Soft content fingerprinting with bit polarization based on sign-magnitude decomposition

Slava Voloshynovskiy, Taras Holotyak and Fokko Beekhof*
Department of Computer Science, University of Geneva, Switzerland,
* Expedia Inc.
Emails: {svolos, Taras.Holotyak}@unige.ch, fbeekhof@expedia.com

***Abstract*—Content identification based on digital content fingerprinting attracts significant attention in different emerging applications. In this paper, we consider content identification based on the sign-magnitude decomposition of fingerprint codewords and analyze the achievable rates for sign and magnitude components. We demonstrate that the bit robustness in the sign channel, often used in binary fingerprinting, is determined by the value of the corresponding magnitude component. Correspondingly, one can distinguish between two systems depending how the information about the magnitude component is used at the decoding process, i.e., hard fingerprinting when this information is disregarded, and soft fingerprinting when this information is used. To reveal the advantages of soft information at the decoding, we consider a case of soft fingerprinting where the decoder has access to the complete information about the uncoded magnitude component. However, since it requires a lot of extra memory storage or secure communication, the magnitude information is often quantized to a single bit or extracted directly from the noisy observation. To generalize the existing methods and estimate the impact of quantization and noise in the side information about the magnitude components on the achievable rate, we introduce a channel splitting approach and reveal certain interesting phenomena related to channel polarization. We demonstrate that under proper quantization of the magnitude component, one can clearly observe the existence of strong components whose sign is very robust, even to strong distortions. We demonstrate that under certain conditions a great portion of the rate in the sign channel is concentrated in strong channel components. Finally, we demonstrate how to use the channel splitting property in the design of efficient low-complexity identification methods [1].**

I. Introduction

Identification systems are widely used in various emerging applications ranging from human biometrics and physical object security to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on *digital fingerprints* a.k.a. *binary templates* in biometric applications, which represent a short, robust and distinctive content description. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data, thus allowing the introduction of crypto-based security into the noisy analog world [3].

The binary fingerprint extraction can be considered as a *vector quantization* (VQ) problem where the index or label of the VQ centroid is considered as a corresponding binary fingerprint. Such a consideration represents an interesting basis for the information-theoretic analysis of content fingerprinting. However, the design of high-dimensional optimal VQ methods matched with the source distribution, which is rarely known in advance, requires exponential memory and complexity. Some known VQ techniques based on *lattices* [4], [5] are only suited for low-dimensional ($L \leq 24$) and uniformly distributed data. Otherwise, approximate techniques are used based on *product quantization* that produces a list of most likely centroid indices that should be further refined in high-dimensional space [6], [7].

An alternative approach to the binary template/fingerprint extraction is based on dimensionality reduction and binarization; a popular technique in biometrics [8], multimedia fingerprinting/hashing [7], [9], indexing and retrieval [10] and physical object security [11]. The dimensionality reduction is often based on *random projections* (RP) that are optimal under certain conditions in the sense of the Johnson-Lindenstrauss lemma [10]. The binarization is a simple sign-extraction operation. Such an approach does not require knowledge of the source distribution and might be applied to both i.i.d. and correlated data; RP transforms the input data pdf into i.i.d. Gaussian statistics, if the input pdf satisfies certain constraints as discussed in [12], (Proposition 1 with Remarks 1 and 2). This is an attractive property for the analysis and optimization of fingerprinting systems. In addition, RP can be applied to the data of any dimensionality in contrast to lattices. Moreover, the basis vectors of the RP can be generated based on a random key that is a plausible feature for security and privacy [1], [8].

Despite the attractiveness of binary data representations for memory storage, complexity, security and privacy, an important fraction of information is neglected once the data are binarized. At the same time, it is demonstrated that soft information extracted from noisy observations can enhance the overall system performance, complexity and privacy [1]. The soft fingerprinting is closely related to *bit reliability* that leads to a non-uniform treatment of bits in the extracted template.

In addition, several works reported the practical usage of soft information for privacy protection in biometric applications [3], [13], [14] and content identification [15], [16]. However, to our best knowledge, there are no works that theoretically investigate the impact of soft information on the enhancement of the identification rate, the reduction of search complexity and the potential gain for privacy protection.

In this paper, we extend the preliminary results obtained in our previous work [2] and in contrast to known works

[1]Preliminary results from this work were presented in the IEEE Information Theory Workshops 2010 and 2011 [1], [2]

[12], [13], [17], we demonstrate the theoretical impact of soft information on the identification rate using the sign-magnitude decomposition as well as exemplify some interesting effects related to channel polarization and its usage for fast fingerprint matching. More particularly, we investigate the basic decomposition of mutual information between channel input and output into four terms, where two terms are dominating, thus leading to the separate consideration of sign and magnitude channels. We demonstrate that the sign channel can be considered as a *binary symmetric channel* (BSC) where the cross-over bit error probability is determined by the magnitude component of the signal. This is in contrast to many fingerprinting techniques where the parameter of the BSC is considered to be fixed and independent from the magnitude. Such a decomposition makes it possible to reconsider the existing concepts of the optimal fingerprint extraction based on the established phenomena of channel polarization, where the identification rates can be concentrated in several components that are robust to bit errors. The channel polarization effect was also recently discovered in the channel codes known as *polar codes* [18]. However, it is achieved by special code construction whereas identification channel polarization, considered in this paper, concerns random fingerprinting codes. The results are applied to Gaussian random data and an additive white Gaussian noise (AWGN) channel in the projected domain after random projections with a random sensing matrix. In the second part of paper, we apply the developed concept of sign-magnitude decomposition (SMD) to the estimation of identification complexity based on soft fingerprinting.

Notations. We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively, i.e., $\mathbf{x} = (x_1, x_2, ..., x_N)^T$. x_S is used to denote the sign of x and x_M the magnitude of x. We use $X \sim f(x)$ to indicate that a continuous random variable X follows $f_X(x)$ and $X \sim p(x)$ to characterize discrete random variables. $h(\cdot)$, $H(\cdot)$ and $H_2(\cdot)$ denote differential entropy, entropy and binary entropy, respectively, while $I(.;.)$ defines pair-wise mutual information. $\mathbb{E}_{f(x)}[\cdot]$ denotes the expectation with respect to the random variable $X \sim f(x)$. The sign "$\perp$" denotes the statistical independence.

II. Background

In this section, we consider a generalized identification problem based on content fingerprinting.

The input vector $\mathbf{v}_\mathrm{x} \in \mathbb{R}^L$ representing some biometric features, uncloneable physical object function or multimedia data is generated from some distribution $\mathbf{V}_\mathrm{x} \sim f(\mathbf{v}_\mathrm{x})$. A set of M_v objects to be identified is enrolled into a codebook $\mathcal{C}_v = \{\mathbf{v}_\mathrm{x}(m)\}$ with $1 \leq m \leq M_v$. Each object is assigned some index m accordingly. The noisy version $\mathbf{v}_\mathrm{y}$ of $\mathbf{v}_\mathrm{x}$ is observed via some channel $f(\mathbf{v}_\mathrm{y}|\mathbf{v}_\mathrm{x})$. Following [19], we will assume that this channel pdf is known and can be represented as a memoryless channel (MC) $f(\mathbf{v}_\mathrm{y}|\mathbf{v}_\mathrm{x}) = \prod_{j=1}^{L} f(v_{\mathrm{y}_j}|v_{\mathrm{x}_j})$. The maximum number of objects that can be reliably identified with a negligibly small probability of error ϵ, i.e., $\Pr\{\hat{M} \neq M\} \leq \epsilon$, is $M_v \leq 2^{L(C-\epsilon)}$, where C denotes the *identification capacity* [19]:

$$C = I(V_\mathrm{x}; V_\mathrm{y}), \tag{1}$$

for $L \to \infty$ and ϵ an arbitrarily small constant. Unfortunately, in practice it is difficult to compute the above identification capacity in the general case due to the lack of exact knowledge of source and channel pdfs and absence of perfect synchronization between the sequences. Thus it can be computed only for some classes of assumed models under the assumption of synchronization between the sequences. The extension of (1) to the Gaussian source and noisy enrollment and identification in the Gaussian channels as well as for the binary source and binary symmetric channels is given in [3]. Besides that, the actual identification capacity remains unknown.

Moreover, such a Shannon-type result [19] requires exponential search or decoding complexity and memory to store all enrolled objects in the codebook. In addition, such a public storage of data is not always desirable for biometric applications due to privacy concerns. Finally, to practically achieve the theoretical limit (1), perfect knowledge of the observation channel pdf $f(\mathbf{v}_\mathrm{y}|\mathbf{v}_\mathrm{x})$ is required for optimal decoder design.

Therefore, as highlighted in the Introduction, a practical approach consists in the transformation of data to a lower dimensional representation via some dimensionality reduction transform (Figure 1) as:

$$\begin{cases} \mathbf{x} = \mathbb{W}\mathbf{v}_\mathrm{x}, \\ \mathbf{y} = \mathbb{W}\mathbf{v}_\mathrm{y}, \end{cases} \tag{2}$$

where the matrix $\mathbb{W} \in \mathbb{R}^{N \times L}$ elements are generated equilikely $W_{i,j} = \{\pm 1/\sqrt{L}\}$ or Gaussian $W_{i,j} \sim \mathcal{N}\{0, 1/\sqrt{L}\}$ based on the secret key k.

The projected vectors $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$ are distributed as $\mathbf{X} \sim f(\mathbf{x})$ and $\mathbf{Y} \sim f(\mathbf{y})$. The observation model $f(\mathbf{v}_\mathrm{y}|\mathbf{v}_\mathrm{x})$ is transformed in the RP domain into $f(\mathbf{y}|\mathbf{x})$ (Figure 1). It was demonstrated in [12] that such a transformation produces i.i.d. Gaussian vectors $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbb{I}_N)$ and $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \sigma_Y^2 \mathbb{I}_N)$ for several popular source/channel models and where the observation model is defined as [2]:

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \tag{3}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2 \mathbb{I}_N)$ and $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$, and σ_X^2 and σ_Z^2 stand for the variance of the input data and the noise, respectively. Since the additive Gaussian noise has the highest entropy among all distributions with a bounded variance [20] and leads to the highest reduction of the achievable rate it is interesting to estimate the limits of identification system performance under such kind of assumption.

Therefore, in the RP domain, the identification rate for the Gaussian counterpart of (1) reduces to the Gaussian case [3]:

$$R^{RP} = I(X;Y) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right). \tag{4}$$

This result should be interpreted with some care regarding the dimensionality N of projected data. From one side N should

[2]In this model, we assume either the invariance of extracted features to the de-synchronization transform or the existence of proper synchronization mechanism.

