Variations on the Guessing Problem

Robert Graczyk and Amos Lapidoth
Signal and Information Processing Laboratory
ETH Zurich, 8092 Zurich, Switzerland
Email: {graczyk, lapidoth}@isi.ee.ethz.ch

***Abstract*—Three variations on the Massey-Arikan guessing problem are considered. Their solutions provide new evidence of the duality between good guessing functions and efficient quantization schemes. They also show how type-covering can be used to provide side-information in the guessing setup.**

I. Introduction

In his seminal paper [1], Arikan related the Rényi Entropy $\mathrm{H}_\alpha(X)$ of a random variable X of finite support set $\mathcal{X}$ to the ρ-th moment of the number of guesses needed to recover its realization. He showed that, using questions of the form "Is $X = x$?",

$$\mathbb{E}[\mathrm{G}^*(X)^\rho] \approx 2^{\rho\,\mathrm{H}_{1/(1+\rho)}(X)}, \tag{1}$$

where G^* denotes the optimal guessing order, i.e., the optimal bijection $\mathcal{X} \to \{1, 2, \ldots, |\mathcal{X}|\}$; ρ is a positive constant; $\mathrm{H}_{1/(1+\rho)}(X)$ is the Rényi Entropy of order $1/(1+\rho)$; and where equality holds up to a factor dominated by $\log|\mathcal{X}|$.

In the IID case, where $X^n \sim P_X^n$ for some PMF P_X on $\mathcal{X}$,

$$\lim_{n\to\infty} \frac{\log \mathbb{E}[\mathrm{G}^*(X^n)^\rho]}{n} = \rho\,\mathrm{H}_{1/(1+\rho)}(X), \tag{2}$$

and the Rényi Entropy thus fully characterizes the exponential growth rate of $\mathbb{E}[\mathrm{G}^*(X^n)^\rho]$.

Together with Merhav [2], the preceding results were generalized to the rate-distortion guessing problem. Here the goal is to minimize the ρ-th moment of the number of guesses required until the guess $\hat{X}^n$ satisfies $d_n(X^n, \hat{X}^n) \leq D$, where $d_n(\cdot,\cdot)$ is some nonnegative distortion function. Under the usual single-letter assumption, i.e., X^n being drawn IID according to P_X and $d_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to \mathbb{R}_{\geq 0}$ being expressible as $d_n(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^n d(x_i, \hat{x}_i)$, Arikan and Merhav showed that

$$\begin{aligned}&\lim_{n\to\infty} \frac{\log \mathbb{E}[\mathrm{G}^*_{d,D}(X^n)^\rho]}{n}\\ &= \sup_{Q_X} [\rho\,\mathrm{R}_{d,D}(Q_X) - \mathrm{D}(Q_X||P_X)].\end{aligned} \tag{3}$$

Here $\mathrm{R}_{d,D}(Q_X)$ denotes the rate-distortion function of a source of law Q_X with respect to the distortion measure d and maximal-allowed distortion D, and $\mathrm{G}^*_{d,D}(\cdot)$ is the optimal guessing function in the rate-distortion setup. It is defined with respect to an implicit optimal guessing order $(\hat{x}_1^n, \hat{x}_2^n, \ldots, \hat{x}^n_{|\hat{\mathcal{X}}^n|})$ on $\hat{\mathcal{X}}^n$, and $\mathrm{G}^*_{d,D}(x^n)$ equals $i \in \{1, 2, \ldots, |\hat{\mathcal{X}}^n|\}$ if i is the lowest index for which $d_n(x^n, \hat{x}_i^n) \leq D$ (if no such i exists, then $\mathrm{G}_{d,D}(x^n)$ is defined as $+\infty$).

Here we present three extensions of these results: The first deals with a setting where X^n is described using nR bits, and the description Z_n is then revealed to the guesser (before the guessing begins). Generalizing an argument from [2], we lower-bound the least ρ-th moment of the number of required guesses. We upper-bound it by proposing a description of X^n that is based on type-covering. Using these bounds, we show that, with the optimal use of the allotted nR bits,

$$\begin{aligned}&\lim_{n\to\infty} \min_{Z_n} \frac{\log \mathbb{E}[\mathrm{G}^*(X^n|Z_n)^\rho]}{n}\\ &= \sup_{Q_X} \inf_{Q_{U|X}:\mathrm{I}(Q_{X;U})\leq R} [\rho\,\mathrm{H}(Q_{X|U}) - \mathrm{D}(Q_X||P_X)],\end{aligned} \tag{4}$$

where $G^*(\cdot|Z_n)$ is the optimal guessing function for X^n given Z_n, $\mathrm{I}(Q_{X;U})$ denotes the mutual information between X and U, and $\mathrm{H}(Q_{X|U})$ is the conditional entropy of X given U. Both $\mathrm{I}(Q_{X;U})$ and $\mathrm{H}(Q_{X|U})$ are computed with respect to $Q_{X,U} = Q_X Q_{U|X}$. By invoking the identity

