

Expurgated Random-Coding Ensembles: Exponents, Refinements and Connections

Jonathan Scarlett¹, Li Peng¹, Neri Merhav², Alfonso Martinez³ and Albert Guillén i Fàbregas^{1,3,4}

¹University of Cambridge, ²Technion, I.I.T., ³Universitat Pompeu Fabra, ⁴ICREA
e-mail: {jms265,lp327}@cam.ac.uk, merhav@ee.technion.ac.il, {guillen,alfonso.martinez}@ieee.org

Abstract—This paper studies expurgated random-coding bounds and exponents for channels with maximum-metric decoding. A simple non-asymptotic bound is shown to attain an exponent which coincides with that of Csiszár and Körner for discrete memoryless channels, while remaining valid for continuous alphabets. Using an alternative approach based on statistical-mechanical methods, an exponent for more general channels and decoding metrics is given.

I. INTRODUCTION

Achievable performance bounds for channel coding are typically obtained by analyzing the average error probability of an ensemble of codebooks with independently generated codewords. At low rates, the error probability of the best code in the ensemble can be significantly smaller than the average. In such cases, better performance bounds are obtained by considering an ensemble in which a subset of the randomly generated codewords are expurgated from the codebook.

The main approaches to obtaining expurgated bounds and exponents are those of Gallager [1, Sec. 5.7] and Csiszár-Körner-Martón [2, Ex. 10.18] [3]. Gallager’s approach is based on simple inequalities such as Markov’s inequality, and has the advantage of being simple and applicable to channels with continuous alphabets. On the other hand, the techniques of [2], [3] are based on the method of types, and are applicable to channels with input constraints. While the exponents of [1]–[3] all coincide after optimizing the input distribution, the exponents of [2], [3] can be higher than that of [1] for a given input distribution [4].

In this paper, we provide techniques that attain the best of each of the above approaches. Our main contributions are as follows:

- 1) We give the precise connection between the exponents of [1]–[3] using Lagrange duality [5], as well as generalizing the exponents of [1], [2] to the setting of mismatched decoding [3], [6].
- 2) We show that variations of Gallager’s techniques can be used to obtain a simple non-asymptotic bound which recovers the exponent of [2], [3], as well as a generalization to the case of continuous alphabets.

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union’s 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03. The work of N. Merhav was partially supported by the Israeli Science Foundation (ISF), grant no. 412/12.

- 3) We present an alternative analysis technique based on statistical-mechanical methods (e.g. see [7], [8]), and use it to derive an achievable exponent for general channels and metrics (e.g. channels with memory).

Due to space constraints, full proofs of the main results are omitted; details can be found in [4].

A. Notation

We use bold symbols for vectors (e.g. \mathbf{x}), and denote the corresponding i -th entry using a subscript (e.g. x_i). The set of all empirical distributions (i.e. types [2, Ch. 2]) on a given alphabet, say \mathcal{X} , is denoted by $\mathcal{P}_n(\mathcal{X})$. For a given type $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of all sequences in \mathcal{X}^n with type Q . For two positive sequences f_n and g_n , we write $f_n \doteq g_n$ if $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, and we write $f_n \dot{\leq} g_n$ if $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} \leq 0$, and analogously for $\dot{\geq}$. All logarithms have base e , and all rates are in units of nats. We define $[c]^+ = \max\{0, c\}$, and denote the indicator function by $\mathbb{1}\{\cdot\}$.

B. System Setup

We consider block coding over a memoryless channel $W^n(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^n W(y_i|x_i)$ with alphabets \mathcal{X} and \mathcal{Y} . The encoder takes as input a message m uniformly distributed on the set $\{1, \dots, M\}$, and transmits the corresponding codeword $\mathbf{x}^{(m)}$ of length n . Given \mathbf{y} , the decoder forms the estimate

$$\hat{m} = \arg \max_{j \in \{1, \dots, M\}} q^n(\mathbf{x}^{(j)}, \mathbf{y}), \quad (1)$$

where $q^n(\mathbf{x}, \mathbf{y}) \triangleq \prod_{i=1}^n q(x_i, y_i)$ for some non-negative function $q(x, y)$. When $q(x, y) = W(y|x)$, (1) is the optimal maximum-likelihood (ML) decoding rule. For other decoding metrics, this setting is that of *mismatched decoding* [3], [6], which is relevant when ML decoding is not feasible.

