Lossy Compression with Privacy Constraints: Optimality of Polar Codes

Farshid Mokhtarinezhad, Jörg Kliewer, Osvaldo Simeone

Helen and John C. Hartmann Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, New Jersey 07102-1982
Email: {fm86, jkliewer, simeone}@njit.edu

***Abstract*—A lossy source coding problem with privacy constraint is studied in which two correlated discrete sources X and Y are compressed into a reconstruction $\hat{X}$ with some prescribed distortion D. In addition, a privacy constraint is specified as the equivocation between the lossy reconstruction $\hat{X}$ and Y. This models the situation where a certain amount of source information from one user is provided as utility (given by the fidelity of its reconstruction) to another user or the public, while some other correlated part of the source information Y must be kept private. In this work, we show that polar codes are able, possibly with the aid of time sharing, to achieve any point in the optimal rate-distortion-equivocation region identified by Yamamoto, thus providing a constructive scheme that obtains the optimal tradeoff between utility and privacy in this framework.**

I. Introduction

An important consequence of the ubiquitous growth of modern information technology is that an increasing amount of private information is shared between different organizations and/or users. This entails a tension between privacy and utility in the sense that disclosing data provides useful information to the receiving entity, while at the same time posing the danger of leaking private information. Examples for such a tension can be found in many real-life systems, e.g., in social networks, smart grids, or databases.

The tradeoff between utility and privacy has been the subject of several recent works as surveyed in [1]. A simple information-theoretic model to analyze this tradeoff is the lossy source coding problem introduced by [2], where utility is measured by the reconstruction fidelity and privacy by an equivocation (i.e., conditional entropy). Reference [2] shows that (vector) quantization, as realized by means of random coding, is optimal in the sense that is achieves any point in the rate-distortion-equivocation region. Several subsequent works [3–7] have addressed related problems in which the introduction of distortion is used to disguise private information. For example, [5] focuses on database privacy in which only certain entries of a database are to be published. Further, the authors in [7] generalize the result in [2] to the case with side information at the decoder. While all these works consider achievability based on random coding, here we focus on the general setup in [2] and provide a constructive coding scheme based on polar codes which achieves the optimal rate-utility-privacy trade-off.

This work was supported in part by NSF grant CCF-1439465.

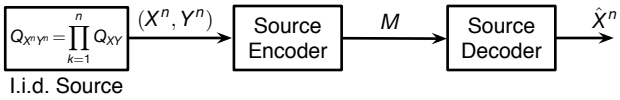

Fig. 1. Illustration of the problem of lossy source coding with privacy constraints [2], in which the privacy is measured by the leakage $H(Y^n|M)/n$ and the utility by the fidelity as gauged with respect to the expected distortion $\mathbb{E}(d(\hat{X}^n, X^n))/n$.

Polar codes, as first proposed in [8], are binary block codes which achieve the capacity of a binary symmetric memoryless channel with efficient encoding and decoding algorithms. The key property of these codes is that they yield virtual channels which either asymptotically converge to an error-free or a completely noisy channel, such that the fraction of asymptotically error-free channels approaches the symmetric capacity of original channel. Polar codes have been generalized to both asymmetric channels [9], [10] and arbitrary alphabets [11]. Moreover, polar codes have been shown to achieve the rate-distortion bound for symmetric binary sources [12] and asymmetric binary sources under Hamming distortion in [9], [13].

In the following, we show that, for the framework in [2] under the assumption of prime source alphabets, polar codes are able, possibly with the aid of time sharing, to achieve any point in the optimal rate-distortion-equivocation region. To the best of our knowledge, this is the first constructive scheme that is provably optimal in terms of the achievable tradeoff between rate, utility, and privacy.

Notation: An upper case letter A denotes a random variable and a denotes its realization. We let A^i denote the random vector $(A_1, ..., A_i)$. For any set $\mathcal{S}$, $|\mathcal{S}|$ denotes its cardinality and $A_{\mathcal{S}}$ denotes the vector $(A_{i_1}, ..., A_{i_{|\mathcal{S}|}})$.