$$\begin{aligned}&\sup_{Q_X} \inf_{Q_{U|X}:\mathrm{I}(Q_{X;U})\leq R} [\rho\,\mathrm{H}(Q_{X|U}) - \mathrm{D}(Q_X||P_X)]\\ &= \rho \max(\mathrm{H}_{1/(1+\rho)}(P_X) - R, 0),\end{aligned} \tag{5}$$

we show that for $X^n \sim P_X^n$ one needs roughly $n\,\mathrm{H}_{1/(1+\rho)}(P_X)$ bits of side-information to guarantee that $\lim_{n\to\infty} \mathbb{E}[\mathrm{G}^*(X^n|Z_n)^\rho] = 1$.

The second extension is presented in section III and has a rate-distortion flavor. We prove that if (X^n, Y^n) are drawn IID according to $P_{X,Y}$ and if after observing Y^n we want to guess X^n to within distortion D as measured by some single-letter distortion measure d, then the optimal rate-distortion guessing exponent is given by

$$\begin{aligned}&\lim_{n\to\infty} \frac{\log \mathbb{E}[\mathrm{G}^*_{d,D}(X^n|Y^n)^\rho]}{n}\\ &= \sup_{Q_{X,Y}} [\rho\,\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|Y}) - \mathrm{D}(Q_{X,Y}||P_{X,Y})].\end{aligned} \tag{6}$$

Here $\mathrm{G}^*_{d,D}(\cdot|\cdot)$ is the optimal conditional rate-distortion guessing function, and $\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|Y})$ denotes the conditional rate-distortion function for a source of law Q_X when side-information Y of conditional law $Q_{Y|X}$ is available to both describer and reconstructor.

The third extension is presented in section IV, where we derive the optimal guessing exponent in a rate-distortion setting where nR bits are allocated for a description Z_n of X^n. We show that the optimal guessing exponent is given by

$$\lim_{n\to\infty} \min_{Z_n} \frac{\log \mathbb{E}[\mathrm{G}^*_{d,D}(X^n|Z_n)^\rho]}{n} \tag{7}$$
$$= \sup_{Q_X} \inf_{Q_{U|X}:\mathrm{I}(Q_{X;U})\le R} [\rho\, \mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U}) - \mathrm{D}(Q_X||P_X)].$$

II. Guessing with Chosen Side-Information

Theorem 1. *The minimal achievable guessing exponent with side-information $Z_n \triangleq \phi_n(X^n)$ over all $\phi_n : \mathcal{X}^n \to \{1,2,\ldots,2^{nR}\}$ is given in* (4).

Proof. We first show that no choice of ϕ_n and no guessing strategy can yield an exponent below (4). To that end we exploit the relationship between guessing strategies and variable-length source coding [3].

We begin by introducing a data-compression setup. A helper is allotted nR bits to produce a description Z_n of X^n. The pair (X^n, Z_n) is observed by an encoder, which generates a binary description W_n of X^n. A reconstructor then recovers X^n from the pair (W_n, Z_n).

For a given guessing tuple (ϕ_n, G) we create the following instance of the above data-compression setup: The helper produces $Z_n = \phi_n(X^n)$, and the encoder uses a binary code for the positive integers $\mathbb{Z}_{>0}$ to describe the positive integer $G(X^n|Z_n)$. The code is such that each $i \in \mathbb{Z}_{>0}$ is described using $l(i)$ bits, where $l(i) = \lceil \log(i^{(1+\delta)}/C(\delta)) \rceil$. Here $\delta > 0$ is arbitrarily small and $C(\delta) = \left(\sum_{i=1}^\infty 1/i^{1+\delta}\right)^{-1}$. (The existence of such a code follows for instance from Kraft's Inequality.) The encoder thus observes (X^n, Z_n) and produces a length-$\lceil \log(\mathrm{G}(X^n|Z_n)^{(1+\delta)}/C(\delta)) \rceil$ string describing $\mathrm{G}(X^n|Z_n)$. From this description and Z_n the reconstructor recovers $\mathrm{G}(X^n|Z_n)$. It then recovers X^n from $\mathrm{G}(X^n|Z_n)$ and Z_n.

