Identification and Lossy Reconstruction in Noisy Databases

Ertem Tuncel and Deniz Gündüz

***Abstract*—A high-dimensional database system is studied where the noisy versions of the underlying feature vectors are observed in both the enrollment and the query phases. The noisy observations are compressed before being stored in the database, and the user wishes to both identify the correct entry corresponding to the noisy query vector and reconstruct the original feature vector within a desired distortion level. A fundamental capacity-storage-distortion tradeoff is identified for this system in the form of single-letter information theoretic expressions. The relation of this problem to the classical Wyner-Ziv rate-distortion problem is shown, where the noisy query vector acts as the correlated side information available only in the lossy reconstruction of the feature vector.**

***Index Terms*—High dimensional databases, identification systems, Wyner-Ziv coding.**

I. Introduction

High-dimensional data, e.g., biometric features such as fingerprints and iris scans, or behavioral patterns such as gait and keystrokes, are replacing classical identification documents for increased security. However, efficient use of such data for sensitive security applications requires building a large database and fast search algorithms for reliable identification of the entries in the database. On top of the storage constraints and search speed requirements, another difficulty arises due to the noisiness of the observation of features in both the enrollment and the identification stages. This might be either due to the random noise in the scanning device as in the case of fingerprinting or iris scanning, or due to the temporal changes in the expression of the underlying feature as in the case of behavioral patterns such as keystrokes.

The first attempts in understanding the fundamental performance limits of retrieval from high-dimensional databases were made in [4] and [10], which characterize the *identification capacity*, i.e., the maximum exponential rate of entries that can be reliably identified in a database. In their model, which all the subsequent work (including this paper) is based upon, the data management system operates in two phases:

1) *Enrollment phase:* Noisy vectors $Y^n(m)$, $m = 1, 2, \ldots, M$, are observed and recorded in the database. It is assumed that the underlying feature vectors $X^n(m)$ are independent and identically distributed (i.i.d.) with a known distribution P_X, and pass through a memoryless channel $P_{Y|X}$ to produce $Y^n(m)$.

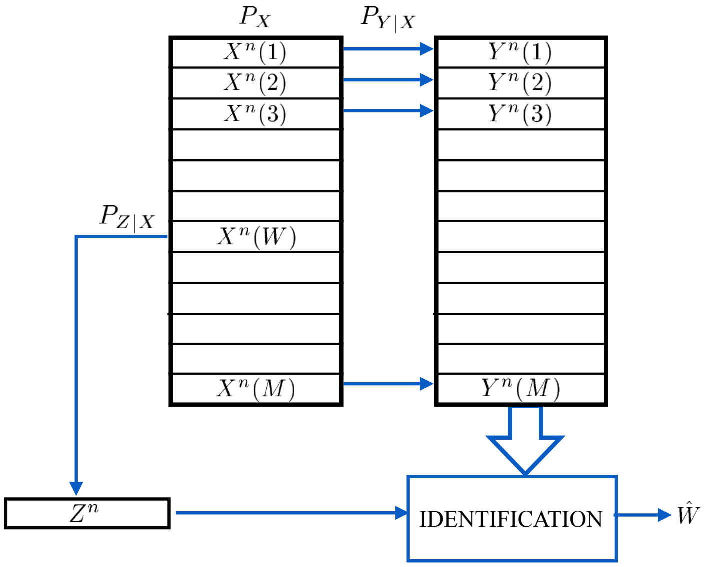

Fig. 1. The block diagram of the data management system of [4] and [10].

2) *Identification phase:* Nature chooses W uniformly from $1, 2, \ldots, M$, and the corresponding $X^n(W)$ passes through another memoryless channel $P_{Z|X}$, producing the query vector Z^n. The goal is then to identify W with high probability by using only the query Z^n and the enrolled noisy vectors $Y^n(1), \ldots, Y^n(M)$.

Figure 1 shows these two stages in one diagram. It was shown in [4] and [10] that for large n, $M \approx 2^{nR^i}$ objects can be reliably identified if and only if $R^i < C$, where C has a single-letter characterization given by

$$C = I(Y; Z) .$$

To reduce the storage space (thereby speeding up the identification process), it may be desirable to store only a compressed version of the observed feature vectors rather than the whole noisy observation, as shown in Figure 3. Obviously, this enhancement is not free, but rather comes at a cost of reduced identification capability. That is because some feature vectors distinguishable in the uncompressed database scenario would now be mapped to the same quantization index J, and cannot be disambiguated. In other words, the compression at the enrollment stage introduces a tradeoff between the identification capacity and the compression rate. This trade-

Ertem Tuncel is with the Department of Electrical Engineering, University of California, Riverside, CA 92521. E-mail: ertem.tuncel@ucr.edu.

Deniz Gündüz is with the Department of Electrical and Electronic Engineering, Imperial College London, London, UK. E-mail: d.gunduz@imperial.ac.uk.

This work was in part presented in IEEE International Symposium on Information Theory, Austin, TX, June 2010.

off was independently characterized in [9] and [6][1]: A compression/identification rate pair (R^c, R^i) is achievable if and only if there exists an auxiliary random variable U such that $Z - X - Y - U$ forms a Markov chain and

$$\begin{aligned} I(Y;U) &\leq R^c \\ I(Z;U) &\geq R^i , \end{aligned}$$

where U is distributed over some discrete alphabet $\mathcal{U}$ satisfying $|\mathcal{U}| \leq |\mathcal{Y}| + 1$. Equivalently, $M \approx 2^{nR^i}$ objects are reliably identified if and only if $R^i < C(R^c)$, where $C(R^c)$ is the storage-constrained capacity of the system given by

$$C(R^c) = \max_{\substack{Z - X - Y - U \\ I(Y;U) \leq R^c}} I(Z;U) .$$

In this work, we consider another dimension of the problem. Suppose that in the identification stage, we require not only a reliable identification of the index W, but also a lossy reconstruction of the underlying feature vector $X^n(W)$. In a sense, this general problem combines the capacity-storage tradeoff problem studied in [9] and [6] with the classical Wyner-Ziv rate-distortion problem in [11]. That is because the noisy query vector serves as correlated side information available only at the identification/reconstruction stage.

Interestingly, the behavior of optimal schemes significantly differ in the two problems we combine. More specifically, while *binning* is an essential component of optimal Wyner-Ziv coding, it is not utilized at all in order to achieve the optimal capacity-storage tradeoff in biometric databases. So, how can we unify the two coding approaches and obtain a general scheme that always achieves the capacity-storage-distortion tradeoff?

The answer lies in understanding what binning brings about in the capacity-storage tradeoff for a fixed reconstruction distortion: While it reduces the compression rate, say by ΔR, it compromises the identification rate also by ΔR. That is because having access to only the bin indices of the entries in the identification phase, the system has no choice but to check the query vector Z^n against every possible codevector in the encoded bin for a "match" (which comes in the form of joint typicality). Since there are $\approx 2^{n\Delta R}$ randomly created codevectors in each bin, that increases the likelihood of a mismatch by the same factor. Using this and the fact that the (R^c, R^i) tradeoff is always concave with a slope less than unity[2], one can observe that if we start at the boundary of the achievable capacity-storage region and perform binning, we always land on the interior of that region. See Figure 2 for an illustration of this observation on a typical capacity-storage tradeoff region in a high-dimensional data management system. This, in turn, explains why binning need not (and must not) be used for achieving optimal capacity-storage tradeoff. On the other hand, when a distortion constraint is imposed, compression rate cannot be reduced further than a certain rate without recourse to binning, implying that binning is necessary to characterize the entire tradeoff.

