Sign-Magnitude Decomposition of Mutual Information with Polarization Effect in Digital Identification

Svyatoslav Voloshynovskiy, Taras Holotyak, Oleksiy Koval, Fokko Beekhof and Farzad Farhadzadeh
Department of Computer Science, University of Geneva, Switzerland
Emails: {svolos, Taras.Holotyak, Oleksiy.Koval, Fokko.Beekhof, Farzad.Farhadzadeh}@unige.ch

***Abstract*—Content identification based on digital fingerprinting attracts a lot of attention in different emerging applications. In this paper, we consider digital identification based on the sign-magnitude decomposition of fingerprint codewords and analyze the achievable rates for each component. We introduce a channel splitting approach and reveal certain interesting phenomena related to channel polarization. It is demonstrated that under certain conditions almost all rate in the sign channel is concentrated in reliable components, this can be of interest for complexity and security in various content identification applications. The envisioned extensions cover applications where the input and output alphabets of the channel are different at the encoding and decoding stages. Additionally, the reduction of the input data dimensionality at the encoding/enrollment stage can increase the cryptographic protection in terms of privacy leakage and simplify the decoding algorithms in biometric applications.**

I. Introduction

Identification systems are widely used in various emerging applications ranging from identification of physical objects and people to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on digital fingerprints, 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 analog and noisy world [1].

From coding perspectives, the identification problem can be considered as a form of communication with random codewords [2], which can be further implemented based on bounded distance decoding [3]. In contrast to the high attractiveness of binary data representations for memory storage, complexity, security and privacy, an important fraction of information is neglected once the data are binarized. It is demonstrated that soft information extracted from noisy observations can enhance the overall system performance, complexity and privacy [3].

In this paper, we extend this consideration to a general decomposition of real signals into sign and magnitude components in a random projection domain. Then, we investigate the basic decomposition of mutual information between channel input and output into four terms, where two terms are dominating with one degree of freedom for each channel, and two mixed channels that cover the exchange of information across the channels. Such a decomposition makes it possible to use multistage decoding for random codewords as well as to reveal some interesting phenomena linking this consideration with rate concentration or channel polarization, where all rates can be concentrated in several components. 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. This decomposition might reveal some interesting insights into the analysis and design of future identification systems.

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., $\mathbf{x} = \{x(1), x(2), ..., x(N)\}$. 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 $I(\cdot)$ defines pair-wise mutual information.

II. A sign-magnitude decomposition of mutual information

A. General decomposition

Consider a memoryless source $\mathbf{X}$ with some symmetric distribution and a discrete memoryless channel (DMC) $f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} f(y_i|x_i)$ with input $x_i = x_{Si} \cdot x_{Mi}$ and output $y_i = y_{Si} \cdot y_{Mi}$, 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, $x_S \in \mathcal{X}_S$, $y_S \in \mathcal{Y}_S$ ($\mathcal{X}_S, \mathcal{Y}_S = \{-1, +1\}$); $x_M \in \mathcal{X}_M$, $y_M \in \mathcal{Y}_M$ ($\mathcal{X}_M, \mathcal{Y}_M = \mathbb{R}^+$).

Proposition 1 (sign-magnitude decomposition). *The mutual information between channel input and output can be decomposed as shown in Fig. 1 and yields:*

$$\begin{aligned} I(X;Y) &= I(X_S, X_M; Y_S, Y_M) \\ &\overset{(a)}{=} I(X_M; Y_M) + I(X_S; Y_S|X_M) \\ &\overset{(b)}{=} I(X_M; Y_M) + I(X_S; Y_S|Y_M) \\ &+ \underbrace{H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M)}_{(c)}. \end{aligned} \quad (1)$$

Proof: The expansion (a) in (1) follows from the chain rule for mutual information [4] that yields:

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

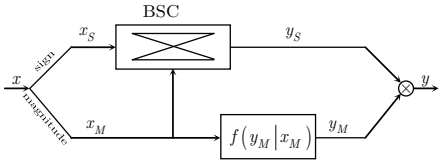


Fig. 1. General sign-magnitude decomposition of DMC $f(y|x)$ into sign $p(y_S|x_S, x_M)$ and magnitude $f(y_M|x_M)$ channels.