Next, let $L_n(P,R)$ denote the least average binary description length for the data-compression setup introduced above, where P denotes the source distribution and R is the rate allotted to the helper. We now relate the performance of the guessing scheme to the performance of the data-compression scheme it instantiates and then use $L_n(\cdot,\cdot)$ to bound the latter:

$$\mathbb{E}_{P_X}[\mathrm{G}(X^n|Z_n)^\rho]$$
$$\overset{(a)}{\ge} \sup_{Q_X} 2^{\mathbb{E}_{Q_X}[\log \mathrm{G}(X^n|Z_n)^\rho] - n\,\mathrm{D}(Q_X||P_X)} \tag{8}$$
$$\overset{(b)}{\ge} \sup_{Q_X} 2^{\rho \frac{\mathbb{E}_{Q_X}[l(\mathrm{G}(X^n|Z_n))]}{1+\delta} + \rho \frac{\log C(\delta) - 1}{1+\delta} - n\,\mathrm{D}(Q_X||P_X)} \tag{9}$$
$$\overset{(c)}{\ge} \sup_{Q_X} 2^{\rho n \frac{L_n(Q_X,R)}{1+\delta} + \rho \frac{\log C(\delta)-1}{1+\delta} - n\,\mathrm{D}(Q_X||P_X)}, \tag{10}$$

where $\mathbb{E}_Q$ denotes expectation with respect to Q, and with the following justification: To obtain the variational inequality (a), we express $p(x^n)$ as $q(x^n) \cdot p(x^n)/q(x^n)$ to arrive at an expectation with respect to q; we then express $p(x^n)/q(x^n) \cdot G(x^n|z_n)$ as $2^{\log \xi}$ and apply Jensen's Inequality to the convex map $\xi \mapsto 2^\xi$; in (b) we restate $\log \mathrm{G}(X^n|Z_n)$ as $\log(\mathrm{G}(X^n|Z_n)^{1+\delta} C(\delta)/C(\delta))/(1+\delta)$, apply the inequality $\xi \ge \lceil \xi \rceil - 1$ and recognize $\lceil \log(\mathrm{G}(X^n|Z_n)^{(1+\delta)}/C(\delta)) \rceil$ as $l(\mathrm{G}(X^n|Z_n))$; (c) follows from the definition of $L_n(\cdot,\cdot)$. To proceed, we state a result from [4].

Lemma 1. *For every $n \in \mathbb{Z}_{>0}$ the least average binary description length $L_n(Q_X,R)$ is lower-bounded by*

$$L_n(Q_X,R) \ge \inf_{Q_{U|X}:\mathrm{I}(Q_{X;U})\le R} \mathrm{H}(Q_{X|U}). \tag{11}$$

From (10) and the preceding lemma (with $\delta > 0$ fixed and n sent to infinity)

$$\liminf_{n\to\infty} \frac{\log \mathbb{E}_{P_X}[\mathrm{G}(X^n|Z_n)^\rho]}{n} \tag{12}$$
$$\ge \sup_{Q_X} \left[\rho \frac{\inf_{Q_{U|X}:\mathrm{I}(Q_{X;U})\le R} \mathrm{H}(Q_{X|U})}{1+\delta} - \mathrm{D}(Q_X||P_X) \right].$$

By letting δ approach 0 from above, we conclude that no guessing exponent below (4) can be achieved.

We next propose a guessing scheme that asymptotically achieves the lower bound. We begin by fixing some small $\delta > 0$ and, for every type class $\mathcal{T}^{(n)}(Q_X)$ on $\mathcal{X}^n$, we select a conditional type $Q_{U|X}$ that–among all those satisfying $\mathrm{I}(Q_{X;U}) \le R - \delta$ and $Q_U \in \mathcal{P}^{(n)}(\mathcal{U})$, i.e., Q_U being a type of denominator n on the alphabet $\mathcal{U}$–minimizes $\mathrm{H}(Q_{X|U})$. The derivation of the type-covering lemma (see for instance [5, Chapter 6, p. 152 – 153]) shows that for large enough n there exists a codebook $\mathcal{C}_{Q_X}$, such that $\log |\mathcal{C}_{Q_X}|/n \le R - \delta/2$ and such that for every $x^n \in \mathcal{T}^{(n)}(Q_X)$ we can find some $u^n \in \mathcal{C}_{Q_X}$ satisfying $(x^n, u^n) \in \mathcal{T}^{(n)}(Q_X Q_{U|X})$.

The side-information Z_n that we propose comprises two parts. The first is of length at most $(\delta/2)n$ and describes the type of X^n, which requires distinguishing between a polynomial number of outcomes. The second part is the index of some codeword $U^n \in \mathcal{C}_{Q_X}$ for which (X^n, U^n) is in $\mathcal{T}^{(n)}(Q_X Q_{U|X})$ and is thus at most of length $(R - \delta/2)n$ bits.

