The Error Exponent of Generalized Random-Gilbert Varshamov Codes

Anelia Somekh-Baruch
Bar-Ilan University
somekha@biu.ac.il

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—We introduce a random code construction for channel coding in which the codewords are constrained to be well-separated according to a given distance function, analogously to an existing construction attaining the Gilbert-Varshamov bound. We derive an achievable error exponent for this construction, and prove its tightness with respect to the ensemble average. We show that the exponent recovers the Csiszár and Körner exponent as a special case by choosing the distance function to be the negative of the empirical mutual information. We further establish the optimality of this distance function with respect to the exponent of the random coding scheme.**

I. Problem Setup

We consider transmission over a discrete memoryless channel (DMC) described by a conditional probability mass function $W(y|x)$, with input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$ for finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. We define $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{k=1}^n W(y_k|x_k)$ for input/output sequences $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathcal{X}^n, \boldsymbol{y} = (y_1, \dots, y_n) \in \mathcal{Y}^n$. The corresponding random variables are denoted by $\boldsymbol{X}, \boldsymbol{Y}$.

An encoder maps a message $m \in \{1, \dots, M_n\}$ to a channel input sequence $\boldsymbol{x}_m \in \mathcal{X}$, where the number of messages is denoted by M_n. The message, represented by the random variable S, is assumed to take values on $\{1, \dots, M_n\}$ equiprobably. This mapping induces an (n, M_n)-codebook $\mathcal{M}_n = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{M_n}\}$ with rate $R_n = \frac{1}{n} \log M_n$.

Upon observing the channel output $\boldsymbol{y}$, the decoder produces an estimate of the transmitted message $\hat{m} \in \{1, \dots, M_n\}$. We consider the family of type-dependent maximum-metric decoders, for which the transmitted message is estimated as

$$\hat{m} = \operatorname*{arg\,max}_{\boldsymbol{x}_i \in \mathcal{M}_n} q(\hat{P}_{\boldsymbol{x}_i, \boldsymbol{y}}), \tag{1}$$

where $\hat{P}_{\boldsymbol{x}, \boldsymbol{y}}$ is the joint empirical distribution (or type [1, Ch. 2]) of the pair $(\boldsymbol{x}, \boldsymbol{y})$. We assume that $q : \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ is bounded and continuous on the probability simplex. Maximum-likelihood (ML) decoding is a special case of (1), but more generally, the decoder may be *mismatched* [2], [3].

Denoting the random variable corresponding to the decoded message by $\hat{S}$, we define the probability of error as $P_e = \Pr\big[\hat{S} \neq S\big]$. A rate-exponent pair (R, E) is said to be achievable for channel W if, for all $\epsilon > 0$, there exists a sequence of $(n, \mathrm{e}^{n(R-\epsilon)})$-codebooks such that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \Pr\big[\hat{S} \neq S\big] \geq E - \epsilon. \tag{2}$$

Equivalently, we say that E is an achievable error exponent at rate R if (R, E) is an achievable rate-exponent pair.

Using random selection and graph decomposition techniques, Csiszár and Körner [4] studied the error exponent of constant-composition codes under a decoder that uses a type-dependent decoding metric, and derived the following achievable exponent for an arbitrary input distribution P:

$$E_q(R, P, W) = \min_{V \in \mathcal{T}_I} D(V_{Y|X} \| W | P) + \big| I(\widetilde{X}; Y, X) - R \big|_+, \tag{3}$$

where

$$\mathcal{T}_I \triangleq \Big\{ V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : V_X = V_{\widetilde{X}} = P, \; q(V_{\widetilde{X}Y}) \geq q(V_{XY}), I(X; \widetilde{X}) \leq R \Big\} \tag{4}$$

with $\mathcal{P}(\cdot)$ denoting the set of distributions on a given alphabet. This exponent was shown to be at least as high as the maximum of the expurgated and the random coding exponents.

In this paper, we introduce an alternative random code construction that achieves $E_q(R, P, W)$. Our construction constrains the codewords to be well-separated according to a general distance function, and thus, it can be viewed as a randomized Gilbert-Varshamov construction [5], [6]. Setting the distance function to be the negative mutual information (which turns out to maximize the exponent), we recover the exponent $E_q(R, P, W)$ in (3). While the analysis of [4] establishes the existence of codes attaining the exponent using a decomposition lemma, our scheme provides an explicit randomized construction whose ensemble average directly achieves the exponent. In addition, our construction provides insight into general distance functions, and among the constructions in the literature that are known to achieve the expurgated exponent, it is the first for which ensemble tightness is established, i.e., the derived error exponent of the ensemble-average error probability is exact.

This work was supported in part by the Israel Science Foundation under grant 631/17, the European Research Council under Grant 725411, the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3-1-R, and an NUS Startup Grant.

II. Random Codebook and Properties

Codes that attain the Gilbert-Varshamov bound on the Hamming space [5], [6] ensure that all codewords are at least at a certain Hamming distance Δ from each other. Our construction is a randomized constant-composition version of such codes for arbitrary DMCs and more general distance functions. We therefore refer to the resulting construction as the generalized random Gilbert-Varshamov (RGV) construction.

