Identification rate, search and memory complexity trade-off

Citation for published version (APA):

Farhadzadeh, F., & Willems, F. M. J. (2016). Identification rate, search and memory complexity trade-off: fundamental limits. *IEEE Transactions on Information Theory*, 62(11), 6173 - 6188.
<https://doi.org/10.1109/TIT.2016.2612234>

DOI:

[10.1109/TIT.2016.2612234](https://doi.org/10.1109/TIT.2016.2612234)

Document status and date:

Published: 21/09/2016

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Identification Rate, Search and Memory Complexity Trade-off: Fundamental Limits

Farzad Farhadzadeh, *Member, IEEE*, Frans M.J. Willems, *Fellow, IEEE*

Abstract—In an information-theoretic framework, we introduce a two-stage decoding scheme capable of achieving identification capacity to address search and memory complexities in large-scale identification systems. This two-stage decoding procedure is accomplished as follows. For a given query, at the first stage, a list of cluster indices is estimated. Then, at the second stage, refinement checks are performed for all members of the clusters to produce a single index. The first result this paper presents is the achievable rate quadruple region that specifies necessary conditions that the two-stage decoding scheme should satisfy to be able to achieve the identification capacity. The rest of paper is designated to investigate various achievable rate quadruples in which the proposed two-stage identification setup can reduce the search complexity with respect to conventional identification setups.

Index Terms—Identification systems, identification capacity, list decoding, databases.

I. INTRODUCTION

Identification by nearest neighbor search (NNS) is a research problem emerging in vast forensic applications including, but not limited to, human biometrics [1], multimedia security (copy detection, content identification and tracking) [2] and physical object security [2].

An identification system [1] usually consists of two main phases: *enrollment* and *identification*. In the enrollment phase, feature vectors representing digital contents, humans or physical objects are extracted and stored in a database. In the identification phase, a query, i.e., a noisy and degraded counterpart of an enrolled item, is presented for identification, which is accomplished by comparing the query with feature vectors stored in the database and selecting the best match. The performance of an identification system can be evaluated in terms of “identification rate”, “memory complexity” and “search complexity”. The exponential rate of the total number of

items enrolled at the enrollment phase is referred to as the “*identification rate*”. The overall amount of memory that should be allocated to the enrolled items and their indices is referred to as the “*memory complexity*”. The overall time complexity, which is commonly estimated by counting the number of elementary operations performed by an algorithm, in the identification phase to identify a received query is referred to as the “*search complexity*”.

Willems et al. [1] investigated the capacity C_{id} of an identification system. C_{id} is defined as the maximum achievable rate of the number of distinguishable items in a database, where the feature vectors extracted from items have constant and known statistics. They showed that approximately 2^{NR} items can be distinguished from each other, if and only if $R < C_{id}$ when N , the dimensionality of the feature space, becomes very large. They presented a single-letter characterization for the identification capacity C_{id} , which is equal to the mutual information between the enrollment and identification observations. See also [3]. A crucial methodology in this development is to regard the set of enrolled feature vectors as random channel codes. In the identification system there is no control neither on the probability distribution of enrolled items nor on the enrolled items in the database. As such, the identification phase can be accomplished using a NNS scheme in which the search complexity is proportional to the number of checks.

In modern applications, the size of a database can be billions. Therefore, theoretical investigation and the development of practical methods for achieving identification capacity [1] is of great interest. As mentioned above, an efficient approach should satisfy several important requirements. First, users should be able to reliably identify objects or individuals, i.e., find the most similar and related objects or individuals in a database with a low probability of error (high reliability). Second, the decoding method should be as fast as possible (low search complexity). Finally, it should require the least possible amount of memory for both the items and the indexing structure (low memory complexity). In general, there is a tradeoff between these three conditions, reliability, search complexity and memory complexity, and

F. Farhadzadeh and F. M.J. Willems are with the department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands (Email: {f.farhadzadeh, f.m.j.willems}@tue.nl).

This work was in part presented in IEEE International Symposium on Information Theory, Seoul, Korea, June 2009 and Istanbul, Turkey, July 2013.

the fundamental information-theoretical characterization of this triple trade-off is an open and emerging research problem, which we deal with in this paper.

To achieve the identification capacity, an identification system can perform an exhaustive search on all database entries to find the best match. The search complexity of this method is in order of the total number of items in the database, i.e., $\mathcal{O}(2^{N_{C_{id}}})$. Chavez et al. [4] gave an extensive overview of techniques to reduce search complexity in metric spaces by reducing the number of entries to be checked. The conventional approach to reduce the search complexity is to use a multidimensional index structure. Space-partitioning methods like grid file [5], K-D-B-tree [6] and quad-tree [7] divide the data space along predefined planes. Data-partitioning methods like R-tree [8], R^* -tree [9], R^+ -tree [10], X-tree [11], SR-tree [12], M-tree [13], TV-tree [14] and hB-tree [15] divide the data space according to the distribution of data. Although these access methods generally work well for low-dimensional spaces, their performance degrades as the number of dimensions increases. This phenomenon has been reported for the R^* -tree [11], the X-tree [16] and the SR-tree [12].

Weber et al. [17] compared indexing techniques to methods based on vector-approximations (VA). Similar to these VA methods are fingerprinting techniques used in content-based audio identification, Haitsma and Kalker [18], and Cano et al. [19]. Weber et al. [17] showed that for searching in high-dimensional spaces quantization methods outperform indexing methods.

Tuncel et al. [20] are among the first who investigated rate-distortion approaches in database searching. They applied quantization during enrollment and considered a fundamental trade-off between compression rate and reconstruction distortion. Later Tuncel [21] considered the trade-off between enrollment compression rate and identification rate, an extension of [1].

In this paper, we introduce a two-stage decoding scheme that can achieve identification capacity C_{id} with lower search complexity than exhaustive search. In contrast to [20], [21], we do not compress the database's entries. In an information-theoretical framework, we evaluate the achievable region for this two-stage decoding scheme.

Upon observing a query, the two-stage decoding works as follows. At the first stage, a cluster, which may include the related item, is detected. At the second stage, the decision concerning the item is taken. If there are M items, an ideal system to minimize the search complexity, i.e., the total number of checks, has $\sqrt{M}$ disjoint clusters, each containing $\sqrt{M}$ items. To determine the cluster index, $\sqrt{M}$ clusters has to be checked. To determine

Fig. 1. Nine individual entries organized into three clusters. Three cluster-checks and five refinement-checks, eight checks in total, are sufficient for identification.

the item within the detected cluster, $\sqrt{M}$ refinement-checks are needed. This results in $2\sqrt{M}$ checks in total, which is considerably less than M checks required for exhaustive search.

Due to imperfections of an observation channel, a query may fall in a cluster different from the one including the true item. This causes an error. To tackle this, one can enlarge clusters to have overlap with each other. This means each item can be in more than one cluster. Correspondingly, it is reasonable to detect a list of clusters instead of a single cluster. We investigate the fundamental trade-off between cluster-check rate and refinement-check rate of this two-stage identification that can achieve identification capacity.

This paper can be seen as a generalization of the schemes introduced by Willems [22] and Tuncel [23]. The main differences between the current manuscript and [22] and [23] are: (a) generalization of cluster centroids, considered as auxiliary random variables, based on statistics of both entries of the database and queries, while in [22] and [23] the cluster centroids have been generated based on either the statistics of the queries, or the entries of the database, respectively; (b) while identification capacity and search complexity are explained in the above examples, memory complexity is not explained at all; (c) a new result on the search-memory complexity region of capacity achieving identification systems. The main goal of this sequel is to introduce a search strategy based on a two-stage decoding scheme, so as to achieve the identification capacity and to reduce the search complexity.

The rest of this paper is organized as follows. In Section II we introduce notations and definitions exploited through this paper. Section III-A introduces the generalized two-stage identification system. Section IV represents some properties of the introduced two-stage decoding scheme. In Section V, we investigate search-

memory complexity of the proposed two-stage decoding schemes. We present necessary conditions to reduce search complexity in Section VI. In Section VII we consider binary symmetric and Gaussian systems and present their corresponding numerical results. Concluding remarks follow in Section VIII.

II. DEFINITION AND PRELIMINARIES

We give the definition of jointly typical sequences and their main properties, since typicality is an important notion in our proofs. For more details and the proofs of the properties see e.g. [24].

Definition 1 (Set of jointly typical sequences, [24]). Let $(X_1, X_2, \dots, X_K)$ be a finite collection of K discrete random variables with some joint distribution $P(x_1, x_2, \dots, x_K)$, $(x_1, x_2, \dots, x_K) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_K$. The set $\mathcal{A}_\epsilon^{(N)}(X_1, X_2, \dots, X_K)$ of ϵ -typical N -sequence $(x_1^N, x_2^N, \dots, x_K^N)$ is defined as

$$\mathcal{A}_\epsilon^{(N)}(X_1, X_2, \dots, X_K) \triangleq \left\{ (x_1^N, x_2^N, \dots, x_K^N) : \left| -\frac{1}{N} \log P(v^N) - H(V) \right| < \epsilon, \forall V \subseteq \{X_1, X_2, \dots, X_K\} \right\}, \quad (1)$$

where v^N is a subset of N -sequence $(x_1^N, x_2^N, \dots, x_K^N)$ corresponding to V , and $P(V^N) = \prod_{n=1}^N P(v_n)$. Moreover, let $V, W \subseteq \{X_1, X_2, \dots, X_K\}$, then for a given w^N we define

$$\mathcal{A}_\epsilon^{(N)}(V | w^N) \triangleq \left\{ v^N : (v^N, w^N) \in \mathcal{A}_\epsilon^{(N)}(V, W) \right\}. \quad (2)$$

Lemma 1 (Properties of jointly typical sequences, [24]). For any $\epsilon > 0$

- 1) $\forall V \subseteq \{X_1, X_2, \dots, X_K\}$ and N large enough

$$\Pr\{\mathcal{A}_\epsilon^{(N)}(V)\} \geq 1 - \epsilon. \quad (3)$$

- 2) $\forall V \subseteq \{X_1, X_2, \dots, X_K\}$

$$|\mathcal{A}_\epsilon^{(N)}(V)| \leq 2^{N(H(V)+\epsilon)}. \quad (4)$$

- 3) Let $V, W \subseteq \{X_1, X_2, \dots, X_K\}$, then

$$|\mathcal{A}_\epsilon^{(N)}(V | w^N)| \leq 2^{N(H(V|W)+\epsilon)}. \quad (5)$$

- 4) Let $K = 2$. If (X_1^N, X_2^N) are drawn according to $P(x_1^N)P(x_2^N)$, i.e. X_1^N and X_2^N are independent sequences with the same marginals as $P(x_1^N, x_2^N)$, then

$$\Pr \left\{ (X_1^N, X_2^N) \in \mathcal{A}_\epsilon^{(N)}(X_1, X_2) \right\} \leq$$

$$2^{-N(I(X_1, X_2) - 3\epsilon)}. \quad (6)$$

Notations

Throughout this paper, we adopt the convention that a scalar random variable is denoted by a capital letter X , a specific value it may take is denoted by the lower case letter x , and its alphabet is designated by the script letter $\mathcal{X}$. As for vectors, a capital letter X^N with a corresponding superscript will denote an N -dimensional random vector $X^N = (X_1, \dots, X_N)$. A lower case letter x^N will represent its particular realization $x^N = (x_1, \dots, x_N)$. The vector x_a^b denotes a sequence $(x_a, x_{a+1}, \dots, x_b)$, where a and b are integers, and $b \geq a$. The expectation operator is designated by $E[\cdot]$. Furthermore, $H_2(a) = -a \log_2(a) - (1-a) \log_2(1-a)$ for $0 \leq a \leq 1/2$ denotes the binary entropy function and $H_2^{-1} : [0, 1] \rightarrow [0, 0.5]$ is its inverse mapping.