In the last decomposition of (2), 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 binary symmetrical channel (BSC) $p(y_S|x_S, x_M)$, whose state is defined by the magnitude component x_M:

$$
\begin{aligned}
I(X_S; Y_S|X_M) &= \int_{\mathcal{X}_M} I(X_S; Y_S|x_M) f(x_M) dx_M \\
&= \mathbb{E}_{f(x_M)}[I(X_S; Y_S|x_M)],
\end{aligned} \tag{3}
$$

where $\mathbb{E}_{f(x_M)}[\cdot]$ denotes the expectation with respect to the random variable $X_M \sim f(x_M)$. To proceed with the analysis of the cross-terms, it should be noted that the sign and magnitude components of input and output are independent, i.e., $X_S \perp X_M$ and $Y_S \perp Y_M$. Then, due to the independence of these components the cross-term 1: $I(X_M; Y_S|Y_M) = H(Y_S|Y_M) - H(Y_S|X_M, Y_M) = 0$. For the cross-term 2: $I(X_S; Y_M|X_M, Y_S) = H(X_S|X_M, Y_S) - H(X_S|X_M, Y_S, Y_M) = H(X_S|X_M, Y_S) - H(X_S|X_M, Y_S) = 0$.

The expansion (b) in (1) follows from the chain rule decomposition:

$$
\begin{aligned}
&I(X_S, X_M; Y_S, Y_M) \\
&= I(X_S, X_M; Y_M) + I(X_S, X_M; Y_S|Y_M) \\
&= \underbrace{I(X_M; Y_M)}_{\text{magnitude term, } R_M} + \underbrace{I(X_S; Y_S|Y_M)}_{\text{sign term, conditioned on } Y_M,\ R_{S|y_M}} \\
&+ \underbrace{I(X_S; Y_M|X_M)}_{\langle \text{cross-term 1} \rangle = 0} + \underbrace{I(X_M; Y_S|X_S, Y_M)}_{\langle \text{cross-term 2} \rangle}.
\end{aligned} \tag{4}
$$

The magnitude term coincides with one in (2), 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:

$$
\begin{aligned}
I(X_M; Y_S|X_S, Y_M) &= H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M, Y_M) \\
&= H(Y_S|X_S, Y_M) - H(Y_S|X_S, X_M).
\end{aligned} \tag{5}
$$

Finally, $I(X_S; Y_M|X_M) = H(X_S|X_M) - H(X_S|X_M, Y_M) = 0$. ■

Remark 1. Under the proper knowledge of BSC state x_M, the mutual information $I(X;Y)$ can be decomposed and achieved by independent processing of the magnitude and sign channels. Alternatively, one can apply multistage decoding (MSD) by first decoding the magnitude x_M and then the sign. If for some technical or security reasons x_M is unavailable and the channel state x_M is approximated by y_M, one observes a rate loss proportional to the entropy of the mismatch between the exact channel state and its approximation.

Remark 2. The decomposition (1) can be very helpful to understand "soft" fingerprinting schemes, where only binary information is stored, but soft information about bit reliabilities is extracted from the noisy magnitudes [3].

In some applications such as biometrics and content identification, where the identification is based on random codebooks, the mutual information between channel input and output defines the identification capacity $C_{id} = I(X;Y)$ [2]. The difference with the design of data transmission codebook consists in the absence of the maximization with respect to the input distribution $f(x)$. In the following Sections we will consider the sign-magnitude decomposition of mutual information $I(X;Y)$ and analyze some interesting dependencies for Gaussian input X and AWGN channel.

B. Decomposition of AWGN identification channel

In this section we apply decomposition (1) to the Gaussian input X and AWGN channel. Besides the great importance of the Gaussian case, it can be demonstrated that any i.i.d. pdf $f(x)$ or correlated signals following a Gauss-Markov model of the first order can be transformed into an approximately i.i.d. Gaussian signal of lower dimensionality using random projections [5]. Therefore, under these conditions we assume that $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y = X + Z$, where $Z \sim \mathcal{N}(0, \sigma_Z^2)$.