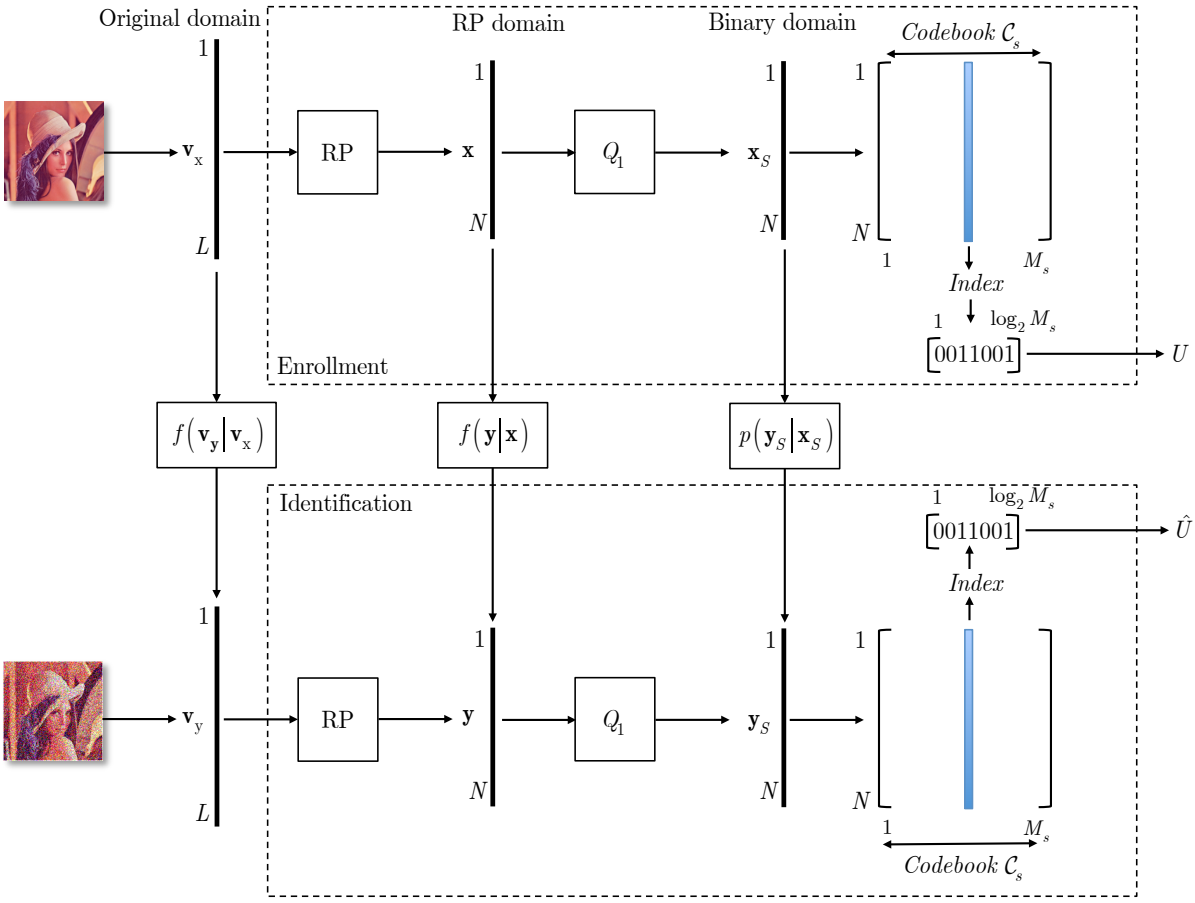


Fig. 1. Identification problem in the original domain, RP domain and binary domain.

be sufficiently large to satisfy the conditions of the asymptotic equipartition property used for the achievability part of proof [19], [20] based on random sequences and to achieve the Shannon identification limit. At the same time, some results of recent studies indicate that the maximum achievable rate close to the Shannon capacity can be tightly approximated already for block length as short as 100 [21]. Accordingly, in our overview we will assume that a proper finite dimensional random Gaussian block code $(2^{NR^{RP}}, N)$ with the rate R^{RP} is used.

The projected data $\mathbf{x}$ are stored in a codebook $\mathcal{C}_x = \{\mathbf{x}(w)\}$, $1 \le w \le M_x$. The maximum number of reliably recognizable objects is bounded as $M_x \le 2^{N(R^{RP}-\epsilon)}$ for $\Pr\{\hat{W} \ne W\} \le \epsilon$.

The further reduction of memory storage and search complexity is obtained by compressing $\mathbf{x}$ and $\mathbf{y}$. In the case of binary fingerprints and templates, a binary scalar quantizer $Q_1(.)$ is applied to each component of $\mathbf{x}$ and $\mathbf{y}$ that results into the vectors $\mathbf{x}_S \in \{-1,+1\}^N$ and $\mathbf{y}_S \in \{-1,+1\}^N$:

$$\begin{cases} \mathbf{x}_S = Q_1(x_1) \circ Q_1(x_2) \circ \cdots \circ Q_1(x_N), \\ \mathbf{y}_S = Q_1(y_1) \circ Q_1(y_2) \circ \cdots \circ Q_1(y_N), \end{cases} \tag{5}$$

where $Q_1(a) = sign(a)$, where $sign(a) = +1$, for $a \ge 0$ and -1, otherwise. It should be also pointed out that in this interpretation the scalar quantizer just extracts the sign of each component of the vector. That is why the notation has the subindex "S". At this point, we introduce a difference between the *hard* and *soft* content fingerprinting. In the scope of this paper, we assume that the binarized version $\mathbf{x}_S$ is stored in the database $\mathcal{C}_s$. Hard content fingerprinting assumes the usage of the binarized probe $\mathbf{y}_S$ while the soft one refers to the real-valued vector $\mathbf{y}$. The binarized projected data $\mathbf{x}_S$ are stored in a codebook $\mathcal{C}_s = \{\mathbf{x}_S(u)\}$, $1 \le u \le M_s$ (Figure 1). The maximum number of reliably recognizable objects is bounded as $M_s \le 2^{L(R^B-\epsilon)}$ for $\Pr\{\hat{U} \ne U\} \le \epsilon$. For the quantized data, one can characterize a link between $\mathbf{y}_S$ and $\mathbf{x}_S$ as a BSC $p(\mathbf{y}_S|\mathbf{x}_S)$ with a cross-over probability $P_b = \frac{1}{\pi}\arccos\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_Z^2}\right)$ [1]. The corresponding identification rate of this system is [3]:

$$R^B = I(X_S; Y_S) = 1 - H_2(P_b). \tag{6}$$

Obviously, such a compression represents a rate loss in comparison to the RP domain such that $R^B \le R^{RP}$. The equality is only true for the regime of very strong distortions, i.e., when $\frac{\sigma_X^2}{\sigma_Z^2} \ll 1$ [1]. Therefore, considerably smaller number of objects can be recognized based on the binarized data.

Although, one can now operate with the binary data, the identification search complexity is still exponential in N, i.e., $\mathcal{O}(2^{NR^B})$. Therefore, in practice N should be relatively small to enable even the storage of such a binary codebook. This

motivates the usage of approximate methods such as a product of vector quantizers where a long sequence of length N is split in blocks of manageable length and the results of decoding in each block are then fused to produce a final estimate [6], [7]. Under such conditions efficient search methods based on look up tables can be used resulting in some additional memory increase.

As previously mentioned, in many identification setups, a real vector $\mathbf{y}$ is available that represents a sort of soft information, but it is only used in the binarized form. At the same time, this soft information might be very useful to analyze the increase of the identification rate, investigate enhanced search methods and study new privacy-protection schemes. In the next section, we provide a rigorous analysis of identification capacity with the soft information, i.e., $I(X_S;Y)$ between the quantized X_S and real-valued Y and compare it to (6).

III. A SIGN-MAGNITUDE DECOMPOSITION OF MUTUAL INFORMATION

A. General decomposition

Consider a memoryless source $\mathbf{X}$ with some symmetric distribution $f(\mathbf{x}) = \prod_{i=1}^{N} f(x_i)$ that is observed via the memoryless channel with a symmetric distribution $f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} f(y_i|x_i)$ with input $x_i \in \mathbb{R}$ and output $y_i \in \mathbb{R}$. The channel input and output can be decomposed as $x_i = x_{Si}\cdot x_{Mi}$ and $y_i = y_{Si}\cdot y_{Mi}$, where $x_{Si} = \text{sign}(x_i)$ and $y_{Si} = \text{sign}(y_i)$ are the sign components and $x_{Mi} = |x_i|$ and $y_{Mi} = |y_i|$ are the magnitudes, $x_S \in \mathcal{X}_S$, $y_S \in \mathcal{Y}_S$ ($\mathcal{X}_S, \mathcal{Y}_S = \{-1, +1\}$); $x_M \in \mathcal{X}_M$, $y_M \in \mathcal{Y}_M$ ($\mathcal{X}_M, \mathcal{Y}_M = \mathbb{R}^+$). This decomposition is shown in Figure 2, where a sign-magnitude channel $f(y_S, y_M|x_S, x_M)$ links the corresponding sign and magnitude components.

The sign component of source X_S and magnitude component of source X_M are distributed as $X_S \sim Bernoulli(\theta_X)$ and $X_M \sim f(x_M)$, respectively. For any symmetric source the parameter of Bernoulli distribution is $\theta_X = 0.5$ and the components X_S and X_M are independent.

Accordingly, the sign component of channel output Y_S and magnitude component Y_M are distributed as $Y_S \sim Bernoulli(\theta_Y)$ and $Y_M \sim f(y_M)$, respectively. For the symmetric channels distribution $f(y|x)$, the channel between the sign components X_S and Y_S corresponds to a binary symmetric channel (BSC) with a probability mass function $p(y_S|x_S, x_M)$ with the cross-over probability $P_{b|x_M}$ defined by the magnitude component x_M (Figure 3). For a given state x_M and corresponding $P_{b|x_M}$, the distribution of Y_S is defined by $\theta_{Y|x_M} = \theta_X * P_{b|x_M}$, where $p * q = p(1-q) + (1-p)q$. If the state of BSC x_M is not given, then the distribution of Y_S is defined by the parameter $\theta_Y = \theta_X * P_b$, where $P_b = \mathbb{E}_{f(x_M)}[P_{b|x_M}]$ corresponds to the average cross-over probability and $\mathbb{E}_{f(x_M)}[\cdot]$ denotes the expectation with respect to the random variable $X_M \sim f(x_M)$. Finally, for a symmetric source, i.e., when $\theta_X = 0.5$, and the considered BSC, the component Y_S is also an equiprobable Bernoulli random variable with $\theta_{Y|x_M} = \theta_Y = 0.5$ for any x_M and $P_{b|x_M}$.

The source component X_S and X_M are independent for the considered source distribution and the output components Y_S and Y_M are independent for the considered channel statistics as well.

One can develop the joint distribution between X_S, X_M, Y_S, Y_M as $f(x_S, x_M, y_S, y_M) = p(x_S)f(x_M)f(y_S, y_M|x_S, x_M)$, where $p(x_S)$ denotes the distribution of input sign component X_S, $f(x_M)$ is the distribution of input magnitude component and the sign-magnitude channel is decomposed as $f(y_S, y_M|x_S, x_M) = p(y_S|x_S, x_M)f(y_M|x_S, x_M)$ with $f(y_M|x_S, x_M) = f(y_M|x_M)$ for the considered symmetric source and channel pdfs as shown in Figure 3.