III. GENERALIZED MODEL DESCRIPTION AND STATEMENT OF RESULT

A. Model Description

In this section, we introduce an identification system based on a generalized two-stage decoding scheme.

The identification system under analysis is shown in Figure 2. In the enrollment phase, a randomly generated sequence (vector) $x^N(w)$ of length N is extracted from each item and stored in a database. The vector $x^N = (x_1, \dots, x_N)$ contains the symbols x_n , $1 \leq n \leq N$, each taking values in the discrete alphabet $\mathcal{X}$. The database $\mathcal{C}$ is a collection of M indexed sequences denoted by

$$\mathcal{C} = \{x^N(1), x^N(2), \dots, x^N(M)\}. \quad (7)$$

Assume the components of $X^N = (X_1, \dots, X_N)$ are independent and identically distributed (i.i.d.) according to $\{Q_s(x); x \in \mathcal{X}\}$. Hence, the probability that sequence x^N occurs for the item indexed by w is

$$\Pr \{X^N(w) = x^N\} = \prod_{n=1}^N Q_s(x_n). \quad (8)$$

Note that this probability does not depend on the index w . We assume that all sequences are generated prior to the identification phase.

In the identification phase, an enrolled item will be presented for identification. It is equally likely that each item w is presented for identification, hence

$$\Pr \{W = w\} = \frac{1}{M}, \quad \forall w \in \mathcal{M} = \{1, 2, \dots, M\}. \quad (9)$$

When item w is presented for identification, its corresponding sequence $x^N(w) = s(w, \mathcal{C})$, is “selected”

Fig. 2. Generalized model of a two-stage identification system.

from the database $\mathcal{C}$, and observed through a memoryless observation channel $\{Q_c(y|x); x \in \mathcal{X}, y \in \mathcal{Y}\}$. $s : \mathcal{M} \times \mathcal{C} \rightarrow \mathcal{X}$ is a selector function. The resulting channel output sequence is $y^N = (y_1, y_2, \dots, y_N)$, where $y_n \in \mathcal{Y}, 1 \leq n \leq N$. Thus,

$$\Pr \{Y^N = y^N | X^N(w) = x^N\} = \prod_{n=1}^N Q_c(y_n | x_n). \quad (10)$$

Upon receiving y^N , the identification starts by constructing a list of indices with cardinality M_3 . This index list with outcome $w_1^{M_3} = (w_1(1), w_1(2), \dots, w_1(M_3))$, $w_1(i) \in \mathcal{M}_1 = \{1, 2, \dots, M_1\}, 1 \leq i \leq M_3$, is constructed by the “first decoder”, $d_1 : \mathcal{Y}^N \rightarrow \mathcal{M}_1^{M_3}$ (see Fig. 2). For convenience, for the rest of the paper we designate the boldface $\mathbf{w}_1$ to refer to the index list $w_1^{M_3}$. Hence

$$\mathbf{w}_1 = d_1(y^N). \quad (11)$$

Now, at the second decoding stage, a decision is made, based on y^N , the index list $\mathbf{w}_1$ and the database $\mathcal{C}$. This decision consisting of $w_1 \in \mathbf{w}_1 = \{w_1(1), \dots, w_1(M_3)\}$ and $w_2 \in \mathcal{M}_2 = \{1, \dots, M_2\}$ is taken by the “second decoder”, $d_2 : \mathcal{Y}^N \times \mathcal{M}_1^{M_3} \times \mathcal{C} \rightarrow \mathcal{M}_2 \times \mathcal{M}_1$ (see Fig. 2). Hence

$$(w_2, w_1) = d_2(y^N, \mathbf{w}_1, \mathcal{C}). \quad (12)$$

Finally, a combiner based on the second decoder outputs w_1 and w_2 forms an estimate of the index of the item presented for identification, $c : \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{M}$. Hence

$$\hat{w} = c(w_1, w_2). \quad (13)$$

We assume that $\hat{w} \in \{1, \dots, M\}$. The reliability of our identification system is measured by the error probability

$$P_{\mathcal{E}} = \Pr \{\hat{W} \neq W\}. \quad (14)$$

B. Statement of The Main Result

Definition 2. The rate quadruple (R_1, R_2, R_3, R) with $R \geq 0$ is *achievable*, if for all $\epsilon > 0$ and N large enough, there exist mappings $d_1(\cdot)$, $d_2(\cdot, \cdot, \cdot)$, and $c(\cdot, \cdot)$ such that

$$\begin{aligned} \log_2(M_1) &\leq N(R_1 + \epsilon), \\ \log_2(M_2) &\leq N(R_2 + \epsilon), \\ \log_2(M_3) &\leq N(R_3 + \epsilon), \\ \log_2(M) &\geq N(R - \epsilon), \end{aligned}$$

$$\text{and } \Pr \{\hat{W} \neq W\} \leq \epsilon. \quad (15)$$

The following theorem states a fundamental trade-off between these rates to achieve the identification capacity.

Theorem 1. The region of achievable rate quadruples of the identification system defined above is given by

$$\begin{aligned} &\{(R_1, R_2, R_3, R) : \\ &R_1 \geq I(X, Y; U), \\ &R_2 \geq \max(0, R - I(X; U)), \\ &R_3 \geq I(X; U|Y), \\ &0 \leq R \leq C_{\text{id}} = I(X; Y), \\ &\text{for } P(x, y, u) = Q_s(x)Q_c(y|x)P(u|x, y)\}. \end{aligned} \quad (16)$$

Proof. Appendix A. $\square$

Remark 1. It should be noted that $P(x, y, u) = Q_s(x)Q_c(y|x)P(u|x, y)$ follows from the chain rule in the probability theory. This implies that contrary to [22], in which the Markov chain condition $X \leftrightarrow Y \leftrightarrow U$ must hold, there is no restriction on $P(x, y, u)$ in (16). And, the auxiliary random variable U depends on both X and Y .

C. Clustering-based identification

The generalized model of the two stage identification scheme introduced above is coming to existence by a

Fig. 3. Memory-complexity of the two-stage decoding scheme. The enrolled items are divided into M_1 clusters, each associated with a cluster centroid $u^N(w_1)$, $1 \leq w_1 \leq M_1$. For each cluster only M_2 blocks of memory is allocated to store their related members.

clustering-based identification system, which can achieve the identification capacity (see Appendix A), with the following implementation.

First, the database $\mathcal{C}$ of the enrolled items is split into different clusters. The total number of clusters is M_1 and every cluster is specified by a cluster index w_1 and a cluster centroid $u^N(w_1)$. Each cluster can include at most M_2 enrolled items (see Fig. 3). The total amount of memory that is allocated to store enrolled items, cluster centroids and their indices is the *memory complexity*.

The clustering-based identification system using above database organization follows the decoding scheme introduced in Section III-A. At the first stage, the first decoder upon receiving y^N , checks all cluster centroids $u^N(w_1)$, $1 \leq w_1 \leq M_1$ and chooses M_3 most similar cluster centroids, using some predefined distance measure, and constructs list $\mathbf{w}_1$. The first decoder can send at most M_3 cluster indices to the second decoder. The total number of checks to be done to construct $\mathbf{w}_1$ is called “first stage search complexity”. This search complexity is proportional to M_1 , number of clusters.

At the second stage, the second decoder, having access to y^N , $\mathbf{w}_1$ and the clustered database, checks the items belonging to the detected clusters (see Fig. 4), and chooses the most similar one to y^N . Similarly, the total number of checks to be done by the second decoder is called the “second stage search complexity”. This search complexity is proportional to M_3 , the number of decoded clusters by the first decoder, multiplied by M_2 , the number of items in each cluster. The total complexity, combining first and second stages, is called the “overall search complexity”.

Theorem 1 provides us with a set of achievable rate quadruples that is related to different clustering and decoding schemes. In the rest of the paper, we analyze achievable clustering and decoding schemes that help us reduce the search complexity with respect to exhaustive search and investigate the search-memory complexity trade-off.

Remark 2. In this paper, we only addressed the two-stage identification system, however, it might be of interest to consider multiple-stage identification schemes in which more-than-two-stage decoding can be deployed. We do expect that using multi-stage identification can further reduce the search-complexity.

IV. PROPERTIES OF THE REGION OF ACHIEVABLE RATE QUADRUPLE

In this section we analyze some properties of the achievable region and examine it by utilizing different decoding approaches at the first decoder, i.e., single-output decoder versus list decoder, and using different database organizations.

A. Convexity

Proposition 1. The region of achievable rate quadruples of the generalized two-stage identification system is convex.

Proof. Appendix B. □

B. First decoder as a single-output decoder

As mentioned in Theorem 1 the two-stage identification scheme can achieve the identification capacity if the first decoder outputs a list of indices with rate $R_3 \geq 0$. In this section, we explore a condition in which the first decoder can act as a single-output decoder, i.e., $R_3 = 0$, and achieve the identification capacity.

Corollary 1.1. If the first decoder at the two stage identification setup outputs only a single index, i.e., $R_3 = 0$, then the Markov chain condition $X \leftrightarrow Y \leftrightarrow U$ must hold, i.e., $P(x, y, u) = Q_b(x)Q_c(y|x)P(u|y)$, to achieve C_{id} . The resulting achievable rate region $\mathcal{R}$ reduces to the following set

$$\left\{ (R_1, R_2, R) : R_1 \geq I(Y; U), \right. \\ \left. R_2 \geq \max(0, R - I(X; U)) \right\}$$

Fig. 4. Search complexity of the two-stage decoding scheme. Upon receiving y^N , first all centroids of clusters have been checked to choose the related clusters and construct the list $\mathbf{w}_1$. M_3 blocks of memory is allocated to store the chosen cluster indices. At the second step, the members of the detected clusters have been checked to decode the item that is related to y^N .

$$0 \leq R \leq I(X; Y) \}. \quad (17)$$

Proof. From Theorem 1 if the first decoder outputs a single index, then $0 = R_3 \geq I(X; U|Y)$. Hence, $I(X; U|Y) = 0$. This implies the Markov chain $X \leftrightarrow Y \leftrightarrow U$. Subsequently, we have $R_1 \geq I(X, Y; U) = I(Y; U)$. $\square$

Remark 3. Corollary 1.1 coincides with the results shown by [22].

C. Non-excess rate

Here we introduce the “excess rate” as $\Delta \triangleq R_1 + R_2 - R$. From Theorem 1, the two-stage identification scheme can achieve the identification capacity while the total rate of the clustered items $R_1 + R_2$ is greater than or equal to the total rate of items R , i.e., $R_1 + R_2 \geq R$, i.e., $\Delta \geq 0$. In the following we consider the situation in which there is no excess rate, i.e., $\Delta = 0$.