II. System Model and Preliminary Results

We consider the lossy source coding set-up studied in [2] and depicted in Fig. 1, in which the encoder wishes to communicate a source sequence X^n within some distortion to the decoder, while keeping the receiver's knowledge about a correlated sequence Y^n, also available to the encoder, below some tolerated level. The sources $X^n \in \mathcal{X}^n$ and $Y^n \in \mathcal{Y}^n$

978-1-4673-7852-9/15/$31.00 ©2015 IEEE

take values in discrete alphabets $\mathcal{X}$ and $\mathcal{Y}$, and are memoryless with joint distribution $Q_{X^nY^n}(x^n, y^n) = \prod_{i=1}^n Q_{XY}(x_i, y_i)$ for some joint pmf $Q_{XY}(x, y)$. Encoding of these pairs leads to an index $M = m$ with $m \in \{1, 2, \ldots, \lfloor q^{nR} \rfloor\}$, where n is the blocklength and q is a prime number. Finally, the reconstruction of X^n at the decoder is given by a sequence $\hat{X}^n \in \hat{\mathcal{X}}^n$, with $\hat{\mathcal{X}} = \{0, 1, ..., q-1\}$. The goal in designing the system in Fig. 1 is to obtain a desired tradeoff between the rate R, the expected distortion $\mathbb{E}(d(\hat{X}^n, X^n))/n$, and the information leakage $H(Y^n|M)/n$ about the source Y^n that can be obtained from observing M. For simplicity, in the following we will identify pmfs by their arguments only and drop any subscripts.

We now define the operation of both encoder and decoder and the notion of the rate-distortion-equivocation region. To this end, we introduce a standard bounded distortion metric $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{max}]$, where $d_{max} < \infty$ is the maximal distortion.

Definition 1 (Code)**.** An (n, R, D, Δ) code consists of an encoding function that maps each sequence $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ to an index $m\,(x^n, y^n) \in \{1, 2, \ldots, \lfloor q^{nR} \rfloor\}$ and a decoding function that maps each index m to an estimate $\hat{x}^n\,(m) \in \hat{\mathcal{X}}^n$, such that the average distortion $\frac{1}{n}\mathbb{E}(d(X^n, \hat{X}^n)) \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{E}(d(X_i, \hat{X}_i))$ satisfies the inequality

$$\frac{1}{n}\mathbb{E}(d(X^n, \hat{X}^n)) \leq D, \tag{1}$$

and the equivocation rate guarantees the inequality[1]

$$\frac{1}{n}H(Y^n|M) \geq \Delta. \tag{2}$$

Definition 2 (Rate-distortion-equivocation region)**.** A triple (R, D, Δ) is said to be achievable, if, for any $\epsilon > 0$ and n sufficiently large, there exists an $(n, R, D + \epsilon, \Delta - \epsilon)$ code. The closure of all achievable triples $\mathcal{R}^*$ is referred to as the *rate-distortion-equivocation region*.

Remark 1. The distortion D can be constrained without loss of generality to lie in the interval $[0, d_{\max}]$, while the equivocation Δ may range in the interval $[H(Y|X), H(Y)]$.

A. Preliminaries

Lemma 1 ([2])**.** *The rate-distortion-equivocation region $\mathcal{R}^*$ is given by the closure of the union of all tuples (R, D, Δ) such that the inequalities*

$$R \leq I(XY; \hat{X}) \tag{3a}$$
$$D \leq \mathbb{E}(d(X, \hat{X})), \tag{3b}$$
$$\Delta \geq H(Y|\hat{X}), \tag{3c}$$

hold for some pmf $P\,(x, y, \hat{x})$ that satisfies

$$\sum_{\hat{x} \in \hat{\mathcal{X}}} P\,(x, y, \hat{x}) = Q\,(x, y)\,, \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{4}$$

[1]All the entropies will be computed with base q logarithms and all summations are done modulo q.

Remark 2. From a pmf $P(x, y, \hat{x})$, the test channel

$$W(x, y|\hat{x}) = \frac{P(x, y, \hat{x})}{\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} P(x, y, \hat{x})} \tag{5}$$

can be calculated. In [2], two specific binary examples are worked out, namely a source in which the correlation between the binary variables X and Y is a Z-channel, and a doubly symmetric binary source, both under Hamming distortion. From the results in [2], it can be inferred that test channels (5) that yield boundary points on the rate-distortion-equivocation region for the former case are generally asymmetric, while for the latter they can be assumed to be symmetric with no loss of optimality. We recall that a channel $W(x, y|\hat{x})$ is said to be symmetric if there exists a permutation $\pi(x, y)$ of the output alphabet $\mathcal{X} \times \mathcal{Y}$ such that, the identity $\pi(x, y) = \pi^{-1}(x, y)$ holds and the equality $W(x, y|1) = W(\pi(x, y)|0)$ is satisfied for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