1) Magnitude term: The first term of both decompositions (2) and (4) 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 information. The conditional term can be rewritten as:

$$
h(Y_M|X_M) = h(Z) - H(Y_S|X_S, X_M) \tag{6}
$$

that follows from the decomposition:

$$
\begin{aligned}
h(Z) = h(Y|X) &= h(Y_S, Y_M|X_S, X_M) \\
&= h(Y_M|X_S, X_M) + H(Y_S|X_S, X_M, Y_M) \\
&= h(Y_M|X_M) + H(Y_S|X_S, X_M)
\end{aligned} \tag{7}
$$

with differential entropy $h(Z) = 1/2 \log_2(2\pi e \sigma_Z^2)$. 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 that can be developed as:

$$
\begin{aligned}
H(Y_S|X_S, X_M) &= \int_{\mathcal{X}_M} H(Y_S|X_S, x_M) f(x_M) dx_M \\
&= \mathbb{E}_{f(x_M)}[H(Y_S|X_S, x_M)] \\
&= \mathbb{E}_{f(x_M)}[H_2(Pr[Y_S \neq X_S|x_M])] \\
&= \mathbb{E}_{f(x_M)}[H_2(P_{b|x_M})],
\end{aligned} \tag{8}
$$

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

$$\begin{aligned}
P_{b|x_M} =& \Pr[Y_S \neq X_S | x_M] \\
=& \Pr[Y_S = -1 | X_S = +1, x_M] \Pr[X_S = +1] \\
&+ \Pr[Y_S = +1 | X_S = -1, x_M] \Pr[X_S = -1] \\
=& \Pr[Y_S = -1 | X_S = +1, x_M] \\
=& \int_{-\infty}^{0} p(y|x_M) dy \\
=& \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left[-\frac{(y - x_M)^2}{2\sigma_Z^2} \right] dy \\
=& Q\left(\frac{x_M}{\sigma_Z} \right). \qquad (9)
\end{aligned}$$

Substituting (9) into (8) yields:

$$H(Y_S|X_S, X_M) = \int_{\mathcal{X}_M} H_2 \left[Q\left(\frac{x_M}{\sigma_Z} \right) \right] f(x_M) dx_M. \quad (10)$$

Alternatively, one can find the conditional entropy (6) as:

$$\begin{aligned}
& h(Y_M|X_M) \\
&= -\int_{\mathcal{X}_M} \int_{\mathcal{Y}_M} f(y_M, x_M) \log_2 \left(f(y_M|x_M) \right) dy_M dx_M, \qquad (11)
\end{aligned}$$

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 folded normal distribution.

Remark 3. The conditional entropy $h(Y_M|X_M)$ converges to the entropy of $h(Z)$ for the regime of small degradations, when the term $H(Y_S|X_S, X_M)$ tends to zero in (6). For high degradations, the term $H(Y_S|X_S, X_M)$ tends to 1 and $h(Y_M|X_M)$ converges to the entropy of channel noise magnitude $h(Z_M) = h(Z) - 1$. 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 Fig. 2. The conditional entropy $h(Y_M|X_M)$ was computed according to (6) and (11) by numerical integration.

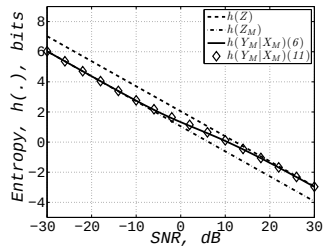


Fig. 2. Behavior of conditional entropy $h(Y_M|X_M)$ with respect to $h(Z)$ and $h(Z_M)$.

2) Sign term: In the decomposition (1), there are two sign terms conditioned on X_M and Y_M. The sign term in the decomposition (1(a)) is:

$$\begin{aligned}
I(X_S; Y_S|X_M) =& H(Y_S|X_M) - H(Y_S|X_S, X_M) \\
=& H(Y_S) - H(Y_S|X_S, X_M). \qquad (12)
\end{aligned}$$

The entropy $H(Y_S) = 1$ due to the symmetry of the Gaussian pdf of Y; the second term is given by (10). Thus, (12) yields:

$$\begin{aligned}
R_{S|x_M} =& I(X_S; Y_S|X_M) \\
=& 1 - \int_{\mathcal{X}_M} H_2 \left[Q\left(\frac{x_M}{\sigma_Z} \right) \right] f(x_M) dx_M. \qquad (13)
\end{aligned}$$

The sign term in the decomposition (1(b)) is:

$$\begin{aligned}
I(X_S; Y_S|Y_M) =& H(Y_S|Y_M) - H(Y_S|X_S, Y_M) \\
=& 1 - H(Y_S|X_S, Y_M). \qquad (14)
\end{aligned}$$