Except where stated otherwise, we assume that the codebooks are unconstrained. However, in some cases we will consider input constraints of the form

$$\frac{1}{n} \sum_{i=1}^n c(x_i) \leq \Gamma, \quad (2)$$

where $c(\cdot)$ is referred to as a cost function, and Γ is a constant.

C. Expurgated Exponents and Duality

We will primarily be interested in the following exponent, which was derived in [3] for the case of finite alphabets:¹

$$E_{\text{ex}}^{\text{cc}}(Q, R) \triangleq \min_{\substack{P_{X\bar{X}Y} \in \mathcal{T}^{\text{cc}}(Q) \\ I_P(X; \bar{X}) \leq R}} D(P_{X\bar{X}Y} \| Q \times Q \times W) - R, \quad (3)$$

where

$$\mathcal{T}^{\text{cc}}(Q) \triangleq \left\{ P_{X\bar{X}Y} : P_X = Q, P_{\bar{X}} = Q, \mathbb{E}_P[\log q(\bar{X}, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \right\} \quad (4)$$

and Q is an arbitrary input distribution. The following theorem links this exponent with those given in [1], [2].

Theorem 1. *For any input distribution Q and rate R , we have*

$$E_{\text{ex}}^{\text{cc}}(Q, R) = \sup_{s \geq 0} \min_{\substack{P_{X\bar{X}} : P_X = Q, P_{\bar{X}} = Q, \\ I_P(X; \bar{X}) \leq R}} \mathbb{E}_P[d_s(X, \bar{X})] + I_P(X; \bar{X}) - R \quad (5)$$

$$= \sup_{\rho \geq 1} E_x^{\text{cc}}(Q, \rho) - \rho R, \quad (6)$$

where

$$d_s(x, \bar{x}) \triangleq -\log \sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \quad (7)$$

$$E_x^{\text{cc}}(Q, \rho) \triangleq \sup_{s \geq 0, a(\cdot)} -\rho \sum_x Q(x) \log \sum_{\bar{x}} Q(\bar{x}) \frac{e^{a(\bar{x})}}{e^{a(x)}} e^{-d_s(x, \bar{x})/\rho}. \quad (8)$$

The right-hand side of (5) can be considered a generalization of the exponent in [2] to the setting of mismatched decoding; the exponent for ML decoding is recovered by setting $s = \frac{1}{2}$. Theorem 1 shows that the exponents of [2] and [3] are equivalent even when Q is fixed; the equivalence for the optimal Q is well-known [3].

The right-hand side of (8) resembles Gallager's E_x function, which can be extended to the mismatched setting to obtain

$$E_{\text{ex}}^{\text{iid}}(Q, R) \triangleq \sup_{\rho \geq 1} E_x^{\text{iid}}(Q, \rho) - \rho R, \quad (9)$$

where

$$E_x^{\text{iid}}(Q, \rho) \triangleq \sup_{s \geq 0} -\rho \log \sum_{x, \bar{x}} Q(x) Q(\bar{x}) e^{-d_s(x, \bar{x})/\rho}. \quad (10)$$

We immediately see that $E_{\text{ex}}^{\text{cc}} \geq E_{\text{ex}}^{\text{iid}}$. While equality holds under the optimal Q for ML decoding [3], the inequality can be strict for a suboptimal Q and/or a suboptimal decoding rule.

In this paper, we seek alternative derivations of the stronger exponent $E_{\text{ex}}^{\text{cc}}$ which are not sensitive to the assumption of finite alphabets, and which remain valid for channels with input constraints.

¹The notation $Q \times Q \times W$ denotes the distribution $Q(x)Q(\bar{x})W(y|x)$.