III. Optimality of Polar Codes

Let us define as $P(x, y, \hat{x})$ a pmf that achieves an operating point of interest in the rate-distortion-equivocation region $\mathcal{R}^*$ in Lemma 1. Let us also define as $R^* = I(XY; \hat{X}), D^* = \mathbb{E}(d(X, \hat{X}))$, and $\Delta^* = H(Y|\hat{X})$ the rate, distortion and equivocation attained under such distribution $P(x, y, \hat{x})$, respectively. In this section, we demonstrate that polar codes can achieve any such triple (R^*, D^*, Δ^*) in $\mathcal{R}^*$. As mentioned, we focus in the following on the case of a prime size alphabet $\hat{\mathcal{X}} = \{0, 1, ..., q-1\}$, although extensions to alphabets of arbitrary cardinality are possible by following [11].

A. Lossy source coding via polar codes

We consider a polar coding scheme that is a variant of the approach proposed in [9] for asymmetric sources, which is in turn inspired by [10], [13], and extended to prime alphabets by applying results of [11]. We fix a joint distribution $P(x, y, \hat{x})$ that achieves a desired point (R^*, D^*, Δ^*) in $\mathcal{R}^*$. To start, let us define the following joint distribution on the set $\mathcal{X}^n \times \mathcal{Y}^n \times \hat{\mathcal{X}}^n \times \mathcal{U}^n$ where $\mathcal{U} = \{0, 1, ..., q-1\}$:

$$P(x^n, y^n, \hat{x}^n, u^n) = \prod_{i=1}^n Q(x_i, y_i) P(\hat{x}_i|x_i, y_i) \mathbf{1}\{u^n = \hat{x}^n G_n\}, \tag{6}$$

with $n = 2^k$ for some integer k, $G_n = G^{\otimes k}$ is the polarizing transform with $G = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $G^{\otimes k}$ denotes the k-times Kronecker power, and $P(\hat{x}|x, y) = P(x, y, \hat{x}) \diagup Q(x, y)$. The distribution (6) can be interpreted as providing the target joint distribution over variables $(X^n, Y^n, \hat{X}^n)$ since, under (6), it is easy to see that the desired distortion D^* and equivocation Δ^* are attained (see [2]). The challenge is to construct a coding scheme that mimics (6) without having to transmit a message u^n of n symbols and hence of rate $R = 1$ from encoder to decoder. Note that the matrix G_n satisfies $G_n = G_n^{-1}$ and hence, from u^n, one can recover $\hat{x}^n$ as $\hat{x}^n = G_n u^n$ [14].

As explained in the following, the encoder maps the sources (x^n, y^n) into a vector u^n, which is divided into two subvectors,

namely the information vector $u_{\mathcal{I}}$, indexed by the set $\mathcal{I}$ of size $|\mathcal{I}| = nR$ symbols and the complementary vector $u_{\mathcal{I}^c}$. The information vector $u_{\mathcal{I}}$ constitutes the message M sent by the encoder to the decoder. We partition the set $\mathcal{I}^c$ into two sets, namely, the set $\mathcal{F}$ that identifies the "frozen" symbols $u_{\mathcal{F}}$ and the set $\mathcal{D}$ that identifies the "computable" symbols $u_{\mathcal{D}}$. These sets are defined as

$$\mathcal{F} \triangleq \left\{ i \in [1:n] : Z(U_i|U^{i-1}, X^n, Y^n) \geq 1 - 2^{-n^\beta} \right\} \tag{7}$$

$$\text{and } \mathcal{D} \triangleq \left\{ i \in [1:n] : Z(U_i|U^{i-1}) \leq 2^{-n^\beta} \right\}, \tag{8}$$

where $\beta < \frac{1}{2}$ is a parameter of the underlying polar coding scheme. Further, the source Bhattacharyya parameter Z for two random variables $A \in \{0, 1, \ldots, q-1\}$ and $B \in \mathcal{B}$ is defined as

$$Z(A|B) \triangleq \frac{1}{q-1} \sum_{\substack{a,a' \in \mathcal{A}: \\ a \neq a'}} \sum_{b \in \mathcal{B}} \sqrt{P_{A,B}(a,b) P_{A,B}(a',b)}. \tag{9}$$

The Bhattacharyya parameters in (7) and (8) are calculated based on the joint distribution $P(x^n, y^n, \hat{x}^n, u^n)$ given in (6). From [9, Theorem 1] and [11, Theorem 4.3], the size nR of the set $\mathcal{I} = \{1, 2, \ldots, n\} \setminus (\mathcal{F} \cup \mathcal{D})$ is such that the rate R is arbitrarily close to R^* as n grows large.