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 (8) can be rewritten as:

$$\begin{aligned}
H(Y_S|X_S, Y_M) =& \mathbb{E}_{f(y_M)}[H(Y_S|X_S, y_M)] \\
=& \mathbb{E}_{f(y_M)}[H_2(\Pr[Y_S \neq X_S|y_M])] \\
=& \mathbb{E}_{f(y_M)}[H_2(P_{b|y_M})], \qquad (15)
\end{aligned}$$

where

$$\begin{aligned}
P_{b|y_M} =& \Pr[Y_S \neq X_S | y_M] \\
\overset{(a)}{=}& \mathbb{E}_{f(x_M|y_M)}[\Pr[Y_S \neq X_S|x_M]] \\
\overset{(b)}{=}& \mathbb{E}_{f(x_M|y_M)} \left[Q\left(\frac{x_M}{\sigma_Z} \right) \right] \\
=& \int_{\mathcal{X}_M} Q\left(\frac{x_M}{\sigma_Z} \right) f(x_M|y_M) dx_M, \qquad (16)
\end{aligned}$$

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) follow from (9). Substituting (16) into (15) yields:

$$\begin{aligned}
& R_{S|y_M} = 1 - H(Y_S|X_S, Y_M) = \\
& 1 - \int_{\mathcal{Y}_M} H_2 \left[\int_{\mathcal{X}_M} Q\left(\frac{x_M}{\sigma_Z} \right) f(x_M|y_M) dx_M \right] f(y_M) dy_M. \qquad (17)
\end{aligned}$$

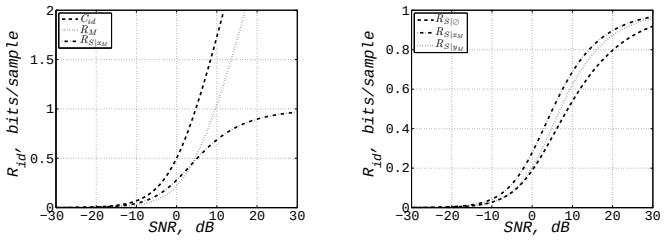
Remark 4. In the case of no channel state information (CSI) about the state x_M of channel $p(y_S|x_S, x_M)$, the rate is:

$$R_{S|\oslash} = I(X_S; Y_S|\oslash) = 1 - H_2(P_b), \qquad (18)$$

where $P_b = \mathbb{E}_{f(x_M)}[P_{b|x_M}]$ corresponds to the average probability of bit error.

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 denoted as $C_{id} = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_X^2}{\sigma_Z^2} \right)$ in Fig.3a and $R_{S|x_M}$, $R_{S|y_M}$ and $R_{S|\oslash}$ in Fig.3b.

Remark 5 *(Sign-magnitude decomposition of the mutual information for AWGN).* The decomposition of C_{id} 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



(a) (b)

Fig. 3. 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, partial side information based on y_M and zero side information about the channel $p(y_S|x_S, x_M)$.

use that exactly corresponds to the difference between C_{id} and R_M.

Remark 6. The impact of side information about the BSC state x_M:

- 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 $R_{S|y_M}$ approaches rate $R_{S|x_M}$ at high SNR that indicates that the decoding in the sign channel can be performed independently from the magnitude channel;
- blind decoding without any information about x_M, often used in digital fingerprinting, represents about 0.15 bit loss with respect to $R_{S|x_M}$ ($SNR \approx 10dB$).

III. Channel splitting and polarization

In this Section we introduce a practical model that achieves the above theoretical limits based on *channel splitting*. Along this way, we will demonstrate the effect of rate concentration in a few coefficients that we will refer to as *channel polarization*. The channel splitting and polarization will be demonstrated for the sign channel.

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, 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 of the magnitude coefficients with a threshold T. Equivalently, the same proportion of coefficients can be chosen based on sorting of the N magnitude coefficients of $\mathbf{x}_M$ or $\mathbf{y}_M$ and selecting the L largest ones. The L bits related to the large magnitude coefficients are considered as those belonging to the *strong* BSC with the cross-over probability P_b^S and the remaining to the *weak* one characterized by P_b^W.