Corollary 1.2. In the two stage identification setup, if the excess rate is zero, i.e., $\Delta = R_1 + R_2 - R = 0$, then $U \leftrightarrow X \leftrightarrow Y$ must form a Markov chain, i.e., $P(x, y, u) = Q_s(x)Q_c(y | x)P(u | x)$, to achieve C_{id} . The resulting achievable rate region $\mathcal{R}$ is given by

$$\begin{aligned} \{ (R_1, R_2, R_3, R) : & R_1 \geq I(X; U), \\ & R_2 \geq \max(0, R - I(X; U)), \\ & R_3 \geq I(U; X|Y), \\ & 0 \leq R \leq I(X; Y) \}. \end{aligned} \quad (18)$$

Proof. From Theorem 1, we have

$$\begin{aligned} \Delta = R_1 + R_2 - R \geq & \begin{cases} I(U; X, Y) + R - I(U; X) - R = I(U; Y | X) & \text{if } R \geq I(U; X) \\ I(U; Y | X) + I(U; X) - R \geq I(U; Y | X) & \text{if } R < I(U; X). \end{cases} \end{aligned} \quad (19)$$

Thus $\Delta \geq I(U; Y | X)$ and the excess rate is always bigger than or equal to zero, i.e., $\Delta \geq 0$. If the excess rate is zero, then we have $I(U; Y | X) = 0$ that implies $U \leftrightarrow X \leftrightarrow Y$. Following the Markov chain $U \leftrightarrow X \leftrightarrow Y$ condition or equivalently $P(x, y, u) = Q_s(x)Q_c(y | x)P(u | x)$, we have from (16), that $R_1 \geq I(X, Y; U) = I(X; U)$. $\square$

Remark 4. From Corollary 1.2, the excess rate satisfies $\Delta \geq I(U; Y | X) \geq 0$ equality holds iff $U \leftrightarrow X \leftrightarrow Y$. Moreover, in the single-output decoder case the Markov chain $X \leftrightarrow Y \leftrightarrow U$ must hold, which implies that the excess rate can not be equal to zero in this case, unless $U = \emptyset$ or $U = X = Y$. Therefore, one can conclude that if the first decoder outputs only a single index, except some special cases, the excess rate is always positive.

Remark 5. The Markov chain $U \leftrightarrow X \leftrightarrow Y$ is also utilized in an information retrieval setting introduced in [25], [26], where, to reduce the search complexity, the authors proposed clustering based on k -means over the database entries. For a given query, to find the most related contents they investigated clusters whose centroids are among the closest ones to the query. However, contrary to our identification setup, in which the identification’s output is only a single index of an enrolled item, their retrieval system outputs a list of contents.

V. SEARCH-MEMORY COMPLEXITY

In this section, we consider the search-memory complexity of the generalized two-stage decoding scheme introduced in Section III-A.

A. Memory-Complexity Exponent

As mentioned in the proof of Theorem 1 (Appendix A), using the proposed scheme, we try to cover the space $\mathcal{X}^N \times \mathcal{Y}^N$. Fixing $P(u | x, y)$, we generate a collection of $M_1 \approx 2^{NI(U; X, Y)}$ cluster centroids $\mathcal{C}_u = \{u^N(1), \dots, u^N(M_1)\}$, where $u^N(w_1), 1 \leq w_1 \leq M_1$, is generated according to (42).

The memory-complexity (illustrated in Figure 3) related to the memory allocation of $\mathcal{C}_u$, using naïve memory allocation, which is assigning $N \log_2 |\mathcal{U}|$ bits for each $u^N(w_1) \in \mathcal{C}_u$ is given by

$$M_{\text{comp}_u} = N \log_2 |\mathcal{U}| 2^{NI(U;X,Y)}. \quad (20)$$

In the enrollment phase, for each cluster centroid $u^N(w_1) \in \mathcal{C}_u$ the cluster $\mathcal{K}(u^N(w_1))$ is constructed. From Theorem 1

- if $R > I(U; X)$ there are approximately $2^{N[R-I(U;X)]}$ sequences $x^N(w) \in \mathcal{C}$ jointly typical with a single $u^N(w_1) \in \mathcal{C}_u$. Therefore, the cardinality of the cluster $\mathcal{K}(u^N(w_1))$ is approximately $2^{N[R-I(U;X)]}$.
- if $R \leq I(U; X)$ there is at most a single sequence $x^N(w) \in \mathcal{C}$ jointly typical with a single sequence $u^N(w_1) \in \mathcal{C}_u$. The cardinality of the cluster $\mathcal{K}(u^N(w_1))$ is at most 1.

Consequently, the overall memory storage needed for this two-stage identification using $N \log_2 |\mathcal{X}|$ bits for each $x^N(w) \in \mathcal{C}$, is given by

$$M_{\text{comp}} = \begin{cases} M_{\text{comp}_u} + N \log_2 |\mathcal{X}| 2^{N[I(U;X,Y)+R-I(U;X)]} & \text{if } R > I(U; X), \\ M_{\text{comp}_u} + N \log_2 |\mathcal{X}| 2^{NI(U;X,Y)} & \text{if } R \leq I(U; X). \end{cases} \quad (21)$$

Therefore, the memory-complexity exponent M_e of the scheme is

$$\begin{aligned} M_e &= \limsup_{N \rightarrow \infty} \frac{1}{N} \log M_{\text{comp}} \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \log \left(2^{NI(U;X,Y)} + 2^{N[\max\{R+I(U;Y|X), I(U;X,Y)\}]} \right) \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \log \left(2^{N[\max\{I(U;X,Y), R+I(U;Y|X)\}]} \right) \\ &= \max \{ R + I(U; Y|X), I(U; X, Y) \}. \end{aligned} \quad (22)$$

B. Search-Complexity Exponent

As mentioned in the proof of Theorem 1 (Appendix A), for a given query y^N , the first decoder constructs the list $\mathbf{w}_1$ with a cardinality of $M_3 \approx 2^{NI(U;X|Y)}$ using the following strategy:

- checking all cluster centroids $u^N(w_1) \in \mathcal{C}_u$ and finding out which $u^N(w_1)$ is jointly typical with y^N , i.e., $(y^N, u^N(w_1)) \in \mathcal{A}_\epsilon^{(N)}(UY)$.

Consequently, the first stage search complexity (illustrated in Figure 4) is given by

$$S_{\text{comp}_1} = \mathcal{O} \left(N 2^{NI(U;X,Y)} \right), \quad (23)$$

where $\mathcal{O}(N)$ indicates the required complexity to check each cluster centroid $u^N(w_1) \in \mathcal{C}_u$.

Next, the second decoder for each $w_1 \in \mathbf{w}_1$ considers the items belong to the cluster $\mathcal{K}(u^N(w_1))$. All these sequences should be checked to determine the one that $(x^N(w), y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)$. This requires,

- if $R > I(U; X)$, $2^{NI(U;X|Y)} 2^{N[R-I(U;X)]} = 2^{N[R+I(U;X|Y)-I(U;X)]}$ checks,
- if $R \leq I(U; X)$, $2^{NI(U;X|Y)}$ checks.

Therefore, the second stage search complexity turns out to be

$$S_{\text{comp}_2} = \begin{cases} \mathcal{O} \left(N 2^{N[R+I(U;X|Y)-I(U;X)]} \right) & \text{if } R > I(U; X) \\ \mathcal{O} \left(N 2^{NI(U;X|Y)} \right) & \text{if } R \leq I(U; X) \end{cases}. \quad (24)$$

Consequently, it can be shown that the overall search-complexity exponent S_e related to the two-stage decoding scheme is

$$\begin{aligned} S_e &= \limsup_{N \rightarrow \infty} \frac{1}{N} \log(S_{\text{comp}_1} + S_{\text{comp}_2}) \\ &= \max \left\{ I(U; X, Y), \right. \\ &\quad \left. \max\{R + I(U; X|Y) - I(U; X), I(U; X|Y)\} \right\} \\ &= \max \{ I(U; X, Y), R + I(U; X|Y) - I(U; X) \}. \end{aligned} \quad (25)$$

VI. REDUCING THE SEARCH COMPLEXITY

In this section, we investigate necessary conditions to reduce the search complexity of the two-stage scheme with respect to an identification system using the exhaustive search approach.

Assume that the collection of cluster centroids $\mathcal{C}_u$ has been generated randomly according to the distribution given in (42). The following two propositions state necessary conditions to reduce the search-complexity and its lower bound.

Remark 6. *The main objective of the two-stage identification setup is to reduce the search complexity in identification systems, and to achieve the identification capacity. To achieve this goal, two-stage decoding based on clustering should satisfy the following conditions. First, the constraints in (16) should be fulfilled. Secondly, the number of items M should be larger than the number of clusters $M_1 \approx 2^{NI(U;X,Y)}$, equivalently $R > I(U; X, Y)$. In other words, if $R \leq I(U; X, Y)$, then $\mathcal{O}(N 2^{NI(U;X,Y)})$, search complexity of the first decoder, exceeds $\mathcal{O}(N 2^{NR})$, the overall search complexity of exhaustive search. Consequently, when using the clustering scheme, we assume $R > I(U; X, Y)$.*

Proposition 2. *The search complexity in an identification system using the two-stage decoding scheme can be reduced with respect to the exhaustive search scheme if one of the two following conditions is satisfied:*

$$S_e = I(U; X, Y) \text{ if } I(U; X, Y) < R \leq I(U; X) + I(U; Y) \quad (26a)$$

$$S_e = R + I(U; X|Y) - I(U; X) \text{ if } I(U; X, Y) \leq I(U; X) + I(U; Y) < R. \quad (26b)$$

Proof. Appendix C. $\square$

Proposition 3. *The search complexity in an identification system using the two-stage decoding scheme is lower bounded as $S_e \geq R/2$. Equality occurs if the condition $I(U; X, Y) = I(U; X) = I(U; Y) = R/2$ holds. As a special case this condition holds when $X = Y$, i.e., there is no noise in the identification stage.*

Proof. Appendix D. $\square$

Remark 7. *When the first decoder acts as a single-output decoder, from Corollary 1.1 we have the Markov chain $X \leftrightarrow Y \leftrightarrow U$, and thus $R > I(U; X, Y) = I(U; Y) > I(U; X)$. Subsequently, the search complexity exponent S_e and the memory complexity exponent M_e are given by*

$$S_e = \max\{I(U; Y), R - I(U; X)\}, \quad (27a)$$

$$M_e = \max\{R + I(U; Y|X), I(U; Y)\} = R + I(U; Y|X). \quad (27b)$$

Remark 8. *When the excess rate is zero, from Corollary 1.2 we have the Markov chain $U \leftrightarrow X \leftrightarrow Y$, and thus $R > I(U; X, Y) = I(U; X)$. Subsequently, the search complexity exponent S_e and the memory complexity exponent M_e are given by*

$$S_e = \max\{I(U; X), R - I(U; Y)\}, \quad (28a)$$

$$M_e = \max\{R, I(U; X)\} = R. \quad (28b)$$

From (28b) one can conclude that generalized two-stage identification under the Markov chain condition $U \leftrightarrow X \leftrightarrow Y$, increases the memory complexity linearly, since it has the same memory complexity exponent as exhaustive search, which is equal to R . However, it reduces the search complexity exponentially.

VII. EXAMPLES

A. Binary Source

We consider here an identification system with binary uniform sequences. Thus $Q_s(x) = 1/2$ for $x \in \{0, 1\}$. We also, assume a binary symmetric observation chan-

nel, which implies $Q_c(y|x) = q$ if $y \neq x$ and $Q_c(y|x) = 1 - q$ if $y = x$, where $y \in \{0, 1\}$.

