-Esseen Theorem [30] as $O\left(n^{-\frac{1}{2}}\right)$. In our case we apply a recent, remarkable result of [17] that improves and extends this convergence rate to inner products of i.i.d. RVs (*i.e.* any row of A) and vectors (*i.e.* plaintexts x) uniformly distributed on $S_{e_x}^{n-1}$.

Proposition 3 (Rate of convergence of random measurements)**.** *Let X, Y be RVs as in (m_1) with A a random matrix of i.i.d. zero mean, unit variance, finite fourth moment entries. For any $\rho \in (0,1)$, there exists a subset $\mathcal{F} \subseteq S_{e_x}^{n-1}$ with probability measure $\sigma^{n-1}(\mathcal{F}) \geq 1-\rho$ such that if $x \in \mathcal{F}$ then all Y_j in Y verify*

$$\sup_{\alpha<\beta} \left| \int_\alpha^\beta f_{Y_j|X}(\nu|x)\mathrm{d}\nu - \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-\frac{t^2}{2e_x}} \mathrm{d}t \right| \leq \frac{C(\rho)}{n} \quad (13)$$

for $C(\rho)$ a non-increasing function of ρ.

Proposition 3 with ρ sufficiently small means that it is most likely (actually, with probability exceeding $1-\rho$) to observe an $O(n^{-1})$ convergence between $f_{Y_j|X}$ and the limiting distribution $\mathcal{N}(0, e_x)$. The function $C(\rho)$ is loosely bounded in [17], so to complete this analysis we performed a thorough Monte Carlo evaluation of its possible values. In particular,



(a) $\rho \in (0,1)$ (b) $\rho \in (0, 0.01)$

Fig. 4. Empirical evaluation of $C(\rho)$ in the convergence rate (13) based on a large number of plaintexts x on the sphere S_1^{n-1} and $n = 2^4, 2^5, \ldots, 2^{10}$.

we have taken 10^4 instances of a RV X uniformly distributed on S_1^{n-1} for each $n = 2^4, 2^5, \ldots, 2^{10}$. The PDF $f_{Y_j|X}(y_j|x)$ is estimated with the following procedure: we generate $5 \cdot 10^7$ rows of an i.i.d. Bernoulli random matrix and perform the linear combination in (1), thus yielding the same number of instances of Y_j for each x and n. On this large sample set we are able to accurately estimate the previous PDF on 4096 equiprobable intervals, and compare it to the same binning of the normal distribution as in the LHS of (13) for each (x, n). This method yields sample values for (13), allowing an empirical evaluation of the quantity $C(\rho)$. In this example, when $\rho \geq 10^{-3}$ Proposition 3 holds with $C(\rho) = 1.34 \cdot 10^{-2}$.

Hence, straightforward statistical attacks on a CS-encoded ciphertext may only extract very limited information from the plaintext. Yet, other attacks may rely on a larger amount of information, the next level of threat being known-plaintext attacks [29]. These attacks are based on the availability of some plaintext-ciphertext pairs and aim at the extraction of information on the encoding process that can be reused to decode future ciphertexts. Due to its criticality and theoretical depth, the robustness of multiclass CS w.r.t. this class of attacks will be tackled in a separate contribution.

IV. Application Examples

A. Experimental Framework

In this section we detail some example applications for the multiclass CS scheme we propose. For each example we study the recovery quality attained by first-class receivers against second-class ones in the two-class scheme (Section II-D1). These results encompass the multiclass setting since high-class receivers correspond to first-class recovery performances (*i.e.* $\eta = 0$), while lower-class users attain the performances of a second-class receiver at a fixed $\eta > 0$.

For each plaintext $x = Ds$ being reconstructed the recovery signal-to-noise ratio $\mathrm{RSNR} = \frac{\|x\|_2^2}{\|x-\hat{x}\|_2^2}$ with $\hat{x} = D\hat{s}$ denoting the recovered approximation is a common recovery quality index; its average[2], $\mathrm{ARSNR}\,[\mathrm{dB}] = 10\log_{10}\hat{\mathbb{E}}\left(\frac{\|x\|_2^2}{\|x-\hat{x}\|_2^2}\right)$ is then used as an average performance index, and compared against some best- and worst-case curves with the purpose of choosing a suitable perturbation density η so that lower-class recovery performances are set to the desired quality level.

[2] $\hat{\mathbb{E}}(\cdot)$ denotes the sample average over a set of realisations of the argument.

We complement the previous evidence with an automated assessment of the information content intelligible from $\hat{x}$ by means of feature extraction algorithms. These are equivalent to partially informed attacks attempting to expose the sensitive content inferred from the recovered signal. More specifically, we will try to recover an English sentence from a speech segment, the location of the PQRST peaks in an electrocardiographic (ECG) signal, and printed text in an image. The simulation framework reproducing these tests is available at http://securecs.googlecode.com.

1) Recovery Algorithms: While fundamental sparse signal recovery guarantees are well-known from [21] when solving BP and BPDN, these convex problems are often replaced in practice by a variety of high-performance algorithms (see *e.g.* [31]). As reference cases for most common algorithmic classes we tested the solution of BPDN as implemented in SPGL_1 [32], [33] against the greedy algorithm CoSaMP [34] and the generalised approximate message-passing algorithm (GAMP, [35]). To optimise these algorithms' performances the tests were optimally tuned in a "genie-aided" fashion: BPDN was solved as in (P_2) with the noise parameter $\gamma = \|\Delta A x\|_2$ as if $(\Delta A, x)$ were known beforehand; CoSaMP was initialised with the exact sparsity level k for each case; GAMP was run with the sparsity-enforcing, i.i.d. Bernoulli-Gaussian prior (see *e.g.* [36]) broadly applicable in the well-known message passing framework [37] and initialised with the exact sparsity ratio $\frac{k}{n}$ of each instance, and the exact mean and variance of each considered test set. Moreover, signal-independent parameters were hand-tuned in each case to yield optimal recovery performances.