Remark 7. The crossover probabilities for strong and weak channels based on the perfect CSI (given $\mathbf{x}_M$) are:

$$P^S_{b|x_M} = \frac{1}{\Pr^S}\int_{T_{\rm x}}^{+\infty} P_{b|x_M} f(x_M)dx_M, \tag{19}$$

$$P^W_{b|x_M} = \frac{1}{\Pr^W}\int_{0}^{T_{\rm x}} P_{b|x_M} f(x_M)dx_M, \tag{20}$$

where $\Pr^S = \int_{T_{\rm x}}^{+\infty} f(x_M)dx_M$ and $\Pr^W = \int_0^{T_{\rm x}} f(x_M)dx_M$ correspond to probabilities of observing the strong and weak channels, respectively. The corresponding identification rates are:

$$R^S_{S|x_M} = \Pr^S\left[1 - H_2\left(P^S_{b|x_M}\right)\right], \tag{21}$$

$$R^W_{S|x_M} = \Pr^W\left[1 - H_2\left(P^W_{b|x_M}\right)\right], \tag{22}$$

and the total rate is:

$$R^{\rm 2Ch}_{S|x_M} = R^S_{S|x_M} + R^W_{S|x_M}. \tag{23}$$

Remark 8. The crossover probabilities for strong and weak channels based on the degraded CSI (given $\mathbf{y}_M$) are:

$$P^S_{b|y_M} = \frac{1}{\Pr^S}\int_{T_{\rm y}}^{+\infty} P_{b|y_M} f(y_M)dy_M, \tag{24}$$

$$P^W_{b|y_M} = \frac{1}{\Pr^W}\int_{0}^{T_{\rm y}} P_{b|y_M} f(y_M)dy_M, \tag{25}$$

where $T_{\rm y}$ is selected to satisfy above $\Pr^S = \int_{T_{\rm y}}^{+\infty} f(y_M)dy_M$ and $\Pr^W = \int_0^{T_{\rm y}} f(y_M)dy_M$. The corresponding identification rates are:

$$R^S_{S|y_M} = \Pr^S\left[1 - H_2\left(P^S_{b|y_M}\right)\right], \tag{26}$$

$$R^W_{S|y_M} = \Pr^W\left[1 - H_2\left(P^W_{b|y_M}\right)\right], \tag{27}$$

and the total rate yields:

$$R^{\rm 2Ch}_{S|y_M} = R^S_{S|y_M} + R^W_{S|y_M}. \tag{28}$$

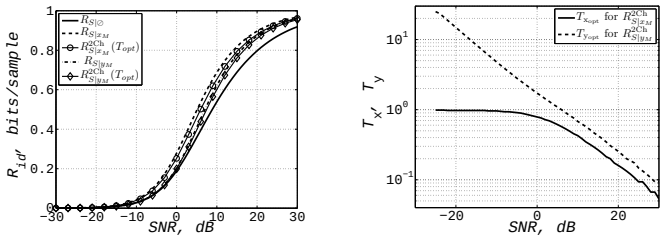
Remark 9. The total crossover probability P_b remains the same as for the case of no CSI:

$$P_b = \int_{\mathcal{X}_M} P_{b|x_M} f(x_M)dx_M = \Pr^S P^S_{b|x_M} + \Pr^W P^W_{b|x_M}. \tag{29}$$

The channel splitting by the selection of the threshold T can be performed according to the several strategies:

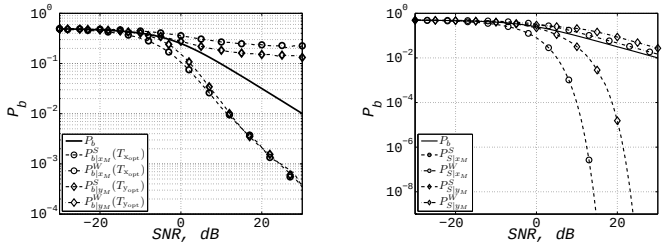
- **Strategy 1**: maximize the total rates $R^{\rm 2Ch}_{S|x_M}$ or $R^{\rm 2Ch}_{S|y_M}$ to approach upper theoretical limits $R_{S|x_M}$ or $R_{S|y_M}$, respectively, that gives optimal values of thresholds $T_{\mathbf{x}_{\rm opt}}$ and $T_{\mathbf{y}_{\rm opt}}$ for each SNR;
- **Strategy 2**: minimize probabilities $P^S_{b|x_M}$ or $P^S_{b|y_M}$ for complexity reasons and privacy amplification [3].

According to strategy 1, the binary channel splitting approaches theoretical performance limits under the optimal threshold selection as shown in Fig. 4a. The remaining gap is easily compensated by more accurate models using more channels in splitting. In fact, two-channel splitting can be considered as the first level decomposition of $x_M = (x_{M_1}, x_{M_2}, ..., x_{M_k})$ according to the multilevel coding framework [6]. The optimal thresholds $T_{\mathbf{x}_{\rm opt}}$ and $T_{\mathbf{y}_{\rm opt}}$ are shown in Fig. 4b. To exemplify strategy 2 we show in Fig. 5a the pairs of crossover probabilities for strong and weak channels under perfect and degraded CSI that resulted from the strategy 1. Fig. 5b shows the same pairs for the fixed thresholds $T_{\rm x}$ and $T_{\rm y}$, where one can clearly observe the significant reduction of $P^S_{b|x_M}$ or $P^S_{b|y_M}$ that asymptotically goes to zero for $SNR > 15$ dB.