Proposition 1 (sign-magnitude decomposition). *The mutual information between channel input $X = (X_S, X_M)$ and output $Y = (Y_S, Y_M)$ can be decomposed as shown in Figure 3 and equivalently represented as in Figure 4 and yields:*

$$\begin{aligned} I(X;Y) &= I(X_S, X_M; Y_S, Y_M) \\ &\overset{(a)}{=} I(X_M; Y_M) + I(X_S; Y_S|X_M) \\ &\overset{(b)}{=} I(X_M; Y_M) + I(X_S; Y_S|Y_M) \\ &\quad + \underbrace{H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M)}_{(c)}. \end{aligned} \tag{7}$$

Proof: The expansion (a) in (7) follows from the chain rule for mutual information [20] that yields:

$$\begin{aligned} &I(X_S, X_M; Y_S, Y_M) \\ &= I(X_M; Y_S, Y_M) + I(X_S; Y_S, Y_M|X_M) \\ &= \underbrace{I(X_M; Y_M)}_{\text{magnitude term, } R_M} + \underbrace{I(X_S; Y_S|X_M)}_{\text{sign term, conditioned on } X_M,\ R_{S|x_M}} \\ &+ \underbrace{I(X_M; Y_S|Y_M)}_{\langle \text{cross-term 1} \rangle = 0} + \underbrace{I(X_S; Y_M|X_M, Y_S)}_{\langle \text{cross-term 2} \rangle = 0}. \end{aligned} \tag{8}$$

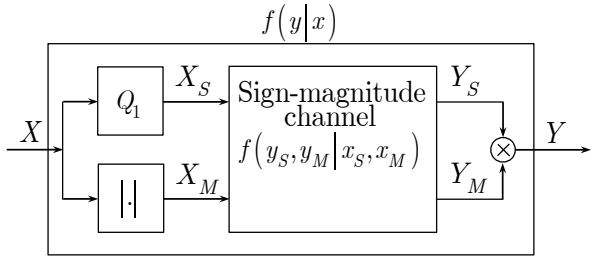


Fig. 2. General decomposition of channel $f(y|x)$ between the input X and output Y via a sign-magnitude channel $f(y_S, y_M|x_S, x_M)$.

In the last decomposition of (8), the first term corresponds to the mutual information between the magnitude components communicated via the channel $f(y_M|x_M)$. The second term represents the mutual information between the sign components communicated through the state dependent BSC $p(y_S|x_S, x_M)$, whose state is defined by the magnitude component x_M:

$$\begin{aligned} I(X_S; Y_S|X_M) &= \int_{\mathcal{X}_M} I(X_S; Y_S|x_M) f(x_M) dx_M \\ &= \mathbb{E}_{f(x_M)}[I(X_S; Y_S|x_M)]. \end{aligned} \tag{9}$$

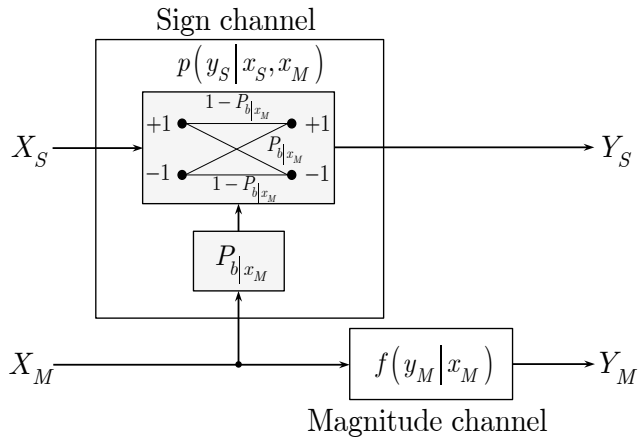


Fig. 3. Statistical model of the sign-magnitude channel linking the input (X_S, X_M) with the output (Y_S, Y_M) via the sign $p(y_S|x_S, x_M)$ and magnitude $f(y_M|x_M)$ channels. The input magnitude component X_M determines the cross-over probability $P_{b|x_M}$ of BSC.

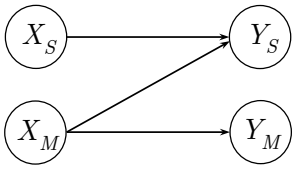


Fig. 4. Equivalent probabilistic graphical model of the sign-magnitude channel for the considered assumptions.

In (8), the first cross-term $I(X_M; Y_S|Y_M) = 0$ and the second cross-term $I(X_S; Y_M|X_M, Y_S) = 0$. The proofs are given in Appendices A and B, respectively.

The expansion (b) in (7) follows from the chain rule decomposition:

$$\begin{aligned} &I(X_S, X_M; Y_S, Y_M) \\ &= I(X_S, X_M; Y_M) + I(X_S, X_M; Y_S|Y_M) \\ &= \underbrace{I(X_M; Y_M)}_{\text{magnitude term, } R_M} + \underbrace{I(X_S; Y_S|Y_M)}_{\text{sign term, conditioned on } Y_M,\ R_{S|y_M}} \\ &+ \underbrace{I(X_S; Y_M|X_M)}_{\langle\text{cross-term 1}\rangle = 0} + \underbrace{I(X_M; Y_S|X_S, Y_M)}_{\langle\text{cross-term 2}\rangle}. \end{aligned} \tag{10}$$

The magnitude term coincides with one in (8), while the sign term is conditioned by y_M, i.e., degraded version of x_M that causes a rate loss $R_{S|y_M} \leq R_{S|x_M}$. This rate loss becomes more evident from the analysis of cross-term 2:

$$\begin{aligned} I(X_M; Y_S|X_S, Y_M) &= H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M, Y_M) \\ &= H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M). \end{aligned} \tag{11}$$

Finally, the first cross-term in (10) is $I(X_S; Y_M|X_M) = 0$ as demonstrated in Appendix C. ■

Remark 1. The decomposition (7) can be very helpful to understand soft fingerprinting schemes, where only binary information is stored, but soft information on bit reliabilities is extracted from the noisy magnitudes. The total mutual information between the real data X and Y is decomposed on the magnitude term and sign term where the sign term is conditioned by the magnitude term. The second term in (7a) represents the mutual information between the input and output sign components under the perfect channel state x_M. It is interesting to point out that the second term in (7b) corresponds to the same mutual information under the noisy information y_M on the channel state x_M. In this case, the mismatch between the exact state x_M and its noisy counterpart y_M manifests itself as a term (7c).

B. Decomposition of AWGN identification channel

In this Section we apply decomposition (7) to the Gaussian input X and AWGN channel (3) and analyze the mutual information for the magnitude term and sign term under the perfect knowledge of channel state x_M and noisy one y_M. These last two cases correspond to soft fingerprinting. We also consider a case of hard fingerprinting which disregards any information on the channel state.

1) Magnitude term: The first term of both decompositions (8) and (10) is $R_M = I(X_M; Y_M) = h(Y_M) - h(Y_M|X_M)$. The differential entropy $h(Y_M) = 1/2\log_2(1/2\pi e(\sigma_X^2 + \sigma_Z^2)) = 1/2\log_2(2\pi e(\sigma_X^2 + \sigma_Z^2)) - 1$ corresponds to the entropy of the half-normal distribution, where "-1" reflects the absence of the sign, which requires 1 bit of information. The conditional term can be rewritten as:

$$h(Y_M|X_M) = h(Z) - H(Y_S|X_S, X_M), \tag{12}$$

that follows from the decomposition:

$$\begin{aligned} h(Z) = h(Y|X) &\overset{(a)}{=} h(Y_S, Y_M|X_S, X_M) \\ &= h(Y_M|X_S, X_M) + H(Y_S|X_S, X_M, Y_M) \\ &\overset{(b)}{=} h(Y_M|X_M) + H(Y_S|X_S, X_M), \end{aligned} \tag{13}$$

with differential entropy $h(Z) = 1/2\log_2(2\pi e\sigma_Z^2)$. Equality (a) follows from the representation $X = X_S \cdot X_M$ and $Y = Y_S \cdot Y_M$. The first term in the equality (b) is obtained as $h(Y_M|X_S, X_M) = h(Y_M|X_M)$ since $Y_M \perp X_S|X_M$ for the model $Y_M \leftarrow X_M \rightarrow Y_S \leftarrow X_S$ and the second one $H(Y_S|X_S, X_M, Y_M) = H(Y_S|X_S, X_M)$ due to $Y_S \perp Y_M|X_M$.

The term $H(Y_S|X_S, X_M)$ corresponds to the entropy of the event related to the mismatch of the signs of Y_S and X_S under X_M for the considered BSC that can be developed as:

$$\begin{aligned} H(Y_S|X_S, X_M) &= \int_{\mathcal{X}_M} H(Y_S|X_S, x_M) f(x_M) dx_M \\ &= \mathbb{E}_{f(x_M)}[H(Y_S|X_S, x_M)] \\ &= \mathbb{E}_{f(x_M)}[H_2(Pr[Y_S \neq X_S|x_M])] \\ &= \mathbb{E}_{f(x_M)}[H_2(P_{b|x_M})], \end{aligned} \tag{14}$$

where $f(x_M) = \frac{2}{\sqrt{2\pi\sigma_X^2}} \exp[-\frac{x_M^2}{2\sigma_X^2}]$ and:

$$\begin{aligned}P_{b|x_M}=&\Pr[Y_S\neq X_S|x_M]\\ =&\Pr[Y_S=-1|X_S=+1,x_M]\Pr[X_S=+1]\\ &+\Pr[Y_S=+1|X_S=-1,x_M]\Pr[X_S=-1]\\ =&\Pr[Y_S=-1|X_S=+1,x_M]\\ =&\int_{-\infty}^{0}p(y|x_M)dy\\ =&\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi\sigma_Z^2}}\exp\left[-\frac{(y-x_M)^2}{2\sigma_Z^2}\right]dy\\ =&Q\left(\frac{x_M}{\sigma_Z}\right).\end{aligned}\tag{15}$$

Substituting (15) into (14) yields:

$$H(Y_S|X_S,X_M)=\int_{\mathcal{X}_M}H_2\left[Q\left(\frac{x_M}{\sigma_Z}\right)\right]f(x_M)dx_M.\tag{16}$$

Alternatively, one can find the conditional entropy (12) as:

$$\begin{aligned}&h(Y_M|X_M)\\ &=-\int_{\mathcal{X}_M}\int_{\mathcal{Y}_M}f\left(y_M,x_M\right)\log_2\left(f\left(y_M|x_M\right)\right)dy_Mdx_M,\end{aligned}\tag{17}$$

where the conditional pdf $f(y_M|x_M)=\frac{1}{\sqrt{2\pi\sigma_Z^2}}\exp\left[-\frac{(-y_M-x_M)^2}{2\sigma_Z^2}\right]+\frac{1}{\sqrt{2\pi\sigma_Z^2}}\exp\left[-\frac{(y_M-x_M)^2}{2\sigma_Z^2}\right]$ is characterized by a folded normal distribution.

Remark 2. Denoting the signal-to-noise ratio as $SNR=10\log_{10}\frac{\sigma_X^2}{\sigma_Z^2}$, we present the behavior of $h(Y_M|X_M)$, $h(Z)$, and $h(Z_M)$ in Figure 5. The conditional entropy $h(Y_M|X_M)$ was computed according to (12) and (17) by numerical integration that both give the same result. The conditional entropy $h(Y_M|X_M)$ approaches the entropy of $h(Z)$ for the regime of small degradations, i.e., with the increase of SNR, when the term $H(Y_S|X_S,X_M)$ tends to zero in (12). For high degradations, i.e., with the decrease of SNR, the term $H(Y_S|X_S,X_M)$ tends to 1 and $h(Y_M|X_M)$ approaches the entropy of channel noise magnitude $h(Z_M)=h(Z)-1$.