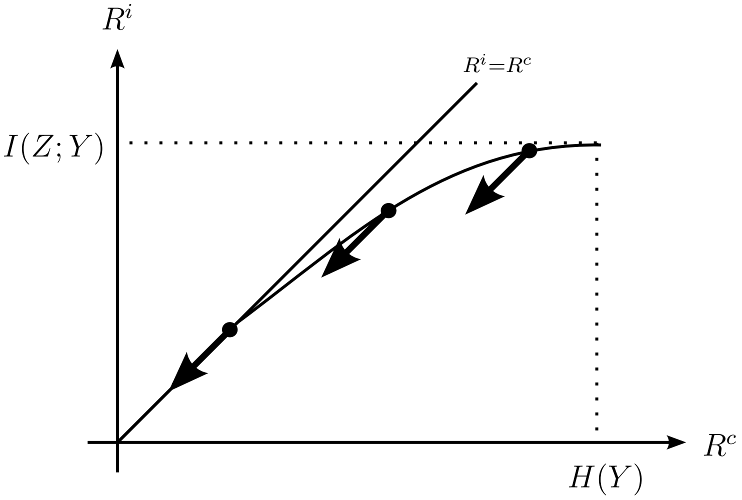

Fig. 2. Illustration of a typical capacity-storage tradeoff in a noisy database system. Here R^c denotes that compression rate while R^i denotes the identification rate, whose values are bounded by the entropy of the noisy enrollment distribution, and the mutual information between the noisy query and enrollment distributions, respectively. Since the optimal tradeoff is concave with slope < 1, the point $(R^c - \Delta R, R^i - \Delta R)$ always lands on the interior of the achievability region.

In light of these observations, we derive a single-letter information theoretic expression for the set of achievable capacity-storage-distortion triplets. We also compute the tradeoff for two examples. Although these examples are simple, they are instrumental in understanding the behavior of optimal codes with respect to binning.

The rest of the paper is organized as follows. We introduce the system model and the necessary definitions in Section II. The main result of the paper is presented in Section III, and Sections IV and V are dedicated to its proof. In Section VI we study a binary symmetric feature vector and identify the capacity-storage-distortion tradeoff assuming noiseless observation in the enrollment phase and an erasure channel in the query phase. Section VII concludes the paper.

II. System Model

Our system model is depicted in Figure 3. We assume that the feature vectors $\{X^n(m)\}_{m=1}^M$ are generated independently with the identical distribution of

$$P[X^n(m) = x^n] = \prod_{i=1}^n P_X(x_i)$$

over the finite feature alphabet $\mathcal{X}$.

The database is formed by an enrollment phase, in which the noisy versions of the feature vectors of individuals are observed and recorded to the database. We denote the observed noisy feature vector of individual m by $Y^n(m)$, $m \in \mathcal{M} = \{1, \ldots, M\}$, which are assumed to be the output of a discrete memoryless

[1] We note here that, in [9], in addition to the enrolled vectors, noisy queries are also subject to finite-rate compression, which adds another dimension to the tradeoff.

[2] The only exception is when there is no noise, in which case $R^i = R^c$ is the tradeoff.

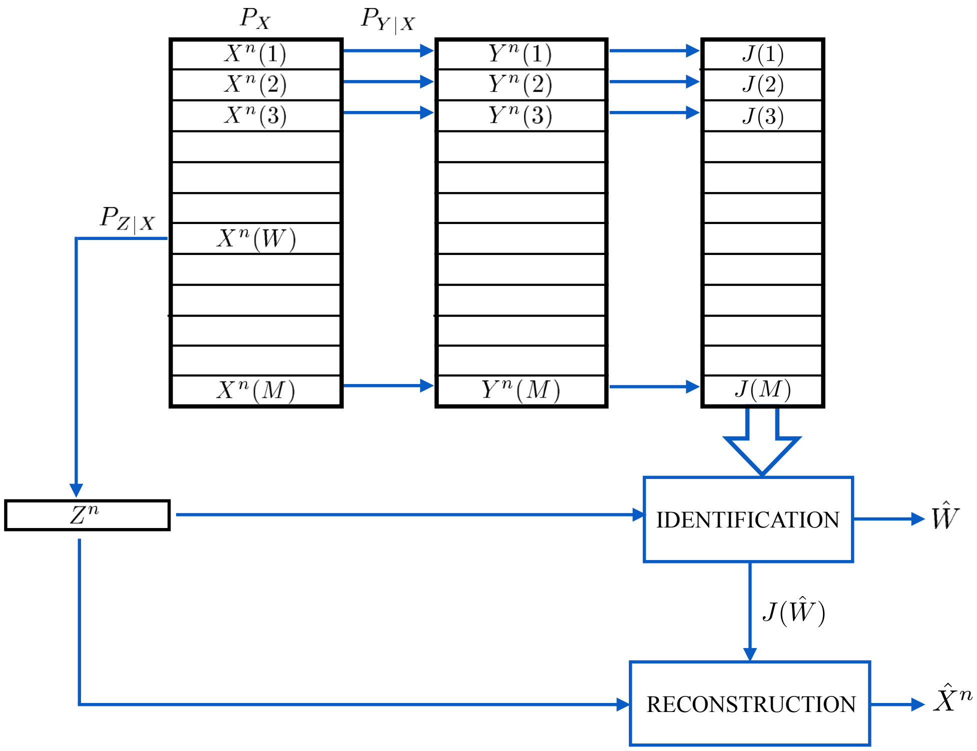

Fig. 3. The diagram of the proposed data management system with the added reconstruction phase.

channel (DMC) characterized by $P_{Y|X}$, where $\mathcal{Y}$ is the finite observation alphabet. We have

$$P[Y^n(m) = y^n | X^n(m) = x^n] = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$$

for $m \in \mathcal{M}$.

In the enrollment phase, each entry is compressed before it is recorded to the database, and only the compressed descriptions of the observed feature vectors are stored in the database. We consider a deterministic compression function

$$f : \mathcal{Y}^n \to \mathcal{L} = \{1, \ldots, L\} ,$$

where $\mathcal{L}$ denotes the index set for the compressed observation vectors. We denote the index for entry $m \in \mathcal{M}$ as $J(m) = f(Y^n(m))$, and define $\mathbf{J} \triangleq J(1), \ldots, J(M)$. These indices refer to length-n codewords from the compression codebook of size L.

In the identification phase, an index W is chosen uniformly from $\mathcal{M}$ and is independent of the database entries. The user of the database observes $X^n(W)$ through a memoryless channel characterized by $P_{Z|X}$ with finite output alphabet $\mathcal{Z}$, i.e.,

$$P[Z^n = z^n | X^n(W) = x^n] = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i) , \qquad (1)$$

where Z^n is the output of the channel. Note that due to the independence of the noisy observation channels in the enrollment and the identification phases, $Y^n(W) - X^n(W) - Z^n$ forms a Markov chain.

The user has two goals. The first goal is to identify W in the database by using the noisy observation vector Z^n and the entries of the database $\{J(m)\}_{m=1}^{M}$. In addition, he/she also wants to reconstruct an estimate of the original feature vector $X^n(W)$ within a desired average distortion requirement.

We define two separate functions for the identification and the reconstruction processes. The identification function is defined as

$$g : \mathcal{L}^M \times \mathcal{Z}^n \to \mathcal{M}$$

and the corresponding estimate is denoted by

$$\hat{W} = g(\mathbf{J}, Z^n) .$$

The average error probability in the identification process is defined as

$$P_e^n \triangleq \frac{1}{M} \sum_{w \in \mathcal{M}} \Pr[\hat{W} \neq W | W = w] .$$

The lossy reconstruction function is defined as

$$h : \mathcal{L} \times \mathcal{Z}^n \to \hat{\mathcal{X}}^n ,$$

where $\hat{\mathcal{X}}$ is the finite reconstruction alphabet. The corresponding reconstruction is denoted by

$$\hat{X}^n = h(J(\hat{W}), Z^n)$$

and the distortion it incurs is measured by the single-letter measure

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) ,$$

where $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{\max}]$. Though the reconstruction function outputs a legitimate $\hat{X}^n$ even when $\hat{W} \neq W$, we are only interested in upper-bounding the distortion conditioned on $\hat{W} = W$.