It is important to observe that the “backward” channel from Y to X is also binary symmetric with crossover probability q since

$$Q_{X|Y}(x|y) = \frac{Q_s(x)Q_c(y|x)}{\sum_x Q_s(x)Q_c(y|x)} = Q_c(y|x). \quad (29)$$

Therefore $X = Y \oplus Z_b$ where $\oplus$ denotes modulo-2 addition and $Z_b \sim \text{Bernoulli}(q)$ is additive noise independent of Y .

As such, we have

$$\begin{aligned} I(U; Y) &= 1 - H(Y|U), \\ I(U; X) &= 1 - H(X|U). \end{aligned} \quad (30)$$

1) *First decoder as a single-output decoder:* From Corollary 1.1, the Markov chain $X \leftrightarrow Y \leftrightarrow U$ must hold. Since the channel from Y to X is binary additive with crossover probability q , it follows from Mrs. Gerber’s Lemma shown by [27] that if $H(Y|U) = v$ then $H(X|U) \geq H_2(q \star H_2^{-1}(v))$. Now, if $0 \leq p \leq 1/2$ is such that $H_2(p) = v$, then $H(Y|U) = H_2(p)$ and $H(X|U) \geq H_2(q \star p)$, where $(a \star b) = a(1-b) + b(1-a)$.

When we take the “channel” from Y to U as a binary symmetric channel with crossover probability p , the minimum $H(X|U)$ is achieved; and consequently, the region of achievable rate triples for binary uniform sequences and a binary symmetric observation channel is given by [22]

$$\begin{aligned} \left\{ (R_1, R_2, R) : R_1 \geq 1 - h(p), \right. \\ R_2 \geq \max(0, R - 1 + h(p \star q)), \\ \left. 0 \leq R \leq 1 - h(q), \text{ for } 0 \leq p \leq 1/2 \right\}. \end{aligned} \quad (31)$$

From (18), (23) and (24) the number of cluster-checks that have to be made by the first decoder is roughly $2^{NR_1} = 2^{N(1-H_2(p))}$ and the number of refinement-checks made by the second decoder is approximately $2^{NR_2} = 2^{N(R-1+H_2(p \star q))}$ for $0 \leq p \leq 1/2$.

2) *Non-excess rate:* From Corollary 1.2, the Markov chain $U \leftrightarrow X \leftrightarrow Y$ must hold. Following from Mrs. Gerber’s Lemma [27], if $H(X|U) = v'$ then $H(Y|U) \geq H_2(q \star H_2^{-1}(v'))$. If now $0 \leq p' \leq 1/2$ is such that $H_2(p') = v'$ then $H(X|U) = H_2(p')$ and $H(Y|U) \geq H_2(q \star p')$.

When we take the “channel” from X to U as a binary symmetric with crossover probability p' the minimum $H(Y|U)$ is achieved. Therefore, the region of achievable quadruple rates for binary uniform sequences and a

Fig. 5. Optimum cluster-refinement rates-pairs for a system with uniform binary sequences and a binary symmetric observation channel with crossover probability $q = 0.1$, for identification rates $R = 0.5310$ (maximum), 0.3540 , and 0.1770 . The solid line indicates the equality of cluster and refinement rates.

binary symmetric observation channel is given by

$$\begin{aligned} & \{(R_1, R_2, R_3, R) : \\ & R_1 \geq 1 - H_2(p'), \\ & R_2 \geq \max(0, R - 1 + H_2(p')), \\ & R_3 \geq H_2(p' \star q) - H_2(p'), \\ & 0 \leq R \leq 1 - H_2(q), \text{ for } 0 \leq p' \leq 1/2\}. \end{aligned} \quad (32)$$

Similarly, using (32), (23) and (24) and the $U \leftrightarrow X \leftrightarrow Y$ condition the complexity of the first and second decoders are approximately $2^{NR_1} = 2^{N(1-H_2(p'))}$ and $2^{N(R_3+R_2)} = 2^{N(R-1+H_2(p' \star q))}$, respectively for $0 \leq p \leq 1/2$.

Remark 9. The overall search complexity for the binary setup is the same under the two Markov conditions.

Fig. 5 shows the optimal cluster-refinement rate-pairs of the two-stage identification setup under the Markov chain conditions $X \leftrightarrow Y \leftrightarrow U$ and (or) $U \leftrightarrow X \leftrightarrow Y$ for three values of the identification rate R and for a BSC observation channel with crossover probability $q = 0.1$. The solid line in Fig. 5 indicates the points that cluster and refinement rates are equal. Fig. 5 implies that the minimum search complexity, which occurs when cluster and refinement rates are equal, is more than $R/2$.

3) *General case:* Now we consider the general two-stage decoding, i.e., $P(x, y, u) = Q_s(x)Q_c(y|x)P(u|x, y)$. Let $\mathcal{U} = \mathcal{X} = \mathcal{Y} = \{0, 1\}$. It should be noted that contrary to the two Markov cases that was shown $\mathcal{U} = \{0, 1\}$ is optimal, it is not possibly enough in the general scheme and $|\mathcal{U}|$ can be bigger than 2. We have generated 10,000 trials of

Fig. 6. Red o's indicate the search-memory complexity exponents of Binary symmetric source using the covering-channel $P(u|x, y)$. Blue x's and black dots the search-memory complexity using the covering-channel $P(u|y)$ under the condition $X \leftrightarrow Y \leftrightarrow U$, and $P(u|x)$ under the condition $U \leftrightarrow X \leftrightarrow Y$, respectively. $u \in \{0, 1\}$

$P(u|x, y)$ for four different combinations of $x \in \{0, 1\}$ and $y \in \{0, 1\}$, uniformly distributed over $[0, 1]$, to generate almost all possible search-memory exponent pairs. Finally, in order to show the ability of the general scheme in reducing search complexity the points that are satisfying either of two conditions in (26) have been chosen to plot. Fig. 6 (red o's) shows search-memory complexity exponents of the binary system with the rate $R = 0.5$ and $q = 0.1$, using (22) and (25). The point $(0.5, 0.5)$ is related to the exhaustive search scheme that we have $R_e = S_e = R$. As it was expected from Proposition 3, Fig. 6 shows that the minimum search complexity exponent even for the general scheme is bigger than $R/2 = 0.25$.

We have also generated 10,000 trials of $P(u|y)$ and $P(u|x)$ to specify the search-memory complexity exponents under the Markov chain $X \leftrightarrow Y \leftrightarrow U$ and $U \leftrightarrow X \leftrightarrow Y$, respectively. Fig. 6 (blue x's) and (black dots) show the search-memory complexity exponents under the Markov chain $X \leftrightarrow Y \leftrightarrow U$ and $U \leftrightarrow X \leftrightarrow Y$, respectively. Similar to the general case, we have only chosen the points that are satisfying either of two conditions in (26) to plot. Clearly, the two-stage decoding under the general condition can achieve a lower search-complexity compared to the given Markov conditions, with the price of higher memory-complexity exponent compared to the non-excess rate case, i.e., $U \leftrightarrow X \leftrightarrow Y$.

B. Gaussian Source

A Gaussian identification system is based on *Gaussian source* $\{G_p; x \in \mathcal{R}, y \in \mathcal{R}\}$ that produces a sequence

Fig. 7. Additive equivalent relation between two unit variance correlated variables X and Y .

Fig. 8. Additive equivalent relation between two unit variance correlated variables Y and U .

$x^N = (x_1, x_2, \dots, x_N)$ with N -real valued symbols and a sequence $y^N = \{y_1, y_2, \dots, y_N\}$ also having N real valued components. The density corresponding to the sequence (X^N, Y^N) is given by

$$p(x^N, y^N) = \prod_{n=1}^N G_\rho(x_n, y_n), \quad (33)$$

where

$$G_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right), \quad (34)$$

for $x \in \mathcal{R}$, $y \in \mathcal{R}$ and correlation coefficient $|\rho| < 1$. Observe that the source pairs $\{(X_n, Y_n), n = 1, \dots, N\}$ are independent of each other and identically distributed according to $G_\rho(\cdot, \cdot)$. Also note that scaling can be applied to obtain unit X -variance and unit Y -variance.

Thus, we have

$$\begin{aligned} I(U; Y) &= \frac{1}{2} \log_2(2\pi e) - h(Y|U), \\ I(U; X) &= \frac{1}{2} \log_2(2\pi e) - h(X|U). \end{aligned} \quad (35)$$

1) *First decoder as a single-output decoder:* We are now ready to use Shannon's entropy power inequality shown by [28]. For a simple proof of this inequality see [29]. We use here a conditional version of the entropy power inequality similar to Lemma II in [30]. However, we must first transform the statistical relation between X and Y as described in (34) into an additive version. Note that

$$X = \rho Y + Z_g, \quad (36)$$

where Z_g is Gaussian with zero mean and variance $1-\rho^2$, and independent of Y , see Figure 7.

From the (conditional version of the) entropy power inequality, one can conclude that if $h(Y|U) = \frac{1}{2} \log_2(2\pi e(1-\alpha^2))$ then $h(X|U) \geq \frac{1}{2} \log_2(2\pi e(1-\alpha^2\rho^2))$. Note that we may assume that $-1 < \alpha < 1$ since X has unit variance.

When we take the "channel" from Y to U as an Additive White Gaussian Noise (AWGN) with variance $(1-\alpha^2)$, see Figure 8, the minimum $h(X|U)$ is achieved and consequently the region of achievable triple rates for

Fig. 9. Additive equivalent relation between two unit variance correlated variables X and Y , where $Y = \rho X + Z_g$.

the Gaussian identification system is given by

$$\begin{aligned} \{(R_1, R_2, R) : \\ R_1 &\geq \frac{1}{2} \log_2\left(\frac{1}{1-\alpha^2}\right), \\ R_2 &\geq \max\left(0, R - \frac{1}{2} \log_2\left(\frac{1}{1-\alpha^2\rho^2}\right)\right), \\ 0 &\leq R \leq \frac{1}{2} \log_2\left(\frac{1}{1-\rho^2}\right), \text{ for } -1 < \alpha \leq 1\}. \end{aligned} \quad (37)$$

Using (37), (23) and (24), under the $X \leftrightarrow Y \leftrightarrow U$ conditions the complexity of the first and second decoders are $2^{NR_1} = 2^{-\frac{N}{2} \log_2(1-\alpha^2)}$ and $2^{NR_2} = 2^{N(R + \frac{1}{2} \log_2(1-\alpha^2\rho^2))}$ for $0 < \alpha \leq 1$, respectively.

2) *Non-excess rate:* We consider here a Gaussian identification setup similar to Section VII-B, where i.i.d. joint Gaussian sequences (X^N, Y^N) are distributed according to (34). We transform the statistical relation between X and Y as described in (34) into an additive version. Note that

$$Y = \rho X + Z_g, \quad (38)$$

where Z_g is Gaussian with zero mean and variance $1-\rho^2$, and independent of X , see Figure 9.

From the (conditional version of the) entropy power inequality, one can conclude that if $h(X|U) = \frac{1}{2} \log_2(2\pi e(1-\beta^2))$ then $h(Y|U) \geq \frac{1}{2} \log_2(2\pi e(1-\beta^2\rho^2))$. Note that we may assume that $-1 < \beta < 1$ since Y has unit variance.