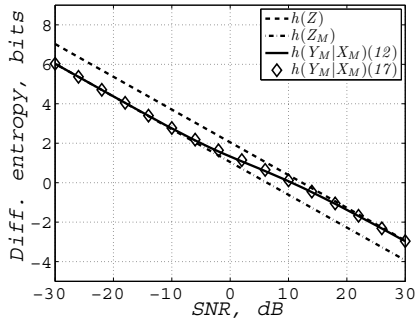


Fig. 5. Behavior of conditional entropy $h(Y_M|X_M)$ with respect to $h(Z)$ and $h(Z_M)$.

2) Sign term: In the decomposition (7), there are two sign terms conditioned on X_M and Y_M. The sign term in the decomposition (7(a)) is:

$$\begin{aligned}I(X_S;Y_S|X_M)=&H(Y_S|X_M)-H(Y_S|X_S,X_M)\\ \overset{(a)}{=}&\mathbb{E}_{f(x_M)}[H(Y_S|x_M)]-H(Y_S|X_S,X_M)\\ \overset{(b)}{=}&1-H(Y_S|X_S,X_M),\end{aligned}\tag{18}$$

where in (a) $H(Y_S|X_M)=\mathbb{E}_{f(x_M)}[H(Y_S|x_M)]$ and in (b) $\mathbb{E}_{f(x_M)}[H(Y_S|x_M)]=\mathbb{E}_{f(x_M)}[H_2(\theta_X*P_{b|x_M})]=\mathbb{E}_{f(x_M)}[H_2(0.5)]=1$ as shown in Appendix A for the assumed conditions on source and channel. The second term is given by (16). Thus, (18) yields:

$$\begin{aligned}R_{S|x_M}=&I(X_S;Y_S|X_M)\\ =&1-\int_{\mathcal{X}_M}H_2\left[Q\left(\frac{x_M}{\sigma_Z}\right)\right]f(x_M)dx_M.\end{aligned}\tag{19}$$

This rate is very important for further consideration and represents the *identification rate with perfect side information* on the channel state x_M.

The sign term in the decomposition (7(b)) is:

$$\begin{aligned}I(X_S;Y_S|Y_M)=&H(Y_S|Y_M)-H(Y_S|X_S,Y_M)\\ =&1-H(Y_S|X_S,Y_M),\end{aligned}\tag{20}$$

where $H(Y_S|Y_M)=H(Y_S)=1$ which follows from $Y_S\perp Y_M$ for the symmetric $f(y)$ obtained for the assumed source and channel distributions. The term $H(Y_S|X_S,Y_M)$ corresponds to the mismatch of the signs between input and output under Y_M and similar to (14) can be rewritten as:

$$\begin{aligned}H(Y_S|X_S,Y_M)=&\mathbb{E}_{f(y_M)}[H(Y_S|X_S,y_M)]\\ =&\mathbb{E}_{f(y_M)}[H_2(\Pr[Y_S\neq X_S|y_M])]\\ =&\mathbb{E}_{f(y_M)}[H_2(P_{b|y_M})],\end{aligned}\tag{21}$$

where

$$\begin{aligned}P_{b|y_M}=&\Pr[Y_S\neq X_S|y_M]\\ \overset{(a)}{=}&\mathbb{E}_{f(x_M|y_M)}[\Pr[Y_S\neq X_S|x_M]]\\ \overset{(b)}{=}&\mathbb{E}_{f(x_M|y_M)}\left[Q\left(\frac{x_M}{\sigma_Z}\right)\right]\\ =&\int_{\mathcal{X}_M}Q\left(\frac{x_M}{\sigma_Z}\right)f(x_M|y_M)dx_M,\end{aligned}\tag{22}$$

where the conditional pdf $f(x_M|y_M)=\frac{1}{\sqrt{2\pi\sigma_{X|Y}^2}}\exp\left[-\frac{(-x_M-\rho y_M)^2}{2\sigma_{X|Y}^2}\right]+\frac{1}{\sqrt{2\pi\sigma_{X|Y}^2}}\exp\left[-\frac{(x_M-\rho y_M)^2}{2\sigma_{X|Y}^2}\right]$ is represented by a folded normal distribution with $\rho=\sigma_X^2/(\sigma_X^2+\sigma_Z^2)$ and $\sigma_{X|Y}^2=\sigma_X^2\sigma_Z^2/(\sigma_X^2+\sigma_Z^2)$ and (a) corresponds to the MMSE estimation of X for a given Y: $\hat{x}=\mathbb{E}[XY]/\mathbb{E}[X]=\rho y$; (b) follows from (15). Substituting (22) into (21) yields:

$$\begin{aligned}&R_{S|y_M}=1-H(Y_S|X_S,Y_M)=\\ &1-\int_{\mathcal{Y}_M}H_2\left[\int_{\mathcal{X}_M}Q\left(\tfrac{x_M}{\sigma_Z}\right)f(x_M|y_M)dx_M\right]f(y_M)dy_M.\end{aligned}\tag{23}$$

This rate corresponds to the *soft fingerprinting with imperfect or noisy information* about x_M observed via the channel $f(y_M|x_M)$.

Remark 3. In the case of no channel state information (CSI) about the state x_M of channel $p(y_S|x_S,x_M)$, the rate is:

$$\begin{aligned}R_{S|\oslash}=&I(X_S;Y_S|\oslash)\\ =&H(Y_S)-H(Y_S|X_S)\\ =&H_2(\theta_X*P_b)-H_2(P_b)\\ =&1-H_2(P_b),\end{aligned}\tag{24}$$

where the last equality is obtained for symmetric sources by substituting $\theta_X = 0.5$. For the Gaussian source and additive Gaussian channel, the cross-over probability P_b is defined as for (6).

This situation corresponds to identification based on hard fingerprints, where only the sign information is used for decoding. We present the resulting rates R_M and $R_{S|x_M}$ with respect to the capacity of the AWGN channel in Figure 6a and $R_{S|x_M}$, $R_{S|y_M}$ and $R_{S|\oslash}$ in Figure 6b. The differences between the rates explaining the impact of perfect and noisy side information is shown in Figure 6c.

Remark 4 *(Sign-magnitude decomposition of the mutual information for AWGN).* The decomposition of R^{RP} for the AWGN channel results into the rates R_M and $R_{S|x_M}$ that in sum coincide with the capacity of the AWGN channel. At high SNR, the achievable rate $R_{S|x_M}$ converges to 1 bit/channel use that exactly corresponds to the difference between R^{RP} and R_M.

Remark 5. The impact of side information about the BSC state x_M (Figure 6c):

- the presence of perfect side information about the BSC state x_M enhances the rate with respect to the degraded version y_M at low SNR, where y_M is less reliable;
- the rate loss between the perfect and noisy CSI $R_{S|x_M} - R_{S|y_M}$ is bounded by about 0.08 bits/sample;
- the rate $R_{S|y_M}$ approaches rate $R_{S|x_M}$ at high SNR that indicates that the decoding in the sign channel can be performed even based on noisy CSI;
- blind decoding without any information about x_M, often used in digital (hard) fingerprinting, represents about 0.15 bit/sample loss with respect to $R_{S|x_M}$ (at $SNR \approx 9dB$).

IV. Channel splitting and polarization

In the above consideration of upper achievable limit, we assumed that the magnitude component x_M, determining the sign channel state, is completely available at the decoder in the uncompressed form. In practical applications, it is unfeasible to communicate or store this information or even undesirable for privacy reasons. Therefore, this information is quantized just to 1 bit/sample providing a sort of indication whether a component belongs to the class of large or small magnitudes. Therefore, it looks very interesting to investigate the identification rate in the sign channel under the quantized information on the channel state available at the decoder.

In this Section, we introduce a practical model that divides the sign channel into two subchannels corresponding to the large and small magnitude components. We will refer to this model as a *channel splitting*. Along this way, we will demonstrate the effect of rate concentration in a few coefficients which we will refer to as *channel polarization*.

A. Quantized channel state information

Considering the sign channel, one can interpret the magnitude channel as a CSI available at the decoder. If the sequences $\mathbf{x}_M$ or $\mathbf{y}_M$ are available at the decoder completely, one can achieve the rates $R_{S|x_M}$ (19) or $R_{S|y_M}$ (23), respectively. However, it is equivalent to the representation of this CSI with the rates $h(X_M)$ and $h(Y_M)$. In some cases as mentioned above, it is memory costly or might not be desirable for the privacy reasons [14]. That is why partial or quantized information is generally used. We will represent the quantized versions of x_M and y_M as:

$$\hat{x}_M \approx x_{M_0}2^0 + x_{M_1}2^1 + x_{M_2}2^2 + \ldots + x_{M_k}2^{K_Q-1}, \quad (25)$$

$$\hat{y}_M \approx y_{M_0}2^0 + y_{M_1}2^1 + y_{M_2}2^2 + \ldots + y_{M_k}2^{K_Q-1}, \quad (26)$$

where K_Q denotes the number of quantization levels, and $x_{M_\ell}, y_{M_\ell} \in \{0,1\}$, where $1 \leq \ell \leq K_Q$. This results in $H(\hat{X}_M)$ and $H(\hat{Y}_M)$ bit representations.

We will consider only a 1-bit representation, i.e., $K_Q = 1$, using components x_{M_0} and y_{M_0} that correspond to the splitting of magnitude components on two components that will be treated separately. We will demonstrate that one can closely approximate the original rates with the 1-bit CSI and present low-complexity search strategies.

B. Channel splitting

The channel splitting model assumes that the channel input vector $\mathbf{x}_S$ can be transmitted via several BSCs with parameters defined by the state vector $\mathbf{x}_M$ according to the model $p(y_S|x_S, x_M)$. In the most simple case of 2-channel splitting for 1 bit representation of magnitude component, two BSCs are considered. The channel splitting is accomplished based on the available side information $\mathbf{x}_M$ or $\mathbf{y}_M$ and can be implemented by thresholding the magnitude coefficients with the thresholds T_{x} and T_{y}, respectively. Equivalently, the same proportion of coefficients can be chosen based on sorting the N magnitude coefficients of $\mathbf{x}_M$ or $\mathbf{y}_M$ and selecting the J largest ones. The J bits related to the large magnitude coefficients are considered as those belonging to the $strong$ BSC with the cross-over probability $P^S_{b|x_{M_0}}$ and the remaining to the $weak$ one characterized by $P^W_{b|x_{M_0}}$.