Definition 1: (R^c, R^i, D) is an *achievable* compression rate, identification rate, and distortion tuple if, for any $\epsilon > 0$ and sufficiently large n, there exist a deterministic enrollment function f and deterministic identification and reconstruction functions g and h, respectively, such that

$$\frac{1}{n} \log L \leq R^c \tag{2}$$

$$\frac{1}{n} \log M \geq R^i \tag{3}$$

while

$$P_e^n \leq \epsilon \tag{4}$$

$$E[d(X^n(W), \hat{X}^n) | \hat{W} = W] \leq D + \epsilon \; . \tag{5}$$

We also denote by $\mathcal{R}$ the set of all achievable (R^c, R^i, D) triplets.

Remark 1: We would like to point out that the compression rate here, similar to the data compression rate in source coding, quantifies the amount of data that needs to be stored *per* database entry, rather than trying to reduce the total amount of stored data for the whole database. Compression reduces the size of the data that needs to be stored for each entry from $n \log |\mathcal{X}|$ to nR^c, ignoring the sublinear terms. On the other hand, the total amount of stored data depends also on the number of entries M. In the case of a discrete feature alphabet, $\mathcal{X}$, compression reduces the total amount of stored data by a factor of $\frac{\log |\mathcal{X}|}{R^c}$, while the same reduction can be obtained by enrolling $\frac{R^c}{\log |\mathcal{X}|} M$ entries, rather than M, which would still correspond to the same identification rate. However, compression will still be useful when the identification rate is fixed and the user has no control on the number of individuals enrolled in the database, and the memory space is allocated for the maximum number of users corresponding to the specified identification rate. Moreover, in the case of continuous feature alphabets, compression is the only way to store the feature vectors in a finite memory.

III. Capacity-Storage-Distortion Tradeoff

The main result of the paper is stated in the following theorem.

Theorem 1: Define $\mathcal{R}^*$ as the region of triplets (R^c, R^i, D) for which there exist an auxiliary random variable $U \in \mathcal{U}$ with joint distribution p_{UYXZ} and a function $\phi : \mathcal{U} \times \mathcal{Z} \to \hat{\mathcal{X}}$ such that $U - Y - X - Z$ forms a Markov chain and

$$\begin{aligned} R^i &\leq I(U; Z) \\ R^c - R^i &\geq I(U; Y|Z) \\ D &\geq E[d(X, \phi(U, Z))] \; . \end{aligned}$$

Then $\mathcal{R} = \mathcal{R}^*$.

Remark 2: Using arguments that have become folklore, it is straightforward to show that $\mathcal{R}^*$ is convex and it suffices to consider auxiliary alphabets $\mathcal{U}$ with $|\mathcal{U}| \leq |\mathcal{Y}| + 2$. Comparing this cardinality bound with the one in Lemma 1 of [6], we observe that the reconstruction requirement of the original feature vector, which introduces the average distortion constraint in Theorem 1, increases the cardinality bound on the auxiliary random variable by one.

If there is no reconstruction requirement, we obtain the following capacity-storage tradeoff by letting $D = d_{\max}$.

Corollary 1: A compression-identification rate pair (R^c, R^i) is achievable if and only if there exist a random variable $U \in \mathcal{U}$ with joint distribution p_{UYXZ} such that $U - Y - X - Z$ forms a Markov chain and

$$R^i \leq I(U; Z) \tag{6}$$

$$R^c - R^i \geq I(U; Y|Z) \; , \tag{7}$$

where $|\mathcal{U}| \leq |\mathcal{Y}| + 1$.

The equivalence of (6) and (7) to

$$R^i \leq I(U; Z) \tag{8}$$

$$R^c \geq I(U; Y) \; , \tag{9}$$

which characterize the original region derived in [9] and [6], follows from the rate transfer analysis in [7, Theorem 1]. We discuss this subtlety in Appendix A.

Another special case of this setup is obtained if we ignore the identification requirement of the user, i.e., by letting $R^i = 0$. It is not hard to see that the model then reduces to the classical Wyner-Ziv problem of lossy source compression in the presence of receiver side information with the slight difference that the receiver wants to reconstruct the *original* source vector X^n rather than the noisy vector Y^n that is available at the encoder (cf. coding of remote sources [12]). We obtain the following rate-distortion region.

Corollary 2: A compression-distortion pair (R^c, D) is achievable if and only if there exist a random variable $U \in \mathcal{U}$ with joint distribution p_{UYXZ} such that $U - Y - X - Z$ forms a Markov chain and

$$\begin{aligned} R^c &\geq I(U; Y|Z) \\ D &\geq E[d(X, \hat{X})] \end{aligned}$$

with $|\mathcal{U}| \leq |\mathcal{Y}| + 1$.

IV. Achievability

We will first prove the achievability of an (R^c, R^i, D) tuple for which there exist a random variable $U \in \mathcal{U}$ and a function $\phi : \mathcal{U} \times \mathcal{Z} \to \hat{\mathcal{X}}$ satisfying $U - Y - X - Z$ and

$$R^i + \Delta R \leq I(U; Z) \tag{10}$$

$$R^c + \Delta R \geq I(U; Y) \tag{11}$$

$$D \geq E[d(X, \phi(U, Z))] \tag{12}$$

for some $\Delta R \geq 0$ which will be specified later. This auxiliary $-\Delta R$ will play the role of rate transferred from the "second-stage" rate R^c to the "first-stage" rate $-R^i$ of the fictitious source coder mentioned in Appendix A. In the actual scheme we next describe, this rate transfer is concretized through the use of binning, as discussed in the Introduction.

Fix $p_{U|Y}$ and the function ϕ that satisfy the conditions in Theorem 1. We first generate a codebook of size $2^{n(R^c+\Delta R)}$ that consists of i.i.d. codewords U^n. We index the codewords $U^n(j,k)$ for $j=1,\ldots,2^{nR^c}$ and $k=1,\ldots,2^{n\Delta R}$.

Enrollment: For any $y^n \in \mathcal{Y}^n$, we define the enrollment function $f(y^n)$ as the smallest index j for which $(y^n, U^n(j,k)) \in T^n_{[YU]}$[3] for some $k=1,\ldots,2^{n\Delta R}$. We set $f(y^n)=1$ if no such index can be found. Thus, one can think of the collection of all codewords $U^n(j,k)$ for $k=1,\ldots,2^{n\Delta R}$ as "bins," and $f(y^n)$ as a source coder which records only the bin index.

Identification: For any noisy observation $z^n \in \mathcal{Z}^n$ and the given compression indices $j(1)$, …, $j(2^{nR^i})$ of the database entries, the identifier looks for a database entry $m \in \{1,\ldots,2^{nR^i}\}$, such that $(z^n, U^n(j(m),k)) \in T^n_{[ZU]}$ for some $k=1,\ldots,2^{n\Delta R}$. We define the identification function $\hat{w}=g(j(1),\ldots,j(2^{nR^i}),z^n)$ as the smallest such m, and set $g(j(1),\ldots,j(2^{nR^i}),z^n)=1$ if no such m is found.

So far, the only randomness mentioned above is that of the codebook $U^n(j,k)$. We pause here to emphasize that this randomization is for the purpose of creating an ensemble of codebooks over which we compute average probability of error and average distortion. On the other hand, the database is filled with random entries also, but this randomness is inherent to the problem and is independent from codebook generation.

