Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations

Jonathan Scarlett, Alfonso Martinez, *Senior Member, IEEE*,
and Albert Guillén i Fàbregas, *Senior Member, IEEE*

***Abstract*— This paper considers the problem of channel coding with a given (possibly suboptimal) maximum-metric decoding rule. A cost-constrained random-coding ensemble with multiple auxiliary costs is introduced, and is shown to achieve error exponents and second-order coding rates matching those of constant-composition random coding, while being directly applicable to channels with infinite or continuous alphabets. The number of auxiliary costs required to match the error exponents and second-order rates of constant-composition coding is studied, and is shown to be at most two. For independent identically distributed random coding, asymptotic estimates of two well-known non-asymptotic bounds are given using saddlepoint approximations. Each expression is shown to characterize the asymptotic behavior of the corresponding random-coding bound at both fixed and varying rates, thus unifying the regimes characterized by error exponents, second-order rates, and moderate deviations. For fixed rates, novel exact asymptotics expressions are obtained to within a multiplicative $1 + o(1)$ term. Using numerical examples, it is shown that the saddlepoint approximations are highly accurate even at short block lengths.**

***Index Terms*— Mismatched decoding, random coding, error exponents, second-order coding rate, channel dispersion, normal approximation, saddlepoint approximation, exact asymptotics, maximum-likelihood decoding, finite-length performance.**

I. Introduction

INFORMATION-theoretic studies of channel coding typically seek to characterize the performance of coded communication systems when the encoder and decoder can be optimized. In practice, however, optimal decoding rules are often ruled out due to channel uncertainty and implementation constraints. In this paper, we consider the mismatched decoding problem [1]–[8], in which the decoder employs maximum-metric decoding with a metric which may differ from the optimal choice.

The problem of finding the highest achievable rate possible with mismatched decoding is open, and is generally believed to be difficult. Most existing work has focused on achievable rates via random coding; see Section I-C for an outline. The goal of this paper is to present a more comprehensive analysis of the random-coding error probability under various ensembles, including error exponents [9, Ch. 5], second-order coding rates [10]–[12], and refined asymptotic results based on the saddlepoint approximation [13].

A. System Setup

The input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$ respectively. The conditional probability of receiving an output vector $\boldsymbol{y} = (y_1, \dots, y_n)$ given an input vector $\boldsymbol{x} = (x_1, \dots, x_n)$ is given by

$$W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{i=1}^{n} W(y_i|x_i) \tag{1}$$

for some transition law $W(y|x)$. Except where stated otherwise, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite, and thus the channel is a discrete memoryless channel (DMC). The encoder takes as input a message m uniformly distributed on the set $\{1, \dots, M\}$, and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$ from a codebook $\mathcal{C} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}\}$. The decoder receives the vector $\boldsymbol{y}$ at the output of the channel, and forms the estimate

$$\hat{m} = \operatorname*{arg\,max}_{j \in \{1,\dots,M\}} q^n(\boldsymbol{x}^{(j)}, \boldsymbol{y}), \tag{2}$$

where $q^n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \prod_{i=1}^{n} q(x_i, y_i)$. The function $q(x, y)$ is assumed to be non-negative, and is called the *decoding metric*. In the case of a tie, a codeword achieving the maximum in (2) is selected uniformly at random. It should be noted that maximum-likelihood (ML) decoding is a special case of (2), since it is recovered by setting $q(x, y) = W(y|x)$.

An error is said to have occurred if $\hat{m}$ differs from m. A rate R is said to be achievable if, for all $\delta > 0$, there exists a sequence of codes $\mathcal{C}_n$ of length n with $M \geq e^{n(R-\delta)}$ and vanishing error probability $p_e(\mathcal{C}_n)$. An error exponent $E(R)$

Manuscript received March 25, 2013; revised October 24, 2013; accepted February 12, 2014. Date of publication March 11, 2014; date of current version April 17, 2014. This work was supported in part by the European Research Council under Grant 259663, in part by the European Union's 7th Framework Programme under Grant 303633, and in part by the Spanish Ministry of Economy and Competitiveness under Grants RYC-2011-08150 and TEC2012-38800-C03-03. This work was presented at the 2012 Allerton Conference on Communication, Computing and Control and the 2013 and 2014 Information Theory and Applications Workshop.