Definition 1. *Let Ω be the set of symmetric type-dependent functions $d(\cdot,\cdot) : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}$ that are bounded and continuous, i.e., functions that satisfy $d(\boldsymbol{x},\boldsymbol{x}') = d(\boldsymbol{x}',\boldsymbol{x})$ for all $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}^n$, that depend on $(\boldsymbol{x},\boldsymbol{x}')$ only through the joint empirical distribution $\hat{P}_{\boldsymbol{x}\boldsymbol{x}'}$, and that are bounded and continuous as a function of this joint distribution.*

We use the notation $d(\boldsymbol{x},\boldsymbol{x}')$ and $d(\hat{P}_{\boldsymbol{x}\boldsymbol{x}'})$ interchangeably for convenience, and similarly for $q(\boldsymbol{x},\boldsymbol{y})$ and $q(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})$. We refer to $d \in \Omega$ as a *distance function*, though it need not be a distance in the topological sense (e.g., it may be negative).

In the following, we describe how to construct a code $\mathcal{M}_n$ with M_n codewords of length n, such that any two distinct codewords $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{M}_n$ satisfy $d(\boldsymbol{x},\boldsymbol{x}') > \Delta$ for a given function $d(\cdot,\cdot) \in \Omega$ and threshold $\Delta \in \mathbb{R}$. This guarantees that the minimum distance of the codebook exceeds Δ.

Our construction depends on an input distribution $P \in \mathcal{P}(\mathcal{X})$, and throughout the paper, we let P_n denote an arbitrary type [1, Ch. 2] whose entries are $\frac{1}{n}$-close to P. The set of sequences with type P_n is denoted by $\mathcal{T}(P_n)$. For $i < j$, we let $\boldsymbol{x}_i^j$ denote $(\boldsymbol{x}_i,\dots,\boldsymbol{x}_j)$.

Fixing n, M_n, an input distribution $P \in \mathcal{P}(\mathcal{X})$, a distance function $d(\cdot,\cdot) \in \Omega$, and constants $\delta > 0, \Delta \in \mathbb{R}$, the construction is described by the following steps:

1) The first codeword, $\boldsymbol{x}_1$, is drawn uniformly from $\mathcal{T}(P_n)$;
2) The second codeword $\boldsymbol{x}_2$ is drawn uniformly from

$$\mathcal{T}(P_n,\boldsymbol{x}_1) \triangleq \{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) \,:\, d(\bar{\boldsymbol{x}},\boldsymbol{x}_1) > \Delta\} \tag{5}$$
$$= \mathcal{T}(P_n) \backslash \{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) \,:\, d(\bar{\boldsymbol{x}},\boldsymbol{x}_1) \le \Delta\}, \tag{6}$$

the set of sequences of composition P_n whose distance to $\boldsymbol{x}_1$ exceeds Δ;
3) The i-th codeword $\boldsymbol{x}_i$ is drawn uniformly from

$$\mathcal{T}(P_n,\boldsymbol{x}_1^{i-1})$$
$$= \{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) \,:\, d(\bar{\boldsymbol{x}},\boldsymbol{x}_j) > \Delta,\, j = 1\dots,i-1\} \tag{7}$$
$$= \mathcal{T}(P_n,\boldsymbol{x}_1^{i-2}) \backslash \left\{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n,\boldsymbol{x}_1^{i-2}) : d(\bar{\boldsymbol{x}},\boldsymbol{x}_{i-1}) \le \Delta\right\}. \tag{8}$$

Throughout the paper, it will be helpful to generalize the notation $\mathcal{T}(P_n,\boldsymbol{x}_1^{i-1})$ as follows: For any subset $\mathcal{D} \subseteq \mathcal{T}(P_n)$, we define $\mathcal{T}(P_n,\mathcal{D}) \triangleq \{\boldsymbol{x} \in \mathcal{T}(P_n) : d(\boldsymbol{x},\boldsymbol{x}') > \Delta, \forall \boldsymbol{x}' \in \mathcal{D}\}$.

We show in Lemma 1 below that in order to ensure that the above procedure generates the desired number of codewords $M_n = \mathrm{e}^{nR_n}$ (i.e., the sets $\mathcal{T}(P_n,\boldsymbol{x}_1^{i-1})$ are non-empty for all $i = 1,\dots,M_n$), it suffices to choose Δ and δ such that

$$\mathrm{e}^{n(R_n+\delta)}\mathrm{vol}_{\boldsymbol{x}}(\Delta) \le |\mathcal{T}(P_n)| \tag{9}$$

where $\mathrm{vol}_{\boldsymbol{x}}(\Delta) = |\{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n) \,:\, d(\bar{\boldsymbol{x}},\boldsymbol{x}) \le \Delta\}|$ is the volume of a ball of radius Δ according to distance $d(\cdot,\cdot)$ centered at some $\boldsymbol{x} \in \mathcal{T}(P_n)$. Since $d \in \Omega$ is symmetric and type-dependent, $\mathrm{vol}_{\boldsymbol{x}}(\Delta)$ does not depend on the specific choice of $\boldsymbol{x} \in \mathcal{T}(P_n)$. Condition (9) can be rewritten as

$$\sum_{\bar{\boldsymbol{x}} \in \mathcal{T}(P_n)\,:\, d(\bar{\boldsymbol{x}},\boldsymbol{x}) \le \Delta} \frac{1}{|\mathcal{T}(P_n)|} \le \mathrm{e}^{-n(R_n+\delta)}. \tag{10}$$

A. Codebook Properties

Here we provide several lemmas characterizing the key properties of the RGV construction. Due to space constraints, we only provide brief comments on the proofs.