For the sake of brevity, in each example we select and report the algorithm that yields the most accurate recovery quality at a lower-class decoder as the amount of perturbation varies. We found that GAMP achieves the highest ARSNR in all the settings explored in the examples, consistently with the observations in [36] that assess the robust recovery capabilities of this algorithm under a broadly applicable sparsity-enforcing prior. Moreover, as ΔA verifies [38, Proposition 2.1] the perturbation noise $\varepsilon = \Delta A x$ is approximately Gaussian for large (m, n) and thus GAMP tuned as above yields optimal performances as expected. Note that recovery algorithms which attempt to jointly identify x and ΔA [38], [39] can be seen as explicit attacks to multiclass encryption and are thus evaluated in a separate contribution, anticipating that their performances are compatible with those of GAMP.

2) Average Signal-to-Noise Ratio Bounds: The perturbation density η is the main design parameter for the multiclass encryption scheme, and therefore has to be chosen against a reference lower-class recovery algorithm. To provide criteria for the choice of η we adopt two ARSNR bounds derived as follows.

Although rigorous, the lower-class recovery error upper bound of Proposition 1 is only applicable for small values of (k, η). To bound typical recovery performances in a larger range we analyse the behaviour of a lower-class decoder that naively (*i.e.* without attempting any attack) recovers $\hat{x}$ such that $y = A^{(0)}\hat{x} = (A^{(0)} + \Delta A)x$, and thus $A^{(0)}(\hat{x} - x) = \Delta A x$. In most cases, such a recovery produces $\hat{x}$ lying close to x;

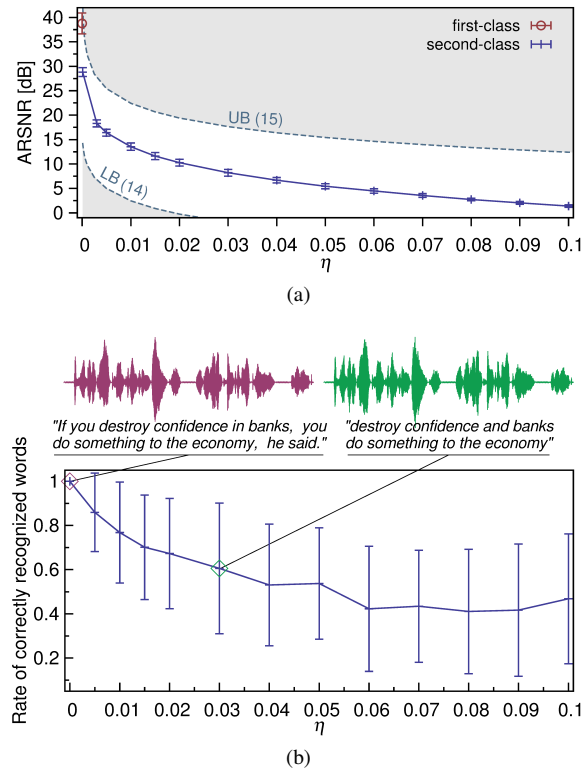


Fig. 5. Multiclass CS of speech signals: (a) Average recovery SNR as a function of the perturbation density $\eta \in [0, 0.1]$ (solid) and second-class RSNR upper bound (dashed); (b) Fraction of words exactly recognised by ASR in $\eta \in [0, 0.1]$ (bottom) and typical decoded signals for $\eta = 0, 0.03$ (top).

we model this by assuming $\|\hat{x} - x\|_2$ is close to be minimum. With this, we may approximate $\hat{x} - x = (A^{(0)})^{+}\Delta A x$, where $\cdot^{+}$ denotes the Moore-Penrose pseudoinverse, that yields $\frac{\|\hat{x}-x\|_2^2}{\|x\|_2^2} \leq \sigma_{\max}((A^{(0)})^{+}\Delta A)^2$. By taking a sample average on both sides, in signal-to-noise ratio our criterion is $\text{ARSNR} > \text{LB}(m, n, \eta)$ where

$$\text{LB}(m, n, \eta) = -10 \log_{10} \hat{\mathbb{E}} \left(\sigma_{\max}((A^{(0)})^{+}\Delta A)^2 \right) \text{ dB} \quad (14)$$

$\text{LB}(m, n, \eta)$ is calculated in each of the following examples by a thorough Monte Carlo simulation of $\sigma_{\max}((A^{(0)})^{+}\Delta A)$ over $5 \cdot 10^3$ cases.

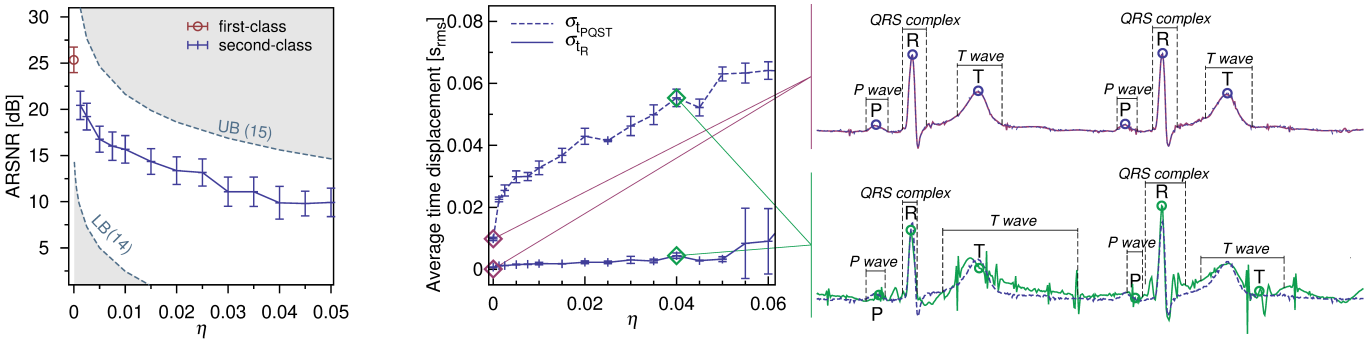
The opposite criterion is found by assuming $\text{ARSNR} < \text{UB}(m, n, \eta)$ where

$$\text{UB}(m, n, \eta) = -10 \log_{10} \frac{4\eta m}{(\sqrt{m} + \sqrt{n})^2} \text{ dB} \quad (15)$$

obtained from a simple rearrangement of (8) with $\theta \simeq 1$. We will see how (14) and (15) fit the ARSNR performances of the examples and provide simple criteria to estimate the range of performances of lower-class receivers from (m, n, η).