When we take the "channel" from U to X as an AWGN with variance $(1-\beta^2)$, see Figure 10, the minimum $h(Y|U)$ is achieved; and consequently, the region of achievable quadruple rates for the Gaussian

Fig. 10. Additive equivalent relation between two unit variance correlated variables X and U .

identification system is given by

$$\begin{aligned} & \{(R_1, R_2, R_3, R) : \\ & R_1 \geq \frac{1}{2} \log_2 \left(\frac{1}{1 - \beta^2} \right), \\ & R_2 \geq \max \left(0, R - \frac{1}{2} \log_2 \left(\frac{1}{1 - \beta^2} \right) \right), \\ & R_3 \geq \frac{1}{2} \log_2 \left(\frac{1 - \beta^2 \rho^2}{1 - \beta^2} \right), \\ & 0 \leq R \leq \frac{1}{2} \log_2 \left(\frac{1}{1 - \rho^2} \right), \text{ for } -1 < \beta < 1 \}. \end{aligned} \quad (39)$$

Similarly, using (39), (23) and the (24), $U \leftrightarrow X \leftrightarrow Y$ condition the complexity of the first and second decoders are $2^{NR_1} = 2^{-\frac{N}{2} \log_2(1 - \beta^2)}$ and $2^{N(R_2 + R_3)} = 2^{N(R + \frac{1}{2} \log_2(1 - \beta^2 \rho^2))}$ for $0 < \beta \leq 1$, respectively.

Remark 10. The overall search complexity for the Gaussian setup is the same under these two Markov conditions.

Fig. 11 shows the optimal cluster-refinement rate-pairs of the two-stage Gaussian identification setup under the Markov chain conditions $X \leftrightarrow Y \leftrightarrow U$ and (or) $U \leftrightarrow X \leftrightarrow Y$ for three values of the identification rate R and for normalized correlation coefficient $\rho = 0.75$. The solid line in Fig. 11 indicates the points that cluster and refinement rates are equal. Fig. 11 implies that the minimum search complexity under these two Markov conditions is more than $R/2$.

3) *General case:* Now we consider the general two-stage scheme. Let (X^N, Y^N, U^N) be i.i.d. jointly Gaussian distributed sequences with covariance matrix

$$K_{XYU} = \begin{bmatrix} 1 & \rho & \beta \\ \rho & 1 & \alpha \\ \beta & \alpha & 1 \end{bmatrix}, \quad (40)$$

where $\rho = 0.95$, β and α are the normalized correlation coefficients between X and Y , X and U and Y and U , respectively. To generate almost all possible search-memory exponent pairs, we have generated 10,000 trials of correlation coefficients α and β at random, which

Fig. 11. Optimum cluster-refinement rates-pairs for a system with correlated Gaussian sequences with normalized correlation coefficient $\rho = 0.75$, for identification rates $R = 0.5367$ (maximum), 0.3975 , and 0.1988 . The solid line indicates the equality of cluster and refinement rates.

are uniformly distributed over $(-1, 1)$. Only α 's and β 's have been chosen that satisfy the positive-definiteness condition of K_{XYU} . And, similar to the binary case, in order to show the ability of the general scheme in reducing the search complexity the points that are satisfying either of two conditions in (26) have been chosen to plot. In Fig. 12 red o's show the achievable search-memory complexity exponent pairs for the Gaussian identification system with rate $R = 0.5$, using (25) and (22). Also in Fig. 12 blue x's and black dots show the search-memory complexity exponent pairs under the Markov chain conditions $X \leftrightarrow Y \leftrightarrow U$ and $U \leftrightarrow X \leftrightarrow Y$, respectively. As it was expected due to Proposition 3, Fig. 12 shows that the minimum search complexity exponent is bigger than $R/2 = 0.25$.

Fig. 12 implies that the generalized two-stage identification scheme can achieve lower search-complexity exponents that are not achievable under these Markov conditions with the price of higher memory-complexity exponent compared to non-excess rate case, i.e., $U \leftrightarrow X \leftrightarrow Y$.

C. Structured covering code

Although our investigation suggests that our random cluster centroids does not contain structure we could use a structured vector quantizer in practice as the first decoder. In such a situation the search complexity of the first decoder is not relevant, and the second decoder plays the role in overall search complexity of the system. Hence, the search-complexity exponent of the two-stage identification system is given by

$$S_e = \limsup_{N \rightarrow \infty} \frac{1}{N} \log(S_{\text{comp}_2})$$

Fig. 12. Red o's indicate the search-memory complexity exponent pairs of Gaussian source using the covering-channel $P(u|x, y)$. (blue $\times$) and (black dots) the search-memory complexity using the covering-channel $P(u|y)$ under the condition $X \leftrightarrow Y \leftrightarrow U$, and $P(u|x)$ under the condition $U \leftrightarrow X \leftrightarrow Y$, respectively.

$$= \max \{R + I(U; X|Y) - I(U; X), I(U; X|Y)\}. \quad (41)$$

Fig. 13 shows the search-memory complexity exponent based on structured covering codes for the binary and Gaussian cases. Comparing Figs. 13a and 13b with Figs. 6 and 12, one can conclude that using structured covering code leads to a more efficient two-stage identification scheme. We need to use list decoders, however.

VIII. CONCLUSIONS

In this paper, we have addressed the search complexity problem in identification systems. We have introduced a generalized two-stage decoding scheme, which is capable of achieving the identification capacity, while reducing the search complexity. We have investigated the achievable rates for this decoding procedure, and analyzed them under various conditions. We have also investigated the search-memory complexity in the two-stage decoding scheme.

In analyzing the search complexity of the first-decoder, i.e., estimation of the “similar” cluster centroids for a given query y^N , we have only considered joint typicality between the query and u^N 's cluster centroids. However, one may instead first determine the set $\mathcal{A}_\epsilon^{(N)}(U|y^N)$, and check which u^N 's from this set are in the collection of cluster centroids $\mathcal{C}_u$, approach known as *shooting* [31], that may reduce the first stage search complexity with respect to the scheme we introduced in this paper. Moreover, we have analyzed the rate quadruples under some Markov conditions for the binary

and Gaussian cases. However, analyzing these achievable regions may also be very interesting for the general case, which might therefore be considered as a further research topic. In addition, we have only considered a two-stage system here, and it may also be interesting to find the fundamental limits for multi-stage systems.

APPENDIX A PROOF OF THEOREM 1

The proof consists of an achievability part and a converse part. We start with the achievability.

A. Achievability Proof for Theorem 1

In this section, we show that there exists some two-stage identification system introduced in Section III-A, that can achieve the identification capacity under the certain conditions shown in (16).

1) *Cluster centroids Generation*: Fix an $0 < \epsilon < 1$, a distribution $P(x, y, u) = Q_s(x)Q_c(y|x)P(u|x, y)$, and an identification rate $0 \leq R \leq I(X; Y)$. Note that $M = 2^{NR}$ sequences extracted from enrolled items are assumed to be generated according to $\{Q_s(x), x \in \mathcal{X}\}$. These sequences are indexed by $w \in \mathcal{M} = \{1, 2, \dots, M\}$, hence we obtain a collection $\{x^N(1), x^N(2), \dots, x^N(M)\}$.

We use a random coding argument to construct a collection of cluster centroids $\{u^N(1), u^N(2), \dots, u^N(M_1)\}$ according to

$$P(u) = \sum_{x, y} Q_s(x)Q_c(y|x)P(u|x, y). \quad (42)$$

These cluster centroids are indexed by $w_1 \in \{1, 2, \dots, M_1\}$.

The system now assigns the enrolled items to different clusters by finding which sequences are jointly typical with the cluster centroid $u^N(w_1)$ for each w_1 . The system creates M_1 clusters $\mathcal{K}(u^N(w_1)) \triangleq \{w : (x^N(w), u^N(w_1)) \in \mathcal{A}_\epsilon^{(N)}(X, U)\}$, including the indices of each cluster member. A problem occurs if the number of items satisfies this condition for some cluster centroid, i.e., $(x^N(w), u^N(w_1)) \in \mathcal{A}_\epsilon^{(N)}(X, U)$ is more than M_2 since there are only M_2 blocks in memory available for each cluster to allocate their members. These clusters stored in the memory can be accessed by the second decoder.

During identification, the first decoder upon receiving y^N forms a list of cluster centroid indices $\mathbf{w}_1 \triangleq \mathcal{L}(y^N) = \{w_1 : (u^N(w_1), y^N) \in \mathcal{A}_\epsilon^{(N)}(Y, U)\}$, where $w_1 \in \{1, 2, \dots, M_1\}$. The list $\mathcal{L}(y^N)$ is conveyed to the second decoder. A problem now occurs if for a given y^N the number of cluster centroids satisfying

Fig. 13. (red o's) indicate the exponent of the search-memory complexity of using structured covering-code: (a) Binary symmetric source, (b) Gaussian source.

$u^N(w_1), y^N \in \mathcal{A}_\epsilon^{(N)}(Y, U)$ is more than M_3 since there is only room for conveying M_3 cluster centroids indices to the second decoder.

The second decoder for all detected cluster centroid indices $w_1 \in \mathcal{L}(y^N)$ and their members $w \in \mathcal{K}(u^N(w_1))$ checks to see if there exists some $x^N(w) \in \mathcal{C}$ such that $(x^N(w), y^N) \in \mathcal{A}_\epsilon^{(N)}(X, Y)$. If it finds a unique w , this w is the decoder output, if not, an error is declared.

2) *Error Probability:* We now want to be more specific about the errors that can occur. Suppose that $x^N(w)$ occurs and y^N results. The first three cases reflect the events that prevent the correct w from being seen by the second decoder. The fourth case corresponds to the event where some other $w' \neq w$ is decoded by the decoder.

- 1) If there exists no w_1 , such that $(x^N, y^N, u^N(w_1)) \in \mathcal{A}_\epsilon^{(N)}(X, Y, U)$ an error occurs. If no error occurs, then from Definition 1 we have $u^N(w_1) \in \mathcal{A}_\epsilon^{(N)}(U|y^N)$ and $x^N \in \mathcal{A}_\epsilon^{(N)}(X|u^N(w_1))$.
- 2) If the list $\mathcal{L}(y^N)$ corresponding to the received y^N is larger than M_3 an error occurs. In this case, the second decoder does not have all w_1 's for which he should check the corresponding clusters $\mathcal{K}(u^N(w_1))$.
- 3) If all clusters $\mathcal{K}(u^N(w_1))$ that contain the correct w are larger than M_2 , then the second decoder can only find part of the indices w that he should check and the correct one may not be included.
- 4) A last problem occurs if there exists a w_1 and corresponding cluster $\mathcal{K}(u^N(w_1))$ that contains a $w' \neq w$ with $(x^N(w'), y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)$. In this case an error occurs.