J. Scarlett is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: jmscarlett@gmail.com).

A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain (e-mail: alfonso.martinez@ieee.org).

A. Guillén i Fàbregas is with the Department of Information and Communication Technologies, Institució Catalana de Recerca i Estudis Avançats, Universitat Pompeu Fabra, Barcelona 08018, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

Communicated by R. F. H. Fischer, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2014.2310453

is said to be achievable if there exists a sequence of codebooks $\mathcal{C}_n$ of length n and rate R such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log p_e(\mathcal{C}_n) \geq E(R). \tag{3}$$

We let $\bar{p}_e(n, M)$ denote the average error probability with respect to a given random-coding ensemble which will be clear from the context. The random-coding error exponent $E_r(R)$ is said to exhibit ensemble tightness if

$$\lim_{n\to\infty} -\frac{1}{n}\log \bar{p}_e(n, e^{nR}) = E_r(R). \tag{4}$$

B. Notation

The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, and the set of all empirical distributions on a vector in $\mathcal{X}^n$ (i.e. types [14, Sec. 2] [15]) is denoted by $\mathcal{P}_n(\mathcal{X})$. The type of a vector $\boldsymbol{x}$ is denoted by $\hat{P}_{\boldsymbol{x}}(\cdot)$. For a given $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of sequences in $\mathcal{X}^n$ with type Q.

The probability of an event is denoted by $\mathbb{P}[\cdot]$, and the symbol $\sim$ means "distributed as". The marginals of a joint distribution $P_{XY}(x, y)$ are denoted by $P_X(x)$ and $P_Y(y)$. We write $P_X = \widetilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. Expectation with respect to a joint distribution $P_{XY}(x, y)$ is denoted by $\mathbb{E}_P[\cdot]$, or $\mathbb{E}[\cdot]$ when the associated probability distribution is understood from the context. Similar notations $I_P(X; Y)$ and $I(X; Y)$ are used for the mutual information. Given a distribution $Q(x)$ and conditional distribution $W(y|x)$, we write $Q \times W$ to denote the joint distribution $Q(x)W(y|x)$.

For two positive sequences f_n and g_n, we write $f_n \doteq g_n$ if $\lim_{n\to\infty} \frac{1}{n}\log\frac{f_n}{g_n} = 0$ and we write $f_n \,\dot{\leq}\, g_n$ if $\limsup_{n\to\infty} \frac{1}{n}\log\frac{f_n}{g_n} \leq 0$. We write $f_n \asymp g_n$ if $\lim_{n\to\infty} \frac{f_n}{g_n} = 1$, and we make use of the standard asymptotic notations $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$, $\Omega(\cdot)$ and $\omega(\cdot)$.

We denote the tail probability of a zero-mean unit-variance Gaussian variable by $\mathsf{Q}(\cdot)$, and we denote its functional inverse by $\mathsf{Q}^{-1}(\cdot)$. All logarithms have base e, and all rates are in units of nats except in the examples, where bits are used. We define $[c]^+ = \max\{0, c\}$, and denote the indicator function by $\mathbb{1}\{\cdot\}$.

C. Overview of Achievable Rates

Achievable rates for mismatched decoding have been derived using the following random-coding ensembles:

1) the i.i.d. ensemble, in which each symbol of each codeword is generated independently;
2) the constant-composition ensemble, in which each codeword is drawn uniformly from the set of sequences with a given empirical distribution;
3) the cost-constrained ensemble, in which each codeword is drawn according to an i.i.d. distribution conditioned on an auxiliary cost constraint being satisfied.

While these ensembles all yield the same achievable rate under ML decoding, i.e. the mutual information, this is not true under mismatched decoding.

The most notable early works on mismatched decoding are by Hui [2] and Csiszár and Körner [1], who used constant-composition random coding to derive the following achievable rate for mismatched DMCs, commonly known as the LM rate:

$$I_{\mathrm{LM}}(Q) = \min_{\widetilde{P}_{XY}} I_{\widetilde{P}}(X; Y), \tag{5}$$

where the minimization is over all joint distributions satisfying

$$\widetilde{P}_X(x) = Q(x) \tag{6}$$

$$\widetilde{P}_Y(y) = \sum_x Q(x)W(y|x) \tag{7}$$

$$\mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \geq \mathbb{E}_{Q\times W}[\log q(X, Y)]. \tag{8}$$

This rate can equivalently be expressed as [7]

$$I_{\mathrm{LM}}(Q) \triangleq \sup_{s\geq 0, a(\cdot)} \mathbb{E}\left[\log\frac{q(X, Y)^s e^{a(X)}}{\mathbb{E}[q(\overline{X}, Y)^s e^{a(\overline{X})} \,|\, Y]}\right], \tag{9}$$

where $(X, Y, \overline{X}) \sim Q(x)W(y|x)Q(\bar{x})$.