B. Speech Signals

We consider a subset of spoken English sentences from the PTDB-TUG database [40] with original sampling frequency $f_s = 48\,\text{kHz}$, variable duration and sentence length. Each speech signal is divided in segments of $n = 512$ samples and encoded by two-class CS with $m = \frac{n}{2}$ measurements. We



(a) (b)

Fig. 6. Multiclass CS of ECG signals: (a) Average recovery SNR as a function of the perturbation density $\eta \in [0, 0.05]$ (solid) and second-class RSNR upper bound (dashed); (b) Time displacement (left) of the R (solid) and P,Q,S,T (dashed) peaks as evaluated by APD for $\eta \in [0, 0.05]$ with typical decoded signals (right) for first-class (top) and second-class (bottom) users.

obtain the sparsity basis D by applying principal component analysis [41] to 500 n-dimensional segments yielding an ONB. The encoding matrix $A^{(1)}$ is generated from an i.i.d. Bernoulli random matrix $A^{(0)}$ by adding to it a sparse random perturbation ΔA chosen as in (3) with density η. The encoding in (5) is simulated in a realistic setting, where each window x of n samples is acquired with a different instance of $A^{(1)}$ yielding m measurements per speech segment. As for the decoding stage, we apply GAMP as specified above to recover $\hat{x}$ given $A^{(1)}$ (first-class) and $A^{(0)}$ (second-class).

For a given encoding matrix a first-class receiver is capable of decoding a clean speech signal with $\mathrm{ARSNR} = 38.76\,\mathrm{dB}$, whereas a second-class receiver is subject to significant ARSNR degradation when η increases, as shown in Fig. 5a. Note that while the RSNR for $\eta = 0$ has a relative deviation of $2.14\,\mathrm{dB}$ around its mean (the ARSNR), as η increases the observed RSNR deviation is less than $0.72\,\mathrm{dB}$. Note how the ARSNR values lie in the highlighted range between (14), (15).

To further quantify the limited quality of attained recoveries, we process the recovered signal with the Google Web Speech API [42], [43] which provides basic Automatic Speech Recognition (ASR). The ratio of words correctly inferred by ASR for different values of η is reported in Fig. 5b. This figure also reports a typical decoding case: a first-class user (*i.e.* $\eta = 0$) recovers the signal with $\mathrm{RSNR} = 36.58\,\mathrm{dB}$, whereas a second-class decoder only achieves a $\mathrm{RSNR} = 8.42\,\mathrm{dB}$ when $\eta = 0.03$. The corresponding ratio of recognised words is $\frac{14}{14}$ against $\frac{8}{14}$. In both cases the sentence is intelligible to a human listener, but the second-class decoder recovers a signal that is sufficiently corrupted to avoid straightforward ASR.

C. Electrocardiographic Signals

We extend the example in [15] by processing a large subset of ECG signals from the MIT PhysioNet database [44] sampled at $f_s = 256\,\mathrm{Hz}$. In particular, we report the case of a typical 25 minutes ECG track (sequence `e0108`) and encode windows of $n = 256$ samples by two-class CS with $m = 90$ measurements, amounting to a dataset of 1500 ECG instances. The encoding and decoding stages are identical to those in Section IV-B and we assume the Symmlet-6 ONB [45] as the sparsity basis D.

In this setting, the first-class decoder is able to reconstruct the original signal with $\mathrm{ARSNR} = 25.36\,\mathrm{dB}$, whereas a second-class decoder subject to a perturbation of density $\eta = 0.03$ achieves an $\mathrm{ARSNR} = 11.08\,\mathrm{dB}$; the recovery degradation depends on η as reported in Fig. 6a.

As an additional quantification of the encryption at second-class decoders we apply PUWave [46], an Automatic Peak Detection algorithm (APD), to first- and second-class signal reconstructions. In more detail, PUWave is used to detect the position of the P,Q,R,S and T peaks, *i.e.* the sequence of pulses whose positions and amplitudes summarise the diagnostic properties of an ECG. The application of this APD yields the estimated peak instants $\hat{t}_{\mathrm{P,Q,R,S,T}}$ for each of $J = 1500$ reconstructed windows and each decoder class, which are afterwards compared to the corresponding peak instants as detected on the original signal prior to encoding. Thus, we define the average time displacement $\sigma_t = \sqrt{\frac{1}{J}\sum_{i=0}^{J-1}(\hat{t}^{(i)} - t^{(i)})^2}$ and evaluate it for t_{R} and t_{PQST}. A first-class receiver is subject to a displacement $\sigma_{t_{\mathrm{R}}} = 0.6\,\mathrm{ms_{rms}}$ of the R-peak and $\sigma_{t_{\mathrm{PQST}}} = 9.8\,\mathrm{ms_{rms}}$ of the remaining peaks w.r.t. the original signal. On the other hand, a second-class user is able to determine the R-peak with $\sigma_{t_{\mathrm{R}}} = 4.4\,\mathrm{ms_{rms}}$ while the displacement of the other peaks is $\sigma_{t_{\mathrm{PQST}}} = 55.3\,\mathrm{ms_{rms}}$. As η varies in $[0, 0.05]$ this displacement increases as depicted in Fig. 6b, thus confirming that a second-class user will not be able to accurately determine the position and amplitude of the peaks with the exception of the R-peak.

D. Sensitive Text in Images

In this final example we consider an image dataset of people holding printed identification text and apply multiclass CS to selectively hide this sensitive content to lower-class users. The $640 \times 512\,\mathrm{pixel}$ images are encoded by CS in $10 \times 8\,\mathrm{blocks}$ each of $64 \times 64\,\mathrm{pixel}$ while the two-class strategy is only applied to a relevant image area of $3 \times 4\,\mathrm{blocks}$. We adopt as sparsity basis the 2D Daubechies-4 wavelet basis [45] and encode each block of $n = 4096\,\mathrm{pixels}$ with $m = 2048$ measurements; the encoding is generated with perturbation density $\eta \in [0, 0.4]$.