Then we can write for the error probability averaged over the random code construction

$$P_{\mathcal{E}} \leq \Pr \left\{ \bigcap_{w_1=1}^{M_1} (U^N(w_1), X^N(W), Y^N) \notin \mathcal{A}_\epsilon^{(N)}(XYU) \right\} \\ + \Pr \{ |\mathcal{L}(Y^N)| > M_3 \} \\ + \Pr \{ |\mathcal{K}(U^N(W_1))| > M_2 \} \\ + \Pr \left\{ \bigcup_{\substack{w' \neq W : (w' = c(w_1(w_3), w_2), \\ w_1(w_3) \in \mathcal{L}(Y^N), w_2 \in \mathcal{K}(U^N(w_1(w_3)))) \\ (X^N(w'), Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)}} \right\} \quad (43)$$

The first term of $P_{\mathcal{E}}$ can be bounded as

$$P_{\mathcal{E},1} = \sum_{x^N \in \mathcal{X}^N, y^N \in \mathcal{Y}^N} p(x^N, y^N) \\ \times \prod_{w_1=1}^{M_1} \Pr \{ U^N(w_1) \notin \mathcal{A}_\epsilon^{(N)}(U|x^N, y^N) \} \\ = \sum_{x^N \in \mathcal{X}^N, y^N \in \mathcal{Y}^N} p(x^N, y^N) \\ \times \prod_{w_1=1}^{M_1} \left(1 - \sum_{u^N \in \mathcal{A}_\epsilon^{(N)}(U|x^N, y^N)} p(u^N) \right) \\ \stackrel{(a)}{\leq} \sum_{x^N \in \mathcal{X}^N, y^N \in \mathcal{Y}^N} p(x^N, y^N) \left(1 - \right. \\ \left. 2^{-N(I(X,Y;U)+4\epsilon)} \sum_{u^N \in \mathcal{A}_\epsilon^{(N)}(U|x^N, y^N)} p(u^N|x^N, y^N) \right)^{M_1}$$

$$\begin{aligned}
& \stackrel{(b)}{\leq} \sum_{x^N \in \mathcal{X}^N, y^N \in \mathcal{Y}^N} p(x^N, y^N) \left(1 - \sum_{u^N \in \mathcal{A}_\epsilon^{(N)}(U|x^N, y^N)} p(u^N|x^N, y^N) + e^{-M_1 2^{-N(I(X, Y; U) + 4\epsilon)}} \right) \\
& = \sum_{(x^N, y^N, u^N) \notin \mathcal{A}_\epsilon^{(N)}(XYU)} p(x^N, y^N, u^N) \\
& + \sum_{x^N \in \mathcal{X}^N, y^N \in \mathcal{Y}^N} p(x^N, y^N) e^{-M_1 2^{-N(I(X, Y; U) + 4\epsilon)}} \\
& \leq 2\epsilon, \tag{44}
\end{aligned}$$

if we take $M_1 = 2^{N(I(X, Y; U) + 5\epsilon)}$, for N large enough. Here step (a) follows from the fact that for $(x^N, y^N, u^N) \in \mathcal{A}_\epsilon^{(N)}(XYU)$

$$\begin{aligned}
p(u^N) &= p(u^N|x^N, y^N) \frac{Q_s(x^N)Q_c(y^N|x^N)p(u^N)}{p(x^N, y^N, u^N)} \\
&\geq p(u^N|x^N, y^N) \\
&\times \frac{2^{-N(H(X) + \epsilon)} 2^{-N(H(Y|X) + \epsilon)} 2^{-N(H(U) + \epsilon)}}{2^{-N(H(X, Y; U) - \epsilon)}} \\
&= p(u^N|x^N, y^N) 2^{-N(I(X, Y; U) + 4\epsilon)}, \tag{45}
\end{aligned}$$

and step (b) follows from the relation $(1 - \alpha\beta)^K \leq 1 - \alpha + e^{-K\beta}$.

For the second term of $P_{\mathcal{E}}$ we obtain

$$\begin{aligned}
P_{\mathcal{E},2} &\leq \Pr\{|\mathcal{L}(Y^N)| > M_3\} \\
&\stackrel{(a)}{\leq} \frac{E[|\mathcal{L}(Y^N)|]}{M_3} \\
&\stackrel{(b)}{\leq} \frac{M_1}{M_3} 2^{-N(I(U; Y) - 3\epsilon)} \\
&\stackrel{(c)}{\leq} 2^{N(I(X, Y; U) + 5\epsilon - I(U; Y) + 3\epsilon - R_3)} \\
&\leq \epsilon \tag{46}
\end{aligned}$$

if we take $M_3 = 2^{N(I(U; X|Y) + 9\epsilon)}$.

Here step (a) follows from Markov's inequality, (b) follows from the fact that

$$\begin{aligned}
E[|\mathcal{L}(Y^N)|] &= \sum_{u^N \in \mathcal{U}^N} p(u^N(1)) \cdots \sum_{u^N \in \mathcal{U}^N} p(u^N(M_1)) \sum_{y^N \in \mathcal{Y}^N} p(y^N) \\
&\times \left(\mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(YU)}(u^N(1), y^N) + \cdots \right. \\
&\left. + \mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(YU)}(u^N(M_1), y^N) \right) \\
&= M_1 \sum_{y^N \in \mathcal{Y}^N, u^N \in \mathcal{U}^N} p(y^N) p(u^N) \mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(YU)}(y^N, u^N)
\end{aligned}$$

$$\begin{aligned}
&\leq M_1 |\mathcal{A}_\epsilon^{(N)}(YU)| 2^{-N(H(Y) - \epsilon)} 2^{-N(H(U) - \epsilon)} \\
&\leq M_1 2^{-N(I(Y; U) - 3\epsilon)}, \tag{47}
\end{aligned}$$

where $\mathbf{1}_{\mathcal{A}}(t)$ denotes the indicator function defined as

$$\mathbf{1}_{\mathcal{A}}(t) = \begin{cases} 1 & \text{if } t \in \mathcal{A}, \\ 0 & \text{if } t \notin \mathcal{A}. \end{cases}$$

And, (c) in (46) follows from (44).

For the third term of $P_{\mathcal{E}}$ we obtain

$$\begin{aligned}
P_{\mathcal{E},3} &\leq \max_{u^N} \Pr\{|\mathcal{K}(u^N)| > M_2\} \\
&\stackrel{(a)}{\leq} \max_{u^N} \frac{E[|\mathcal{K}(u^N)|]}{M_2} \\
&\stackrel{(b)}{\leq} \frac{M}{M_2} 2^{-N(I(U; X) - 3\epsilon)} \\
&\leq \epsilon \tag{48}
\end{aligned}$$

if we take $M_2 = M 2^{-N(I(U; X) - 4\epsilon)} = 2^{N(R - I(U; X) - 4\epsilon)}$ for $R > I(U; X) - 4\epsilon$, otherwise $M_2 = 1$.

Here step (a) follows from Markov inequality, and (b) follows from the fact for every u^N we have that

$$\begin{aligned}
E[|\mathcal{K}(u^N)|] &= \sum_{x^N \in \mathcal{X}^N} Q_s(x^N(1)) \cdots \sum_{x^N \in \mathcal{X}^N} Q_s(x^N(M)) \\
&\times \left(\mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(X|u^N)}(x^N(1)) + \mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(X|u^N)}(x^N(2)) + \cdots \right. \\
&\left. + \mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(X|u^N)}(x^N(M)) \right) \\
&= M \sum_{x^N \in \mathcal{X}^N} Q_s(x^N) \mathbf{1}_{\mathcal{A}_\epsilon^{(N)}(X|u^N)}(x^N) \\
&= M \sum_{x^N \in \mathcal{A}_\epsilon^{(N)}(X|u^N)} p(x^N|u^N) \frac{Q_s(x^N)p(u^N)}{p(x^N, u^N)} \\
&\leq M 2^{-N(I(X; U) - 3\epsilon)} \sum_{x^N \in \mathcal{A}_\epsilon^{(N)}(X|u^N)} p(x^N|u^N) \\
&\leq M 2^{-N(I(X; U) - 3\epsilon)}. \tag{49}
\end{aligned}$$

Finally, for the fourth term of $P_{\mathcal{E}}$ we get

$$\begin{aligned}
P_{\mathcal{E},4} &= \Pr \left\{ \bigcup_{\substack{w' \neq W : (w' = c(w_1(w_3), w_2), \\ w_1(w_3) \in \mathcal{L}(Y^N), w_2 \in \mathcal{K}(U^N(w_1(w_3)))) \\ (X^N(w'), Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)}} \right\} \\
&\leq \sum_{\substack{w=1 \\ w \neq W}}^M p(x^N) p(y^N) \\
&\leq M 2^{-N(H(X) - \epsilon)} 2^{-N(H(Y) - \epsilon)} 2^{N(H(X, Y) + \epsilon)} \\
&\leq \epsilon, \tag{50}
\end{aligned}$$

if we take $M = 2^{N(I(X;Y)+4\epsilon)}$. Note that by (48) we obtain that $M_2 = 2^{N(I(X;Y)-I(X;U))}$ if $I(X;Y) > I(X;U)$ otherwise $M_2 = 1$.

This demonstrates the achievability part corresponding to Theorem 1.

B. Converse Part

For the range M_1 of the first decision we find that

$$\begin{aligned} \log_2(M_1) &\geq H(\widehat{W}_1) \geq I(\widehat{W}_1; X^N, Y^N) \\ &= \sum_{j=1}^N I(X_j, Y_j; \widehat{W}_1 | X^{j-1}, Y^{j-1}) \\ &\stackrel{(a)}{=} \sum_{j=1}^N I(X_j, Y_j; \widehat{W}_1, X^{j-1}, Y^{j-1}) \\ &\stackrel{(b)}{=} \sum_{j=1}^N I(X_j, Y_j; U_j), \end{aligned} \quad (51)$$

where $\widehat{W}_1$ is the refined first decision. Note that it cannot take more than M_1 values. Here (a) follows from the fact that $H(X_j, Y_j | X^{j-1}, Y^{j-1}) = H(X_j, Y_j)$ since $(X_1, Y_1), (X_2, Y_2), \dots$, and (X_N, Y_N) are independent of each other, and (b) from $U_j \triangleq (\widehat{W}_1, X^{j-1}, Y^{j-1})$, $1 \leq j \leq N$. Let J be a random variable taking values in $\{1, 2, \dots, N\}$ with equal probability, and let $X = X_J$ and $Y = Y_J$, when $J = j$. Then

$$\begin{aligned} \sum_{j=1}^N I(X_j, Y_j; U_j) &= NH(X_J, Y_J | J) - NH(X_J, Y_J | U_J, J) \\ &\stackrel{(c)}{=} NH(X, Y) - NH(X, Y | U_J, J) \\ &= NI(X, Y; (U_J, J)) \\ &\stackrel{(d)}{=} NI(X, Y; U), \end{aligned} \quad (52)$$

where step (c) follows from the fact that $(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)$ are identically distributed and $X_J = X, Y_J = Y$ are compound random variables, and (d) follows from $U \triangleq (U_J, J)$.