To determine the vector u^n, the following randomized successive encoding rule is used for $i = 1, 2, \ldots, n$:

$$u_i = \begin{cases} u_i \in \mathcal{U} & \text{with probability } P(u_i|u^{i-1}, x^n, y^n) \text{ if } i \in \mathcal{I}, \\ u_i \in \mathcal{U} & \text{with probability } P(u_i|u^{i-1}) \text{ if } i \in \mathcal{D}, \end{cases} \tag{10}$$

where the probabilities in (10) are obtained from (6). The symbols $u_{\mathcal{F}}$ are predetermined and are available at the decoder prior to encoding. The vector $u_{\mathcal{I}}$ is sent to the decoder, while the decoder obtains the vector $u_{\mathcal{D}}$ according to a maximum likelihood rule as in [10], [13]:

$$\hat{u}_i = \begin{cases} u_i & \text{for } i \in \mathcal{I}, \\ f_i(\hat{u}^{i-1}) \triangleq \arg\max_{u \in \mathcal{U}} P(u|\hat{u}^{i-1}) & \text{for } i \in \mathcal{D}, \\ u_i & \text{for } i \in \mathcal{F}. \end{cases} \tag{11}$$

Finally, the codeword $\hat{x}^n$ is evaluated as $\hat{x}^n = G_n \hat{u}^n$.

Remark 3. Note that the decoding rule (11) does not require encoder and decoder to share the set of Boolean functions needed by the scheme in [9] (see also [10], [13]), hence significantly simplifying the implementation.

Remark 4. If $\hat{X}^n$ is i.i.d. uniformly distributed in $\hat{\mathcal{X}}^n$ under (6), it follows from [9], [14] that the set $\mathcal{D}$ has negligible size as n grows large and hence the encoding and decoding rules (10) and (11) can be simplified by setting $\mathcal{D} = \emptyset$ as done in [12]. This condition applies, for instance, to the doubly symmetric binary source studied in [2] (see Remark 2). Moreover, the encoding rule (10) with $\mathcal{D} = \emptyset$ entails that the set of codewords $\hat{X}^n$ consists of the (approximately) q^{nR} sequences of a block coset code defined by the generator matrix G_n and by the frozen symbols $u_{\mathcal{F}}$.

B. Optimality of polar codes

In this section, we establish the optimality of polar codes for the problem at hand. We start with the following proposition that entails randomization over the frozen bits. The need for randomization is removed in Proposition 2.[2]

Proposition 1. *Fix a triple* $(R^* = I(XY;\hat{X}), D^* = \mathbb{E}(d(X,\hat{X})), \Delta^* = H(Y|\hat{X}))$ *achieved by a joint distribution* $P(x,y,\hat{x})$ *in the rate-distortion-equivocation region* $\mathcal{R}^*$. *For any* $0 < \beta' < \beta < \frac{1}{2}$, *any* $\epsilon > 0$, *and for sufficiently large* n, *the sequence of rates* $R_n = \frac{1}{n}|\mathcal{I}|$, *distortions* $D_n = \frac{1}{n}\mathbb{E}(d^n(X^n, \hat{X}^n))$, *and equivocations* $\Delta_n = \frac{1}{n}H(Y^n|U_{\mathcal{I}})$ *that satisfy*

$$R_n \leq R^* + \epsilon, \tag{12a}$$

$$D_n \leq D^* + O\big(2^{-n^{\beta'}}\big), \tag{12b}$$

$$\Delta_n \geq \Delta^* - O\big(2^{-n^{\beta'}}\big) \tag{12c}$$

is achievable by the polar coding scheme (10)-(11), where the distortion D_n *and the equivocation* Δ_n *are averaged over uniformly distributed frozen symbols* $u_{\mathcal{F}}$.