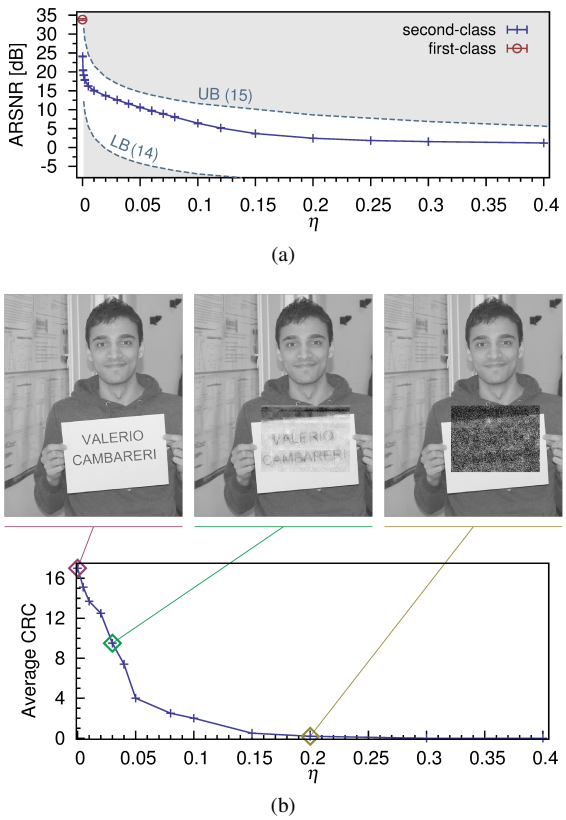


Fig. 7. Multiclass CS of images: (a) Average recovery SNR as a function of the perturbation density $\eta \in [0, 0.4]$ (solid) and second-class RSNR upper bound (dashed); (b) Average consecutive recognised characters by OCR for $\eta \in [0, 0.4]$ (bottom) and typical instances for $\eta = 0, 0.03, 0.2$ (top).

The ARSNR performances of this example are reported in Fig. 7a as averaged on 20 instances per case, showing a rapid degradation of the ARSNR as η is increased. This degradation is highlighted in the typical case of Fig. 7b for $\eta = 0.03, 0.2$.

In order to assess the effect of our encryption method with an automatic information extraction algorithm, we have applied Tesseract [47], an optical character recognition (OCR) algorithm, to the images reconstructed by a second-class user. The text portion in the recovered image data is preprocessed to enhance their quality prior to OCR: the images are first rotated, then we apply standard median filtering to reduce the highpass noise components. Finally, contrast adjustment and thresholding yield the two-level image which is processed by Tesseract. To assess the attained OCR quality we have measured the average number of consecutive recognised characters (CRC) from the decoded text image. In Fig. 7b the average CRC is reported as a function of η: as the perturbation density increases the OCR fails to recognise an increasing number of ordered characters, *i.e.* a second-class user progressively fails to extract text content from the decoded image.

V. Conclusion

Although not perfectly secure, the extremely simple encoding process entailed by CS yields some encryption capabilities with no additional computational complexity, thus providing a limited but zero-cost form of encryption which might be of interest in the design of secure yet resource-limited sensing interfaces. In particular, we have shown that when i.i.d. Bernoulli random matrices are used in this linear encoding scheme the plaintext features that leak into the ciphertext (and therefore retrievable by statistical analysis of the latter) are limited to the power of the plaintext as $n \to \infty$, and thus how an asymptotic definition of secrecy holds for this scheme. In addition, we have given evidence of the $O(\frac{1}{n})$ convergence rate to this limit behaviour. We have also detailed how two plaintexts having the same energy and encoded with an i.i.d. Bernoulli random matrix generate statistically indistinguishable ciphertexts, while even small differences in energy are detected by hypothesis testing on the ciphertext for finite n. The above linear random encoding was modified to envision a multiclass encryption scheme in which all receivers are given the same set of measurements, but are only enabled to reconstruct the original signal with a decoding quality that depends on their class, *i.e.* on the private key they possess. This additional design option amounts to the ability of flipping pseudo-randomly chosen elements of the encoding matrix, and thus represents an appealing alternative to balance the trade-off between the security of the encoded signal and the resources required to provide it.

Finally, the capabilities of multiclass CS were exemplified by simulating the acquisition of sources such as speech segments, electrocardiographic signals and images with the additional security provided by the devised encryption method.

Appendix A
Proofs regarding the Second-Class Recovery Error Lower Bound

We first introduce a Lemma that gives a self-contained probabilistic result on the Euclidean norm of $\varepsilon = \Delta A x$ in (5); this is used in the proofs of Theorem 1 and Corollary 1.

Lemma 1. *Let:*

1) ξ *be a RV with* $\mathcal{E}_\xi = \mathbb{E}[\sum_{j=0}^{n-1} \xi_j^2]$, $\mathcal{F}_\xi = \mathbb{E}[(\sum_{j=0}^{n-1} \xi_j^2)^2]$;
2) ΔA *be the sparse random matrix in* (3) *with i.i.d. entries and density* $\eta = \frac{c}{mn} \leq \frac{1}{2}$.