Now for $m=1,\ldots,2^{nR^i}$, define $J(m)=f(Y^n(m))$ and $K(m)$ as the smallest k found in the process of enrolling $Y^n(m)$. If no (j,k) was found, also set $K(m)=1$. Although $K(m)$ is not recorded, it is useful to define it for analysis purposes. Finally, let $\hat{W}=g(\mathbf{J},Z^n)$.

Reconstruction: For any noisy observation $z^n \in \mathcal{Z}^n$ and a given compression index $j \in \mathcal{L}$, the reconstruction function $h(j,z^n)$ is defined as follows. Find the smallest k such that $(z^n,U^n(j,k)) \in T^n_{[ZU]}$, and output $\phi(U_i(j,k),z_i)$ for the ith component of $h(j,z^n)$. If no such k is found, then output a random vector from the reconstruction alphabet.

Probability of error: We define the following events:

$$\begin{aligned}
E_0(m) &= \left\{(Y^n(m),Z^n)\notin T^n_{[YZ]}\right\}\\
E_1(m) &= \left\{(Y^n(m),U^n(J(m),K(m)))\notin T^n_{[YU]}\right\}\\
E_2(m,k) &= \left\{(Z^n,U^n(J(m),k))\notin T^n_{[ZU]}\right\} .
\end{aligned}$$

The average probability of error for the identification process can then be bounded as

$$\begin{aligned}
\Pr\{\hat{W}\neq W|W=w\} \leq \Pr\{E_0(w)\}&\\
+\Pr\{E_1(w)|E_0(w)^c\}&\\
+\Pr\{E_2(w,K(w))|E_1(w)^c\}&\\
+\sum_{m\neq w}\sum_k \Pr\{E_2(m,k)^c\}&\ . \qquad (13)
\end{aligned}$$

It is straightforward to show that $\Pr\{E_0(w)\}\to 0$. We can also show using standard arguments that $\Pr\{E_1(w)|E_0(w)^c\}$ vanishes with increasing n if

$$R^c+\Delta R > I(U;Y) .$$

That $\Pr\{E_2(w,K(w))|E_1(w)^c\}$ also vanishes with increasing n follows from the Markov lemma [1]. In fact, with high probability,

$$(Z^n,X^n(w),Y^n(w),U^n(J(w),K(w)))\in T^n_{[ZXYU]} , \qquad (14)$$

which will be useful in the distortion analysis. Finally,

$$\sum_{m\neq w}\sum_k \Pr\{E_2(m,k)^c\}\leq 2^{n(R^i+\Delta R)}2^{-nI(U;Z)} ,$$

the right-hand side of which vanishes for large enough n if

$$R^i+\Delta R < I(U;Z) .$$

Next, we consider the average distortion incurred by the reconstruction. A crucial observation at this point is that with probability approaching one, when $\hat{W}=W$,

$$\hat{X}_i=\phi(U_i(J(W),K(W)),Z_i) ,$$

that is, the index k found in the reconstruction process for Z^n and $j=J(W)$ matches $K(W)$. This follows from (14) and the fact that

$$\sum_{k\neq K(W)}\Pr\{E_2(W,k)^c\}\leq 2^{n\Delta R}2^{-nI(U;Z)} ,$$

which vanishes since $\Delta R < I(U;Z)$ is granted. Thus, when $\hat{W}=W$, with high probability

$$\begin{aligned}
&d(X^n(W),\hat{X}^n)\\
&\quad= \frac{1}{n}\sum_{i=1}^n d(X_i(W),\hat{X}_i)\\
&\quad= \frac{1}{n}\sum_{i=1}^n d(X_i(W),\phi(U_i(J(W),K(W)),Z_i))\\
&\quad\leq (1+\epsilon')\sum_{z',x',y',u'} P_{ZXYU}(z',x',y',u')\ d(x',\phi(u',z'))\\
&\quad\leq E[d(X,\phi(U,Z))]+\epsilon' d_{\max}\\
&\quad\leq D+\epsilon .
\end{aligned}$$

Since the ensemble averages satisfy the desired requirements, there must exist a deterministic codebook and functions f, g, and h for which the same requirements are satisfied.

Having shown the sufficiency of the conditions (10)-(12), we can now choose ΔR arbitrarily close to $I(U;Z)-R^i$ to obtain the achievability of the rate-distortion tuples as given in the expression of the theorem.

[3] For a probability distribution P_X, we denote by $T^n_{[X]}$ the set of all strongly typical sequences. For more detail on strong typicality see [2].

V. Converse

Here we prove the converse part of the theorem, i.e., that $\mathcal{R} \subset \mathcal{R}^*$. We assume the achievability of a tuple (R^c, R^i, D), i.e., for any $\epsilon > 0$ there exist deterministic functions f, g and h such that (2)-(5) are satisfied.

We have

$$\begin{aligned} \log M &= H(W) \\ &= H(W|\mathbf{J}, Z^n) + I(W; \mathbf{J}, Z^n) \\ &\leq H(W|\hat{W}) + I(W; \mathbf{J}, Z^n) && (15) \\ &\leq 1 + P_e^n \log M + I(W; \mathbf{J}, Z^n) \ , && (16) \end{aligned}$$

where (15) follows since $\hat{W}$ is a deterministic function of $\mathbf{J}$ and Z^n, and (16) follows from Fano's inequality. From here we can obtain

$$\begin{aligned} (1-\epsilon)\log M - 1 &\leq I(W; \mathbf{J}, Z^n) \\ &= I(W; Z^n|\mathbf{J}) && (17) \\ &= H(Z^n|\mathbf{J}) - H(Z^n|W, \mathbf{J}) \\ &\leq H(Z^n) - H(Z^n|W, \mathbf{J}) \\ &= H(Z^n) - H(Z^n|J(W)) \ , && (18) \end{aligned}$$

where (17) follows since W is independent of the database entries, and hence of $\mathbf{J}$, and (18) follows since Z^n is independent of $J(m)$ if $m \neq W$.

Now, we define

$$U_i \triangleq (Z^{i-1}, Z_{i+1}^n, J(W))$$

and observe that $Z_i - X_i(W) - Y_i(W) - U_i$ forms a Markov chain. Using (3), we can write

$$\begin{aligned} &(1-\epsilon)nR^i - 1 \\ &\leq H(Z^n) - H(Z^n|J(W)) && (19) \\ &= \sum_{i=1}^n \left[H(Z_i|Z^{i-1}) - H(Z_i|Z^{i-1}, J(W))\right] \\ &\leq \sum_{i=1}^n \left[H(Z_i|Z^{i-1}) - H(Z_i|Z^{i-1}, Z_{i+1}^n, J(W))\right] \\ &= \sum_{i=1}^n \left[H(Z_i) - H(Z_i|U_i)\right] \\ &= \sum_{i=1}^n I(Z_i; U_i) \ . \end{aligned}$$

Thus,

$$(1-\epsilon)R^i \leq \frac{1}{n}\sum_{i=1}^n I(Z_i; U_i) + \frac{1}{n} \ . \qquad (20)$$

We also have from (2) that

$$\begin{aligned} nR^c &\geq \log L \\ &\geq H(J(W)) \\ &\geq I(J(W); Y^n(W)) \ . \end{aligned}$$