Proof: We first define the joint distribution induced by the encoding rule (10) under the assumption that the frozen symbols are selected as i.i.d. uniform variables with probability $\frac{1}{q}$ according to

$$P^e(x^n, y^n, u^n, \hat{x}^n) = Q(x^n, y^n)\, q^{-|\mathcal{F}|} \prod_{i \in \mathcal{D}} P\left(u_i|u^{i-1}\right) \cdot \prod_{i \in \mathcal{I}} P\left(u_i|u^{i-1}, x^n, y^n\right) \cdot \mathbf{1}\{\hat{x}^n = u^n G_n\}. \tag{13}$$

Note that in (13) the codeword $\hat{X}^n$ is defined based on the symbols U^n selected by the encoder. We also introduce the joint distribution that includes both (10) and the decoding rule in (11) as

$$P^d(x^n, y^n, u^n, \hat{u}^n, \hat{x}^n) = Q(x^n, y^n)\, q^{-|\mathcal{F}|} \prod_{i \in \mathcal{D}} P\left(u_i|u^{i-1}\right) \cdot \prod_{i \in \mathcal{I}} P\left(u_i|u^{i-1}, x^n, y^n\right) \cdot \prod_{i \in \mathcal{F} \cup \mathcal{I}} \mathbf{1}\{\hat{u}_i = u_i\} \cdot \prod_{i \in \mathcal{D}} \mathbf{1}\{\hat{u}_i = f_i(\hat{u}^{i-1})\} \cdot \mathbf{1}\{\hat{x}^n = \hat{u}^n G_n\}. \tag{14}$$

The rate condition (12a) follows directly by extension of the arguments in [9, Theorem 1] to alphabets of prime size and holds for any choice of the frozen vectors. To prove (12b) for the ensemble of codes inducing the joint distribution by (14), we need to modify the arguments in [9] in order to account for possible decoding errors. To this end, we define the probability of error as $P_e \triangleq \Pr_{P^e}[\hat{U}^n \neq U^n]$.[3] Denoting

[2]The notation $f(n) = O(g(n))$ means that there exist constants n_0 and c such that for all integers $n > n_0$ the inequality $|f(n)| \leq c\,|g(n)|$ holds.

[3]In the following, subscripts are used to identify the distribution with respect to which probabilities, expectations, and information measures are computed.

the decoding error event as $E \triangleq \{\hat{U}^n \neq U^n\}$, the distortion $D_n(u_{\mathcal{F}})$ averaged over the frozen vectors $u_{\mathcal{F}}$ satisfies

$$\begin{aligned}\mathbb{E}_{P^d}[D_n(U_{\mathcal{F}})] &= \frac{1}{n}\Big((1-P_e)\mathbb{E}_{P^d}[d(X^n,\hat{X^n}) \mid E^c] \\ &\quad + P_e\mathbb{E}_{P^d}[d(X^n,\hat{X^n})|E]\Big) \\ &\leq \frac{1}{n}\Big((1-P_e)\mathbb{E}_{P^d}[d(X^n,\hat{X^n})|E^c] \\ &\quad + P_e d_{\max}\Big), \end{aligned} \tag{15}$$

by the law of total probability and the boundedness of the distortion metric. Moreover, we have

$$P^d\left(x^n,y^n,u^n,\hat{u}^n,\hat{x}^n|E^c\right) = \frac{P^e\left(x^n,y^n,u^n,\hat{x}^n\right)\mathbf{1}\{u^n=\hat{u}^n\}}{1-P_e}, \tag{16}$$

and hence the first term in (15) can be computed as

$$\begin{aligned}&\mathbb{E}_{P^d}\big[d(X^n,\hat{X^n})|E^c\big] \\ &= \frac{1}{1-P_e}\cdot\sum_{x^n,y^n,u^n,\hat{u}^n,\hat{x}^n} P^e\left(x^n,y^n,u^n,\hat{x}^n\right) \\ &\qquad\qquad \cdot\mathbf{1}\{u^n=\hat{u}^n\}d(x^n,\hat{x}^n) \\ &= \frac{1}{1-P_e}\mathbb{E}_{P^e}\big[d(X^n,\hat{X^n})\big]. \end{aligned} \tag{17}$$

Furthermore, by [15, Property 2] we have the inequality

$$\begin{aligned}\frac{1}{n}\mathbb{E}_{P^e}\big[d(X^n,\hat{X^n})\big] \leq \frac{1}{n}\mathbb{E}_P\big[d(X^n,\hat{X^n})\big] + \frac{d_{\max}}{n} \\ \cdot\big\|P_{X^n,Y^n,U^n,\hat{X}^n} - P^e_{X^n,Y^n,U^n,\hat{X}^n}\big\|, \end{aligned} \tag{18}$$

where $\|\cdot\|$ denotes the variational distance of two distributions, and, by construction, $\frac{1}{n}\mathbb{E}_P\big[d(X^n,\hat{X^n})\big] = D^*$ holds true. The variational distance in (18) can be characterized by following similar steps as in [9], [13] as (see the Appendix for a sketch)