Since $M_2 \geq 1$ we obtain for the range M_2 of the second decision that:

$$\log_2(M_2) \geq 0. \quad (53)$$

Using Fano's inequality $H(W|\widehat{W}) \leq f$, where $f \triangleq 1 + \Pr\{\widehat{W} \neq W\} \log_2(M)$, we have the following series of (in)equalities

$$\begin{aligned} \log_2(M) &= H(W) \leq H(W) - H(W|\widehat{W}) + f \\ &\leq I(W; \widehat{W}, \widehat{W}_1, W_2) + f \\ &\stackrel{(a)}{=} I(W; \widehat{W}_1) + I(W; W_2 | \widehat{W}_1) + f \end{aligned}$$

$$\begin{aligned} &\leq I(W, X^N; \widehat{W}_1) + \log_2(M_2) + f \\ &\stackrel{(b)}{=} I(X^N; \widehat{W}_1) + \log_2(M_2) + f \\ &= \sum_{j=1}^N I(X_j; \widehat{W}_1 | X^{j-1}) + \log_2(M_2) + f \\ &\leq \sum_{j=1}^N I(X_j; \widehat{W}_1, X^{j-1}, Y^{j-1}) + \log_2(M_2) + f \\ &\stackrel{(c)}{=} \sum_{j=1}^N I(X_j; U_j) + \log_2(M_2) + f \\ &\stackrel{(d)}{=} NI(X; U) + \log_2(M_2) + f. \end{aligned} \quad (54)$$

where (a) follows from the fact that $I(W; \widehat{W}_1, W_2, \widehat{W}) = I(W; \widehat{W}_1, W_2)$, (b) results from the fact that $W \leftrightarrow X^N \leftrightarrow \widehat{W}_1$, (c) results from definition $U_j \triangleq (\widehat{W}_1, X^{j-1}, Y^{j-1})$, and (d) follows the logic of (52).

From $\widehat{W}_1 \in \mathbf{W}_1 = \{W_1(1), W_1(2), \dots, W_1(M_3)\}$, $W_1(i) \in \{1, \dots, M_1\}$, $1 \leq i \leq M_3$, we obtain for the list-size M_3 , assuming that given y^N the variable $\widehat{W}_1$ can not take more than M_3 values, that:

$$\begin{aligned} \log_2(M_3) &\geq H(\widehat{W}_1 | Y^N, \mathbf{W}_1) \\ &\geq I(\widehat{W}_1; X^N | Y^N, \mathbf{W}_1) \\ &= \sum_{j=1}^N I(\widehat{W}_1; X_j | X^{j-1}, Y^N) \\ &= \sum_{j=1}^N H(X_j | X^{j-1}, Y^N) \\ &\quad - H(X_j | \widehat{W}_1, X^{j-1}, Y^{j-1}, Y_j, Y_{j+1}^N) \\ &\stackrel{(a)}{\geq} \sum_{j=1}^N H(X_j | Y_j) - H(X_j | \widehat{W}_1, X^{j-1}, Y^{j-1}, Y_j) \\ &\stackrel{(b)}{=} \sum_{j=1}^N H(X_j | Y_j) - H(X_j | U_j, Y_j) \\ &\stackrel{(c)}{=} NI(X; U | Y) \end{aligned} \quad (55)$$

where (a) follows since $(X_1, Y_1), (X_2, Y_2), \dots$, and (X_N, Y_N) are independent, (b) results from the fact that $U_j \triangleq (\widehat{W}_1, X^{j-1}, Y^{j-1})$ and since conditioning reduces entropy, and (c) follows similar logic as (52).

Finally consider the following inequality for the number of items M :

$$\begin{aligned} \log_2(M) &= H(W) \\ &\leq I(W; \widehat{W}) + f \\ &\stackrel{(a)}{\leq} I(X^N; Y^N) + f \end{aligned}$$

$$\begin{aligned}
 & \stackrel{(b)}{=} \sum_{j=1}^N I(X_j; Y_j) + f \\
 & = NI(X_J; Y_J | J) + f \\
 & \stackrel{(c)}{\leq} NI(X; Y) + f.
 \end{aligned} \tag{56}$$

where (a) follows from $I(W; \widehat{W}) \leq I(W; Y^N, C, \widehat{W}) = I(W; Y^N, C) = I(W; Y^N | C) \leq I(W, X^N; Y^N | C) \leq H(Y^N) - H(Y^N | X^N) = I(X^N; Y^N)$, (b) follows from the fact that $(X_1, Y_1), (X_2, Y_2), \dots, (Y_N, Y_N)$ are independent, and (c) is true since these pairs are identically distributed and $(X, Y) = (X_j, Y_j)$ for $J = j$.

Now, assume that (R_1, R_2, R_3, R) is achievable. Then $f \leq 1 + \epsilon \log_2(M)$. For all blocklengths N and small enough $\epsilon > 0$, we obtain from (51) and (52), (53) and (54), and (55) and (56) that

$$\begin{aligned}
 N(R_1 + \epsilon) & \geq \log_2(M_1) \geq NI(X, Y; U), \\
 N(R_2 + \epsilon) & \geq \log_2(M_2) \geq 0, \\
 N(R_2 + \epsilon) & \geq \log_2(M_2) \geq \log_2(M) - NI(X; U) - f, \\
 & \geq (1 - \epsilon)N(R - \epsilon) - 1 - NI(X; U), \\
 N(R_3 + \epsilon) & \geq \log_2(M_3) \geq NI(X; U|Y), \\
 N(R - \epsilon) & \leq \log_2(M) \leq \frac{1}{1 - \epsilon}(NI(X; Y) + 1), \tag{57}
 \end{aligned}$$

for some $p(x, y, u) = Q_s(x)Q_c(y|x)P(u|x, y)$. If we now let $\epsilon \downarrow 0$ and $N \rightarrow \infty$, then we obtain the converse of Theorem 1 from (57).

C. Bound on the Cardinality of Auxiliary Random Variable U

Note that the achievable regions that we have presented in Section III-A are given in terms of auxiliary random variable U . Therefore they cannot be computed straightforwardly, since the range of U is not specified and can, in principle, be arbitrarily large. In this section we bound the range of U . This makes it possible to characterize our achievable regions.

The problem of region characterization is a typical problem studied in multiuser information theory. It involves the support lemma of [32] and the Fenchel-Eggleston strengthening of the Caratheodory lemma, as in [33]. The arguments that we provide below are similar to those used in [34], to characterize the regions for various privacy leakage problems in biometric secrecy systems. Here, however, four constraints are used.

To find a bound on the cardinality of the auxiliary variable U let $\mathcal{D}$ be the set of probability distributions on $\mathcal{X} \times \mathcal{Y}$ and consider the $|\mathcal{X}| \cdot |\mathcal{Y}| + 2$ continuous functions of $P \in \mathcal{D}$ defined as

$$\phi_{x,y}(P) = P(x, y) \text{ for all but one } (x, y),$$

$$\begin{aligned}
 \phi_{X,Y}(P) & = H_P(X, Y), \\
 \phi_X(P) & = H_P(X), \\
 \phi_Y(P) & = H_P(Y),
 \end{aligned} \tag{58}$$

where in the last two equations we use $\Pr\{X = x\} = \sum_y P(x, y)$ and $\Pr\{Y = y\} = \sum_x P(x, y)$. By the Fenchel-Eggleston strengthening of the Caratheodory lemma [33] there are $|\mathcal{X}| \cdot |\mathcal{Y}| + 2$ elements $P_u \in \mathcal{D}$ and α_u that sum to one, such that

$$\begin{aligned}
 P(x, y) & = \sum_{u=1}^{|\mathcal{X}||\mathcal{Y}|+2} \alpha_u \phi_{x,y}(P_u) \text{ for all but one } (x, y), \\
 H(X, Y|U) & = \sum_{u=1}^{|\mathcal{X}||\mathcal{Y}|+2} \alpha_u \phi_{X,Y}(P_u), \\
 H(X|U) & = \sum_{u=1}^{|\mathcal{X}||\mathcal{Y}|+2} \alpha_u \phi_X(P_u), \\
 H(Y|U) & = \sum_{u=1}^{|\mathcal{X}||\mathcal{Y}|+2} \alpha_u \phi_Y(P_u).
 \end{aligned} \tag{59}$$

The entire probability distribution $\{Q(x, y), x \in \mathcal{X}, y \in \mathcal{Y}\}$ and consequently the entropies $H(X, Y)$, $H(X)$ and $H(Y)$ are now specified, and therefore $I(X, Y; U)$, $I(Y; U)$ and $I(X; U)$ are also specified. This implies that an auxiliary random variable U with cardinality $|\mathcal{X}| \cdot |\mathcal{Y}| + 2$ suffices for the achievable region.

APPENDIX B PROOF OF PROPOSITION 1

Proof. We need to show that region $\mathcal{R}$ of achievable quadruples (R_1, R_2, R_3, R) is convex. That is, if $(R_1^1, R_2^1, R_3^1, R^1) \in \mathcal{R}$ and $(R_1^2, R_2^2, R_3^2, R^2) \in \mathcal{R}$, then for $0 \leq \lambda \leq 1$ it holds that $(\lambda R_1^1 + (1 - \lambda)R_1^2, \lambda R_2^1 + (1 - \lambda)R_2^2, \lambda R_3^1 + (1 - \lambda)R_3^2, \lambda R^1 + (1 - \lambda)R^2) \in \mathcal{R}$.

Let $(R_1^j, R_2^j, R_3^j, R^j) \in \mathcal{R}, j = 1, 2$ be two achievable quadruples, and U_1 and U_2 be two random variables such that

$$\begin{aligned}
 R_1^j & \geq I(U_j; X, Y) \\
 R_2^j & \geq \max\{0, R - I(U_j; X)\} \\
 R_3^j & \geq I(U_j; X|Y) \\
 R^j & \leq I(X; Y), \text{ for } j = 1, 2.
 \end{aligned} \tag{60}$$

We define the random variable $U = (U_J, J)$, such that it takes value U_1 with probability λ , and U_2 with probability $1 - \lambda$, and observe that

$$\begin{aligned}
 I(U; X, Y) & = H(X, Y) - H(X, Y|U_J, J) \\
 & = (1 - \lambda)H(X, Y) - \lambda H(X, Y|U_1) \\
 & \quad - (1 - \lambda)H(X, Y|U_2) + \lambda H(X, Y)
 \end{aligned}$$

$$\begin{aligned} &= \lambda I(U_1; X, Y) + (1 - \lambda)I(U_2; X, Y) \\ &\leq \lambda R_1^1 + (1 - \lambda)R_1^2. \end{aligned} \quad (61)$$

Similarly we have

$$\begin{aligned} I(U; X|Y) &= H(X|Y) - H(X|Y, U_J, J) \\ &= (1 - \lambda)H(X|Y) - \lambda H(X|Y, U_1) \\ &\quad - (1 - \lambda)H(X|Y, U_2) + \lambda H(X|Y) \\ &= \lambda I(U_1; X|Y) + (1 - \lambda)I(U_2; X|Y) \\ &\leq \lambda R_3^1 + (1 - \lambda)R_3^2. \end{aligned} \quad (62)$$

Next observe that

$$\begin{aligned} \lambda R^1 + (1 - \lambda)R^2 &\leq \lambda I(X; Y) + (1 - \lambda)I(X; Y) \\ &= I(X; Y). \end{aligned} \quad (63)$$

Let $R_2^1 > 0$ and $R_2^2 > 0$, then we have

$$\begin{aligned} \lambda R_2^1 + (1 - \lambda)R_2^2 &\geq \lambda R^1 - \lambda I(U_1; X) + (1 - \lambda)R^2 \\ &\quad - (1 - \lambda)I(U_2; X) \\ &= \lambda R^1 + (1 - \lambda)R^2 - I(U; X). \end{aligned} \quad (64)$$

Now, consider $R_2^1 > 0$ and $R_2^2 = 0$, from (60) we have $R^2 - I(U_2; X) < 0$. Hence,

$$\begin{aligned} \lambda R_2^1 + (1 - \lambda)R_2^2 &\geq \lambda R^1 - \lambda I(U_1; X) \\ &\geq \lambda R^1 - \lambda I(U_1; X) + (1 - \lambda)R^2 \\ &\quad - (1 - \lambda)I(U_2; X) \\ &= \lambda R^1 + (1 - \lambda)R^2 - I(U; X). \end{aligned} \quad (65)$$

The same result can be shown similarly for the case $R_2^1 = 0$ and $R_2^2 > 0$. Finally, if $R_2^1 = R_2^2 = 0$, then $\lambda R_2^1 + (1 - \lambda)R_2^2 = 0$ holds.