Combining this with (20), we get

$$\begin{aligned} &n(R^c - R^i + \epsilon R^i + \frac{1}{n}) \\ &\geq I(J(W); Y^n(W)) - I(J(W); Z^n) \\ &= I(J(W); Y^n(W)|Z^n) && (21) \\ &= \sum_{i=1}^n I(J(W); Y_i(W)|Z^n, Y^{i-1}(W)) \\ &= \sum_{i=1}^n \Big[H(Y_i(W)|Z^n, Y^{i-1}(W)) \\ &\quad - H(Y_i(W)|J(W), Z^n, Y^{i-1}(W))\Big] \\ &\geq \sum_{i=1}^n H(Y_i(W)|Z_i) - H(Y_i(W)|J(W), Z^n) \\ &= \sum_{i=1}^n H(Y_i(W)|Z_i) - H(Y_i(W)|U_i, Z_i) \\ &= \sum_{i=1}^n I(Y_i(W); U_i|Z_i) \ , \end{aligned}$$

where (21) follows from the fact that $Z^n - Y^n(W) - J(W)$ forms a Markov chain. Thus,

$$R^c - R^i + \epsilon R^i + \frac{1}{n} \geq \frac{1}{n}\sum_{i=1}^n I(Y_i(W); U_i|Z_i) \ . \qquad (22)$$

As for the distortion constraint, first observe that

$$\begin{aligned} &E[d(X^n(W), h(J(W), Z^n))] \\ &= (1 - P_e^n)E[d(X^n(W), h(J(\hat{W}), Z^n))|\hat{W} = W] \\ &\quad + P_e^n E[d(X^n(W), h(J(W), Z^n))|\hat{W} \neq W] \\ &\leq (1 - P_e^n)(D + \epsilon) + P_e^n d_{\max} \\ &\leq D + \epsilon(1 + d_{\max}) \ . \end{aligned}$$

Thus, denoting by h_i the ith component of h, we have

$$\begin{aligned} &D + \epsilon(1 + d_{\max}) \\ &\geq \frac{1}{n}\sum_{i=1}^n E[d(X_i(W), h_i(J(W), Z^n))] \\ &= \frac{1}{n}\sum_{i=1}^n E[d(X_i(W), h_i(U_i, Z_i))] \ . && (23) \end{aligned}$$

From (20), (22), (23), and convexity of $\mathcal{R}^*$, $\mathcal{R} \subset \mathcal{R}^*$ follows.

VI. Computation of $\mathcal{R}$ for Two Simple Examples

As was discussed in [6] and [11] for the respective special cases, one way of computing the tradeoff is to rewrite the triplet

$$\Big(I(Y;U), I(Z;U), E[d(X, \phi(U,Z))]\Big)$$

as a convex combination of

$$\Big(\Gamma^c(P_{Y|U}(\cdot|u)), \Gamma^i(P_{Y|U}(\cdot|u)), \Delta(P_{Y|U}(\cdot|u))\Big)$$

with coefficients $\lambda_u = P_U(u)$, where for any distribution Q on $\mathcal{Y}$,

$$\Gamma^c(Q) = H(Y) + \sum_{y\in\mathcal{Y}} Q(y)\log Q(y)$$

$$\Gamma^i(Q) = H(Z) + \sum_{y\in\mathcal{Y},z\in\mathcal{Z}} P_{Z|Y}(z|y)Q(y)\log\left(\sum_{y'} P_{Z|Y}(z|y')Q(y')\right)$$

$$\Delta(Q) = \sum_{z\in\mathcal{Z}} \min_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\mathcal{X},y\in\mathcal{Y}} Q(y)P_{X|Y}(x|y)P_{Z|X}(z|x)d(x,\hat{x})$$

assuming that the function $\phi(\cdot,\cdot)$ is chosen optimally. Of course, the coefficients λ_u and vectors $P_{Y|U}(\cdot|u)$ should satisfy

$$\sum_u \lambda_u P_{Y|U}(\cdot|u) = P_Y(\cdot)\ .$$

In other words, to characterize $\mathcal{R}$, it suffices to consider convex combinations of all

$$\begin{bmatrix} Q(\cdot) \\ \Gamma^c(Q) \\ \Gamma^i(Q) \\ \Delta(Q) \end{bmatrix}$$

such that the first $|\mathcal{Y}|$ components of any combination agree with $P_Y(\cdot)$.[4] Note that for simplicity, we focus on $\Gamma^c(Q)$, not $\Gamma^c(Q) - \Gamma^i(Q)$, thus the complete region $\mathcal{R}$ will be obtained only after applying rate transfer through binning.

A. Noiseless Enrollment, Erasure Queries, and Hamming Distortion Measure

Consider binary feature vectors with $P_X(x) = \frac{1}{2}$ for $x \in \mathcal{X} = \{0,1\}$. Let the enrollment channel $P_{Y|X}$ be noiseless (thus $\mathcal{Y} = \mathcal{X}$), and $P_{Z|X}$ be a symmetric erasure channel with $\mathcal{Z} = \{0,?,1\}$ and erasure probability ϵ. Also let $\hat{\mathcal{X}} = \mathcal{X}$ and $d(\cdot,\cdot)$ be the Hamming distortion measure.

This example (without the distortion constraint) was analyzed in [6] and it was shown that $(R^c, R^i, \infty) \in \mathcal{R}$ if and only if

$$R^c \geq \frac{R^i}{1-\epsilon} \tag{24}$$

for $0 \leq R^i \leq 1-\epsilon$. Similarly, the Wyner-Ziv problem with erasure side information was solved before (see [8, Theorem 18] and [5, Theorem 1]), and it was shown that $(R^c, 0, D) \in \mathcal{R}$ if and only if

$$R^c \geq \epsilon\Psi_\epsilon(D) \tag{25}$$

for $0 \leq D \leq \frac{\epsilon}{2}$ with

$$\Psi_\epsilon(D) = 1 - \mathcal{H}\left(\frac{D}{\epsilon}\right)\ ,$$

where $\mathcal{H}(\cdot)$ is the binary entropy function.

Now, letting $Q = [1-\alpha \quad \alpha]^T$, we have

$$\begin{aligned} \Gamma^c(Q) &= \Psi_\epsilon(\alpha) \\ \Gamma^i(Q) &= (1-\epsilon)\Psi_\epsilon(\alpha) \\ \Delta(Q) &= \epsilon\min\{\alpha, 1-\alpha\}\ . \end{aligned}$$

Fig. 4 depicts the set of all triplets

$$\mathcal{L} = \bigcup_Q \left(\Gamma^c(Q), \Gamma^i(Q), \Delta(Q)\right)$$

for $\epsilon = 0.1$. Since the curve $\mathcal{L}$ lies on the plane $\Gamma^i = (1-\epsilon)\Gamma^c$ and its projection on the (Γ^c, Δ)-plane is convex, one can conclude that convex combinations of points on $\mathcal{L}$ cannot yield "better" points, in the sense that for any $(\Gamma^c, \Gamma^i, \Delta)$ obtained by some convex combination, there exists $(\Gamma^c, \Gamma^i, \Delta')$ already on $\mathcal{L}$ with $\Delta' \leq \Delta$. In other words, when we take convex combinations of $|\mathcal{Y}| + 2 = 4$ points on $\mathcal{L}$ corresponding to $\alpha = \alpha_1, \ldots, \alpha_4$ such that

$$\sum_{u=1}^{4} \lambda_u \alpha_u = P_Y(1) = \frac{1}{2}\ ,$$

we can only hope to achieve the points already on $\mathcal{L}$. But this is possible by combining just two points, $Q_1 = [1-\alpha \quad \alpha]^T$ and $Q_2 = [\alpha \quad 1-\alpha]^T$, with weights $\lambda_1 = \lambda_2 = \frac{1}{2}$, simply because $\left(\Gamma^c(Q_1), \Gamma^i(Q_1), \Delta(Q_1)\right) = \left(\Gamma^c(Q_2), \Gamma^i(Q_2), \Delta(Q_2)\right)$.