$$\big\|P_{X^n,Y^n,\hat{X}^n,U^n} - P^e_{X^n,Y^n,\hat{X}^n,U^n}\big\| = O(2^{-n^{\beta'}}) \tag{19}$$

for any $\beta' < \beta$. Finally, we obtain the following bound on the probability of decoding error

$$P_e \overset{(a)}{\leq} \sum_{i\in\mathcal{D}} Z(U_i|U^{i-1}) \overset{(b)}{\leq} |\mathcal{D}|\,2^{-n^\beta} \overset{(c)}{=} O(n2^{-n^\beta}), \tag{20}$$

where (a) follows from [8, Proposition 2], (b) is a consequence of the definition of the set $\mathcal{D}$ in (8), and (c) follows by noting that the cardinality of the set $\mathcal{D}$ is at most linear in n. Using (15), along with (17)-(20), we have

$$\begin{aligned}\mathbb{E}_{P^d}[D_n(U_{\mathcal{F}})] &\leq \frac{1}{n}\big(\mathbb{E}_{P^e}[d(X^n,\hat{X^n})] + P_e d_{\max}\big), \\ &\leq \frac{1}{n}\big[nD^* + d_{\max}(P_e \\ &\quad + \|\,P_{X^n,Y^n,\hat{X}^n,U^n} - P^e_{X^n,Y^n,\hat{X}^n,U^n}\,\|)\big], \\ &= D^* + O(2^{-n^{\beta'}}), \end{aligned} \tag{21}$$

which allows us to conclude that the distortion inequality (12b) is satisfied on average over the choice of the frozen vectors.

To prove (12c), we first observe that, by construction, we have $\frac{1}{n}H_P(Y^n|\hat{X}^n) = \Delta^*$. The achievable average equivocation satisfies the equality

$$\begin{aligned}\mathbb{E}_{P^d}[\Delta_n(U_{\mathcal{F}})] &= \frac{1}{n}H_{P^d}(Y^n|U_{\mathcal{I}},U_{\mathcal{F}}) \\ &= \frac{1}{n}H_{P^e}(Y^n|U_{\mathcal{I}},U_{\mathcal{F}}) = \mathbb{E}_{P^e}[\Delta_n(U_{\mathcal{F}})], \end{aligned} \tag{22}$$

and further we have

$$\begin{aligned}\mathbb{E}_{P^e}[\Delta_n(U_{\mathcal{F}})] &= \frac{1}{n}H_{P^e}(Y^n|U_{\mathcal{I}},U_{\mathcal{F}}) \\ &\geq \frac{1}{n}H_{P^e}(Y^n|U^n) = \frac{1}{n}H_{P^e}(Y^n|\hat{X}^n), \end{aligned} \tag{23}$$

where the inequality in (23) holds since conditioning reduces entropy, and the subsequent equality holds due to the one-to-one correspondence between $\hat{X}^n$ and U^n under P^e, respectively.

Using both the chain rule and the triangle inequality, we obtain

$$\begin{aligned}\big|H_P(Y^n|\hat{X}^n) - H_{P^e}(Y^n|\hat{X}^n)\big| \leq \\ \big|H_P(Y^n,\hat{X}^n) - H_{P^e}(Y^n,\hat{X}^n)\big| \\ + \big|H_P(\hat{X}^n) - H_{P^e}(\hat{X}^n)\big|. \end{aligned} \tag{24}$$

Now, by considering

$$\begin{aligned}\|P_X - Q_X\| &= \sum_x\Big|\sum_y P(x,y) - Q(x,y)\Big| \\ &\leq \sum_{x,y}|P(x,y) - Q(x,y)| = \|P_{X,Y} - Q_{X,Y}\| \end{aligned} \tag{25}$$

and by applying [16, Lemma 2.7] with (19) and (25), we finally obtain the bound

$$\big|H_P(Y^n|\hat{X}^n) - H_{P^e}(Y^n|\hat{X}^n)\big| \leq O(n^{\beta'}2^{n^{\beta'}}).$$

This shows that (12c) is satisfied on average over the choice of the frozen vectors. ∎

We now show that averaging over all frozen vectors is not required to achieve the region (R^*, D^*, Δ^*).