II. ANALYSIS USING FINITE-LENGTH BOUNDS

Let $p_{e,m}(\mathcal{C})$ denote the error probability for a given codebook \mathcal{C} given that the m -th codeword was sent, and let $p_e(\mathcal{C}) \triangleq \max_m p_{e,m}(\mathcal{C})$. We fix an arbitrary codeword distribution $P_{\mathbf{X}}$ and define

$$(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}) \sim P_{\mathbf{X}}(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})P_{\mathbf{X}}(\bar{\mathbf{x}}). \quad (11)$$

Stated in a general form, Gallager's analysis proves the existence of a codebook \mathcal{C} of size $M > M' \frac{\eta}{1+\eta}$ such that

$$f(p_e(\mathcal{C})) \leq (1 + \eta) \mathbb{E}[f(p_{e,m}(\mathcal{C}))] \quad (12)$$

for any $\eta \geq 0$ and non-negative function $f(\cdot)$, where \mathcal{C} is a *random* codebook with M' codewords drawn independently from the distribution $P_{\mathbf{X}}$. In particular, we obtain

$$p_e(\mathcal{C}) \leq \left(2 \mathbb{E}[p_{e,m}(\mathcal{C})^{1/\rho}] \right)^\rho \quad (13)$$

by choosing $\eta = 1$ and $f(\cdot) = (\cdot)^{1/\rho}$, where \mathcal{C} contains $2M - 1$ codewords. The following non-asymptotic bound follows from (13) using the union bound and the inequality

$$\left(\sum_i a_i \right)^{1/\rho} \leq \sum_i a_i^{1/\rho} \quad (\rho \geq 1). \quad (14)$$

Theorem 2. *For any pair (n, M) and codeword distribution $P_{\mathbf{X}}$, there exists a codebook \mathcal{C} with M codewords of length n whose maximal error probability satisfies*

$$p_e(\mathcal{C}) \leq \inf_{\rho \geq 1} \left(4(M-1) \mathbb{E} \left[\mathbb{P} \left[q^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}, \bar{\mathbf{X}} \right]^{1/\rho} \right] \right)^\rho. \quad (15)$$

The bound in (15) extends immediately to general channels and metrics (e.g. channels with memory), and can be considered an analog of the random-coding union (RCU) bound given by Polyanskiy *et al.* [9]. In the remainder of the section, we present the resulting exponents for various ensembles in the memoryless case.

i.i.d. ensemble: Choosing the i.i.d. distribution

$$P_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n Q(x_i), \quad (16)$$

we can use Markov's inequality to weaken (15) and obtain the exponent $E_{\text{ex}}^{\text{iid}}(Q, R)$. This approach does not rely on the alphabets being finite, but it is unsuitable for input-constrained channels, since in all non-trivial cases there is a non-zero probability of violating the constraint.

Constant-composition ensemble: Suppose that $|\mathcal{X}|$ and $|\mathcal{Y}|$ are finite, and consider the constant-composition codeword distribution

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|T^n(Q_n)|} \mathbb{1} \{ \mathbf{x} \in T^n(Q_n) \}, \quad (17)$$

where Q_n is a type with the same support as Q such that $|Q_n(x) - Q(x)| \leq \frac{1}{n}$ for all x . By expanding (15) in terms

of types and applying standard properties [2, Ch. 10], we can derive the exponent

$$\sup_{\rho \geq 1} \min_{P_{X\bar{X}Y} \in \mathcal{T}^{cc}(Q)} D(P_{X\bar{X}Y} \| P_{X\bar{X}} \times W) + \rho(I_P(X; \bar{X}) - R). \quad (18)$$

Using the minimax theorem [10], we recover $E_{\text{ex}}^{\text{cc}}$ in the form given in (3). This provides a simple alternative proof to the one given in [3] based on graph decomposition techniques.

Cost-constrained ensemble: In the case of continuous alphabets and input constraints (see (2)), we can derive $E_{\text{ex}}^{\text{cc}}$ using the cost-constrained codeword distribution

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\mu_n} \prod_{i=1}^n Q(x_i) \mathbb{1}\{\mathbf{x} \in \mathcal{D}_n\}, \quad (19)$$

where

$$\mathcal{D}_n \triangleq \left\{ \mathbf{x} : \frac{1}{n} \sum_{i=1}^n c(x_i) \leq \Gamma, \right. \\ \left. \left| \frac{1}{n} \sum_{i=1}^n a_l(x_i) - \phi_l \right| \leq \frac{\delta}{n}, l = 1, \dots, L \right\}, \quad (20)$$

and where δ is a positive constant, $\{a_l(\cdot)\}_{l=1}^L$ are arbitrary auxiliary cost functions with means $\phi_l \triangleq \mathbb{E}_Q[a_l(X)]$, and μ_n is a normalizing constant.