(a) (b)

Fig. 4. Approaching theoretical rates based on 2-channel splitting model for the optimal threshold selection: (a) achievable identification rates under different CSIs; (b) optimal thresholds for perfect and degraded CSI.



(a) (b)

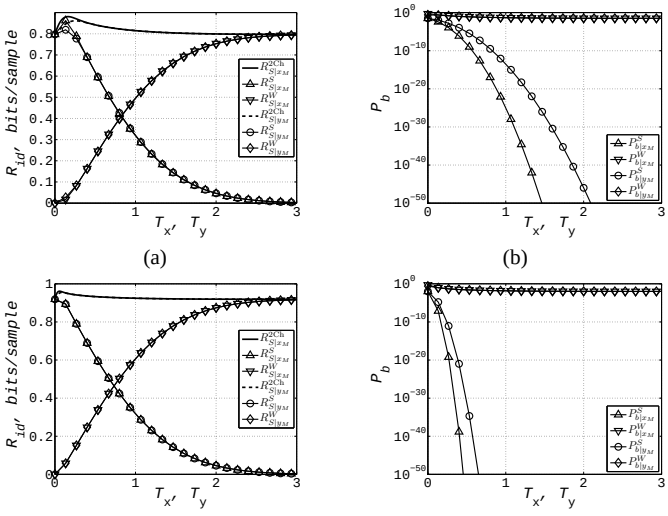
Fig. 5. Probabilities of bit errors: (a) for $T_{\mathrm{x_{opt}}}$ and $T_{\mathrm{y_{opt}}}$; (b) for $T_{\mathrm{x}} = 1$ and corresponding T_{y}.

Remark 10 *(Channel polarization).* An interesting phenomenon is observed for $T_{\mathrm{x_{opt}}}$ and $T_{\mathrm{y_{opt}}}$, when practically all rate is concentrated in the strong channel, i.e., $R_{S|x_M} \approx R^S_{S|x_M}$ and $R_{S|y_M} \approx R^S_{S|y_M}$ after a certain SNR. It means that weak channels can be disregarded from the consideration. Moreover, the positions of the bits belonging to the strong channels can be reliably estimated. We will refer to the effect of rate concentration in strong channels as $channel\ polarization$.

The difference to polar codes, where the relationship between bits are created in such a way that conditional entropy is polarized to either 0 or 1, i.e., "weak" or "strong" bits, should be also noticed. 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 $T > T_{\mathrm{opt}}$ (Fig. 5,b), when the cross-over probability in strong channel is asymptotically equal to 0. Thus, those bits can be directly considered for identification without decoding. However, an unavoidable price for this option is rate loss in the strong channel. Multichannel splitting with multistage decoding can partially resolve this rate-complexity trade-off, that is out of the scope of this paper. The polarization effect is demonstrated in Fig. 6 as the dependence of achievable rates and crossover probabilities on the threshold T for $SNR = 20$ and $30dB$.

The described phenomena can be of interest for:

- design of new search algorithms, when the representation of original content is reduced to the vector of signs of length N, and the decoder searches for the match of the L most reliable components determined based on the noisy observations;
- 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 the extension to [3].



(a) (b) (c) (d)

Fig. 6. Achievable rates and crossover probabilities for channel splitting at $SNR = 20dB$ (a,b) and $SNR = 30dB$ (c,d).

IV. Conclusion

In this paper, we consider 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. This toy model can be of interest for the design of soft multistage decoding algorithms that can trade-off performance and complexity.

Acknowledgment

This paper was partially supported by SNF project 200020-134595. The authors are thankful to V. Balakirsky for his feedback on the early version of this paper.

References

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

[2] 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.

[3] 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.

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

[5] F. Farhadzadeh, S. Voloshynovskiy, O. Koval, T. Holotyak, and F. Beekhof, "Statistical analysis of digital fingerprinting based on random projections," submitted to IEEE 7th International Symposium on Image and Signal Processing and Analysis, ISPA 2011.

[6] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.