What all this means is that it suffices to constrain $P_{U|Y}$ to be a binary symmetric channel (BSC) with crossover probability $\alpha \leq \frac{1}{2}$, and

$$\phi(u,z) = \begin{cases} z & z \neq ? \\ u & z = ? \end{cases}\ .$$

With this choice, we obtain the characterization

$$R^c \geq \frac{R^i}{1-\epsilon}$$

for any $0 \leq D \leq \frac{\epsilon}{2}$ and $(1-\epsilon)\Psi_\epsilon(D) \leq R^i \leq 1-\epsilon$.

As we mentioned in the Introduction, R^c cannot be reduced further than $\Psi_\epsilon(D)$ unless we perform binning and reduce both rates R^i and R^c by the same amount. But since the slope of the (R^c, R^i) tradeoff can never be greater than one (e.g., it is $1-\epsilon$ here), the minimum R^c when $0 \leq R^i \leq (1-\epsilon)\Psi_\epsilon(D)$ will be obtained by applying binning only to the extreme point $R^i = (1-\epsilon)\Psi_\epsilon(D)$, $R^c = \Psi_\epsilon(D)$, yielding the complete characterization

$$R^c \geq \begin{cases} R^i + \epsilon\Psi_\epsilon(D) & 0 \leq R^i \leq (1-\epsilon)\Psi_\epsilon(D) \\ \frac{R^i}{1-\epsilon} & (1-\epsilon)\Psi_\epsilon(D) \leq R^i \leq 1-\epsilon \end{cases} \tag{26}$$

for $0 \leq D \leq \frac{\epsilon}{2}$. See Fig. 5 for the behavior of (26) on the (R^i, R^c)-plane for fixed D.

With the maximum possible distortion $D = \frac{\epsilon}{2}$, (26) reduces to (24), as expected. Similarly, substituting $R^i = 0$ in (26) yields (25).

[4] In fact, one could drop the first components of both Q and P_Y as both vectors lie in a $|\mathcal{Y}| - 1$ dimensional space. This, together with Carathéodory's theorem [3], guarantees that $|\mathcal{U}| = |\mathcal{Y}| + 2$ is sufficient.

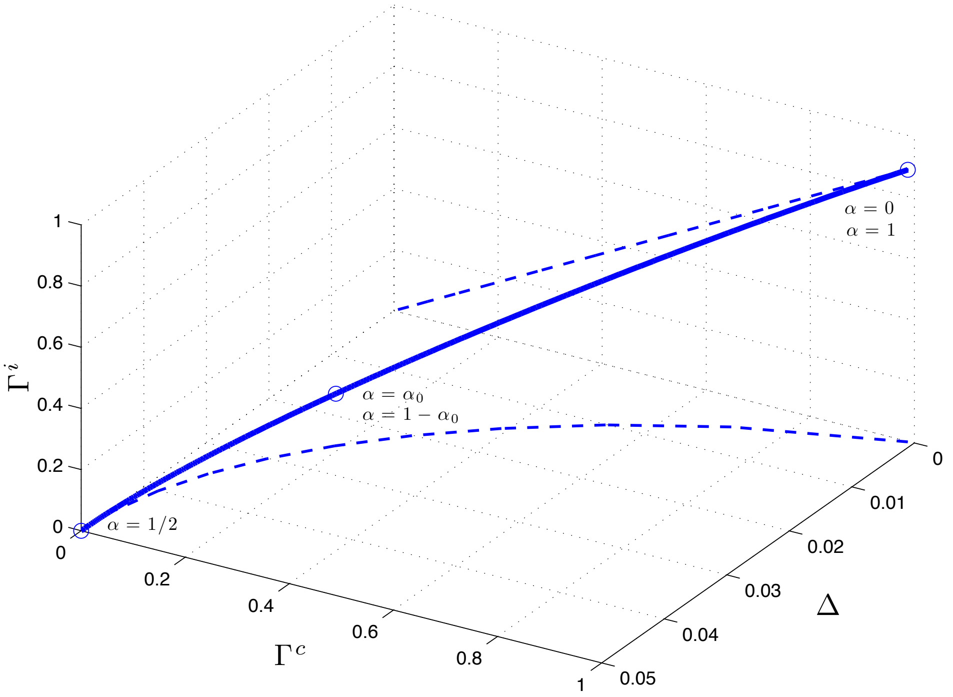

Fig. 4. The solid curve depicts all $(\Gamma^c(Q), \Gamma^i(Q), \Delta(Q))$, whereas the dashed curves represent the respective projections on the (Γ^c, Δ) and (Γ^c, Γ^i) planes. Also shown is the fact that α_0 and $1 - \alpha_0$ correspond to the same point for all $0 \leq \alpha_0 \leq 1/2$.

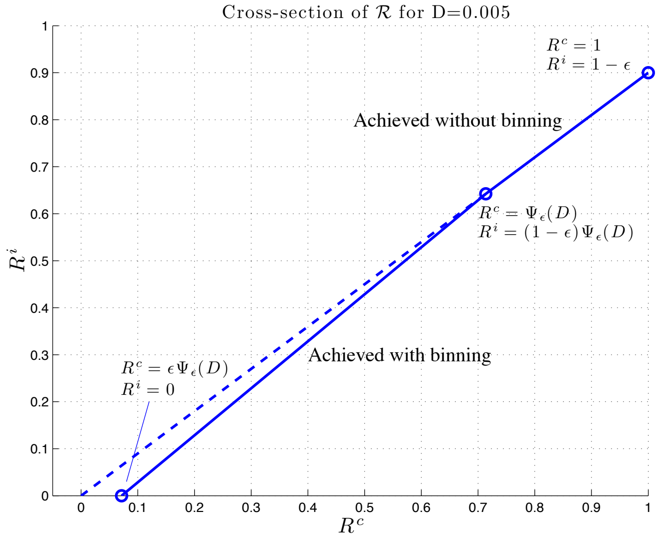

Fig. 5. The behavior of $\mathcal{R}$ for fixed $D = 0.005$. Binning is needed only if $R^c < \Psi_\epsilon(D)$. The dashed line corresponds to the (R^c, R^i) tradeoff without the distortion constraint.

B. Enrollment and Queries Subject to BSC Noise, with Hamming Distortion Measure

This time, $P_X(x) = \frac{1}{2}$ for $x \in \mathcal{X} = \{0, 1\}$, and $P_{Y|X}$ and $P_{Z|X}$ are both binary symmetric channels with crossover probabilities p and q, respectively (implying $\mathcal{Y} = \mathcal{Z} = \mathcal{X}$). Also $\hat{\mathcal{X}} = \mathcal{X}$ and $d(\cdot, \cdot)$ is the Hamming distortion measure. This example was also considered in [6], and it was shown that $(R^c, R^i, \infty) \in \mathcal{R}$ if and only if

$$R^i \leq 1 - \mathcal{H}(p \star q \star \mathcal{H}^{-1}(1 - R^c))$$

for all $0 \leq R^c \leq 1$, where $a \star b = a(1-b) + b(1-a)$ and $\mathcal{H}^{-1}$ returns values between 0 and $\frac{1}{2}$.

Letting $Q = [1 - \alpha \quad \alpha]^T$, we have

$$\begin{aligned} \Gamma^c(Q) &= 1 - \mathcal{H}(\alpha) \\ \Gamma^i(Q) &= 1 - \mathcal{H}(p \star q \star \alpha) \end{aligned}$$

together with

$$\Delta(Q) = q$$

if $p \geq q$, and

$$\Delta(Q) = \begin{cases} p + (1-2p)\alpha & 0 \leq \alpha \leq \frac{q-p}{1-2p} \\ q & \frac{q-p}{1-2p} \leq \alpha \leq \frac{1-p-q}{1-2p} \\ 1 - p - (1-2p)\alpha & \frac{1-p-q}{1-2p} \leq \alpha \leq 1 \end{cases}$$

if $p < q$. Obviously, $p < q$ is the interesting case here because otherwise the tradeoff is the same as that in [6].