Proposition 2. *Any tuple* (R^*, D^*, Δ^*) *in* (12a)*,* (12b)*, and* (12c) *is achievable by time sharing between at most two polar coding schemes defined by* (10) *and* (11) *with different sequences of frozen symbols* $u_{\mathcal{F}}$*.*

Proof: We prove this statement by contradiction. To elaborate, if a sequence of frozen vectors $u_{\mathcal{F}}$ exists such that for any fixed $\epsilon > 0$ both conditions (12b) and (12c) are satisfied, namely $D_n(u_{\mathcal{F}}) \leq D^* + \epsilon$ and $\Delta_n(u_{\mathcal{F}}) \geq \Delta^* - \epsilon$, then the proof is complete. Now, we assume that none of the vectors $u_{\mathcal{F}}$ satisfies both conditions. By the discussion above, we can find a sufficiently large n_0 such that $\mathbb{E}_{P^d}[D_n(U_{\mathcal{F}})] \leq D^* + \epsilon$ and $\mathbb{E}_{P^d}[\Delta_n(U_{\mathcal{F}})] \geq \Delta^* - \epsilon$ for all $n \geq n_0$. Consider a coordinate system with origin at $(D^* + \epsilon, \Delta^* - \epsilon)$ in the distortion-equivocation plane (see Fig. 2). By assumption, for none of the vectors $u_{\mathcal{F}}$ the point

$(D_n(u_{\mathcal{F}}), \Delta_n(u_{\mathcal{F}}))$ is in the second (upper left) quadrant, while the average $(\mathbb{E}_{P^d}[D_n(U_{\mathcal{F}})], \mathbb{E}_{P^d}[\Delta_n(U_{\mathcal{F}})])$ lies in the second quadrant. Moreover, the average is in the convex hull of the points $(D_n(u_{\mathcal{F}}), \Delta_n(u_{\mathcal{F}}))$, which is a polytope. By simple geometric arguments, one of the edges of this polytope must cross the second quadrant. Therefore, if the vertices of this crossing edge are denoted as $(D_n(u_{\mathcal{F}1}), \Delta_n(u_{\mathcal{F}1}))$ and $(D_n(u_{\mathcal{F}2}), \Delta_n(u_{\mathcal{F}2}))$, then we can find $0 \leq \alpha \leq 1$ such that $D^\dagger = \alpha D_n(u_{\mathcal{F}1}) + (1-\alpha) D_n(u_{\mathcal{F}2})$, and $\Delta^\dagger = \alpha \Delta_n(u_{\mathcal{F}1}) + (1-\alpha)\Delta_n(u_{\mathcal{F}2})$, and $(D^\dagger, \Delta^\dagger)$ lies in the second quadrant, hence completing the proof. ∎

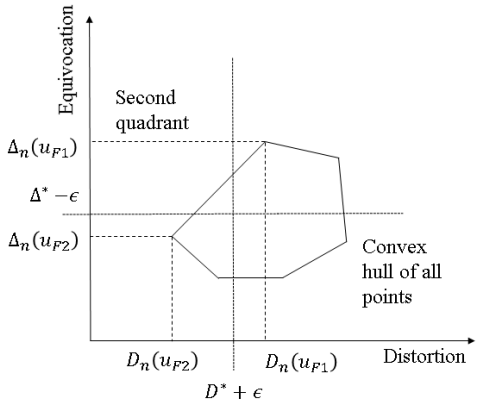

Fig. 2. Convex hull of points in the equivocation-distortion plane.

Remark 5. For the important case of the doubly symmetric source and Hamming distortion, time sharing is not necessary. Hence, there exists a single polar coding scheme defined by (10) and (11) with a specific choice for the sequence of frozen bits $u_{\mathcal{F}}$ (and $\mathcal{D} = \emptyset$, see Remark 4) that achieves the desired point (R^*, D^*, Δ^*). This can be seen from the fact that for each vector $u_{\mathcal{F}}$ we have $\mathbb{E}_{P^d}[D_n(U_{\mathcal{F}})] = D_n(u_{\mathcal{F}})$ [12]. Now, since we know that $\mathbb{E}_{P^d}[\Delta_n(U_{\mathcal{F}})] \geq \Delta^* - \epsilon$, there must be at least one frozen vector $u_{\mathcal{F}}$ such that $\Delta_n(u_{\mathcal{F}}) \geq \Delta^* - \epsilon$, which completes the proof.

REFERENCES

[1] L. Sankar, W. Trappe, K. Ramchandran, H. V. Poor, and M. Debbah, "The role of signal processing in meeting privacy challenges: An overview," *IEEE Signal Processing Magazine*, vol. 30, pp. 95–1096, Sep. 2013.

[2] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretapper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, Nov. 1983.