If ξ *and* ΔA *are independent, then for any* $\theta \in (0, 1)$

$$\mathbb{P}\left(\|\Delta A \xi\|_2^2 \geq 4mn\eta\, \mathcal{E}_\xi \theta\right) \geq \zeta \tag{16}$$

with

$$\zeta = \left\{1 + (1-\theta)^{-2}\left[\left(1 + \frac{1}{m}\left(\frac{3}{2\eta} - 1\right)\right)\frac{\mathcal{F}_\xi}{\mathcal{E}_\xi^2} - 1\right]\right\}^{-1} \tag{17}$$

Proof of Lemma 1. Consider

$$\|\Delta A \xi\|_2^2 = \sum_{j=0}^{m-1}\sum_{l=0}^{n-1}\sum_{i=0}^{n-1} \Delta A_{j,l} \Delta A_{j,i} \xi_l \xi_i$$

We now derive the first and second moments of this positive RV. ΔA is a random matrix of i.i.d. RVs with mean $\mu_{\Delta A_{j,l}} = 0$, variance $\sigma^2_{\Delta A_{j,l}} = 4\eta$ and $\mathbb{E}[\Delta A_{j,l}^4] = 16\eta$. Using the independence between ξ and ΔA, and the fact that ΔA is i.i.d. we have

$$\mathbb{E}\left[\|\Delta A \xi\|_2^2\right] = \sum_{j=0}^{m-1}\sum_{l=0}^{n-1}\sum_{i=0}^{n-1} \mathbb{E}[\Delta A_{j,l}\Delta A_{j,i}]\mathbb{E}[\xi_l \xi_i] =$$

$$\sum_{j=0}^{m-1}\sum_{l=0}^{n-1}\sum_{i=0}^{n-1} \sigma^2_{\Delta A}\delta_{l,i}\mathbb{E}[\xi_l\xi_i] = \sum_{j=0}^{m-1}\sigma^2_{\Delta A_{j,l}}\sum_{l=0}^{n-1}\mathbb{E}[\xi_l^2] = 4mn\eta\,\mathcal{E}_\xi$$

For the aforementioned properties of ΔA we also have

$$\mathbb{E}[\Delta A_{j,l}\Delta A_{j,i}\Delta A_{v,h}\Delta A_{v,o}] = \\ = \begin{cases} \sigma^4_{\Delta A}, & \begin{cases} j \neq v, l = i, h = o \\ j = v, l = i, h = o, l \neq h \\ j = v, l = h, i = o, l \neq i \\ j = v, l = o, i = h, l \neq i \end{cases} \\ \mathbb{E}[\Delta A^4_{j,l}], & j = v, l = i = h = o \\ 0, & \text{otherwise} \end{cases}$$

that can be used in some cumbersome but straightforward calculations yielding

$$\mathbb{E}\left[(\|\Delta A\xi\|_2^2)^2\right] = 16m\eta(\eta(m-1)\mathcal{F}_\xi + 3\eta(\mathcal{F}_\xi - \mathcal{G}_\xi) + \mathcal{G}_\xi)$$

where $\mathcal{G}_\xi = \mathbb{E}\left[\sum_{j=0}^{n-1}\xi_j^4\right]$. We are now in the position of using a one-sided version of Chebyshev's inequality for positive RVs[3] to say that, for any $\theta \in (0,1)$,

$$\mathbb{P}\left(\|\Delta A\xi\|_2^2 \geq \theta\mathbb{E}[\|\Delta A\xi\|_2^2]\right) \geq \\ \geq \left\{1 + (1-\theta)^{-2}\left[\frac{\mathbb{E}[(\|\Delta A\xi\|_2^2)^2]}{\mathbb{E}[\|\Delta A\xi\|_2^2]^2} - 1\right]\right\}^{-1} = \\ = \left\{1 + (1-\theta)^{-2}\left[\left(1 - \frac{1}{m}\right)\frac{\mathcal{F}_\xi}{\mathcal{E}_\xi^2} + \frac{3\eta(\mathcal{F}_\xi - \mathcal{G}_\xi) + \mathcal{G}_\xi}{\eta m \mathcal{E}_\xi^2} - 1\right]\right\}^{-1}$$

which yields (17) by considering that when $\eta \leq \frac{1}{2}$, $3\eta(\mathcal{F}_\xi - \mathcal{G}_\xi) + \mathcal{G}_\xi \leq \frac{3}{2}\mathcal{F}_\xi$. □

Proof of Theorem 1. Since all decoders receive in absence of other noise sources the same measurements $y = A^{(1)}x$, a second-class decoder would naively assume $y = A^{(0)}\hat{x}$, with $\hat{x}$ an approximation of x obtained by a recovery algorithm that satisfies this equality. Since $A^{(1)} = A^{(0)} + \Delta A$, if we define $\Delta x = \hat{x} - x$ we may write $A^{(0)}x + \Delta Ax = A^{(0)}\hat{x}$ and thus $A^{(0)}\Delta x = \Delta Ax$. $\|\Delta x\|_2^2$ can then be bounded straightforwardly as $\sigma_{\max}(A^{(0)})^2\|\Delta x\|_2^2 \geq \|\Delta Ax\|_2^2$ yielding

$$\|\hat{x} - x\|_2^2 \geq \frac{\|\Delta Ax\|_2^2}{\sigma_{\max}(A^{(0)})^2} \tag{18}$$

By applying the probabilistic lower bound of Lemma 1 on $\|\Delta Ax\|_2^2$ in (18), we have that $\|\Delta Ax\|_2^2 \geq 4m\eta\,\mathcal{E}_x\theta$ for $\theta \in (0,1)$ and a given probability value exceeding ζ in (17). Plugging the RHS of this inequality in (18) yields (6). □

The following Lemma applies to finding the asymptotic result (8) of Corollary 1.

Lemma 2. *Let* $\mathbf{X}$ *be an* α*-mixing RP with uniformly bounded fourth moments* $\mathbb{E}[X_j^4] \leq m_x$ *for some* $m_x > 0$. *Define ms* $\mathcal{E}_x = \mathbb{E}\left[\sum_{j=0}^{n-1}X_j^2\right]$, $\mathcal{F}_x = \mathbb{E}\left[\left(\sum_{j=0}^{n-1}X_j^2\right)^2\right]$.

If $\mathcal{W}_x = \lim_{n\to\infty}\frac{1}{n}\mathcal{E}_x > 0$ *then* $\lim_{n\to\infty}\frac{\mathcal{F}_x}{\mathcal{E}_x^2} = 1$.