We begin with the key fact that the construction is well-defined, in the sense that the procedure described above always produces the desired number of codewords M_n.

Lemma 1. *The RGV codebook construction with condition (10) is such that for all $i \in \{1,\dots,M_n\}$, all $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{i-1})$ occurring with non-zero probability, and any $\delta > 0$,*

$$(1-\mathrm{e}^{-n\delta})|\mathcal{T}(P_n)| \le |\mathcal{T}(P_n,\boldsymbol{x}_1^{i-1})| \le |\mathcal{T}(P_n)|. \tag{11}$$

The upper bound is immediate, and the lower bound follows from (9) and the fact that at most $e^{nR_n} \cdot \mathrm{vol}_{\boldsymbol{x}}(\Delta)$ codewords are removed from consideration throughout the construction.

Lemma 2. *A codebook $\mathcal{M}_n = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_{M_n}\}$ occurs with positive probability under the generalized RGV construction if and only if $\boldsymbol{x}_i \in \mathcal{T}(P_n)$ for all $i \in \{1,\dots,M_n\}$, and $d(\boldsymbol{x}_i,\boldsymbol{x}_j) > \Delta$ for all $i,j \in \{1,\dots,M_n\}$ such that $i \ne j$.*

Moreover, for any disjoint index sets $\mathcal{I},\mathcal{J} \subseteq \{1,\dots,M_n\}$ (i.e., $\mathcal{I} \cap \mathcal{J} = \emptyset$), the set of codewords $\boldsymbol{x}_{\mathcal{I}} = \{\boldsymbol{x}_i\}_{i\in\mathcal{I}}$ occurring with non-zero probability given $\boldsymbol{x}_{\mathcal{J}} = \{\boldsymbol{x}_j\}_{j\in\mathcal{J}}$ (with $\Pr(\boldsymbol{x}_{\mathcal{J}}) > 0$) is given by

$$\mathcal{T}'_{\mathcal{I}}(P_n,\boldsymbol{x}_{\mathcal{J}}) \triangleq \Big\{\boldsymbol{x}_{\mathcal{I}}\,;\, d(\boldsymbol{x}_i,\boldsymbol{x}_{i'}) > \Delta, \forall i,i' \in \mathcal{I}, i \ne i', \text{ and } d(\boldsymbol{x}_i,\boldsymbol{x}_j) > \Delta, \forall i \in \mathcal{I}, j \in \mathcal{J}\Big\}. \tag{12}$$

More generally, for possibly-overlapping $\mathcal{I}$ and $\mathcal{J}$, the corresponding set is

$$\mathcal{T}_{\mathcal{I}}(P_n,\boldsymbol{x}_{\mathcal{J}}) \triangleq \Big\{\bar{\boldsymbol{x}}_{\mathcal{I}}\,;\, \bar{\boldsymbol{x}}_i = \boldsymbol{x}_i, \forall i \in \mathcal{I}\cap\mathcal{J}, \text{ and } \bar{\boldsymbol{x}}_{\mathcal{I}\backslash\mathcal{J}} \in \mathcal{T}'_{\mathcal{I}\backslash\mathcal{J}}(P_n,\boldsymbol{x}_{\mathcal{J}})\Big\}. \tag{13}$$

We will often make use of the set (13) in the special case that $\mathcal{I} = \{1,\dots,i\}$ for some index i, in which case we will adopt the shorthand $\mathcal{T}_1^i(P_n,\boldsymbol{x}_{\mathcal{J}}) \triangleq \mathcal{T}_{\{1,\dots,i\}}(P_n,\boldsymbol{x}_{\mathcal{J}})$. Moreover, we will use notation such as $\mathcal{T}_1^{i-1}(P_n,\boldsymbol{x}_i,\boldsymbol{x}_m)$ to mean $\mathcal{T}_1^{i-1}(P_n,\{\boldsymbol{x}_i,\boldsymbol{x}_m\})$. We also define $\delta_n \triangleq \frac{\mathrm{e}^{-n\delta}}{1-\mathrm{e}^{-n\delta}}$.

Lemma 3. *Under the generalized RGV construction, for any message m and $k < m$, it holds for all $\boldsymbol{x}_m \in \mathcal{T}(P_n,\boldsymbol{x}_1^k)$ that*

$$\frac{1-\delta_n^2}{\mathrm{e}^{\delta_n}} \cdot \frac{1}{|\mathcal{T}(P_n)|} \le \Pr(\boldsymbol{x}_m|\boldsymbol{x}_1^k) \le \frac{1}{1-\mathrm{e}^{-n\delta}} \cdot \frac{1}{|\mathcal{T}(P_n)|}, \tag{14}$$

while $\Pr(\boldsymbol{x}_m|\boldsymbol{x}_1^k) = 0$ for all $\boldsymbol{x} \notin \mathcal{T}(P_n,\boldsymbol{x}_1^k)$.

In the following, we consider the probability $\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1},\boldsymbol{x}_m)$ conditioned on the previous codewords

$\boldsymbol{x}_1^{i-1}$ along with another codeword $\boldsymbol{x}_m$ (which will later represent the transmitted codeword). It may be the case that $m < i$, in which case $\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m)$ can equivalently be viewed as $\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1})$.