[3] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *ACM Sigmod Record*, vol. 29, no. 2, pp. 439–450, Jun. 2000.

[4] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Lecture Notes in Computer Science*. Springer, 2006, vol. 3876, pp. 265–284.

[5] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics and Security*, vol. 8, no. 6, pp. 838–852, Jun. 2014.

[6] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *Proc. IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 502–506.

[7] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Lossy source coding with reconstruction privacy," in *Proc. IEEE Int. Symposium on Inform. Theory*, Honululu, HI, Jul. 2014, pp. 386–390.

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

[9] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[10] M. Mondelli, S. H. Hassani, I. Sason, and R. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," [Online]. Available at http://arxiv.org/abs/1401.6060, Jan. 2014.

[11] E. Sasoglu, "Polarization and polar codes," in *Foundations and Trends in Communications and Information Theory*. NOW Publishers, 2012, vol. 8, no. 4, pp. 259–381.

[12] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[13] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," [Online]. Available at http://arxiv.org/abs/1410.3542, Oct. 2014.

[14] E. Arikan, "Source polarization," in *Proc. IEEE Int. Symposium on Inform. Theory*, Austin, TX, Jun. 2010, pp. 899–903.

[15] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," [Online]. Available at http://arxiv.org/abs/1305.3905, May 2013.

[16] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[17] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," [Online]. Available at http://arxiv.org/abs/1301.3905, Jan. 2013.

APPENDIX

Proof sketch of (19):

$$
\begin{aligned}
&\left\| P_{X^n,Y^n,U^n,\hat{X}^n} - P^e_{X^n,Y^n,U^n,\hat{X}^n} \right\| \\
&= \sum_{u^n,x^n,y^n} |P(u^n,x^n,y^n) - P^e(u^n,x^n,y^n)| \\
&\overset{(a)}{=} \sum_{u^n,x^n,y^n} \Bigg| \sum_{i=1}^{n} [P(u_i|u^{i-1},x^n,y^n) - P^e(u_i|u^{i-1},x^n,y^n)] \cdot \\
&\quad Q(x^n,y^n) \prod_{j=1}^{i-1} P(u_j|u^{j-1},x^n,y^n) \prod_{j'=i+1}^{n} P^e(u_{j'}|u^{j'-1},x^n,y^n) \Bigg| \\
&\overset{(b)}{\leq} \sum_{i\in\mathcal{F}} \sum_{u^i,x^n,y^n} \left| P(u_i|u^{i-1},x^n,y^n) - P^e(u_i|u^{i-1},x^n,y^n) \right| \cdot \\
&\quad Q(x^n,y^n) \prod_{j=1}^{i-1} P(u_j|u^{j-1},x^n,y^n) \\
&= \sum_{i\in\mathcal{F}} \sum_{u^{i-1},x^n,y^n} P(u^{i-1},x^n,y^n) \left\| P_{U_i|u^{i-1},x^n,y^n} - P^e_{U_i|u^{i-1},x^n,y^n} \right\| \\
&\overset{(c)}{\leq} \sum_{i\in\mathcal{F}} \sum_{u^{i-1},x^n,y^n} P(u^{i-1},x^n,y^n) \cdot \\
&\quad \sqrt{(2\ln 2) D(P_{U_i|u^{i-1},x^n,y^n} \,\|\, P^e_{U_i|u^{i-1},x^n,y^n})} \\
&= \sum_{i\in\mathcal{F}} \sqrt{(2\ln 2) D(P_{U_i} \,\|\, P^e_{U_i}|U^{i-1},X^n,Y^n)} \\
&\overset{(d)}{=} \sum_{i\in\mathcal{F}} \sqrt{(2\ln 2)(1 - H_P(U_i|U^{i-1},X^n,Y^n))} \\
&\overset{(e)}{\leq} \sum_{i\in\mathcal{F}} \sqrt{(2\ln 2)(1 - (Z(U_i|U^{i-1},X^n,Y^n))^2)} \\
&\overset{(f)}{\leq} n\sqrt{(4\ln 2) 2^{-n^\beta}} = O(2^{-n^{\beta'}})
\end{aligned}
$$

Here, the equalities and inequalities follow from (a) a telescopic expansion, (b) the fact that the distributions P and P^e are the same for $i \notin \mathcal{F}$, (c) Pinsker's inequality where $D(\cdot\|\cdot)$ is the relative entropy, (d) [17, Lemma 10], (e) [11, Proposition 4.8], (f) (7).