Proof of Lemma 2. Note first that from Jensen's inequality $\mathcal{F}_x \geq \mathcal{E}_x^2$, so $\lim_{n\to\infty}\frac{1}{n}\mathcal{E}_x > 0$ also implies that $\lim_{n\to\infty}\frac{1}{n^2}\mathcal{E}_x^2 > 0$ and $\lim_{n\to\infty}\frac{1}{n^2}\mathcal{F}_x > 0$. Since $\lim_{n\to\infty}\frac{1}{n^2}\mathcal{E}_x^2 = \mathcal{W}_x^2 > 0$ we may write

$$\lim_{n\to\infty}\frac{\mathcal{F}_x}{\mathcal{E}_x^2} = 1 + \frac{\lim_{n\to\infty}\frac{1}{n^2}\mathcal{F}_x - \frac{1}{n^2}\mathcal{E}_x^2}{\mathcal{W}_x^2} \tag{19}$$

[3]If a r.v. $Z \geq 0$ then $\forall\theta \in (0,1)$, $\mathbb{P}(Z \geq \theta\mathbb{E}[Z]) \geq \frac{(1-\theta)^2\mathbb{E}[Z]^2}{(1-\theta)^2\mathbb{E}[Z]^2 + \sigma_Z^2}$.

and observe that $\left|\frac{1}{n^2}\mathcal{F}_x - \frac{1}{n^2}\mathcal{E}_x^2\right| \leq \frac{1}{n^2}\sum_{j=0}^{n-1}\sum_{l=0}^{n-1}|\mathcal{X}_{j,l}|$ where $\mathcal{X}_{j,l} = \mathbb{E}[X_j^2X_l^2] - \mathbb{E}[X_j^2]\mathbb{E}[X_l^2] = \mathbb{E}[(X_j^2 - \mathbb{E}[X_j^2])(X_l^2 - \mathbb{E}[X_l^2])]$. From the α-mixing assumption we know that $|\mathcal{X}_{j,l}| \leq \alpha(|j-l|) \leq m_x$ and a sequence $\alpha(h)$ vanishing as $h \to \infty$. Hence,

$$\left|\frac{1}{n^2}\mathcal{F}_x - \frac{1}{n^2}\mathcal{E}_x^2\right| \leq \frac{1}{n^2}\sum_{j=0}^{n-1}|\mathcal{X}_{j,j}| + \frac{2}{n^2}\sum_{h=1}^{n-1}\sum_{j=0}^{n-h-1}|\mathcal{X}_{j,j+h}| \leq \\ \leq \frac{n\,m_x}{n^2} + \frac{2}{n^2}\sum_{h=1}^{n-1}(n-h)\alpha(h) \leq \frac{m_x}{n} + \frac{2}{n}\sum_{h=1}^{n-1}\alpha(h) \tag{20}$$

The thesis follows from the fact that the upper bound in (20) vanishes as $n \to \infty$. This is obvious when $\sum_{h=0}^{+\infty}\alpha(h)$ is convergent. Otherwise, if $\sum_{h=0}^{+\infty}\alpha(h)$ is divergent we may resort to the Stolz-Cesàro theorem to find $\lim_{n\to\infty}\frac{1}{n}\sum_{h=1}^{n-1}\alpha(h) = \lim_{n\to\infty}\alpha(n) = 0$. □

Proof of Corollary 1. The inequality (18) in the proof of Theorem 1 is now modified for the asymptotic case of a RP $\mathbf{X}$. Note that $A^{(0)}$ is an i.i.d. random matrix with zero mean, unit variance entries; thus, when $m, n \to \infty$ with $\frac{m}{n} \to q$ the value $\sqrt{n}\sigma_{\max}(A^{(0)})$ is known from [48] since all the singular values belong to the interval $[1-\sqrt{q}, 1+\sqrt{q}]$. We therefore assume $\sigma_{\max}(A^{(0)}) \simeq \sqrt{m} + \sqrt{n}$ and take the limit of (18) normalised by $\frac{1}{n}$ for $m, n \to \infty$, yielding

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}(\hat{x}_j - x_j)^2 \geq \lim_{m,n\to\infty}\frac{\|\Delta A\frac{x^{(n)}}{\sqrt{n}}\|_2^2}{\left(\sqrt{m}+\sqrt{n}\right)^2} \tag{21}$$

with $x^{(n)}$ the n-th finite-length term in a plaintext $x = \{x^{(n)}\}_{n=0}^{+\infty}$ of $\mathbf{X}$. We may now apply Lemma 1 in $\xi = \frac{x^{(n)}}{\sqrt{n}}$ for each $\|\Delta A\xi\|_2^2$ at the numerator of the RHS of (21) with $\mathcal{F}_\xi = \frac{1}{n^2}\mathcal{F}_x$, $\mathcal{E}_\xi = \frac{1}{n}\mathcal{E}_x$ and $\mathcal{E}_x, \mathcal{F}_x$ as in Lemma 1. For $m, n \to \infty$ and $\eta \leq \frac{1}{2}$, the probability in (17) becomes

$$\lim_{m,n\to\infty}\zeta = \left\{1 + (1-\theta)^{-2}\left[\lim_{n\to\infty}\frac{\frac{1}{n^2}\mathcal{F}_x}{\frac{1}{n^2}\mathcal{E}_x^2} - 1\right]\right\}^{-1}$$

Since $\mathbf{X}$ satisfies by hypothesis the assumptions of Lemma 2, then $\lim_{n\to\infty}\frac{\mathcal{F}_\xi}{\mathcal{E}_\xi^2} = 1$ and $\lim_{m,n\to\infty}\zeta = 1$. Hence, with $\frac{m}{n} \to q$ and probability 1 the RHS of (21) becomes

$$\lim_{m,n\to\infty}\frac{\|\Delta A\xi\|_2^2}{n(1+\sqrt{\frac{m}{n}})^2} = \lim_{m,n\to\infty}\frac{4\frac{m}{n}\eta\frac{\mathcal{E}_x}{n}}{(1+\sqrt{\frac{m}{n}})^2}\theta,\ \forall\theta \in (0,1)$$

and the recovery error power satisfies (8). □

Appendix B
Proofs regarding the Spherical Secrecy of Compressed Sensing

Proof of Proposition 2. The proof is given by simple verification of the Lindeberg-Feller central limit theorem (see [23, Theorem 27.4]) for Y_j in $\mathbf{Y}$ conditioned on a plaintext x of $\mathbf{X}$ in (m_2). By the hypotheses, the plaintext $x = \{x_l\}_{l=0}^{n-1}$ has power $0 < W_x < \infty$ and $x_l^2 \leq M_x$ for some finite $M_x > 0$. Any $Y_j|\mathbf{X} = \lim_{n\to\infty}\sum_{l=0}^{n-1}Z_{j,l}$, $Z_{j,l} = A_{j,l}\frac{x_l}{\sqrt{n}}$ where $Z_{j,l}$ is a sequence of independent, non-identically distributed random variables of moments $\mathbb{E}[Z_{j,l}] = 0, \mathbb{E}[Z_{j,l}^2] = \frac{x_l^2}{n}$.