Lemma 4. *Under the generalized RGV construction, for any pair $(\boldsymbol{x}_m, \boldsymbol{x}_i)$ $(m \neq i)$ satisfying $d(\boldsymbol{x}_m, \boldsymbol{x}_i) > \Delta$, and any $\boldsymbol{x}_1^{i-1} \in \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m, \boldsymbol{x}_i)$, it holds that*

$$\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m) \geq \frac{(1-\mathrm{e}^{-n\delta})^2(1-\delta_n^2)}{\mathrm{e}^{\delta_n}} \cdot \frac{1}{|\mathcal{T}(P_n)|} \quad (15)$$

$$\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m) \leq \frac{\mathrm{e}^{\delta_n}}{(1-\delta_n^2)(1-\mathrm{e}^{-n\delta})} \cdot \frac{1}{|\mathcal{T}(P_n)|}, \quad (16)$$

while $\Pr(\boldsymbol{x}_i|\boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m) = 0$ whenever $\boldsymbol{x}_1^{i-1} \notin \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m, \boldsymbol{x}_i)$.

Finally, using a basic symmetry argument, we have that the marginal distribution of any given codeword $\boldsymbol{X}_m$ (without any conditioning on other codewords) is uniform over $\mathcal{T}(P_n)$.

Lemma 5. *For any message index m, the marginal distribution of codeword $\boldsymbol{X}_m$ is $\Pr(\boldsymbol{x}_m) = \frac{1}{|\mathcal{T}(P_n)|}$.*

III. Main Result

We now state our main result, namely, a single-letter lower bound for the error exponent of the RGV construction. We show in Section V that it reduces to the exponent of [4] when the distance function $d(\cdot,\cdot)$ is optimized. Let

$$E_{\mathrm{RGV}}(R, P, W, q, d, \Delta) = \min_{V \in \mathcal{T}_{d,q,P}(\Delta)} D(V_{Y|X}\|W|P) + \big|I(\widetilde{X}; Y, X) - R\big|_+, \quad (17)$$

where

$$\mathcal{T}_{d,q,P}(\Delta) \triangleq \Big\{V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : V_X = V_{\widetilde{X}} = P, \; q(V_{\widetilde{X}Y}) \geq q(V_{XY}),\, d(P_{X\widetilde{X}}) \geq \Delta\Big\}. \quad (18)$$

Subsequently, the notations $\doteq$, $\dot{\leq}$, and $\dot{\geq}$ denote (in)equalities up to subexponential factors.

Theorem 1. *For all $P \in \mathcal{P}(\mathcal{X})$, $\delta > 0$, $\Delta \in \mathbb{R}$, $d \in \Omega$, and $R > 0$ satisfying*

$$R \leq \min_{P_{X\widetilde{X}} : d(P_{X\widetilde{X}}) \leq \Delta,\, P_X = P_{\widetilde{X}} = P} I(X;\widetilde{X}) - 2\delta, \quad (19)$$

the ensemble average error probability $\bar{P}_{\mathrm{e}}^{(n)}$ of the RGV construction with parameters $(n, R, P, d, \Delta, \delta)$ and the continuous type-dependent decoding metric $q(\cdot)$ over DMC W satisfies

$$\bar{P}_{\mathrm{e}}^{(n)} \dot{\leq} \mathrm{e}^{-nE_{\mathrm{RGV}}(R,P,W,q,d,\Delta)}. \quad (20)$$

In addition, if q is an additive decoding metric, then

$$\bar{P}_{\mathrm{e}}^{(n)} \dot{\geq} \mathrm{e}^{-nE_{\mathrm{RGV}}(R,P,W,q,d,\Delta+\epsilon)} \quad (21)$$

for arbitrarily small $\epsilon > 0$.

As will be shown in the sequel, the rate constraint (19) is not restrictive in the sense that if the distance function is chosen appropriately, the generalized RGV construction achieves the exponent $E_q(R, P, W)$ in (3), which in turn shows the achievability of capacity for ML decoding or the LM rate in the mismatched case [4].

While Theorem 1 states the error exponent, the central part of the analysis is in arriving at the following asymptotic expression for the ensemble average probability of error:

$$\bar{P}_{\mathrm{e}}^{(n)} \doteq \sum_{\boldsymbol{x}\in\mathcal{T}(P_n),\boldsymbol{y}} \frac{1}{|\mathcal{T}(P_n)|} W^n(\boldsymbol{y}|\boldsymbol{x}) \cdot \min\Bigg\{1, (M_n-1) \sum_{\substack{\boldsymbol{x}'\in\mathcal{T}(P_n):\, q(\boldsymbol{x}',\boldsymbol{y}) \geq q(\boldsymbol{x},\boldsymbol{y}) \\ d(\boldsymbol{x}',\boldsymbol{x}) \geq \Delta}} \frac{1}{|\mathcal{T}(P_n)|}\Bigg\}. \quad (22)$$

This can be interpreted as a stronger (albeit asymptotic) analog of the *random coding union* bound [7] that achieves not only the random coding exponent, but also the low-rate improvements of the expurgated exponent.