Another well-known rate in the literature is the generalized mutual information (GMI) [3], [7], given by

$$I_{\mathrm{GMI}}(Q) = \min_{\widetilde{P}_{XY}} D\big(\widetilde{P}_{XY} \| Q \times \widetilde{P}_Y\big), \tag{10}$$

where the minimization is over all joint distributions satisfying (7) and (8). This rate can equivalently be expressed as

$$I_{\mathrm{GMI}}(Q) \triangleq \sup_{s\geq 0} \mathbb{E}\left[\log\frac{q(X, Y)^s}{\mathbb{E}[q(\overline{X}, Y)^s \,|\, Y]}\right]. \tag{11}$$

Both (10) and (11) can be derived using i.i.d. random coding, but only the latter has been shown to remain valid in the case of continuous alphabets [3].

The GMI cannot exceed the LM rate, and the latter can be strictly higher even after the optimization of Q. Motivated by this fact, Ganti *et al.* [7] proved that (9) is achievable in the case of general alphabets. This was done by generating a number of codewords according to an i.i.d. distribution Q, and then discarding all of the codewords for which $\big|\frac{1}{n}\sum_{i=1}^n a(x_i) - \mathbb{E}_Q[a(X)]\big|$ exceeds some threshold. An alternative proof is given in [16] using cost-constrained random coding.

In the terminology of [7], (5) and (10) are primal expressions, and (9) and (11) are the corresponding dual expressions. Indeed, the latter can be derived from the former using Lagrange duality techniques [5], [17].

For binary-input DMCs, a matching converse to the LM rate was reported by Balakirsky [6]. However, in the general case, several examples have been given in which the rate is strictly smaller than the mismatched capacity [4], [5], [8]. In particular, Lapidoth [8] gave an improved rate using multiple-access coding techniques. See [18], [19] for more recent studies of multiuser coding techniques, [20] for a study of expurgated exponents, and [21] for multi-letter converse results.

D. Contributions

Motivated by the fact that most existing work on mismatched decoding has focused on achievable rates, the main goal of this paper is to present a more detailed analysis of the random-coding error probability. Our main contributions are as follows.

1) In Section II, we present a generalization of the cost-constrained ensemble in [9, Ch 7.3], [16] to include multiple auxiliary costs. This ensemble serves as an alternative to constant-composition codes for improving the performance compared to i.i.d. codes, while being applicable to channels with infinite or continuous alphabets.
2) In Section III, an ensemble-tight error exponent is given for the cost-constrained ensemble. It is shown that the exponent for the constant-composition ensemble [1] can be recovered using at most two auxiliary costs, and sometimes fewer.
3) In Section IV, an achievable second-order coding rate is given for the cost-constrained ensemble. Once again, it is shown that the performance of constant-composition coding can be matched using at most two auxiliary costs, and sometimes fewer. Our techniques are shown to provide a simple method for obtaining second-order achievability results for continuous channels.
4) In Section V, we provide refined asymptotic results for i.i.d. random coding. For two non-asymptotic random-coding bounds introduced in Section II, we give saddlepoint approximations [13] that can be computed efficiently, and that characterize the asymptotic behavior of the corresponding bounds as $n \to \infty$ at all positive rates (possibly varying with n). In the case of fixed rates, the approximations recover the prefactor growth rates obtained by Altuğ and Wagner [22], along with a novel characterization of the multiplicative $O(1)$ terms. Using numerical examples, it is shown that the approximations are remarkably accurate even at small block lengths.

II. Random-Coding Bounds and Ensembles

Throughout the paper, we consider random coding in which each codeword $\boldsymbol{X}^{(i)}$ ($i = 1, \dots, M$) is independently generated according to a given distribution $P_{\boldsymbol{X}}$. We will frequently make use of the following theorem, which provides variations of the random-coding union (RCU) bound given by Polyanskiy *et al.* [11].