The guesser uses the first part of Z_n to recover the type of X_n and from it identifies the codebook $\mathcal{C}_{Q_X}$. The guesser then uses the second part of Z_n to recover U_n from $\mathcal{C}_{Q_X}$. Finally the guesser recovers X^n by sequentially guessing the elements of the conditional type class $\mathcal{T}^{(n)}(Q_{X|U}|U^n)$ in an arbitrary order. The ρ-th moment of the number of guesses can be upper-bounded as follows:

$$\mathbb{E}[\mathrm{G}(X^n|Z_n)^\rho]$$
$$= \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \mathbb{E}[\mathrm{G}(X^n|Z_n)^\rho | X^n \in \mathcal{T}^{(n)}(Q_X)] \tag{13}$$
$$\mathbb{P}[X^n \in \mathcal{T}^{(n)}(Q_X)]$$

$$\overset{(a)}{\leq} \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \mathbb{E}[\mathrm{G}(X^n|Z_n)^\rho | X^n \in \mathcal{T}^{(n)}(Q_X)] \quad 2^{-n\,\mathrm{D}(Q_X||P_X)} \tag{14}$$

$$\overset{(b)}{\leq} \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} 2^{n \min^*_{Q_{U|X}:\mathrm{I}(Q_X;U)\leq R-\delta} \rho\, \mathrm{H}(Q_{X|U})} \quad 2^{-n\,\mathrm{D}(Q_X||P_X)} \tag{15}$$

$$\overset{(c)}{\leq} \max_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \Big[2^{n \min^*_{Q_{U|X}:\mathrm{I}(Q_X;U)\leq R-\delta} \rho\, \mathrm{H}(Q_{X|U})} \quad 2^{-n\,\mathrm{D}(Q_X||P_X)+n\delta_n} \Big], \tag{16}$$

where (a) follows from Sanov's Theorem; (b) follows from the fact that in the worst case we go through all the elements of the conditionally typical set $\mathcal{T}^{(n)}(Q_{X|U}|U^n)$, the size of which is determined by the entropy of the auxiliary conditional type $Q_{X|U}$. This type is in turn induced by the choice of $Q_{U|X}$, where the notation $\min^*$ in (15) denotes that the optimization is with respect to types; and (c) follows by maximizing over the set of all types $\mathcal{P}^{(n)}(\mathcal{X})$, where the overhead of the sum is absorbed into the exponent δ_n, with the property that $\delta_n \downarrow 0$, as there are at most polynomially many types.

To recover (4) from (16), we first observe that $\mathrm{H}(Q_{X|U})$ is a continuous function with respect to $Q_{U|X}$. Since the set of types is dense in the set of all probability distributions, we may allow the minimization $\min^*_{Q_{U|X}} \mathrm{H}(Q_{X|U})$ to be carried out without the restriction to types at the expense of some small deviation δ'_n satisfying $\delta'_n \downarrow 0$ for $n \to \infty$. Therefore

$$\limsup_{n\to\infty} \frac{\log \mathbb{E}_{P_X}[\mathrm{G}(X^n|Z_n)^\rho]}{n} \leq \sup_{Q_X} \inf_{Q_{U|X}:\mathrm{I}(Q_X;U)\leq R-\delta} \left[\rho\,\mathrm{H}(Q_{X|U}) - \mathrm{D}(Q_X||P_X)\right]. \tag{17}$$

And since the above holds for for any $\delta > 0$ and $\inf \mathrm{H}(Q_{X|U})$ is a continuous function of the rate constraint R, there is indeed a choice of Z_n and a guessing scheme achieving (4). □

Before moving on, we briefly point out a consequence of this result. It has been shown [6, Corollary 7] that for any $\delta > 0$, a judicious length-$(\mathrm{H}_{1/(1+\rho)}(P_X)+\delta)n$ description of X^n suffices to drive the the ρ-th moment associated with guessing X^n to one. This is congruous with Theorem 1, which, in combination with the identity (5) implies that the guessing exponent is zero if and only if $R \geq \mathrm{H}_{1/(1+\rho)}(P_X)$. For a derivation of (5) see [4]. Also note that our choice of Z_n does not necessarily minimize $\mathbb{E}[\mathrm{G}^*(X^n|Z_n)^\rho]$; for $\rho = 1$, an explicit construction of a minimizing Z_n can be found in [7].

III. Rate-Distortion Guessing with Side-Information

We next consider a setting where $(X^n, Y^n) \sim P^n_{X,Y}$. For a given pair (d, D), the goal is to guess X^n to within distortion D after observing Y^n in as few guesses as possible. Our result is summarized in the following theorem.

Theorem 2. *With access to the side-information Y^n, the minimal achievable rate-distortion guessing exponent is given in* (6).

Proof. To see why no smaller exponent is achievable, we again use the duality between guessing and data-compression. For this guessing setup, the corresponding data-compression problem is the lossy description of X^n, where the side-information Y^n is revealed to both the encoder and the reconstructor. Every guessing function $\mathrm{G}_{d,D}(\cdot|y^n)$ induces, along with its guessing order $(\hat{x}^n_1, \hat{x}^n_2, \ldots, \hat{x}^n_{|\hat{\mathcal{X}}^n|})$, a data-compression scheme as follows: Upon observing the pair (X^n, Y^n), the encoder describes the approximation $\hat{X}^n$ of X^n by producing the length-$l(i)$ string describing the positive integer $\mathrm{G}_{d,D}(X^n|Y^n)$, where $l(i) = \lceil \log(i^{(1+\delta)}/C(\delta)) \rceil$. Using this string and Y^n, the reconstructor recovers $\mathrm{G}_{d,D}(X^n|Y^n)$. Finally $\hat{X}^n$ is obtained from $\mathrm{G}_{d,D}(X^n|Y^n)$, Y^n, and the implicit guessing order of $\mathrm{G}_{d,D}$.