Remark 6. The crossover probabilities for strong and weak channels based on the errorless 1-bit CSI (given x_{M_0}) are:

$$P^S_{b|x_{M_0}} = \frac{1}{\Pr^S} \int_{T_{\mathrm{x}}}^{+\infty} P_{b|x_M} f(x_M) dx_M, \quad (27)$$

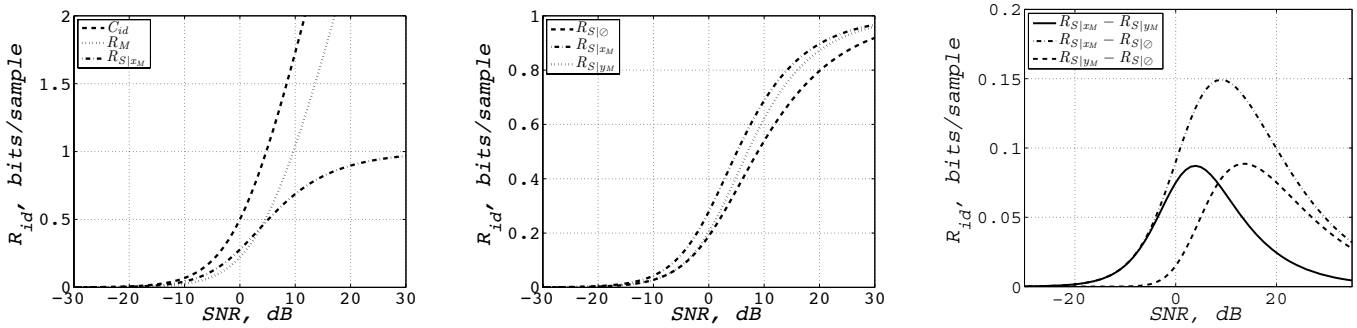
$$P^W_{b|x_{M_0}} = \frac{1}{\Pr^W} \int_{0}^{T_{\mathrm{x}}} P_{b|x_M} f(x_M) dx_M, \quad (28)$$

where $\Pr^S = \int_{T_{\mathrm{x}}}^{+\infty} f(x_M) dx_M$ and $\Pr^W = \int_0^{T_{\mathrm{x}}} f(x_M) dx_M$ correspond to probabilities of observing the strong and weak channels, respectively. The threshold T_{x} is computed to satisfy the chosen probability of observing the strong channel component $\Pr^S$. Therefore, in following we will not deal with the absolute value of the threshold but rather express all threshold values in terms of the probability of the strong channel $\Pr^S$.

The corresponding identification rates are:

$$R^S_{S|x_{M_0}} = \Pr^S \left[1 - H_2\left(P^S_{b|x_{M_0}}\right)\right], \quad (29)$$

$$R^W_{S|x_{M_0}} = \Pr^W \left[1 - H_2\left(P^W_{b|x_{M_0}}\right)\right], \quad (30)$$



(a) (b) (c)

Fig. 6. The rates for sign-magnitude decomposition: (a) achievability of AWGN capacity by sum of the rates for magnitude R_M and sign $R_{S|x_M}$ channels; (b) achievable rates under perfect x_M, degraded side information based on y_M and zero side information on the channel $p(y_S|x_S, x_M)$; (c) differences between the rates from (b).

and the total rate for the two channel splitting corresponding to the 1-bit CSI is:

$$R^{\text{2Ch}}_{S|x_{M_0}} = R^{S}_{S|x_{M_0}} + R^{W}_{S|x_{M_0}}. \tag{31}$$

Remark 7. The cross-over probabilities for strong and weak channels based on the quantized degraded CSI (given y_{M_0}) are:

$$P^{S}_{b|y_{M_0}} = \frac{1}{\Pr^S} \int_{T_\text{y}}^{+\infty} P_{b|y_M} f(y_M) dy_M, \tag{32}$$

$$P^{W}_{b|y_{M_0}} = \frac{1}{\Pr^W} \int_{0}^{T_\text{y}} P_{b|y_M} f(y_M) dy_M, \tag{33}$$

where T_y is selected to satisfy the conditions stated above, i.e. $\Pr^S = \int_{T_\text{y}}^{+\infty} f(y_M) dy_M$ and $\Pr^W = \int_0^{T_\text{y}} f(y_M) dy_M$. The corresponding identification rates are:

$$R^{S}_{S|y_{M_0}} = \Pr^S \left[1 - H_2\left(P^{S}_{b|y_{M_0}}\right)\right], \tag{34}$$

$$R^{W}_{S|y_{M_0}} = \Pr^W \left[1 - H_2\left(P^{W}_{b|y_{M_0}}\right)\right], \tag{35}$$

and the total rate yields: $R^{\text{2Ch}}_{S|y_{M_0}} = R^{S}_{S|y_{M_0}} + R^{W}_{S|y_{M_0}}$.

Remark 8. The total cross-over probability P_b remains the same as for the case of no CSI:

$$P_b = \int_{\mathcal{X}_M} P_{b|x_M} f(x_M) dx_M = \Pr^S P^{S}_{b|x_{M_0}} + \Pr^W P^{W}_{b|x_{M_0}}. \tag{36}$$

Therefore, we split the sign channel into two subchannels with different probabilities of bit error and corresponding rates. We consider both cases, when perfect and degraded CSI are available for practical reasons since it might be communicated together with the template [3] or extracted on-the-fly [1]. Interesting results are obtained for different threshold selection strategies in the next Section.

C. Channel polarization

The channel splitting by the selection of the thresholds T_x and T_y can be performed according to two strategies:

- **Strategy 1**: maximize the total rates $R^{\text{2Ch}}_{S|x_{M_0}}$ or $R^{\text{2Ch}}_{S|y_{M_0}}$ to approach upper theoretical limits $R_{S|x_M}$ or $R_{S|y_M}$, respectively, which gives optimal values of thresholds T_{x_opt} and T_{y_opt} for each SNR;
- **Strategy 2**: minimize probabilities $P^{S}_{b|x_{M_0}}$ or $P^{S}_{b|y_{M_0}}$ to minimize search complexity and enhance privacy amplification.

According to the first strategy of total rate maximization for the 2-channel splitting, the binary channel splitting approaches theoretical performance limits under the optimal threshold selection as shown in Figure 7a with the magnified version in Figure 7b. The threshold optimization is performed numerically. The remaining gap is easily compensated by more accurate models using a larger number K_Q of channels in splitting (25). The optimal thresholds T_{x_opt} and T_{y_opt}, expressed in terms of $(1 - \Pr^S)$ are shown in Figure 7c. It should be noticed that the optimal thresholds do not coincide.

To exemplify strategy 2 we show in Figure 8a the pairs of cross-over probabilities for strong and weak channels under perfect and degraded CSI for the optimal thresholds T_{x_opt} and T_{y_opt}. Figure 8b shows the same pairs for the fixed thresholds T_x and T_y, where one can clearly observe the significant reduction of $P^{S}_{b|x_{M_0}}$ or $P^{S}_{b|y_{M_0}}$ that asymptotically goes to zero for $SNR > 15$ dB.

Remark 9 *(Channel polarization).* An interesting phenomenon is observed for T_{x_opt} and T_{y_opt}, when practically all rate is concentrated in the strong channel, i.e., $R_{S|x_M} \approx R^{S}_{S|x_{M_0}}$ and $R_{S|y_M} \approx R^{S}_{S|y_M}$ after a certain SNR. We will refer to the effect of rate concentration in strong channels as *channel polarization*. In addition, it means that two channels might be treated separably and the decoding rules for the weak and strong channels might be optimized to reduce the decoding complexity. Moreover, the positions of the bits belonging to the strong channels can be reliably estimated from the magnitude component.

One can point out the difference to polar codes [18], where a similar effect exists and the relationship between bits is created in such a way that conditional entropy is polarized to either 0 or 1, i.e., "weak" or "strong" bits. This property has a considerable impact on the complexity. In our case, when the codewords are random, the effect of polarization can be achieved by selecting $T > T_\text{opt}$ (Figure 8b), when the cross-over probability in the strong channel is asymptotically equal to 0. Thus, those bits can be considered to be reliable and can be used directly for identification without decoding. However,

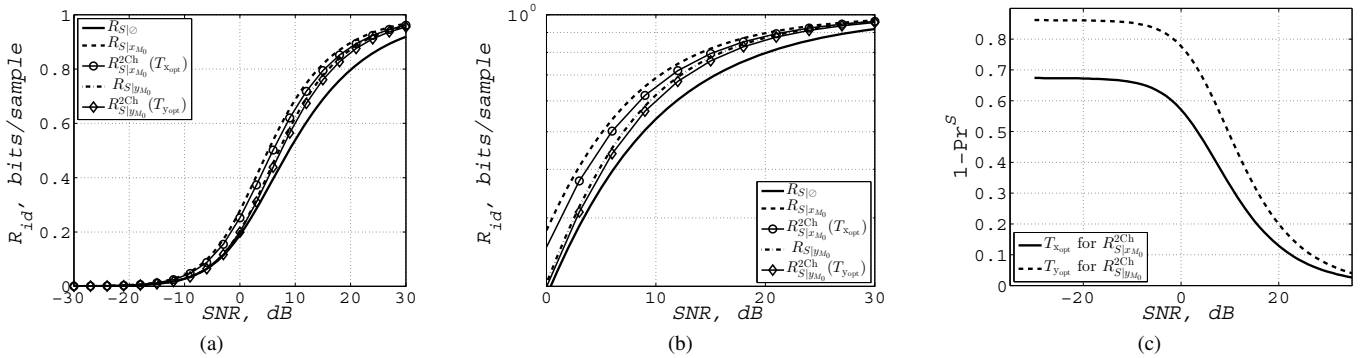


Fig. 7. Approaching theoretical rates based on 2-channel splitting model for the optimal threshold selection: (a) achievable identification rates under different CSIs; (b) magnified region of (a); (c) optimal thresholds for perfect and imperfect CSI.

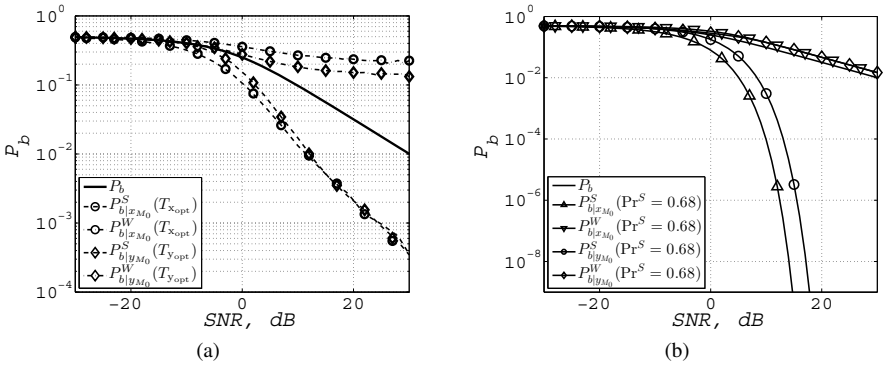


Fig. 8. Probabilities of bit errors: (a) for $T_{\mathrm{x}_{\mathrm{opt}}}$ and $T_{\mathrm{y}_{\mathrm{opt}}}$; (b) for $T_{\mathrm{x}}=1$ and corresponding T_{y}.

an unavoidable price for this option is a rate loss in the strong channel. Multichannel splitting (25) with multistage decoding can partially resolve this rate-complexity trade-off, but that is out of the scope of this paper. The polarization effect is demonstrated in Figure 9 as the dependence of achievable rates and cross-over probabilities on the threshold T or equivalently $(1-\Pr^S)$ for $SNR=10,20,30dB$.

The described phenomena can be of interest for:

- design of new search algorithms, when the representation of original content is reduced to a vector of signs $\mathbf{x}_S$ of length N, and the decoder searches for a match between the query $\mathbf{y}_S$ and $\mathbf{x}_S$ assuming that the reliable bits in both vectors determined by the large magnitude components are preserved with high probability;
- joint multistage search in the random Gaussian codebooks, where the search is performed over the magnitude and signs codebooks;
- security and privacy amplification of biometrics and privacy-preserving content identification as a extension to [1].

We will only analyze possible benefits for the first use case in the next section.