By letting the partial sum $S_j^{(n)} = \sum_{l=0}^{n-1} Z_{j,l}$, its mean $\mathbb{E}[S_j^{(n)}] = 0$ and $\mathbb{E}[(S_j^{(n)})^2] = \frac{1}{n}\sum_{l=0}^{n-1} x_l^2$. Thus, we verify the necessary and sufficient condition [23, (27.19)]

$$\lim_{n\to\infty} \max_{l=0,\dots,n-1} \frac{\mathbb{E}[Z_{j,l}^2]}{\mathbb{E}[(S_j^{(n)})^2]} = 0$$

by straightforwardly observing

$$\lim_{n\to\infty} \max_{l=0,\dots,n-1} \frac{\frac{x_l^2}{n}}{\frac{1}{n}\sum_{l=0}^{n-1} x_l^2} \leq \frac{M_x}{W_x} \lim_{n\to\infty} \frac{1}{n} = 0$$

The verification of this condition guarantees that $Y_j|\mathbf{X} = \lim_{n\to\infty} S_j^{(n)}$ is normally distributed with variance $\mathbb{E}[(Y_j|\mathbf{X})^2] = \lim_{n\to\infty} \mathbb{E}[(S_j^{(n)})^2] = W_x$, *i.e.* $f_{Y_j|\mathbf{X}} \underset{\mathcal{D}}{\to} \mathcal{N}(0, W_x)$. □

Proof of Proposition 3. We start by considering Y_j in Y of model (m_1) conditioned on a given x with finite energy e_x. Each of such variables is a linear combination (1) of n i.i.d. RVs $A_{j,l}$ with zero mean, unit variance and finite fourth moments. The coefficients of this linear combination are $x = (x_0, \cdots, x_{n-1})$ which by now we assume to have $e_x = 1$, *i.e.* to lie on the unit sphere S_1^{n-1} of $\mathbb{R}^n$. Define $\delta = \left(\frac{1}{n}\sum_{l=0}^{n-1} \mathbb{E}[A_{j,l}^4]\right)^{\frac{1}{4}} < \infty$, which for i.i.d. Bernoulli random matrices is $\delta = 1$, whereas for standard $\mathcal{N}(0,1)$ random matrices $\delta = 3^{\frac{1}{4}}$. This setting verifies [17, Theorem 1.1]: for any $\rho \in (0,1)$ there exists a subset $\mathcal{F} \subseteq S_1^{n-1}$ with measure $\mu(\mathcal{F})$ such that $\frac{\mu(\mathcal{F})}{\mu(S_1^{n-1})} \geq 1-\rho$ and if $x \in \mathcal{F}$, then

$$\sup_{\substack{(\alpha,\beta)\in\mathbb{R}^2 \\ \alpha<\beta}} \left| \mathbb{P}\left(\alpha \leq \sum_{l=0}^{n-1} A_{j,l} x_l \leq \beta\right) - \right.$$
$$\left. -\frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-\frac{t^2}{2}} \mathrm{d}t \right| \leq \frac{C(\rho)\delta^4}{n} \tag{22}$$

with $C(\rho)$ a positive, non-increasing function. An application of this result to x with energy e_x, *i.e.* on the sphere of radius $\sqrt{e_x}$, $\delta = 1$ (A i.i.d. Bernoulli) can be done by straightforwardly scaling the standard normal PDF in (22) to $\mathcal{N}(0, e_x)$, thus yielding the statement of Proposition 3. □

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] Y. Wang, G. Attebury, and B. Ramamurthy, "A survey of security issues in wireless sensor networks," *Communications Surveys & Tutorials, IEEE*, vol. 8, no. 2, pp. 2–23, 2006.

[3] J. Daemen and V. Rijmen, *The design of Rijndael: AES-the advanced encryption standard*. Springer, 2002.

[4] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E. J. Candès and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[6] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520–544, 2010.

[7] M. Mishali, Y. C. Eldar, and A. J. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4719–4734, 2011.

[8] J. Haboba, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A pragmatic look at some compressive sensing architectures with saturation and quantization," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 2, no. 3, pp. 443–459, Sept.

[9] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, 2008.

[10] F. Chen, A. Chandrakasan, and V. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 3, pp. 744–756, 2012.

[11] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[12] Y. Rachlin and D. Baron, "The Secrecy of Compressed Sensing Measurements," in *46th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2008, pp. 813–817.

[13] M. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.

[14] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "On the security and robustness of encryption via compressed sensing," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*. IEEE, 2008, pp. 1–7.

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

[16] M. Zhang, M. Kermani, A. Raghunathan, and N. Jha, "Energy-efficient and secure sensor data transmission using encompression," in *VLSI Design and 2013 12th International Conference on Embedded Systems (VLSID), 2013 26th International Conference on*, 2013, pp. 31–36.

[17] B. Klartag and S. Sodin, "Variations on the Berry – Esseen Theorem," *Theory of Probability & Its Applications*, vol. 56, no. 3, pp. 403–419, 2012.

[18] P.-L. Loh and M. J. Wainwright, "Corrupted and missing predictors: Minimax bounds for high-dimensional linear regression," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 2601–2605.

[19] ——, "High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity," *Annals of Statistics*, vol. 40, no. 3, p. 1637, 2012.