A notable difficulty in the analysis is that unlike ordinary random coding with pairwise codeword independence, the conditional distribution of the m-th codeword depends on the previously drawn $m-1$ codewords. Lemma 4 is key to overcoming this challenge.

IV. Proof of Theorem 1

Here, we give details of the proof of (20). Details of the proof of (21) can be found in [8].

Step 1: Characterizing the permitted rates

For convenience, we define

$$R' = \min_{P_{X\widetilde{X}} \in \mathcal{P}(\mathcal{X}^2) :\, d(P_{X\widetilde{X}}) \leq \Delta,\, P_X = P_{\widetilde{X}} = P} I(X;\widetilde{X}) - 2\delta. \quad (23)$$

Moreover, we let $\mathcal{T}(P_{\widetilde{X}|X})$ denote a conditional type class [1, Ch. 2] corresponding to $\boldsymbol{x} \in \mathcal{T}(P_n)$, and let $\mathcal{P}_n(\mathcal{X}|\boldsymbol{x})$ be the set of all conditional types. For n sufficiently large, we have

$$\sum_{\bar{\boldsymbol{x}}\in\mathcal{T}(P_n) :\, d(\bar{\boldsymbol{x}},\boldsymbol{x}) \leq \Delta} \frac{1}{|\mathcal{T}(P_n)|} \leq (n+1)^{|\mathcal{X}|^2} \max_{\substack{P_{\widetilde{X}|X} \in \mathcal{P}_n(\mathcal{X}|\boldsymbol{x}) :\, P_{\widetilde{X}} = P_X = P_n \\ d(P_{X\widetilde{X}}) \leq \Delta}} \frac{|\mathcal{T}(P_{\widetilde{X}|X})|}{|\mathcal{T}(P_n)|} \quad (24)$$

$$\leq \exp\Bigg(-n\Bigg(\min_{\substack{P_{X\widetilde{X}} \in \mathcal{P}_n(\mathcal{X}^2) :\, d(P_{X\widetilde{X}}) \leq \Delta \\ P_X = P_{\widetilde{X}} = P}} I(X;\widetilde{X}) - \delta\Bigg)\Bigg) \quad (25)$$

$$\leq \mathrm{e}^{-n(R'+\delta)}, \quad (26)$$

where (24) follows since the number of joint types is upper bounded by $(n+1)^{|\mathcal{X}|^2}$, (25) holds for n sufficiently large because $|\mathcal{T}(P_{\widetilde{X}|X})| \doteq \mathrm{e}^{nH_P(\widetilde{X}|X)}$ and $|\mathcal{T}(P_n)| \doteq \mathrm{e}^{nH(P)}$, and (26) follows from (23). Hence, if the rate of the RGV construction satisfies $R_n \leq R'$, we have

$$\sum_{\bar{\boldsymbol{x}}\in\mathcal{T}(P_n) : d(\bar{\boldsymbol{x}},\boldsymbol{x}) \leq \Delta} \frac{1}{|\mathcal{T}(P_n)|} \leq \mathrm{e}^{-n(R_n+\delta)}, \quad (27)$$

which is precisely the condition assumed in (10).

We henceforth assume that the number of codewords of the RGV construction is such that $R_n \le R'$, and we calculate the resulting average probability of error.

Step 2: Conditional probability of correct decoding

The ensemble average error probability is

$$\bar{P}_e^{(n)} = \frac{1}{M_n} \sum_{m=1}^{M_n} \bar{P}_{e,m}^{(n)} \tag{28}$$

where the error probability for the m-th codeword is

$$\bar{P}_{e,m}^{(n)} = 1 - \mathbb{E}[\Pr(\text{no error} \,|\, \boldsymbol{X}_m, \boldsymbol{Y})], \tag{29}$$

and where $\Pr(\text{no error} \,|\, \boldsymbol{x}_m, \boldsymbol{y})$ is the probability of correct decoding for the m-th codeword assuming that the realizations of the codeword and received sequences are $\boldsymbol{x}_m$ and $\boldsymbol{y}$. We perform the analysis conditioned on the transmitted and received sequences being $\boldsymbol{x}_m$ and $\boldsymbol{y}$ (and implicitly on m being transmitted), and later we duly average over these choices.

We define the i-th pairwise correct decoding event given $(\boldsymbol{X}_m, \boldsymbol{Y}) = (\boldsymbol{x}_m, \boldsymbol{y})$ as

$$\mathcal{C}_i = \{q(\boldsymbol{X}_i, \boldsymbol{y}) < q(\boldsymbol{x}_m, \boldsymbol{y})\}, \tag{30}$$

meaning that $\boldsymbol{x}_m$ is favored over the random codeword $\boldsymbol{X}_i$ (recall that ties are counted as errors). We further define $\mathcal{C}_i^j \triangleq \{\mathcal{C}_i, \mathcal{C}_{i+1}, \ldots, \mathcal{C}_j\}$ for $j > i$, that is, the intersection of the events $\mathcal{C}_i, \mathcal{C}_{i+1}, \ldots, \mathcal{C}_j$.