V. Estimation of Search Complexity for Soft Fingerprints

An efficient decoding of random binary codes, i.e., the content identification based on random or unstructured fingerprints, represents a challenging computational problem. Given a codebook $\mathcal{C}_s$ of M_s binary codewords, the identification system should find a unique estimate $\hat{u}$ that satisfies the *bounded distance decoding* (BDD) rule:

$$d^H(\mathbf{y}_S,\mathbf{x}_S(u))\le\gamma N, \tag{37}$$

for all $1\le u\le M_s$ where $d^H(.)$ denotes the Hamming distance and a threshold γ is chosen to minimize the identification error probability as [1]:

$$\gamma_{\text{opt}}=\frac{1-R_{S|\varnothing}+\log_2(1-P_b)}{\log_2(1-P_b)-\log_2(P_b)}, \tag{38}$$

to communicate with the rate $R_{S|\varnothing}$ close to the capacity of BSC (6).

This decoder is schematically shown in Figure 10.

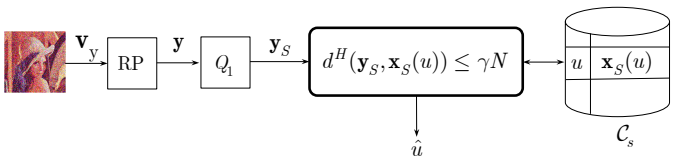


Fig. 10. The identification based on the BDD.

Remark 10. (*Optimal threshold for the capacity achieving rate*): For the identification rate satisfying $R_{S|\varnothing}\le 1-H_2(P_b)$, the above optimal threshold yields $\gamma_{\text{opt}}\le P_b$. This means that the decoding region around each codeword is defined by the radius close to P_bN.

An exhaustive implementation of the decoding rule (37) to verify all candidates $1\le u\le M_s$ in the codebook $\mathcal{C}_s$, with

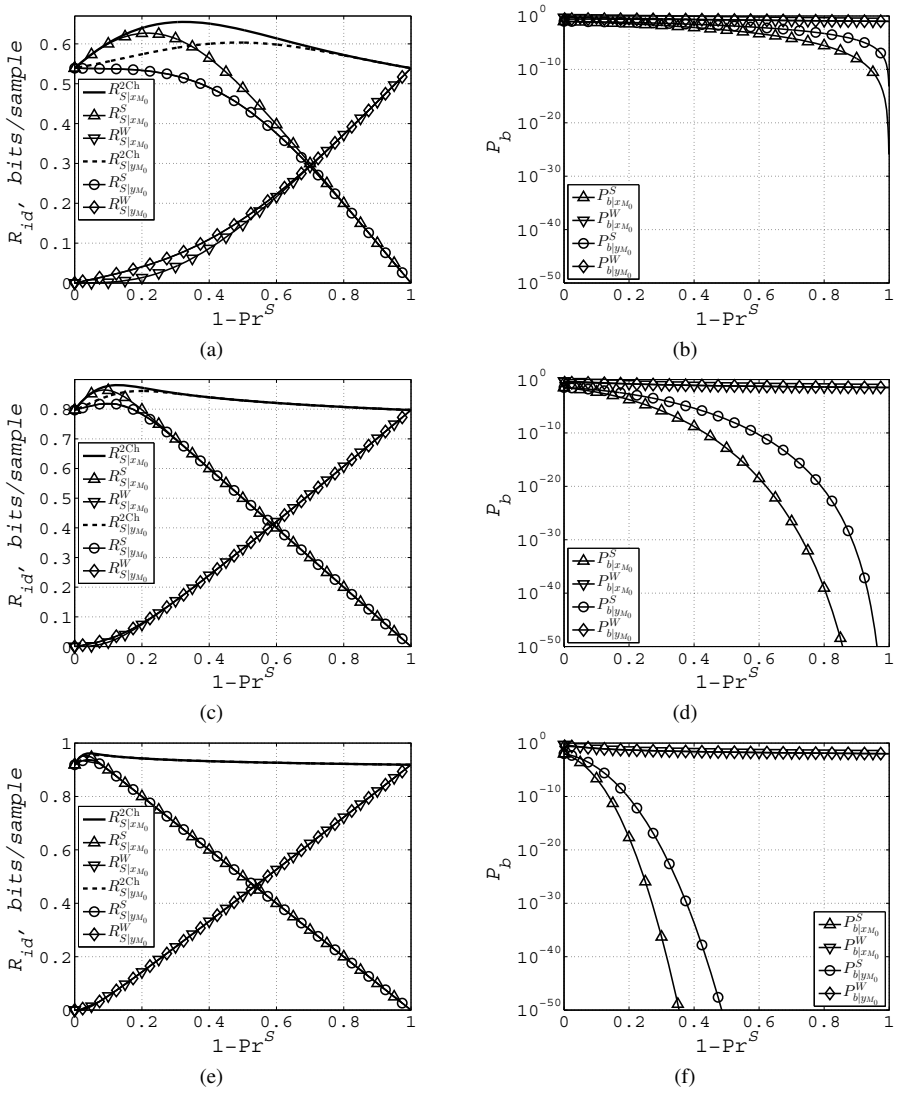


Fig. 9. Achievable rates and cross-over probabilities for channel splitting at $SNR = 10dB$ (a,b), $SNR = 20dB$ (c,d) and $SNR = 30dB$ (e,f).

$M_s \leq 2^{NR_{S|\oslash}}$ and with $R_{S|\oslash}$ defined by (24) is a NP-hard problem.

Alternatively, in the next Section we present an approach that targets the content identification at the rate $R_{S|\oslash}$ based on the Hamming similarity measure (37) with a complexity lower than those of the exhaustive search for sufficiently high SNR.

We assume that the binary fingerprints $\mathbf{x}_S(u)$, $1 \leq u \leq M_s$ are stored in the database and the real probe $\mathbf{y} = (\mathbf{y}_S, \mathbf{y}_M)$ is presented to the decoder. We consider two versions of the algorithm using the BDD, based on hard and soft decoding. The BDD based on hard decoding uses only the binary version of probe $\mathbf{y}_S$ while the soft counterpart uses both $(\mathbf{y}_S, \mathbf{y}_M)$. It should be also pointed out that knowledge of $\mathbf{y}_M$ might increase the achievable rate $R_{S|y_M}$ under proper similarity metric. In our analysis, we will use $\mathbf{y}_M$ not to increase the rate from $R_{S|\oslash}$ to $R_{S|y_M}$ but solely targeting the reduction of search complexity in (37). In addition, the considered methods might be of interest for fast list decoding.

A. BDD based on hard decoding

The memory storage problem can be relaxed with acceptable complexity by an alternative interpretation of the BDD (37). One can consider the BDD even without the necessity to compute the Hamming distances as an exact matching rule assuming that the searched sequence $\mathbf{x}_S(\hat{u})$ is within the Hamming sphere of radius P_bN around the codeword $\mathbf{y}_S$ (Figure 11).

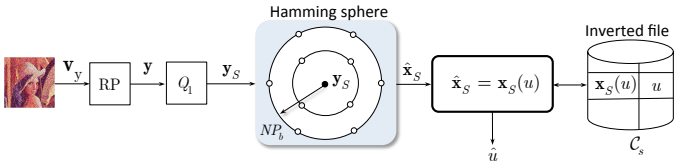


Fig. 11. The BDD implementation as the Hamming sphere decoding. The Hamming sphere is represented schematically as the concentration circles with the different radii around $\mathbf{y}_S$ that form a search space.

We will refer to this type of BDD as a *Hamming sphere decoder* (HSD). The decoder exhaustively generates all possi-

ble "virtual" codewords $\hat{\mathbf{x}}_S$ within the Hamming sphere and sequentially validates their presence by querying the database for the exact match $\hat{\mathbf{x}}_S = \mathbf{x}(u)$. If such a match exists, the index $\hat{u}$ of the corresponding codeword is declared as the searched result. Otherwise, an erasure or no match is declared. It should be pointed out that the HSD does not require any computation of Hamming distances similarly to the methods based on look-up-tables (LUT) and approximate search using product of vector quantizers [6], [7]. However, contrarily to the LUT approach that requires a lot of memory to store the tables, the HSD generates the sequences on-the-fly. Finally, the matching complexity for each query from the Hamming sphere is logarithmic in N if the database is sorted.

1) HSD implementation: The HSD can be implemented as a recursive function that searches through a space explored by the decoder, a design that has been inspired by DPLL-solvers for the Satisfyability (SAT) problem [22]. The decoder has a state that it updates continuously, and an assignment of the binary variables. In the first state, the assignment of variables is equal to $\mathbf{y}_S$, and the decoder then changes its state in each step. In each state, a so-called *branch variable* is selected that can either be flipped, or kept, thus creating a binary search tree. Hence, the assignment of the binary variables is only changed in some state changes of the decoder. A variable cannot be chosen as a branch variable if it already has been a branch variable in a node at a higher level in the search tree. The states of the decoder are depicted in Figure 12 where black nodes indicate a new assignment, and transparent nodes indicate a new state of the decoder where the assignment is unchanged. Each new assignment of the variables can be seen as a probe that is tested in the database with respect to the correspondence to an existing fingerprint. The resulting assignments cover the search space, which should span the fingerprint of item u, if that item was observed. The result is schematically shown in Figure 13, where each black node represents a different assignment of the N binary variables. In reality, many different states can be derived from a particular state, but for simplicity, only two derived states are drawn. On average, the Hamming distance between the channel output $\mathbf{y}_S$ and the fingerprint of the observed item $\mathbf{x}(u)$ should equal $P_b N$.

Therefore, the algorithm uses: (a) *similarity measure* based on Hamming distance: which indicates the quality of a particular assignment; (b) *stop condition*: that decides if branching should stop or continue and (c) *branch rule*: the procedure that selects the next branch variable.

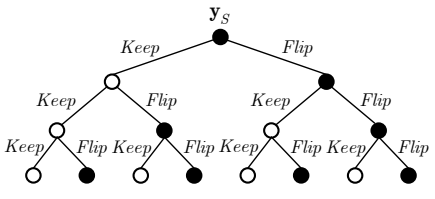


Fig. 12. Search algorithm organization in the structure of tree.

2) HSD complexity: The complexity of HSD is defined by the total number of codewords in the Hamming sphere that should be verified to find a match with the fingerprint in the database. It is important to point out that contrary to the exhaustive search method, this complexity is not determined by the size of the database.

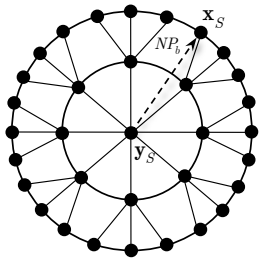


Fig. 13. Spanning the search space by flipping bits.

Proposition 2: The number of codewords in the Hamming sphere of radius γN, where $0 \leq \gamma \leq \frac{1}{2}$, is defined by the partial sum of the binomial coefficients that can be bounded as:

$$\sum_{t=0}^{\gamma N} \binom{N}{t} \leq 2^{NH_2(\gamma)}, \tag{39}$$

where $0 \leq t \leq \gamma N$. Furthermore, asymptotically we have:

$$\lim_{N\to\infty} \sum_{t=0}^{\gamma N} \binom{N}{t} \doteq 2^{NH_2(\gamma)}. \tag{40}$$

This result is a direct consequence of the weak law of large numbers when all sequences will be concentrated within a thin shell of sphere of radius $2^{NH_2(\gamma)}$.