[20] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[21] E. J. Candès, 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, Aug. 2006.

[22] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[23] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.

[24] M. Rudelson and R. Vershynin, "Non-asymptotic theory of random matrices: extreme singular values," in *Proceedings of the International Congress of Mathematicians*. Citeseer, 2010.

[25] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4658–4671, 2012.

[26] R. Latała, "Some estimates of norms of random matrices," *Proceedings of the American Mathematical Society*, vol. 133, no. 5, pp. 1273–1282, 2005.

[27] D. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, june 2010.

[28] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[29] L. C. Washington and W. Trappe, *Introduction to cryptography: with coding theory*. Prentice Hall PTR, 2002.

[30] A. C. Berry, "The accuracy of the gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941.

[31] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[32] E. van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM Journal on Optimization*, vol. 21, no. 4, pp. 1201–1229, 2011.
[33] ——, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.
[34] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
[35] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 2168–2172.
[36] J. Vila and P. Schniter, "Expectation-maximization bernoulli-gaussian approximate message passing," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 799–803.
[37] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
[38] J. T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 804–808.
[39] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2002–2016, 2011.
[40] G. Pirker, M. Wohlmayr, S. Petrik, and F. Pernkopf, "A Pitch Tracking Corpus with Evaluation on Multipitch Tracking Scenario," in *Interspeech 2011*, Florence (Italy), Aug. 27-31, 2011, pp. 1509–1512.
[41] K. Karhunen, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. Universitat Helsinki, 1947, vol. 37.
[42] G. Shires and H. Wennborg, "Web Speech API Specification," Oct. 2012, http://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html.
[43] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 82–97, 2012.
[44] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).
[45] S. Mallat, *A wavelet tour of signal processing*. Access Online via Elsevier, 1999.
[46] R. Jane, A. Blasi, J. García, and P. Laguna, "Evaluation of an automatic threshold based detector of waveform limits in holter ecg with the qt database," in *Computers in Cardiology 1997*. IEEE, 1997, pp. 295–298.
[47] R. Smith, "An Overview of the Tesseract OCR Engine," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2, 2007, pp. 629–633.
[48] S. Geman, "A limit theorem for the norm of random matrices," *The Annals of Probability*, vol. 8, no. 2, pp. 252–261, 1980.

Valerio Cambareri (S'13) received the B.S. and M.S. degree (*summa cum laude*) in Electronic Engineering from the University of Bologna, Italy, in 2008 and 2011 respectively. Since 2012 he is a Ph.D. student in Electronics, Telecommunications and Information Technologies at DEI – University of Bologna, Italy. In 2014 he was a visiting Ph.D. student in the Integrated Imagers team at IMEC, Belgium. His current research activity focuses on statistical and digital signal processing, compressed sensing and computational imaging.

Mauro Mangia (S'09-M'13) received the B.S. and M.S. degree in Electronic Engineering from the University of Bologna, Italy, in 2004 and 2009 respectively; he received the Ph.D. degree in Information Technology from the University of Bologna in 2013. He is currently a post-doc researcher in the statistical signal processing group of ARCES – University of Bologna, Italy. In 2009 and 2012 he was a visiting Ph.D. student at the École Polytechnique Fédérale de Lausanne (EPFL). His research interests are in nonlinear systems, compressed sensing, ultra-wideband systems and system biology. He was recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and the best student paper award at ISCAS2011.

Fabio Pareschi (S'05-M'08) received the Dr. Eng. degree (with honours) in Electronic Engineering from University of Ferrara, Italy, in 2001, and the Ph.D. in Information Technology under the European Doctorate Project (EDITH) from University of Bologna, Italy, in 2007. He is currently an Assistant Professor in the Department of Engineering (ENDIF), University of Ferrara. He is also a faculty member of ARCES – University of Bologna, Italy. He served as Associate Editor for the IEEE Transactions on Circuits and Systems – Part II (2010-2013). His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, random number generation and testing, and electromagnetic compatibility. He was recipient of the best paper award at ECCTD2005 and the best student paper award at EMCZurich2005.

Riccardo Rovatti was born in 1969. He received the M.S. degree in Electronic Engineering and the Ph.D. degree in Electronics, Computer Science, and Telecommunications both from the University of Bologna, Italy in 1992 and 1996, respectively. He is now a Full Professor of Electronics at the University of Bologna. He is the author of approximately 300 technical contributions to international conferences and journals, and of two volumes. His research focuses on mathematical and applicative aspects of statistical signal processing and on the application of statistics to nonlinear dynamical systems. He received the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as the best paper award at ECCTD2005, and the best student paper award at EMCZurich2005 and ISCAS2011. He was elected IEEE Fellow in 2012 *for contributions to nonlinear and statistical signal processing applied to electronic systems*.

Gianluca Setti (S'89,M'91,SM'02,F'06) received the Ph.D. degree in Electronic Engineering and Computer Science from the University of Bologna in 1997. Since 1997 he has been with the School of Engineering at the University of Ferrara, Italy, where he is currently a Professor of Circuit Theory and Analog Electronics and is also a permanent faculty member of ARCES – University of Bologna, Italy. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, electromagnetic compatibility, statistical signal processing and biomedical circuits and systems. Dr. Setti received the 2013 IEEE CAS Society Meritorious Service Award and co-recipient of the 2004 IEEE CAS Society Darlington Award, of the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as of the best paper award at ECCTD2005, and the best student paper award at EMCZurich2005 and at ISCAS2011. He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE Transactions on Circuits and Systems – Part II (2006-2007) and of the IEEE Transactions on Circuits and Systems – Part I (2008-2009). Dr. Setti was the Technical Program Co-Chair at ISCAS2007, ISCAS2008, ICECS2012, BioCAS2013 as well as the General Co-Chair of NOLTA2006. He was Distinguished Lecturer of the IEEE CAS Society (2004-2005), a member of its Board of Governors (2005-2008), and he served as the 2010 President of CASS. He held several other volunteer positions for the IEEE and in 2013-2014 he was the first non North-American Vice President of the IEEE for Publication Services and Products.