We write the correct decoding probability given $(\boldsymbol{x}_m, \boldsymbol{y})$ as

$$\begin{aligned}&\Pr(\text{no error} \,|\, \boldsymbol{x}_m, \boldsymbol{y}) \\ &= \prod_{i<m} \Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \prod_{i>m} \Pr(\mathcal{C}_i | \mathcal{C}_1^{m-1}, \mathcal{C}_{m+1}^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}).\end{aligned} \tag{31}$$

For brevity, we use the shorthand notation $\Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y})$ also for $i > m$ (i.e., for $\Pr(\mathcal{C}_i | \mathcal{C}_1^{m-1}, \mathcal{C}_{m+1}^{i-1}, \boldsymbol{x}_m, \boldsymbol{y})$).

We proceed by lower bounding $\Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y})$. Since only sequences $\boldsymbol{x}_1$ such that $d(\boldsymbol{x}_1, \boldsymbol{x}_m) > \Delta$ have positive probability conditioned on $\boldsymbol{X}_m = \boldsymbol{x}_m$, we have

$$\Pr(\mathcal{C}_1 | \boldsymbol{x}_m, \boldsymbol{y}) = \sum_{\substack{\boldsymbol{x}_1 \,:\, q(\boldsymbol{x}_1, \boldsymbol{y}) < q(\boldsymbol{x}_m, \boldsymbol{y}) \\ d(\boldsymbol{x}_1, \boldsymbol{x}_m) > \Delta}} \Pr(\boldsymbol{x}_1 | \boldsymbol{x}_m, \boldsymbol{y}). \tag{32}$$

Recall Lemma 2, and define

$$\begin{aligned}\mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y}) \triangleq \big\{ &\boldsymbol{x}_1^{i-1} \in \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m) : \\ &q(\boldsymbol{x}_j, \boldsymbol{y}) < q(\boldsymbol{x}_m, \boldsymbol{y}),\, \forall j \le i-1, j \ne m \big\}.\end{aligned} \tag{33}$$

Also recalling $\mathcal{C}_i$ defined in (30), we have for $i > 1$ that

$$\begin{aligned}\Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) &= \sum_{\boldsymbol{x}_1^{i-1} \in \mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\boldsymbol{x}_1^{i-1} | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \\ &\quad \times \Pr(\mathcal{C}_i | \boldsymbol{x}_1^{i-1}, \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (34) \\ &= \sum_{\boldsymbol{x}_1^{i-1} \in \mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\boldsymbol{x}_1^{i-1} | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \Pr(\mathcal{C}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (35) \\ &\ge \min_{\boldsymbol{x}_1^{i-1} \in \mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\mathcal{C}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}), \qquad (36)\end{aligned}$$

where (35) follows since given $(\boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y})$, the event $\mathcal{C}_i$ does not depend on $\mathcal{C}_1^{i-1}$, and (36) follows since the summation in (35) includes all $\boldsymbol{x}_1^{i-1}$ for which the probability $\Pr(\boldsymbol{x}_1^{i-1} | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y})$ is non-zero (see Lemma 2). Continuing, we rewrite (36) as

$$\begin{aligned}&\Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \\ &\ge 1 - \max_{\boldsymbol{x}_1^{i-1} \in \mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\mathcal{C}_i^c | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (37) \\ &= 1 - \max_{\boldsymbol{x}_1^{i-1} \in \mathcal{S}_{m,i}(\boldsymbol{x}_m, \boldsymbol{y})} \sum_{\boldsymbol{x}_i : q(\boldsymbol{x}_i, \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\boldsymbol{x}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (38) \\ &\ge 1 - \max_{\boldsymbol{x}_1^{i-1} \in \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m)} \sum_{\boldsymbol{x}_i : q(\boldsymbol{x}_i, \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y})} \Pr(\boldsymbol{x}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (39) \\ &= 1 - \max_{\boldsymbol{x}_1^{i-1} \in \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m)} \sum_{\substack{\boldsymbol{x}_i : q(\boldsymbol{x}_i, \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y}) \\ d(\boldsymbol{x}_i, \boldsymbol{x}_m) > \Delta}} \Pr(\boldsymbol{x}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \qquad (40) \\ &= 1 - \max_{\boldsymbol{x}_1^{i-1} \in \mathcal{T}_1^{i-1}(P_n, \boldsymbol{x}_m)} \sum_{\substack{\boldsymbol{x}_i : q(\boldsymbol{x}_i, \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y}) \\ d(\boldsymbol{x}_i, \boldsymbol{x}_m) > \Delta}} \Pr(\boldsymbol{x}_i | \boldsymbol{x}_1^{i-1}, \boldsymbol{x}_m), \qquad (41)\end{aligned}$$

where (39) follows by enlarging the set over which the maximization takes place, (40) follows from the fact that given $\boldsymbol{x}_m$, only codewords $\boldsymbol{x}_i$ such that $d(\boldsymbol{x}_i, \boldsymbol{x}_m) > \Delta$ may have positive probability, and (41) follows since $\boldsymbol{X}_i - (\boldsymbol{x}_1^{i-1}, \boldsymbol{X}_m) - \boldsymbol{Y}$ forms a Markov chain.