Combining (61), (62), (63), (64) and (65), we conclude that $\mathcal{R}$ is convex. $\square$

APPENDIX C

PROOF OF PROPOSITION 2

Proof. As mentioned in Section V-B the search complexity exponent at the first stage of the proposed scheme is $\approx I(U; X, Y)$. Consequently, the two-stage decoding using random covering-code can reduce search complexity if $R > I(U; X, Y)$ is valid and thus $S_e = \max\{I(U; X, Y), R + I(U; X|Y) - I(U; X)\}$.

- The search complexity is $S_e = I(U; X, Y)$ if $I(U; X, Y) \geq R + I(U; X|Y) - I(U; X)$. Consequently, we have $I(U; X, Y) < R \leq I(U; X, Y) + I(U; X) - I(U; X|Y) = I(U; X) + I(U; Y)$ which implies condition (26a).
- The search complexity is $S_e = R + I(U; X|Y) - I(U; X)$ if (a) $R + I(U; X|Y) - I(U; X) >$

$I(U; X, Y)$ or equivalently $R > I(U; X) + I(U; Y)$. Moreover, we can reduce the search complexity if (b) $S_e = R + I(U; X|Y) - I(U; X) < R$. From condition (b) we have $I(U; X) > I(U; X|Y)$ and equivalently $I(U; Y) > I(U; Y|X)$. Combining conditions (a) and (b), we have $R > I(U; X) + I(U; Y) > I(U; X) + I(U; Y|X) = I(U; X, Y)$ which implies condition (26b). $\square$

APPENDIX D

PROOF OF PROPOSITION 3

Proof. In the followings, we show that it is not feasible to have $S_e < R/2$ using the proposed two-stage decoding scheme.

- From Proposition 2, the search complexity exponent is equal to $S_e = I(U; X, Y)$ if the condition $I(U; X, Y) < R \leq I(U; X) + I(U; Y)$ holds. Under this condition, we show that it is not feasible to have $S_e = I(U; X, Y) < R/2$. Let $2I(U; X, Y) < R$ or equivalently, $I(U; X) + I(U; Y|X) + I(U; Y) + I(U; X|Y) < R$. Since $I(U; X|Y) \geq 0$ and $I(U; Y|X) \geq 0$, then $I(U; X) + I(U; Y) < R$. However, the condition $I(U; X) + I(U; Y) \geq R$ must hold, as such $S_e = I(U; X, Y) < R/2$ is not feasible.
- From Proposition 2, the search complexity exponent is equal to $S_e = R + I(U; X|Y) - I(U; X)$ if the condition $I(U; X, Y) \leq I(U; X) + I(U; Y) < R$ holds. Consequently, we should check if it is possible to have $S_e = R + I(U; X|Y) - I(U; X) < R/2$ under the mentioned condition. We have $2I(U; X) - 2I(U; X|Y) > R$, thus $2I(U; X) + 2I(U; Y) - 2I(U; X, Y) > R$, or equivalently $I(U; X) + I(U; Y) - [I(U; X|Y) + I(U; Y|X)] > R$. As a result, $I(U; X) + I(U; Y) > R$ since $I(U; X|Y) + I(U; Y|X) \geq 0$. However, from the condition $I(U; X) + I(U; Y) < R$, thus $S_e = R + I(U; X|Y) - I(U; X) < R/2$ is not feasible.

Now, we consider conditions in which we can have $S_e = R/2$.

- We have $I(U; X, Y) = R/2$ or equivalently $I(U; X) + I(U; Y|X) + I(U; Y) + I(U; X|Y) = R$ under the condition $I(U; X, Y) < R \leq I(U; X) + I(U; Y)$. Since, $I(U; Y|X) \geq 0$ and $I(U; X|Y) \geq 0$, the equality can hold if $I(U; X) + I(U; Y) = R$ and $I(U; Y|X) = I(U; X|Y) = 0$. Moreover, from $I(U; Y|X) = I(U; X|Y) = 0$ we have $I(U; Y) = I(U; X)$, and thus $I(U; X, Y) = I(U; Y) = I(U; X) = R/2$ must hold to have equality.

- For the other condition it can be similarly shown that $I(U; X, Y) = I(U; Y) = I(U; X) = R/2$ must hold to have the equality $S_e = R/2$. $\square$

REFERENCES

- [1] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *IEEE Int. Symp. Inform. Th.*, Jun. 2003, p. 82.
- [2] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Trans. on Inf. Foren. and Sec.*, vol. 7, no. 5, pp. 1652–1667, 2012.
- [3] J. O'Sullivan and N. Schmidt, "Large deviations performance analysis for biometrics recognition," in *Proc. 40th Ann. Allerton Conf. Comm. Control, and Comput.*, Oct. 2-4 2002.
- [4] E. Chávez, G. Navarro, R. Baeza-Yates, and J. L. Marroquín, "Searching in metric spaces," *ACM Comput. Surv.*, vol. 33, no. 3, pp. 273–321, Sep. 2001.
- [5] J. Nievergelt, H. Hinterberger, and K. C. Sevcik, "The grid file: An adaptable, symmetric multikey file structure," *ACM Transactions on Database Systems*, vol. 9, pp. 38–71, 1984.
- [6] J. T. Robinson, "The k-d-b-tree: a search structure for large multidimensional dynamic indexes," in *Proceedings of the 1981 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '81. New York, NY, USA: ACM, 1981, pp. 10–18.
- [7] R. Finkel and J. Bentley, "Quad trees a data structure for retrieval on composite keys," *Acta Informatica*, vol. 4, no. 1, pp. 1–9, 1974.
- [8] A. Guttman, "R-trees: a dynamic index structure for spatial searching," *SIGMOD Rec.*, vol. 14, no. 2, pp. 47–57, Jun. 1984.
- [9] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "The r*-tree: an efficient and robust access method for points and rectangles," in *Proceedings of the 1990 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '90. New York, NY, USA: ACM, 1990, pp. 322–331.
- [10] T. Sellis, N. Roussopoulos, and C. Faloutsos, "The r+-tree: A dynamic index for multi-dimensional objects," in *Proc. 13th VLDB Conf.*, 1987, pp. 507–518.
- [11] S. Berchtold, D. A. Keim, and H.-P. Kriegel, "The x-tree: An index structure for high-dimensional data," in *Proceedings of the 22th International Conference on Very Large Data Bases*, 1996, pp. 28–39.
- [12] N. Katayama and et al., "The sr-tree: An index structure for high-dimensional nearest neighbor queries," 1997.
- [13] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An efficient access method for similarity search in metric spaces," in *Proceedings of the 23rd International Conference on Very Large Data Bases*, ser. VLDB '97. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997, pp. 426–435.
- [14] K. ip Lin, H. V. Jagadish, and C. Faloutsos, "The TV-tree – an index structure for high-dimensional data," *VLDB Journal*, vol. 3, pp. 517–542, 1994.
- [15] D. B. Lomet and B. Salzberg, "The hb-tree: a multiattribute indexing method with good guaranteed performance," *ACM Trans. Database Syst.*, vol. 15, no. 4, pp. 625–658, Dec. 1990.
- [16] S. Berchtold, C. Böhm, B. Braunmüller, D. A. Keim, and H.-P. Kriegel, "Fast parallel similarity search in multimedia databases," *SIGMOD Rec.*, vol. 26, no. 2, pp. 1–12, Jun. 1997.
- [17] R. Weber, H.-J. Schek, and S. Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces," in *24th VLDB Conf.*, Aug. 24-27 1998, pp. 194–205.
- [18] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *3rd Int. Conf. on Music Inform. Retrieval, ISMIR*, Oct. 13-17 2002, pp. 107 – 115.
- [19] P. Cano, E. Batle, T. Kalker, and J. Haitsma, "A review of algorithms for audio fingerprinting," in *5th IEEE Workshop MMSP*, Dec. 2002, pp. 169 – 173.
- [20] E. Tuncel, P. Koulgi, and K. Rose, "Rate-distortion approach to databases: Storage and content-based retrieval," *IEEE Trans. Inform. Th.*, vol. 50, no. 6, pp. 953 – 967, Jun. 2004.
- [21] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," in *IEEE Int. Symp. Inform. Th.*, Jul. 2006, pp. 1929 –1933.
- [22] F. Willems, "Searching methods for biometric identification systems: Fundamental limits," in *IEEE Int. Symp. Inform. Th.*, Jul. 2009, pp. 2241 –2245.
- [23] E. Tuncel, "Recognition capacity versus search speed in noisy databases," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, Jul 2012, pp. 2566–2570.
- [24] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.
- [25] B. Girod, V. Chandrasekhar, R. Grzeszczuk, and Y. A. Reznik, "Mobile visual search: Architectures, technologies, and the emerging MPEG standard," *IEEE Multimedia*, vol. 18, no. 3, pp. 86–94, 2011.
- [26] H. Jégou, M. Douze, and C. Schmid, "Improving bag-of-features for large scale image search," *International Journal of Computer Vision*, vol. 87, no. 3, pp. 316–336, May 2010.
- [27] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications-i," *IEEE Trans. Inform. Th.*, vol. 19, no. 6, pp. 769 – 772, Nov. 1973.
- [28] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.
- [29] S. Verdú and D. Guo, "A simple proof of the entropy-power inequality," *Information Theory, IEEE Transactions on*, vol. 52, no. 5, pp. 2165–2166, 2006.
- [30] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (corresp.)," *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 279–280, 1974.
- [31] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE Inf. Th. Work.*, Aug. 30-Sep. 3 2010.
- [32] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *Information Theory, IEEE Transactions on*, vol. 21, no. 6, pp. 629–637, 1975.
- [33] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Th.*, vol. 22, no. 1, pp. 1 – 10, Jan. 1976.
- [34] T. Ignatenko and F. Willems, "Biometric systems: Privacy and secrecy aspects," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 4, pp. 956 –973, Dec. 2009.

Farzad Farhadzadeh Farzad Farhadzadeh was born in Yazd, Iran, on February, 1983. He received the M.Sc. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran, and the Ph.D. degree from the University of Geneva, Geneva, Belgium, in 2009 and 2014 respectively. From 2014 to 2015 he was a postdoctoral fellow at the Eindhoven University of Technology.

His research contributions are in the areas of multi-user information theory and noiseless source coding. He joined ASML Holding in The Netherlands as a design engineer.

Frans Willems Frans M.J. Willems was born in Stein, The Netherlands, on June 26, 1954. He received the M.Sc. degree in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, and the Ph.D. degree from the Catholic University of Louvain, Louvain, Belgium, in 1979 and 1982 respectively. Since 1982, he is a staff member at the Electrical Engineering Department of

Eindhoven University of Technology. His research contributions are in the areas of multi-user information theory and noiseless source coding. He is co-recipient of the 1996 IEEE Information Theory Society Paper Award. From 2001 to 2004 he served as an Associate Editor for Information Theory for the European Transactions on Telecommunications. Prof. Willems is a Fellow of the IEEE since 2005.