Key is that the average string length in the above data-compression problem is bounded from below by the conditional rate-distortion function. With this idea in mind, we alter (8)–(10) as follows:

$$\mathbb{E}_{P_X}[\mathrm{G}_{d,D}(X^n|Y^n)^\rho]$$

$$\geq \sup_{Q_X} 2^{\mathbb{E}_{Q_X}[\log \mathrm{G}_{d,D}(X^n|Y^n)^\rho] - n\,\mathrm{D}(Q_X||P_X)} \tag{18}$$

$$\geq \sup_{Q_X} 2^{\rho \frac{\mathbb{E}_{Q_X}[l(\mathrm{G}_{d,D}(X^n|Y^n))]}{1+\delta} + \rho\frac{\log C(\delta)-1}{1+\delta} - n\,\mathrm{D}(Q_X||P_X)} \tag{19}$$

$$\geq \sup_{Q_X} 2^{\rho n \frac{\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|Y})}{1+\delta} + \rho\frac{\log C(\delta)-1}{1+\delta} - n\,\mathrm{D}(Q_X||P_X)}. \tag{20}$$

The justification for the above inequalities is analogous to the justification of (8)–(10). Observe that as mentioned above, the conditional rate-distortion function has been introduced as a lower bound in the last inequality. To recover (6) as a lower bound on $\liminf_{n\to\infty} \log \mathbb{E}_{P_X}[\mathrm{G}_{d,D}(X^n|Y^n)^\rho]/n$, we again let $n \to \infty$ and observe that (20) holds for any $\delta > 0$.

To show that there exists a guessing scheme achieving (6), we need the following lemma from [4].

Lemma 2. *For every $\delta \geq 0$, $D \geq 0$ and distortion measure d, there exists a positive integer n_0, such that for all $n \geq n_0$ and every length-n sequence y^n of type $Q_Y \in \mathcal{P}^{(n)}(\mathcal{Y})$ and every conditional type $Q_{X|Y}$ satisfying $Q_X \in \mathcal{P}^{(n)}(\mathcal{X})$, there exists a codebook $\mathcal{C}_{y^n} \subset \hat{\mathcal{X}}^n$ satisfying $|\mathcal{C}_{y^n}| \leq 2^{n(\mathrm{R}^{cond}_{d,D}(Q_{X|Y})+\delta)}$ and such that for every $x^n \in \mathcal{T}^{(n)}(Q_{X|Y}|y^n)$ there is some*

$\hat{x}^n \in \mathcal{C}_{y^n}$ *satisfying* $1/n \sum_{i=1}^n d(x_i, \hat{x}_i) \leq D$.

With Lemma 2 at hand, we can follow Arikan's universal guessing approach [2]. After observing Y^n and determining its type Q_Y, the guesser generates, for every conditional type $Q_{X|Y}$ satisfying $Q_X \in \mathcal{P}^{(n)}(\mathcal{X})$, a codebook $\mathcal{C}_{Y^n,Q_{X|Y}}$ such that for every $X^n \in \mathcal{T}^{(n)}(Q_{X|Y}|Y^n)$ there is some $\hat{X}^n \in \mathcal{C}_{Y^n,Q_{X|Y}}$ satisfying $1/n\sum_{i=1}^n d(X_i,\hat{X}_i) \leq D$ and such that the number of entries in the codebook satisfies $|\mathcal{C}_{Y^n,Q_{X|Y}}| \leq 2^{n(\mathrm{R}^{\text{cond}}_{d,D}(Q_{X|Y})+\delta)}$. The existence of such a codebook is guaranteed by Lemma 2, and $\delta > 0$ is some small constant. Since the size of $\mathcal{C}_{Y^n,Q_{X|Y}}$ only depends on Y^n via its type Q_Y, we use the notation $|\mathcal{C}_{Q_Y Q_{X|Y}}|$ whenever we refer to the cardinality of $\mathcal{C}_{Y^n,Q_{X|Y}}$.