In Fig. 6, the set of all triplets is shown for $p = 0.15$, $q = 0.2$. In contrast with the previous example, convex combinations do generate better $(\Gamma^c, \Gamma^i, \Delta)$, as can be seen from the figure. However, similar to that example, it can be seen by close inspection that for every convex combination of $|\mathcal{Y}| + 2 = 4$ points achieving $(\Gamma^c, \Gamma^i, \Delta)$, one can find just 2 points $Q_1 = [1 - \alpha \quad \alpha]^T$ and $Q_2 = [1 - \beta \quad \beta]^T$ with

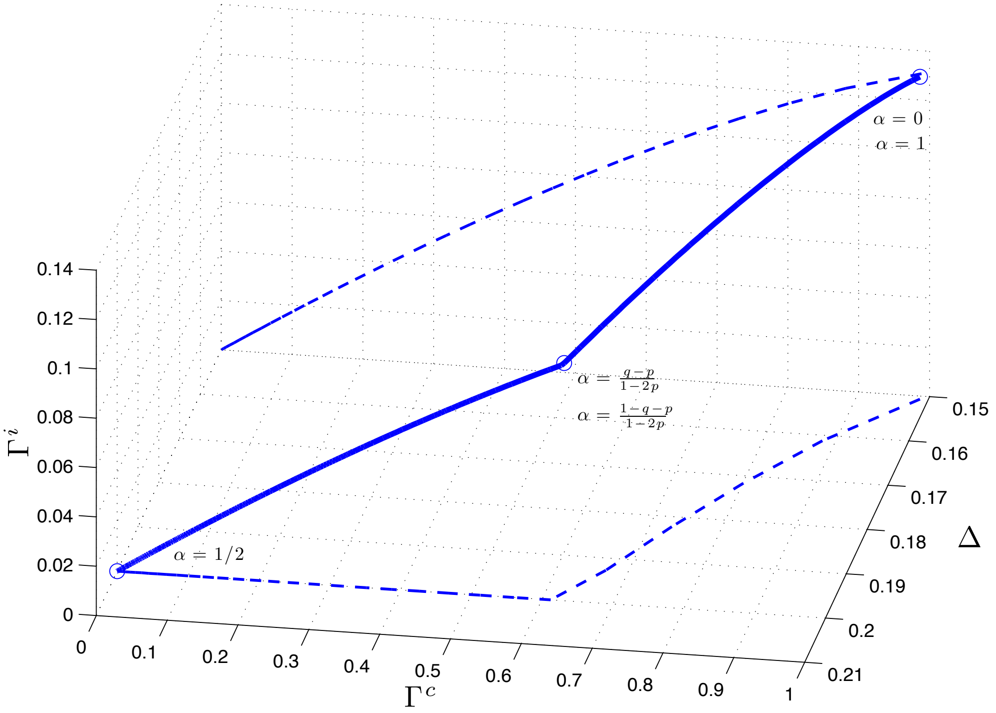

Fig. 6. The solid curve depicts all $(\Gamma^c(Q), \Gamma^i(Q), \Delta(Q))$, whereas the dashed curves represent the respective projections on the (Γ^c, Δ) and (Γ^c, Γ^i) planes.

$0 \le \alpha \le \frac{q-p}{1-2p} \le \beta \le \frac{1}{2}$, some convex combination of which[5] achieves $(\Gamma^c, \Gamma^{i'}, \Delta')$ with $\Delta' \le \Delta$ and $\Gamma^{i'} \ge \Gamma^i$. To be consistent with $P_Y(y)$, these two points can then be paired up with $Q_3 = [\alpha \quad 1-\alpha]^T$ and $Q_4 = [\beta \quad 1-\beta]^T$ to form the most general convex combination with weights $\lambda_1 = \lambda_3 = \frac{r}{2}$ and $\lambda_2 = \lambda_4 = \frac{\bar{r}}{2}$ with $\bar{r} = 1 - r$. Translating these, we obtain the optimal forward test channel $P_{U|Y}$ as shown in Fig. 7. As in the previous example, the reconstruction function ϕ is simple:

$$\phi(u, z) = \begin{cases} z & u = 2, 4 \\ 0 & u = 1 \\ 1 & u = 3 \end{cases} .$$

Despite the fact that we know the optimal $P_{U|Y}$ and $\phi(\cdot, \cdot)$, it proved very difficult to optimally choose the parameters (α, β, r). Therefore, for any fixed D, we numerically computed the (R^i, R^c, D)-tradeoff for the same instance $p = 0.15$, $q = 0.2$, as shown in Fig. 8. Although as shown in Fig.7 we need an alphabet $\mathcal{U}$ of size 4 in general, $|\mathcal{U}|$ could be taken smaller in two distinct regimes, as indicated in Fig. 8:

1) Up to a certain value of α (or above a certain R^c), the optimal $r = 1$, so $|\mathcal{U}|$ could be limited to 2.
2) In the binning regime, where all the (R^c, R^i) pairs are obtained by rate transfer applied to the minimum achievable (R^c, R^i) without binning, optimal $\beta = \frac{1}{2}$. Therefore the symbols $u = 2$ and $u = 4$ can be collapsed together without changing anything.

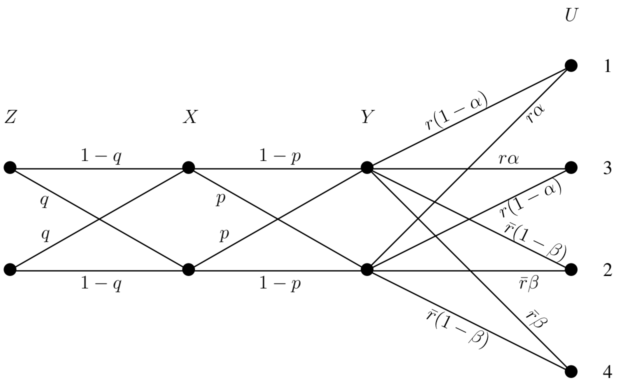

Fig. 7. The optimal test channel $P_{U|Y}$.

VII. Conclusions

We have studied a noisy database system where both the enrollment and the query vectors are noisy versions of the underlying feature vectors. The noisy enrollment vectors are compressed before being stored in the database to reduce the storage requirement and increase the search speed. The user of the database wishes not only to identify the correct entry corresponding to a noisy query vector, but also to reconstruct the original feature vector of the queried entry within a desired distortion requirement. This problem combines and generalizes

[5] Intuitively, if there are more than one point in either interval $0 \le \alpha \le \frac{q-p}{1-2p}$ or $\frac{q-p}{1-2p} \le \beta \le \frac{1}{2}$, one can first take the average of those multiple points with weights proportional to those in the original convex combination. From the concavity of the (Γ^c, Γ^i) tradeoff everywhere and convexity of the (Γ^c, Δ) tradeoff for $0 \le \alpha \le \frac{q-p}{1-2p}$, it follows that for the Γ^c's this new pair of points achieves, the original curve has points with better (Γ^i, Δ) values. So, we might as well choose the pair of points on the curve.