Remark 11. According to Remark 10, $\gamma_{opt} \leq P_b$ that results in the identification complexity $\mathcal{O}(2^{NH_2(P_b)})$. Therefore, this complexity critically depends on P_b.

B. BDD based on soft decoding

The information about the bit reliability extracted from $\mathbf{y}$ based on the analysis of the magnitude component $\mathbf{y}_M$ can be used to increase the identification rate from $R_{S|\oslash}$ defined by (24) to $R_{S|y_M}$ defined by (23) as well as to reduce the search complexity. Figure 14 schematically explains a concept of bit reliability in the RP domain. The coefficients with the small values of magnitude $\mathbf{y}_{M_i}$ have a high probability of bit flipping according to (15). As pointed out in Section IV, the

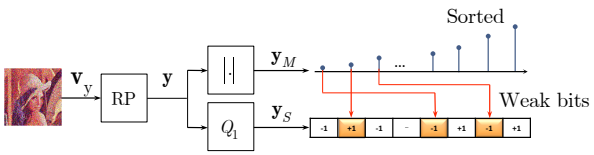


Fig. 14. The concept of bit reliability based on the magnitude component $\mathbf{y}_{M_i}$. The small magnitude components have a high probability of bit error $P_{b|y_{M_i}}$.

components with the small magnitudes do not significantly contribute to the total sum rate under the proper selection of the threshold. In fact, as shown, one can practically assume that the total channel is split on two channels with the negligibly small probability of error that represents the strong

channel and the probability of bit error close to 0.2-0.5 that corresponds to the weak channel. Since the communication via the weak channel is practically equivalent to a bit guessing, the exhaustive generation of all possible combinations of bits in the positions of weak components corresponds to the HSD strategy.

The knowledge of the positions of weak bits can be efficiently used for the reduction of ambiguity about the searched estimate $\hat{\mathbf{x}}_S$ within the Hamming sphere. Therefore, one can assume with high probability that the strong bits can be kept as observed in $\mathbf{y}_S$ while the decoder generates all possible codewords within the Hamming sphere of radius $P_b N$ in the positions of weak bits. This decoder will be called a *reliability based* HSD (RHSD) and its block diagram is shown in Figure 15.

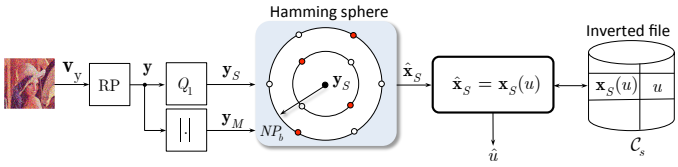


Fig. 15. The BDD implementation as the reliability based Hamming sphere decoding.

1) RHSD implementation: The implementation of RHSD follows the same principle as the HSD. The similarity measure is based on the Hamming distance. The initial decoder state is set up to be $\hat{\mathbf{x}}_S = \mathbf{y}_S$. The decoder updates its state until the correspondence to any stored fingerprint $\mathbf{x}_S(u)$, $1 \leq u \leq M_s$, i.e., $\hat{\mathbf{x}}_S = \mathbf{x}_S(u)$, is found or the stop condition is met. The decoder state update is accomplished by the bit flipping in the nodes whose visiting priority is determined by the bit reliability computed according to the sorted magnitudes $\mathbf{y}_S$ (Figure 14). The bits with the smallest reliabilities are visited first. The stop condition is met when the Hamming distance between the probe $\mathbf{y}_S$ and the current state $\mathbf{x}_S$ exceeds a chosen threshold NP_b. In this sense, the algorithm still represents a sort of hard decoding according to the similarity measure [3]. The branch rule is to pick the node with the closest level of reliability to the current one based on the sorted list of magnitudes.

2) RHSD complexity: We will consider the complexity of RHSD with the assumption that the positions of weakest bits might be determined with the high probability. The number of error bits in a fingerprint of length N corresponds to the mean of binomial distribution and is NP_b. Assuming that the positions of NP_b bits are known, one should perform $\mathcal{O}(2^{NP_b})$ verifications. The obtained complexity is still exponential in N but it critically depends on P_b. This complexity is lower than those of HSD since $P_b < H_2(P_b)$.

We will exemplify the RHSD complexity reduction for $SNR = 30dB$. The average number of error bits can be decomposed using (36) as:

$$NP_b = N\Pr^S P^S_{b|x_{M_0}} + N\Pr^W P^W_{b|x_{M_0}}. \quad (41)$$

[3]A soft version of similarity measure can be also used that leads to the faster search due to the priority of branch selection and higher achievable rate [23].

According to *strategy II* considered in section IV-C, one can select such a threshold corresponding to $\Pr^W = (1 - \Pr^S)$ for the magnitude-based bit classification to the group of strong or weak bits such that the probability $P^S_{b|x_{M_0}}$ is negligibly small and almost all bit error probability is defined by the term $P^W_{b|x_{M_0}}$. For example, according to Figure 9f, setting up $\Pr^W = (1 - \Pr^S) = 0.3$, one obtains $P^S_{b|x_{M_0}} = 10^{-40}$ while $P^W_{b|x_{M_0}} = 0.02$. Practically, it means that $NP_b \approx N\Pr^W P^W_{b|x_{M_0}}$. Hence, it is indicative that for $\Pr^S = 0.7$, 70% of all bits in the fingerprint of length N are communicated almost errorless and only a fraction of 30% has a non-negligible error that leads to bit flipping that should be verified at the decoding. Obviously, one can set up such a threshold $(1 - \Pr^S)$ to minimize the number of bits to be verified at the decoding. However, such a reduction in turn increases the probability $P^S_{b|x_{M_0}}$ and thus leads to the probability that the strong bits might contain some fraction of errors.

At the same time, it is interesting to point out that the above selection splits the total achievable rate $R^{2\text{Ch}}_{S|x_{M_0}} = 0.93$ bits/sample into the corresponding strong and weak rates according to (29), i.e., $R^S_{S|x_{M_0}} = 0.7$ bits/sample and $R^W_{S|x_{M_0}} = 0.23$ bits/sample as shown in Figure 9e. It means that the rate 0.7 bits/sample can be achieved without any additional "decoding" while the rate 0.23 bits/sample should be achieved via the matching of the observed bits in the weak positions with the corresponding fingerprint in the database.

Finally, it is important to emphasize the impact of the operational SNR regime on the channel splitting and decoding. The above considered example concerned the high SNR regime. In the low SNR regime of $10dB$, as for example shown in Figures 9a and 9b, the majority of bits will be affected by the noise and belong to the weak class. In this case, it is known that the advantages of soft decoding diminish and the achievable rate of soft decoding is close to those of hard decoding (Figure 6a). Therefore, the RHSD is expected to be more advantageous for high SNR regimes.

In the above analysis, we have assumed that the positions of weak bits are known with the high probability. In practice, it is the case when either $\mathbf{x}_M$ or its quantized version $\mathbf{x}_{M_0}$ is used for the bit selection and is shared between the encoder and decoder. However, when $\mathbf{y}_M$ is only available at the decoder, the bit selection should be performed with some care since there is a probability that the strong bits close to the index NP_b in the sorted list of magnitudes or magnitude values close to the threshold T_y might be wrongly flipped to the weak bits and vice versa. The extended channel splitting model considered in [24] might be of help to estimate such a probability.

C. Comparison of decoding complexities

To compare the complexity of the considered decoding strategies, we express the complexity in the form $\mathcal{O}(M_s^{\alpha})$, where $\alpha_1 = 1$, $\alpha_2 = \frac{H_2(P_b)}{1-H_2(P_b)}$ and $\alpha_3 = \frac{P_b}{1-H_2(P_b)}$ for the exhaustive search (ES), HSD and RHSD, respectively, with $M_s = 2^{N(1-H_2(P_b))}$ to be the recognizable number of objects for the BSC. The exponents α_1, α_2 and α_3 are shown in Figure 16.

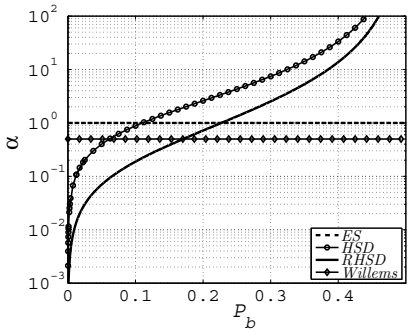


Fig. 16. Complexity exponents for the ES, HSD, RHSD and Willems [25].

As expected, the main gain in the reduction of decoding complexity based on the Hamming sphere decoding strategies can be achieved for small values of P_b when the radius of sphere P_bN is small. The increase of P_b leads to the exponential increase of the number of codewords needed to be checked that exceeds the size of codebook $\mathcal{C}_s$ defined by M_s after certain value. Therefore, to keep the complexity of HSD and RHSD smaller than those of ES, we obtain the conditions for the HSD $H_2(P_b) \leq 1 - H_2(P_b)$ and $P_b \leq H_2^{-1}(0.5) \approx 0.11$ and for the RSH $P_b \leq 1 - H_2(P_b)$ and $P_b \leq 0.22$. These points correspond to the crossings with the "ES" exponent in Figure 16.

It is interesting to point out that the information-theoretic scheme proposed by Willems [25] achieves the complexity $\mathcal{O}(M_s^{0.5})$, i.e., 2 times smaller complexity exponent with respect to the ES for all values of P_b but for the price of memory increase for the storage of an extra indexing structure. The decoding strategies considered in this paper do not require any memory extension and achieve significantly lower complexity reduction for small values of P_b. In fact, another limiting case of the Hamming sphere decoding strategy might even achieve the complexity $\mathcal{O}(1)$ for any P_b. However, in this case the memory storage would include the storage of an extra $2^{NH_2(P_b)}$ codewords for each codeword of the codebook $\mathcal{C}_s$ which might be infeasible in practice. Therefore, we have constrained our consideration only for memory non-increasing strategies.

Finally, to exemplify a practical situation we computed the search complexity as a function of the SNR in the projected domain for $N = 128$ for the ES, HSD and RHSD as $\mathcal{O}(2^{N(1-H_2(P_b))})$, $\mathcal{O}(2^{NH_2(P_b)})$ and $\mathcal{O}(2^{NP_b})$, respectively. The results are shown in Figure 17. It is interesting to point out that the search complexity of ES growths with SNR due to the increase of the maximum number of uniquely distinguishable sequences in the codebook of size M_s caused by the corresponding decrease of the cross-over probability P_b. At the same time, the search complexities of HSD and RHSD decrease with SNR due to the decrease of the cardinality of the Hamming sphere despite the actual increase of the codebook size. As expected, the best performance is achieved by the RHSD. The achievability of these theoretical bounds for different N in physical object microstructure fingerprinting was demonstrated in [26]. Obviously, the efficiency of these methods is expected for the SNR values satisfying the above conditions of HSD and RHSD superiority over the ES for the corresponding $P_b \leq 0.11$ and $P_b \leq 0.22$. As a practical trade-off solution one can choose the ES strategy for the low SNR and RHSD for the high one.