After generating the codebooks, the guesser defines the binary relation "$\preceq$", where $Q'_{X|Y} \preceq Q_{X|Y} \implies \mathrm{R}^{\text{cond}}_{d,D}(Q'_{X|Y}) \leq \mathrm{R}^{\text{cond}}_{d,D}(Q_{X|Y})$ and arranges the elements of $\{Q_{X|Y}\}$ in ascending order of "$\preceq$". Picking an arbitrary guessing order for every codebook, the guesser then sequentially guesses elements in $\mathcal{C}_{Y^n,Q^1_{X|Y}}, \mathcal{C}_{Y^n,Q^2_{X|Y}}, \ldots, \mathcal{C}_{Y^n,Q^{|\mathcal{P}_n(X|Y)|}_{X|Y}}$. The index i on $Q^i_{X|Y}$ denotes the position of $Q_{X|Y}$ in the ascending arrangement with respect to "$\preceq$". In the worst case we go through all codebooks until and including the one corresponding to the actual joint type of (X^n, Y^n), so the ρ-th moment of the number of guesses can be bounded by

$$\mathbb{E}[\mathrm{G}^\rho_{d,D}(X^n|Y^n)]$$
$$= \sum_{Q_Y} \mathbb{E}[\mathrm{G}^\rho_{d,D}(X^n|Y^n)|Y^n \in \mathcal{T}^{(n)}(Q_Y)]\,\mathbb{P}[Y^n \in \mathcal{T}^{(n)}(Q_Y)] \tag{21}$$
$$\leq \sum_{Q_Y} \mathbb{E}[\mathrm{G}^\rho_{d,D}(X^n|Y^n)|Y^n \in \mathcal{T}^{(n)}(Q_Y)]2^{-n\,\mathrm{D}(Q_Y||P_Y)} \tag{22}$$
$$\overset{(a)}{\leq} \sum_{Q_Y}\sum_{Q_{X|Y}} \left(\sum_{Q'_{X|Y} \preceq Q_{X|Y}} |\mathcal{C}_{Q_Y Q'_{X|Y}}|\right)^\rho 2^{-n\,\mathrm{D}(Q_{X|Y}||P_{X|Y})}2^{-n\,\mathrm{D}(Q_Y||P_Y)} \tag{23}$$
$$\overset{(b)}{\leq} \sum_{Q_Y}\sum_{Q_{X|Y}} \left(\sum_{Q'_{X|Y} \preceq Q_{X|Y}} 2^{n(\mathrm{R}^{\text{cond}}_{d,D}(Q_Y Q'_{X|Y})+\delta)}\right)^\rho 2^{-n\,\mathrm{D}(Q_Y Q_{X|Y}||P_{X,Y})} \tag{24}$$
$$\overset{(c)}{\leq} \sum_{Q_Y}\sum_{Q_{X|Y}} 2^{n\rho(\mathrm{R}^{\text{cond}}_{d,D}(Q_Y Q_{X|Y})+\delta+\delta_n)}2^{-n\,\mathrm{D}(Q_Y Q_{X|Y}||P_{X,Y})} \tag{25}$$
$$\overset{(d)}{\leq} \sup_{Q_{X,Y}} 2^{n\rho(\mathrm{R}^{\text{cond}}_{d,D}(Q_Y Q_{X|Y})+\delta+\delta_n)}2^{-n\,\mathrm{D}(Q_Y Q_{X,Y}||P_{X,Y})}, \tag{26}$$

where the sums $\sum_{Q_Y}$ and $\sum_{Q_{X|Y}}$ are read as $\sum_{Q_Y \in \mathcal{P}^{(n)}(\mathcal{Y})}$ and $\sum_{Q_{X|Y} \in \mathcal{P}^{(n)}(\mathcal{X}|\mathcal{Y})}$, respectively, and with the following justification: To recover (a), observe that the size of all codebooks up to and including the one corresponding to the actual type of (X^n, Y^n) constitutes an upper bound on the expected number of guesses; (b) follows from Lemma 2; (c) is a result of the guessing order induced by "$\preceq$" and further follows from absorbing the sum overhead into the exponent δ_n; and (d) is due to a maximization over all types where the sum overhead is again included in δ_n. By letting δ approach 0 from above and $\delta_n \downarrow 0$, it follows that (6) is indeed achievable. □

IV. Rate-Distortion Guessing with a Helper

We consider a rate-distortion guessing problem where, as in section II, nR bits are alotted to create a description Z_n to help the guesser.

Theorem 3. *The minimal achievable rate-distortion guessing exponent with side-information* $Z_n \triangleq \phi_n(X^n)$ *over all* $\phi_n : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$ *is given in* (7).

Proof. In order to prove that no choice of Z_n allows for a guessing exponent below (7), we begin by reintroducing the data-compression setup from section II. However, instead of requiring that the reconstructor recovers X^n from (W_n, Z_n), we content ourselves with an approximation $\hat{X}^n$ that satisfies $\frac{1}{n}\sum_{i=1}^n d(X_i, \hat{X}_i) \leq D$.