One can show that $\mu_n \doteq 1$ provided that $\mathbb{E}_Q[c(X)] \leq \Gamma$, $\mathbb{E}_Q[c(X)^2] < \infty$ and $\mathbb{E}_Q[a_l(X)^2] < \infty$ for $l = 1, \dots, L$ [11]. Assuming these conditions are satisfied, we can analyze (15) similarly to the case of random coding without expurgation [11] to obtain the exponent

$$E_{\text{ex}}^{\text{cost}}(Q, R, \{a_l\}) \triangleq \sup_{\rho \geq 1} E_x^{\text{cost}}(Q, \rho, \{a_l\}) - \rho R, \quad (21)$$

where²

$$E_x^{\text{cost}}(Q, R, \{a_l\}) \triangleq \sup_{s \geq 0, \{r_l\}, \{\bar{r}_l\}} \\ - \rho \log \sum_{x, \bar{x}} Q(x) Q(\bar{x}) \frac{e^{\sum_{l=1}^L \bar{r}_l (a_l(\bar{x}) - \phi_l)}}{e^{\sum_{l=1}^L r_l (a_l(x) - \phi_l)}} e^{-d_s(x, y)/\rho}, \quad (22)$$

and the constants $\{r_l\}$ and $\{\bar{r}_l\}$ are arbitrary. Roughly speaking, the additional factor in (22) compared to (10) is obtained using the fact that the empirical mean of each auxiliary cost is close to the true mean.

Finally, we claim that (22) reduces to (8) when $L = 2$ and $a_1(\cdot), a_2(\cdot)$ are optimized. This is easily shown by setting $r_l = \bar{r}_l = 1$ for $l = 1, 2$, choosing $a_2(\cdot)$ such that Jensen's inequality holds with equality when $\sum_x Q(x)$ is taken outside the logarithm, and using the definition of ϕ_l to write

$$- \sum_x Q(x) \log \frac{e^{-\phi_1}}{e^{a_2(x) - \phi_2}} = - \sum_x Q(x) \log e^{-a_1(x)}. \quad (23)$$

In summary, this derivation shows that, under mild technical assumptions, (6) is an achievable exponent even in the case of infinite or continuous alphabets, provided that Q satisfies $\mathbb{E}_Q[c(X)] \leq \Gamma$ in accordance with (2).

²In the case of continuous alphabets, the summations should be replaced by integrals.

III. ANALYSIS USING ENUMERATOR FUNCTIONS

In this section, we present an alternative method for deriving expurgated exponents which is based on statistical-mechanical methods (e.g. see [7], [8]). In [4], we provide two variations of this approach depending on whether the alphabets are discrete or continuous. We begin here by discussing the discrete case.

Applying the union bound to (13), we obtain

$$p_e \leq \left(2\mathbb{E} \left[\left(\sum_{\bar{m} \neq m} \mathbb{P} \left[\frac{q^n(\mathbf{X}^{\bar{m}}, \mathbf{Y})}{q^n(\mathbf{X}^m, \mathbf{Y})} \geq 1 \mid \mathcal{C} \right] \right)^{1/\rho} \right] \right)^\rho, \quad (24)$$

where $\{\mathbf{X}^{(j)}\}_{j=1}^{2M-1} \sim \prod_{j=1}^M P_{\mathbf{X}}(\mathbf{x}^{(j)})$ are the random codewords in \mathcal{C} . For any codeword distribution $P_{\mathbf{X}}(\mathbf{x})$ depending only on the type of \mathbf{x} , we can perform an exponentially tight analysis of (24) using type enumerators [7]. For the constant-composition ensemble (see (17)), we obtain $E_{\text{ex}}^{\text{cc}}$ in the form given in (3). Although the exponent is the same as that obtained via Theorem 2, the type enumerator approach guarantees exponential tightness starting from an earlier step.

On the other hand, for the i.i.d. ensemble (see (16)), we show in [4] that (24) yields an exponent which is strictly higher in general than that obtained via Theorem 2. It follows that the inequality in (14) is not exponentially tight in general, thus motivating the more refined analysis of (24).