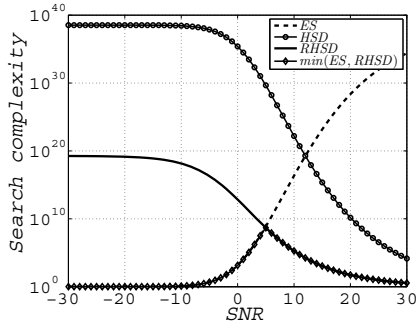


Fig. 17. Search complexity for the ES, HSD and RHSD for $N = 128$.

Therefore, the bit reliability makes it possible to achieve a significant reduction of the search complexity while ensuring the identification rates close to the identification capacity without any memory extensions. Further complexity reduction is possible for the rates close to the capacity at the price of a memory increase to store the indexing structure [27].

VI. Conclusion

In this paper, we consider the sign-magnitude decomposition of mutual information for identification applications. We consider the sign channel and demonstrate the effect of rate concentration under proper channel splitting parameters. We also demonstrate that the cross-over probability in the strong channel might be made arbitrary small under a proper threshold selection. This creates an interesting basis for the development of fast identification algorithms, efficiently resolving the memory-complexity trade-off. In this paper, we considered only a two-channel splitting problem. However, the introduced framework can be easily extended to the multilevel framework. Our preliminary results also indicate that generally 3 levels of decomposition suffice to approach the rate with the unquantized soft information. The developed framework might be also of interest for privacy-preserving applications. In this case, the side information about the bit reliability is used to preserve the strong bits while the weak bits are randomized since they do not carry any useful information. The attacker, who does not possess the information on the bit reliability, is unable to find the strong bits in the fingerprint with the manageable complexity. We plan to compare this approach with several state-of-the-art methods based on helper data and fuzzy commitment. Finally, another interesting extension concerns the possibility to store both the sign components and quantized magnitude components in the codebook and to perform the search jointly. Such a method might also increase the identification rate of the binary sign channel by the rate of magnitude components.

Acknowledgment

The authors are thankful to the former and current SIP group members for very fruitful and stimulating discussions and all anonymous referees for their constructive comments.

APPENDIX A

We consider the first cross-term $I(X_M; Y_S|Y_M)$ in (8), which can be decomposed as:

$$I(X_M; Y_S|Y_M) = H(Y_S|Y_M) - H(Y_S|Y_M, X_M). \quad (42)$$

The first term $H(Y_S|Y_M) \stackrel{(a)}{=} H(Y_S) = H_2(\theta_Y) \stackrel{(b)}{=} H_2(0.5) = 1$, where (a) is due to the independence of $Y_S \perp Y_M$ for the assumed symmetric source and channel distributions, (b) follows from the fact that $\theta_Y = \theta_X * P_b = 0.5$ for $\theta_X = 0.5$ for the above distributions and $H_2(.)$ denotes the binary entropy [20].

The second component $H(Y_S|Y_M, X_M) = H(Y_S|X_M)$ since $Y_S \perp Y_M|X_M$ for the model $X_S \rightarrow Y_S \leftarrow X_M \rightarrow Y_M$. Additionally, $H(Y_S|X_M) = \mathbb{E}_{f(x_M)}[H(Y_S|x_M)] = \mathbb{E}_{f(x_M)}[H_2(\theta_X * P_{b|x_M})] = \mathbb{E}_{f(x_M)}[H_2(0.5)] = H_2(0.5) = 1$ for $\theta_X = 0.5$.

Finally, the first cross-term is $I(X_M; Y_S|Y_M) = 1 - 1 = 0$.

APPENDIX B

In this part, we consider the second cross-term $I(X_S; Y_M|X_M, Y_S)$ in (8), which can be decomposed as:

$$I(X_S; Y_M|X_M, Y_S) = h(Y_M|X_M, Y_S) - h(Y_M|X_M, Y_S, X_S). \quad (43)$$

The first term $h(Y_M|X_M, Y_S) = h(Y_M|X_M)$ since $Y_M \perp Y_S|X_M$ for $Y_M \leftarrow X_M \rightarrow Y_S \leftarrow X_S$. The second term $h(Y_M|X_M, Y_S, X_S) = h(Y_M|X_M)$ since $Y_M \perp (Y_S, X_S)|X_M$. Therefore, $I(X_S; Y_M|X_M, Y_S) = h(Y_M|X_M) - h(Y_M|X_M) = 0$.

APPENDIX C

In this part, we consider the first cross-term $I(X_S; Y_M|X_M)$ in (10), which can be decomposed as:

$$I(X_S; Y_M|X_M) = h(Y_M|X_M) - h(Y_M|X_M, X_S). \quad (44)$$

Since the second term $h(Y_M|X_M, X_S) = h(Y_M|X_M)$ due to $Y_M \perp X_S|X_M$ for the model $Y_M \leftarrow X_M \rightarrow Y_S \leftarrow X_S$, one obtains $I(X_S; Y_M|X_M) = h(Y_M|X_M) - h(Y_M|X_M) = 0$.

REFERENCES

[1] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *Proceedings of IEEE Information Theory Workshop, ITW2010*, Dublin, Ireland, Aug.30-Sep.3 2010.

[2] S. Voloshynovskiy, T. Holotyak, O. Koval, F. Beekhof, and F. Farhadzadeh, "Sign-magnitude decomposition of mutual information with polarization effect in digital identification," in *Proceedings of IEEE Information Theory Workshop, ITW2011*, Paraty, Brazil, October, 16-20 2011.

[3] P. Tuyls, B. Skoric, and T. Kevenaar, *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer-Verlag, 2007.

[4] O. Amrani, Y. Be'ery, A. Vardy, F.-W. Sun, and H. van Tilborg, "The leech lattice and the golay code: bounded-distance decoding and multilevel constructions," *Information Theory, IEEE Transactions on*, vol. 40, no. 4, pp. 1030 –1043, jul 1994.

[5] A. Vardy, "Even more efficient bounded-distance decoding of the hexacode, the golay code, and the leech lattice," *Information Theory, IEEE Transactions on*, vol. 41, no. 5, pp. 1495 –1499, sep 1995.

[6] H. Jégou, M. Douze, and C. Schmid, "Product Quantization for Nearest Neighbor Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 117–128, Jan. 2011.

[7] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system." in *ICM Information Retrieval,*, 2002.

[8] A. B. J. Teoh, Y. W. Kuan, and S. Lee, "Cancellable biometrics and annotations on biohash," *Pattern Recogn.*, vol. 41, no. 6, pp. 2034–2044, Jun. 2008.

[9] J. Fridrich, "Robust bit extraction from images," in *MCS*, vol. 2, july 1999, pp. 536 –540.

[10] P. Indyk and A. Naor, "Nearest-neighbor-preserving embeddings," *ACM Trans. Algorithms*, vol. 3, no. 3, Aug. 2007.

[11] S. Voloshynovskiy, M. Diephuis, F. Beekhof, O. Koval, and B. Keel, "Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos)," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[12] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1652 –1667, oct. 2012.

[13] P. Tuyls, A. H. Akkermans, T. A. Kevenaar, G.-J. Schrijen, A. M. Bazen, and R. N. Veldhuis, "Practical biometric authentication with template protection," in *Audio-and Video-Based Biometric Person Authentication*. Springer, 2005, pp. 436–446.

[14] S. Voloshynovskiy, O. Koval, T. Holotyak, F. Beekhof, and F. Farhadzadeh, "Privacy amplification of content identification based on fingerprint bit reliability," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Seattle, The USA, December 12–15 2010.

[15] M. Jin and C. D. Yoo, "Quantum hashing for multimedia," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 4, pp. 982–994, 2009.

[16] J. Seo, "An asymmetric matching method for a robust binary audio fingerprinting," *Signal Processing Letters, IEEE*, vol. 21, no. 7, pp. 844–847, July 2014.

[17] A. L. Varna and M. Wu, "Modeling and analysis of correlated binary fingerprints for content identification," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1146–1159, 2011.

[18] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3051–3073, 2009.

[19] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *Proceedings of IEEE International Symposium on Information Theory 2003*, Yokohama, Japan, Jun 2003.

[20] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 1991.

[21] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.

[22] M. Davis, G. Logemann, and D. Loveland, "A machine program for theorem-proving," *Commun. ACM*, vol. 5, no. 7, pp. 394–397, Jul. 1962.

[23] F. Beekhof, S. Voloshynovskiy, O. Koval, and T. Holotyak, "Fast identification algorithms for forensic applications," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, London, UK, December 6–9 2009.

[24] V. Balakirsky, S. Voloshynovskiy, O. Koval, and T. Holotyak, "Information theoretic analysis of privacy protection for noisy identification based on soft fingerprinting," in *CSIT 2011*, Yerevan, Armenia, September, 26-30 2011.

[25] F. M. Willems, "Searching methods for biometric identification systems: Fundamental limits," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 2241–2245.

[26] F. Beekhof, "Physical object protection based on digital micro-structure fingerprinting," *PhD Dissertation, University of Geneva*, 2012.

[27] F. Farhadzadeh, F. M. Willems, and S. Voloshynovskiy, "Fundamental limits of identification: Identification rate, search and memory complexity trade–off," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, July 7–12 2013.

Slava Voloshynovskiy (IEEE SM11) received a radio engineer degree from Lviv Polytechnic Institute, Lviv, Ukraine, in 1993 and a Ph.D. degree in electrical engineering from the State University Lvivska Polytechnika, Lviv, Ukraine, in 1996. From 1998 to 1999, he was a visiting scholar with the University of Illinois at Urbana-Champaign. Since 1999, he has been with the University of Geneva, Switzerland, where he is currently an Associate Professor with the Department of Computer Science and head of the Stochastic Information Processing group. His current research interests are in information-theoretic aspects of digital data hiding, content fingerprinting, physical object security, stochastic image modeling and machine learning. He has coauthored over 200 journal and conference papers in these areas and holds ten patents. He served as Associate Editor for IEEE Transactions on Information Forensics and Security (2013-2015). S. Voloshynovskiy was an elected member of the IEEE Information Forensics and Security Technical Committee (2011-2013) where he was area chair in information-theoretic security and an associated member since 2015. He was a general chair of ACM Multimedia Security Conference, 2006 and technical co-chair of Workshop on Information Forensics and Security WIFS15, 2015. He has served as a consultant to private industry in the above areas. He was a recipient of the Swiss National Science Foundation Professorship Grant in 2003.

Taras Holotyak received MSc and PhD degrees in electrical engineering from the Lviv Polytechnic National University, Lviv, Ukraine, in 1997 and 2001. From 2003 to 2005, he was a post-doctoral fellow at the State University of New York at Binghamton, the USA. From 2006 to 2012, he pursued a PhD degree in computer sciences with the Stochastic Information Processing (SIP) Group, Department of Computer Sciences, the University of Geneva. Since 2012, he has been with the SIP group as a post doctoral and senior researcher. His research interests are mostly in privacy preserving identification/authentication protocols and algorithms, information-theoretic aspects of digital watermarking and steganography technologies, stochastic modeling of multimedia and digital data hiding.

Fokko Beekhof is currently a Manager of Analytics at Expedia, Inc., after obtaining a PhD in computer science from the University of Geneva, Switzerland under supervision of S. Voloshynovskiy, which was awarded a Latsis prize in 2013. Earlier, he obtained a Master of Science in parallel and distributed computing from the Technical University of Delft in the Netherlands. His interests are in probability, statistics, signal processing and high-performance computing.