For a given guessing tuple $(\phi_n, \mathrm{G}_{d,D})$, we instantiate this data-compression setup as follows: The helper generates $Z_n = \phi_n(X^n)$ and the encoder, observing Z_n, describes X^n by the string for the positive integer $\mathrm{G}_{d,D}(X^n|Z_n)$. This string has length $l(i)$, where $l(i) = \lceil \log(i^{(1+\delta)}/C(\delta)) \rceil$. From this description and Z_n the reconstructor recovers $\mathrm{G}_{d,D}(X^n|Z_n)$. It then recovers $\hat{X}^n$ from $\mathrm{G}_{d,D}(X^n|Z_n)$, Z_n, and the implicit guessing order of $\mathrm{G}_{d,D}$.

To continue, we need a lower bound for the above data-compression setup. The bound is stated in the following lemma from [4].

Lemma 3. *Suppose* $X^n \sim Q_X^n$ *and let* $Z_n = \phi_n(X^n)$ *denote the chosen side-information about* X^n*, where for some positive constant* R *the side-information is generated by applying a helper* $\phi_n : \mathcal{X}^n \to \{1, 2, \ldots, 2^{nR}\}$*. An encoder* $\varphi_n : \mathcal{X}^n \times \{1, 2, \ldots, 2^{nR}\} \to \{1, 2, \ldots, 2^{nR_0}\}$ *produces a description of* X^n *based on* Z_n*. This description is revealed to a reconstructor* ψ_n *along with the side-information* Z_n*. From the description and* Z_n *the reconstructor produces* $\hat{X}^n = \psi_n(\varphi_n(X^n, Z_n), Z_n)$ *satisfying*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}[d(X_i, \hat{X}_i)] \leq D. \tag{27}$$

For every $n \in \mathbb{Z}_{>0}$ *the least achievable* R_0 *in this setup is lower-bounded by*

$$R_0 \geq \inf_{Q_{U|X}: \mathrm{I}(Q_{X;U}) \leq R} \mathrm{R}^{cond}_{d,D}(Q_{X|U}). \tag{28}$$

With (28) we can lower-bound the ρ-th moment of the number of guesses by

$$\mathbb{E}_{P_X}[\mathrm{G}_{d,D}(X^n|Z_n)^\rho]$$
$$\geq \sup_{Q_X} 2^{\mathbb{E}_{Q_X}[\log \mathrm{G}_{d,D}(X^n|Z_n)^\rho] - n\,\mathrm{D}(Q_X||P_X)} \tag{29}$$
$$\geq \sup_{Q_X} 2^{\rho \frac{\mathbb{E}_{Q_X}[l(\mathrm{G}_{d,D}(X^n|Z_n))]}{1+\delta} + \rho \frac{\log C(\delta)-1}{1+\delta} - n\,\mathrm{D}(Q_X||P_X)} \tag{30}$$
$$\geq \sup_{Q_X} \left[2^{\rho n \frac{\inf_{Q_{U|X}: \mathrm{I}(Q_{X;U}) \leq R} \mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U})}{1+\delta} + \rho \frac{\log C(\delta)-1}{1+\delta}} \right. \tag{31}$$
$$\left. 2^{-n\,\mathrm{D}(Q_X||P_X)} \right].$$

The above arguments differ from those of the preceding two setups only in the last inequality, which is now due to Lemma 3. We recover the exponent in (7) as a lower bound on $\liminf_{n\to\infty} \log \mathbb{E}_{P_X}[\mathrm{G}_{d,D}(X^n|Z_n)^\rho]/n$ by again letting n tend to infinity and observing that (31) holds for any $\delta > 0$.

To derive a guessing scheme that asymptotically achieves the optimal exponent (7), we combine the ideas introduced in sections II and III. We begin by fixing two small constants $\delta > 0$, $\delta' > 0$ and, for every type class $\mathcal{T}^{(n)}(Q_X)$ on $\mathcal{X}^n$, select a conditional type $Q_{U|X}$ that–among all those satisfying $\mathrm{I}(Q_{X;U}) \leq R - \delta$ and $Q_U \in \mathcal{P}^{(n)}(\mathcal{U})$–minimizes $\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U})$. We observe that for large enough n there exists a codebook $\mathcal{C}_{Q_X}$, such that $\log|\mathcal{C}_{Q_X}|/n \leq R - \delta/2$ and such that for every $x^n \in \mathcal{T}^{(n)}(Q_X)$ we can find some $u^n \in \mathcal{C}_{Q_X}$ satisfying $(x^n, u^n) \in \mathcal{T}^{(n)}(Q_X Q_{U|X})$.

The side-information Z_n is again made up of two parts. The first is of length at most $(\delta/2)n$ and describes the type of X^n. The second part is the index of some $U^n \in \mathcal{C}_{Q_X}$ satisfying $(X^n, U^n) \in \mathcal{T}^{(n)}(Q_X Q_{U|X})$. This description requires no more than $(R - \delta/2)n$ bits.