In the remainder of the section, we consider a more general approach which remains applicable in the continuous case. We assume that each codeword must satisfy (2), and that

$$\lim_{\gamma \rightarrow \infty} \frac{1}{\gamma} \log \log \frac{1}{\pi(\gamma)} = 0, \quad (25)$$

where

$$\pi(\gamma) \triangleq \min_{(x, \bar{x}) : c(x) \leq \gamma, c(\bar{x}) \leq \gamma} \mathbb{P}[Y_x \in \mathcal{E}(x, \bar{x})] \quad (26)$$

$$\mathcal{E}(x, \bar{x}) \triangleq \{y : q(\bar{x}, y) \geq q(x, y)\}, \quad (27)$$

and in (26) we define $Y_x \sim W(\cdot | x)$. This assumption is mild and generally easily to verify. For example, for the power-constrained additive white Gaussian noise channel with ML decoding, $\pi(\cdot)$ only decays exponentially in γ , whereas (25) allows for nearly double-exponential rates of decay. See [4] for further discussion and examples.

The following theorem follows by applying (12) with a function of the form $f(\cdot) = f_n(\cdot) = \log(\cdot) + c_n$, where c_n is chosen such that $f_n(p_{e,m})$ is non-negative for all values of $p_{e,m}$ which can occur when (25) holds.

Theorem 3. Fix $R > 0$ and consider a sequence of codebooks \mathcal{C}_n containing $M'_n = \lfloor \exp(nR) \rfloor$ codewords which are generated independently according to $P_{\mathbf{X}}$. Under the assumption in (25), there exists a sequence of codebooks \mathcal{C}_n with M_n codewords such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log M_n = R \quad (28)$$

and

$$p_e(\mathcal{C}_n) \stackrel{\dot{\leq}}{\leq} \exp\left(\mathbb{E}[\log p_{e,m}(\mathcal{C}_n)]\right) \quad (29)$$

$$\leq \exp\left(\rho \mathbb{E}\left[\log \mathbb{E}[p_{e,m}(\mathcal{C}_n)^{1/\rho} \mid \mathbf{X}^{(m)}]\right]\right), \quad (30)$$

where (30) holds for any $\rho > 0$.

Equation (30) can be thought of as improving on (13) to the fact that the expectation with respect to the transmitted codeword is outside the logarithm.

Applying the union bound and Markov's inequality to (30), we conclude that there exists a sequence of codebooks \mathcal{C}_n of rate approaching R such that

$$p_e(\mathcal{C}_n) \stackrel{\dot{\leq}}{\leq} \exp\left(\mathbb{E}[\log A_n(R, \rho, \mathbf{X}^{(m)})]\right), \quad (31)$$

where

$$A_n(R, \rho, \mathbf{X}^{(m)}) \triangleq \mathbb{E}\left[\left(\sum_{\bar{m} \neq m} e^{-d_s^n(\mathbf{X}^{(m)}, \mathbf{X}^{(\bar{m})})} \right)^{1/\rho} \mid \mathbf{X}^{(m)}\right]^\rho \quad (32)$$

and $d_s^n(\mathbf{x}, \bar{\mathbf{x}}) \triangleq \sum_{i=1}^n d_s(x_i, \bar{x}_i)$. We fix $\delta > 0$ and write

$$\sum_{\bar{m} \neq m} e^{-d_s^n(\mathbf{x}, \mathbf{X}^{(\bar{m})})} \leq \sum_{k=0}^{\infty} e^{-nk\delta} N_m(k, \mathbf{x}), \quad (33)$$

where

$$N_m(k, \mathbf{x}) \triangleq \sum_{\bar{m} \neq m} \mathbf{1}\{nk\delta \leq d_s^n(\mathbf{x}, \mathbf{X}^{(\bar{m})}) < n(k+1)\delta\}. \quad (34)$$

The key observation which permits the subsequent analysis is that the maximum value of k for which $N_m(k, \mathbf{x}) \neq 0$ grows subexponentially in n ; this can easily be verified using the assumption in (25). Applying this observation to (32) multiple times, we obtain

$$A_n(R, \rho, \mathbf{x}) \stackrel{\dot{\leq}}{\leq} \max_{k \geq 0} \left(\mathbb{E}[N_m(k, \mathbf{x})^{1/\rho}]\right)^\rho e^{-nk\delta}. \quad (35)$$