Step 3: Applying the technical lemmas

Combining (41) and Lemma 4, we obtain

$$\begin{aligned}&\Pr(\mathcal{C}_i | \mathcal{C}_1^{i-1}, \boldsymbol{x}_m, \boldsymbol{y}) \\ &\ge 1 - \frac{\mathrm{e}^{\delta_n}}{(1-\delta_n^2)(1-\mathrm{e}^{-n\delta})} \cdot \sum_{\substack{\boldsymbol{x}' \in \mathcal{T}(P_n) : \\ q(\boldsymbol{x}', \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y}), \\ d(\boldsymbol{x}_i, \boldsymbol{x}_m) > \Delta}} \frac{1}{|\mathcal{T}(P_n)|}.\end{aligned} \tag{42}$$

Using Lemma 5 to take the expectation over $(\boldsymbol{X}_m, \boldsymbol{Y})$ gives

$$\begin{aligned}\bar{P}_{e,m}^{(n)} \dot{\le} &\sum_{\boldsymbol{x}_m \in \mathcal{T}(P_n), \boldsymbol{y}} \frac{1}{|\mathcal{T}(P_n)|} W^n(\boldsymbol{y} | \boldsymbol{x}_m) \\ &\cdot \left[1 - \left(1 - \sum_{\substack{\boldsymbol{x}' \in \mathcal{T}(P_n) : \\ q(\boldsymbol{x}', \boldsymbol{y}) \ge q(\boldsymbol{x}_m, \boldsymbol{y}) \\ d(\boldsymbol{x}', \boldsymbol{x}_m) > \Delta}} \frac{1}{|\mathcal{T}(P_n)|} \right)^{M_n - 1} \right],\end{aligned} \tag{43}$$

where the $\dot{\le}$ follows from $\frac{\mathrm{e}^{\delta_n}}{(1-\delta_n^2)(1-\mathrm{e}^{-n\delta})} \doteq 1$. Since the above bound does not depend on m, it is an upper bound on $\bar{P}_{\mathrm{e}}^{(n)}$. Applying the inequality $1 - (1-a)^M \le \min\{1, Ma\}$ for $a \in [0,1]$, and slightly enlarging the set of summands by replacing $d(\boldsymbol{x}', \boldsymbol{x}) > \Delta$ by $d(\boldsymbol{x}', \boldsymbol{x}) \ge \Delta$, we recover the asymptotic upper bound corresponding to (22). The matching lower bound that establishes (22) can be found in [8], though it is not needed for proving (20).

Step 4: Deducing the error exponent

Similarly to (24), the inner sum in (22) satisfies

$$\sum_{\substack{\boldsymbol{x}'\in\mathcal{T}(P_n):\\ q(\boldsymbol{x}',\boldsymbol{y})\geq q(\boldsymbol{x},\boldsymbol{y}),\\ d(\boldsymbol{x}',\boldsymbol{x})\geq\Delta}} \frac{1}{|\mathcal{T}(P_n)|} \stackrel{\cdot}{\leq} \max_{\substack{\mathcal{T}(\hat{P}_{\boldsymbol{x}'|\boldsymbol{x}\boldsymbol{y}})\in\mathcal{P}_n(\mathcal{X}|\boldsymbol{x}\boldsymbol{y}):\\ q(\boldsymbol{x}',\boldsymbol{y})\geq q(\boldsymbol{x},\boldsymbol{y})\\ d(\boldsymbol{x}',\boldsymbol{x})\geq\Delta}} \frac{|\mathcal{T}(\hat{P}_{\boldsymbol{x}'|\boldsymbol{x}\boldsymbol{y}})|}{|\mathcal{T}(P_n)|}. \tag{44}$$

We apply the standard properties of types $|\mathcal{T}(\hat{P}_{\boldsymbol{x}'|\boldsymbol{x}\boldsymbol{y}})| \doteq \mathrm{e}^{nH_{\hat{P}}(\widetilde{X}|Y,X)}$ and $|\mathcal{T}(P_n)| \doteq \mathrm{e}^{nH(P_n)}$ [1, Ch. 2], as well as

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \mathrm{e}^{n(D(\hat{P}_{\boldsymbol{y}|\boldsymbol{x}}\|W|P_n)+H(\hat{P}_{\boldsymbol{y}|\boldsymbol{x}}))}, \tag{45}$$

which implies that $(\boldsymbol{X}_m, \boldsymbol{Y})$ has a given conditional type $V_{Y|X}$ with probability $\mathrm{e}^{-nD(V\|W|P_n)}$ times a subexponential factor. Combining these properties with the fact that $P_n \to P$ and the assumption that $d(P_{X\widetilde{X}})$ and $q(P_{XY})$ are continuous, we deduce the final single-letter exponent:

$$\bar{P}_{\mathrm{e}}^{(n)} \stackrel{\cdot}{\leq} \mathrm{e}^{-n\min_{V\in\mathcal{T}_{d,q,P}(\Delta)} D(V_{Y|X}\|W|P)+|I(\widetilde{X};Y,X)-R|_+}, \tag{46}$$

where $\mathcal{T}_{d,q,P}(\Delta)$ is defined in (18).

V. Reduction to the Csiszár-Körner Exponent

We next show that when the distance function $d(\cdot,\cdot)$ is optimized, and Δ is chosen appropriately, the exponent in Theorem 1 recovers the exponent $E_q(R,P,W)$ in (3) [4].