The guesser uses the first part of Z_n to recover the type of X_n and from it identifies the codebook $\mathcal{C}_{Q_X}$. Next the guesser uses the second part of Z_n to recover U_n from $\mathcal{C}_{Q_X}$. With U^n and the joint type of (X^n, U^n) at hand, the guesser applies Lemma 2 to generate a codebook $\mathcal{C}_{U^n}$ that satisfies $\log|\mathcal{C}_{U^n}|/n \leq \mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U}) + \delta'$ and such that for every $X^n \in \mathcal{T}^{(n)}(Q_{X|U}|U^n)$ there is some $\hat{X}^n \in \mathcal{C}_{U^n}$ satisfying $1/n \sum_{i=1}^n d(X_i, \hat{X}_i) \leq D$. The guesser then finds a suitable $\hat{X}^n$ by sequentially guessing the elements of $\mathcal{C}_{U^n}$ in an arbitrary order. The ρ-th moment of the number of guesses can be upper-bounded as follows:

$$\mathbb{E}[\mathrm{G}_{d,D}(X^n|Z_n)^\rho]$$
$$= \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \mathbb{E}[\mathrm{G}_{d,D}(X^n|Z_n)^\rho | X^n \in \mathcal{T}^{(n)}(Q_X)] \tag{32}$$
$$\mathbb{P}[X^n \in \mathcal{T}^{(n)}(Q_X)]$$
$$\leq \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \mathbb{E}[\mathrm{G}_{d,D}(X^n|Z_n)^\rho | X^n \in \mathcal{T}^{(n)}(Q_X)] \tag{33}$$
$$2^{-n\,\mathrm{D}(Q_X||P_X)}$$
$$\stackrel{(a)}{\leq} \sum_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} 2^{n \min^*_{Q_{U|X}: \mathrm{I}(Q_{X;U}) \leq R-\delta} \rho(\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U}) + \delta')} \tag{34}$$
$$2^{-n\,\mathrm{D}(Q_X||P_X)}$$
$$\stackrel{(b)}{\leq} \max_{Q_X \in \mathcal{P}^{(n)}(\mathcal{X})} \left[2^{n \min^*_{Q_{U|X}: \mathrm{I}(Q_{X;U}) \leq R-\delta} \rho(\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U}) + \delta')} \right. \tag{35}$$
$$\left. 2^{-n\,\mathrm{D}(Q_X||P_X) + n\delta_n} \right].$$

To see why (a) holds, observe that for X^n of type Q_X the guesser performs at most $2^{(\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U}) + \delta')n}$ many guesses. Here $\mathrm{R}^{\mathrm{cond}}_{d,D}(Q_{X|U})$ is minimized with respect to the type $Q_{U|X}$ under the constraint that $\mathrm{I}(Q_{X;U}) \leq R - \delta$. In (b) we upper-bound the sum over $\mathcal{P}^{(n)}(\mathcal{X})$ by its dominating term and absorb the overhead into the exponent δ_n. We relax the minimization over types to a minimization over all probability distributions at a small surplus in the exponent δ'_n, satisfying $\delta'_n \downarrow 0$. We next let δ and δ' approach 0 from above, and drop the requirement that Q_X must be a type in the first maximization. □

V. Guessing with Correlated Side-Information

The preceding sections present instances of a setup where $(X^n, Y^n) \sim P^n_{X,Y}$, and X^n is to be guessed to within distortion D after observing a rate-R description of Y^n. If X^n is to be guessed exactly, then the optimal guessing exponent E^* satisfies

$$\sup_{Q_Y} \inf_{Q_{U|Y}: I(Q_{U;Y}) \leq R} \sup_{Q_{X|Y,U}} (\rho H(Q_{X|U}) - D(Q_{X|Y,U}||P_{X|Y})$$
$$- D(Q_Y||P_Y)) \geq E^* \tag{36}$$
$$\geq \sup_{Q_Y} \inf_{Q_{U|Y}: I(Q_{U;Y}) \leq R} \sup_{Q_{X|Y}} (\rho H(Q_{X|U}) - D(Q_{X,Y}||P_{X,Y})). \tag{37}$$

References

[1] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99 – 105, Jan. 1996.

[2] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1041 – 1056, May 1998.

[3] R. Sundaresan, "Guessing based on length functions," in *2007 IEEE Int. Symp. Inf. Theory*, June 2007, pp. 716–719.

[4] R. Graczyk, "Guessing with a helper," Master's thesis, ETH Zurich, Aug. 2017.

[5] S. Moser, *Advanced Topics in Information Theory*, 2013, version 2.10 from 12 May 2017.

[6] A. Bracher, E. Hof, and A. Lapidoth, "Guessing attacks on distributed-storage systems," *arXiv:1701.01981*, Jan. 2017.

[7] A. Burin and O. Shayevitz, "Reducing guesswork via an unreliable oracle," *arXiv:1703.01672*, Mar. 2017.