We can further upper bound this expression by removing the lower inequality in the indicator function in (34). Letting $R(D, \mathbf{x})$ be any continuous function such that $\mathbb{P}[d_s^n(\mathbf{x}, \bar{\mathbf{X}}) < nD] \stackrel{\dot{\leq}}{\leq} e^{-nR(D, \mathbf{x})}$ uniformly in \mathbf{x} , it follows by treating the cases $R(D, \mathbf{x}) \leq R$ and $R(D, \mathbf{x}) > R$ separately that

$$A_n(R, \rho, \mathbf{x}) \stackrel{\dot{\leq}}{\leq} e^{-n \min\{E_1(R, \rho, \delta, \mathbf{x}), E_2(R, \delta, \mathbf{x})\}}, \quad (36)$$

where

$$E_1(R, \rho, \delta, \mathbf{x}) \triangleq \min_{k: R((k+1)\delta, \mathbf{x}) \geq R} k\delta + \rho(R((k+1)\delta, \mathbf{x}) - R) \quad (37)$$

$$E_2(R, \delta, \mathbf{x}) \triangleq \min_{k: R((k+1)\delta, \mathbf{x}) \leq R} k\delta + R((k+1)\delta, \mathbf{x}) - R. \quad (38)$$

Finally, we obtain the following by taking $\delta \rightarrow 0$ and $\rho \rightarrow \infty$.

Theorem 4. Under the assumption in (25), the exponent

$$E_{\text{ex}}(R) \triangleq \mathbb{E}\left[\inf_{D: R(D, \mathbf{X}) \leq R} D + R(D, \mathbf{X}) - R\right] \quad (39)$$

is achievable for any continuous function $R(D, \mathbf{x})$ such that $\mathbb{P}[d_s^n(\mathbf{x}, \bar{\mathbf{X}}) < nD] \stackrel{\dot{\leq}}{\leq} e^{-nR(D, \mathbf{x})}$ uniformly in \mathbf{x} .

After a suitable modification of the definition of $d_s^n(\mathbf{x}, \bar{\mathbf{x}})$, (39) extends immediately to general channels and metrics. The ability to simplify the exponent (e.g. to a single-letter expression) depends on the form of $R(D, \mathbf{x})$, which in turn depends strongly on the codeword distribution $P_{\mathbf{X}}$. In some cases, $P_{\mathbf{X}}$ can be chosen in such a way that $R(D, \mathbf{x})$ is the same for all \mathbf{x} with $P_{\mathbf{X}}(\mathbf{x}) > 0$, thus greatly simplifying (39).

Consider the cost-constrained ensemble given in (19) with $L = 1$, and assume analogously to Section III that $\mathbb{E}_Q[c(X)] \leq \Gamma$, $\mathbb{E}[c(X)^2] < \infty$ and $\mathbb{E}[a_1(X)^2] < \infty$. Using standard Chernoff-type bounding techniques, we obtain

$$R(D, \mathbf{x}) = \sup_{t \geq 0, \bar{r}} \bar{r} \phi_1 - tD - \frac{1}{n} \sum_{i=1}^n \theta(x_i, \bar{r}, t), \quad (40)$$

where

$$\theta(x, \bar{r}, t) \triangleq \log \mathbb{E}_Q[e^{\bar{r}a_1(\bar{X}) - td_s(x, \bar{X})}]. \quad (41)$$

Substituting (40) into (39) and performing some manipulations, we obtain $E_{\text{ex}}^{\text{cc}}$ in the form given in (6), with the summations replaced by integrals where necessary. In contrast to Section III, we only require $L = 1$ instead of $L = 2$. However, this comes at the price of requiring (25) to hold.

REFERENCES

- [1] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
- [3] —, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.
- [4] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements and connections," 2013, submitted to *IEEE Trans. Inf. Theory* [Online: <http://arxiv.org/abs/1307.6679>].
- [5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [6] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
- [7] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4439–4447, Oct. 2008.
- [8] —, "Statistical physics and information theory," *Foundations and Trends in Comms. and Inf. Theory*, vol. 6, no. 1–2, pp. 1–212, 2009.
- [9] 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.
- [10] K. Fan, "Minimax theorems," *Proc. Nat. Acad. Sci.*, vol. 39, pp. 42–47, 1953.
- [11] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," submitted to *IEEE Trans. Inf. Theory* [Online: <http://arxiv.org/abs/1303.6166>].