Corollary 1. *Let $\epsilon > 0$ be given, let $q(\cdot)$ be an arbitrary continuous decoding rule, and let R, P be given. The exponent of the ensemble average error probability of the RGV construction with sufficiently small δ, $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$, $\Delta = -(R+2\delta)$, sufficiently large n, and decoding metric $q(\cdot)$ over the DMC W is at least as high as $E_q(R,P,W) - \epsilon$.*

Proof. We first claim that the choices $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$ ad $\Delta = -(R+2\delta)$ are valid for all R. To see this, note that

$$\min_{\substack{P_{X\widetilde{X}}:\, d(P_{X\widetilde{X}})\leq\Delta\\ P_X=P_{\widetilde{X}}=P}} I(X;\widetilde{X})\Big|_{\substack{d(P_{X\widetilde{X}})=-I(X;\widetilde{X})\\ \Delta=-(R+2\delta)}} \tag{47}$$

$$= \min_{\substack{P_{X\widetilde{X}}:\, I(X;\widetilde{X})\geq R+2\delta\\ P_X=P_{\widetilde{X}}=P}} I(X;\widetilde{X}) \tag{48}$$

$$\geq R+2\delta, \tag{49}$$

thus, condition (19), which is imposed in Theorem 1 for the achievability of $E_{\mathrm{RGV}}(R,P,W,q,d,\Delta)$ is met. Now, by setting $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$, and $\Delta = -(R+2\delta)$, we obtain

$$E_{\mathrm{RGV}}(R,P,W,q,d,\Delta)\Big|_{d(P_{X\widetilde{X}})=-I(X;\widetilde{X}),\,\Delta=-(R+2\delta)}$$
$$= \min_{V\in\mathcal{T}_{I,\delta}} D(V_{Y|X}\|W|P) + \big|I(\widetilde{X};Y,X) - R\big|_+, \tag{50}$$

where

$$\mathcal{T}_{I,\delta} \triangleq \Big\{V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : V_X = V_{\widetilde{X}} = P, \\ q(V_{\widetilde{X}Y}) \geq q(V_{XY}), I(\widetilde{X};X) \leq R+2\delta\Big\}. \tag{51}$$

The result follows by letting $\delta \to 0$ and using the continuity of $E_q(R,P,W)$ in R [4]. □

The following proposition reveals that the above choice of (d,Δ) in fact maximizes the general exponent given in Theorem 1.

Proposition 1. *Under the setup of Theorem 1 with*

$$R \leq \min_{\substack{P_{X\widetilde{X}}:\, P_X=P_{\widetilde{X}}=P\\ d(P_{X\widetilde{X}})\leq\Delta}} I(X;\widetilde{X}) - 2\delta, \tag{52}$$

we have

$$E_{\mathrm{RGV}}(R,P,W,q,d,\Delta) \\ \leq E_{\mathrm{RGV}}(R,P,W,q,d,\Delta)\Big|_{d=-I(X;\widetilde{X}),\,\Delta=-(R+2\delta)}. \tag{53}$$

Proof. From (52), we see that among all $P'_{X\widetilde{X}}$ such that $P'_X = P'_{\widetilde{X}} = P$, the condition $d(P'_{X\widetilde{X}}) \leq \Delta$ implies $R+2\delta \leq I_{P'}(X;\widetilde{X})$. The contrapositive statement is that among all $P'_{X\widetilde{X}}$ such that $P'_X = P'_{\widetilde{X}} = P$, the condition $R+2\delta > I_{P'}(X;\widetilde{X})$ implies $d(P'_{X\widetilde{X}}) > \Delta$. Thus, when (52) holds, $\mathcal{T}_{d,q,P}(\Delta)$ defined in (18) satisfies $\mathcal{T}_{d,q,P}(\Delta) \supseteq \mathcal{T}''$, where

$$\mathcal{T}'' \triangleq \Big\{V_{X\widetilde{X}Y} \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{Y}) : V_X = V_{\widetilde{X}} = P, \\ q(V_{\widetilde{X}Y}) \geq q(V_{XY}), R+2\delta \geq I(X;\widetilde{X})\Big\}. \tag{54}$$

Therefore,

$$E_{\mathrm{RGV}}(R,P,W,q,d,\Delta)$$
$$= \min_{V\in\mathcal{T}_{d,q,P}(\Delta)} D(V_{Y|X}\|W|P) + \big|I(\widetilde{X};Y,X) - R\big|_+ \tag{55}$$
$$\leq \min_{V\in\mathcal{T}''} D(V_{Y|X}\|W|P) + \big|I(\widetilde{X};Y,X) - R\big|_+, \tag{56}$$

so the exponent is upper bounded by that corresponding to $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$ and $\Delta = -(R+2\delta)$. □

We note that the choice $d(P_{X\widetilde{X}}) = -I(X;\widetilde{X})$ is not only the one that maximizes the achievable exponent in Theorem 1, but it is *universally* maximizing for all channels, decoding rules, and input distributions.

References

[1] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[2] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[4] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, 1981.

[5] E. N. Gilbert, "A comparison of signalling alphabets," *Bell Labs Tech. J.*, vol. 31, no. 3, pp. 504–522, 1952.

[6] R. R. Varshamov, "Estimate of the number of signals in error correcting codes," in *Dokl. Akad. Nauk SSSR*, vol. 117, no. 5, 1957, pp. 739–741.

[7] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[8] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert-Varshamov codes," submitted to IEEE Trans. Inf. Theory, 2017, arXiv:1805.02515 [cs.IT].