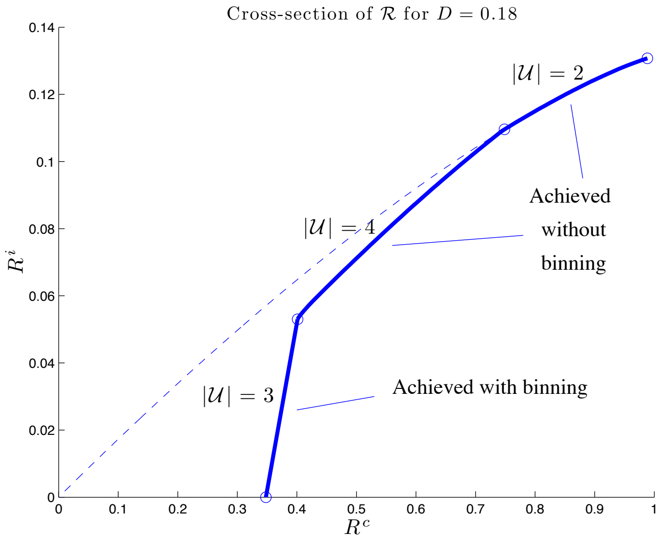

Fig. 8. The behavior of $\mathcal{R}$ for fixed $D = 0.18$. The dashed curve corresponds to the (R^c, R^i) tradeoff without the distortion constraint.

the previously studied capacity/storage tradeoff in databases and the Wyner-Ziv rate distortion function for lossy source compression in the presence of decoder side information.

We have characterized the set of achievable compression rate, identification rate, and distortion tuples in a single-letter form. As examples, we have studied two simple scenarios, and analyzed the behavior of optimal codes with respect to binning. We have demonstrated that there are two regimes, with and without binning. For high compression and identification rates, binning is not needed, whereas the low rate tradeoff is achieved by binning the codewords.

Appendix A
Discussion of Corollary 1 Using Rate Transfer Analysis

In [7], the author discussed under what conditions rate regions described in terms of *marginal rates* are equivalent to those described in terms of *cumulative rates* in a general class of source and channel coding problems. More specifically, the marginal and the cumulative rate regions were respectively defined as

$$\mathcal{R}_{\text{mar}} = \Big\{(R_1, R_2) : \exists \mathbf{X} \in \mathcal{D} \text{ s.t. } I_1(\mathbf{X}) \le R_1, I_2(\mathbf{X}) \le R_2\Big\}$$

and

$$\mathcal{R}_{\text{cum}} = \Big\{(R_1, R_2) : \exists \mathbf{X} \in \mathcal{D} \\ \text{s.t. } I_1(\mathbf{X}) \le R_1, I_1(\mathbf{X}) + I_2(\mathbf{X}) \le R_1 + R_2\Big\},$$

where $\mathbf{X}$ is a random vector, $\mathcal{D}$ is a region in the probability simplex of $\mathbf{X}$, and I_1 and I_2 are information measures intrinsic to the problem. It was shown in [7, Theorem 1] that if $\mathcal{R}_{\text{mar}}$ is a convex region, then

$$\mathcal{R}_{\text{mar}} = \mathcal{R}_{\text{cum}} \Longleftrightarrow (R_{\min}, 0) \in \mathcal{R}_{\text{mar}} \ , \tag{27}$$

where

$$R_{\min} = \min_{\mathbf{X} \in \mathcal{D}} \Big[I_1(\mathbf{X}) + I_2(\mathbf{X})\Big]. \tag{28}$$

We now use (27) to show that the region in Corollary 1 is in fact the same as the original capacity-storage region given by (8) and (9).

First observe that there is no non-negativity restriction on either the rates R_1 and R_2, or the information measures $I_1(\mathbf{X})$ and $I_2(\mathbf{X})$. Therefore, rewriting (8) as

$$-R^i \ge -I(U;Z) \ , \tag{29}$$

we can map our problem to a fictitious two-stage source coding problem using the following:

$$\begin{aligned} R_1 &= -R^i \\ R_2 &= R^c \\ \mathbf{X} &= (U, Y, X, Z) \\ I_1(\mathbf{X}) &= -I(U;Z) \\ I_2(\mathbf{X}) &= I(U;Y) \\ \mathcal{D} &= \{(U, Y, X, Z) : U - Y - X - Z\}. \end{aligned}$$

The minimum cumulative rate in (28) then becomes

$$R_{\min} = \min_{U : U - X - Y - Z} \Big[I(U;Y) - I(U;Z)\Big] \ .$$

Since $I(U;Y) \ge I(U;Z)$ due to the data processing inequality, $R_{\min} \ge 0$. But it is easy to see that $R_{\min} = 0$ with the simple choice of $U = \emptyset$. Thus, using (27), whether the characterization in Corollary 1 is identical to (8) and (9) comes down to whether $(0, 0) \in \mathcal{R}_{\text{mar}}$, i.e., whether there exists U such that

$$\begin{aligned} -I(U;Z) &= 0 \\ I(U;Y) &= 0 \ . \end{aligned}$$

But this is readily achieved also by $U = \emptyset$.

References

[1] T. Berger, "Multiterminal source coding," *Lectures presented at CISM Summer School on the Inform. Theory Approach to Commun.*, July 1977.

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

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: John Wiley and Sons, 1991.

[4] J. A. O'Sullivan and N. A. Schmid, "Large deviations performance analysis for biometrics recognition," *Proc. of Allerton Conf. on Comm., Control, and Computing*, Oct. 2002, Monticello, IL.

[5] E. Perron, S. N. Diggavi, and I. E. Telatar, "Lossy source coding with Gaussian or erased side-information," *Proc. IEEE Int'l Symp. Inform. Theory*, Seoul, S. Korea, 2009, pp. 1035-1039.

[6] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2097-2106, May 2009.

[7] E. Tuncel, "The rate transfer argument in two-stage scenarios: When does it matter?" *Proc. IEEE Int'l Symp. Inform. Theory*, Seoul, S. Korea, July 2009.

[8] S. Verdú and T. Weissman, "The information lost in erasures," *IEEE Trans. Inform. Theory*, vol. 54, no. 11, pp. 5030-5058, Nov. 2008.

[9] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 299-320, Jan. 2008.

[10] F. Willems, T. Kalker, J. Goseling and J.-P. Linnartz, "On the capacity of a biometrical identification system," *Proc. IEEE Int'l Symp. Inform. Theory*, Yokohama, Japan, July 2003.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1-10, Jan. 1976.

[12] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *The Transactions of the IECE of Japan*, Vol.E-63, No.10, pp.700-706, Oct. 1980.

Ertem Tuncel (S'99–M'04) received the B.S. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey in 1995, the M.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey in 1997, and the Ph.D. degree in electrical and computer engineering from University of California, Santa Barbara, in 2002. In 2003, he joined the Department of Electrical Engineering, University of California, Riverside, where he is currently a Professor.

His main research interests are multi-user networks, joint source-channel coding, zero-error information theory, and content-based retrieval in high-dimensional databases.

Dr. Tuncel received the National Science Foundation CAREER Award in 2007.

Deniz Gündüz received the B.S. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Polytechnic Institute of New York University, Brooklyn, NY in 2004 and 2007, respectively.

Currently he is a Lecturer in the Electrical and Electronic Engineering Department of Imperial College London, London, UK. He was a research associate at CTTC in Barcelona, Spain from November 2009 until September 2012. He also held a visiting researcher position at Princeton University from November 2009 until November 2011. Previously he was a consulting assistant professor at the Department of Electrical Engineering, Stanford University, and a postdoctoral Research Associate at the Department of Electrical Engineering, Princeton University.

He is an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, and served as a guest editor of the EURASIP Journal on Wireless Communications and Networking, Special Issue on Recent Advances in Optimization Techniques in Wireless Communication Networks. He is serving as the chair of the IEEE Information Theory Society Student Committee. He is a co-chair of the Network Theory Symposium at the 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP), and also served as a co-chair of the 2012 IEEE European School of Information Theory (ESIT). His research interests lie in the areas of communication theory and information theory with special emphasis on joint source-channel coding, multi-user networks, energy efficient communications and security.