Theorem 1. *For any codeword distribution $P_{\boldsymbol{X}}(\boldsymbol{x})$ and constant $s \geq 0$, the random-coding error probability $\bar{p}_e$ satisfies*

$$\frac{1}{4}\mathrm{rcu}(n, M) \leq \bar{p}_e(n, M) \leq \mathrm{rcu}(n, M) \leq \mathrm{rcu}_s(n, M), \quad (12)$$

where

$$\mathrm{rcu}(n, M) \triangleq \mathbb{E}\big[\min\big\{1, (M-1)\mathbb{P}[q^n(\overline{\boldsymbol{X}}, \boldsymbol{Y}) \geq q^n(\boldsymbol{X}, \boldsymbol{Y}) \,|\, \boldsymbol{X}, \boldsymbol{Y}]\big\}\big] \quad (13)$$

$$\mathrm{rcu}_s(n, M) \triangleq \mathbb{E}\left[\min\left\{1, (M-1)\frac{\mathbb{E}[q^n(\overline{\boldsymbol{X}}, \boldsymbol{Y})^s \,|\, \boldsymbol{Y}]}{q^n(\boldsymbol{X}, \boldsymbol{Y})^s}\right\}\right] \quad (14)$$

with $(\boldsymbol{X}, \boldsymbol{Y}, \overline{\boldsymbol{X}}) \sim P_{\boldsymbol{X}}(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})P_{\boldsymbol{X}}(\bar{\boldsymbol{x}})$.

Proof: Similarly to [11], we obtain the upper bound rcu by writing

$$\bar{p}_e(n, M) \leq \mathbb{P}\bigg[\bigcup_{i \neq m}\big\{q^n(\boldsymbol{X}^{(i)}, \boldsymbol{Y}) \geq q^n(\boldsymbol{X}, \boldsymbol{Y})\big\}\bigg] \quad (15)$$

$$= \mathbb{E}\bigg[\mathbb{P}\bigg[\bigcup_{i \neq m}\big\{q^n(\boldsymbol{X}^{(i)}, \boldsymbol{Y}) \geq q^n(\boldsymbol{X}, \boldsymbol{Y})\big\} \,\bigg|\, \boldsymbol{X}, \boldsymbol{Y}\bigg]\bigg] \quad (16)$$

$$\leq \mathrm{rcu}(n, M), \quad (17)$$

where (15) follows by upper bounding the random-coding error probability by that of the decoder which breaks ties as errors, and (17) follows by applying the truncated union bound. To prove the lower bound in (12), it suffices to show that each of the upper bounds in (15) and (17) is tight to within a factor of two. The matching lower bound to (15) follows since whenever a tie occurs it must be between at least two codewords [23], and the matching lower bound to (17) follows since the union is over independent events [24, Lemma A.2]. We obtain the upper bound rcu_s by applying Markov's inequality to the inner probability in (13). ■

We consider the cost-constrained ensemble characterized by the following codeword distribution:

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbb{1}\big\{\boldsymbol{x} \in \mathcal{D}_n\big\}, \quad (18)$$

where

$$\mathcal{D}_n \triangleq \left\{\boldsymbol{x} \,:\, \left|\frac{1}{n}\sum_{i=1}^{n} a_l(x_i) - \phi_l\right| \leq \frac{\delta}{n},\, l = 1, \dots, L\right\}, \quad (19)$$

and where μ_n is a normalizing constant, δ is a positive constant, and for each $l = 1, \dots, L$, $a_l(\cdot)$ is a real-valued function on $\mathcal{X}$, and $\phi_l \triangleq \mathbb{E}_Q[a_l(X)]$. We refer to each function $a_l(\cdot)$ as an auxiliary cost function, or simply a cost. Roughly speaking, each codeword is generated according to an i.i.d. distribution conditioned on the empirical mean of each cost function $a_l(x)$ being close to the true mean. This generalizes the ensemble studied in [9, Sec. 7.3], [16] by including multiple costs.

The cost functions $\{a_l(\cdot)\}_{l=1}^{L}$ in (18) should not be viewed as being chosen to meet a system constraint (e.g. power limitations). Rather, they are introduced in order to improve the performance of the random-coding ensemble itself. However, system costs can be handled similarly; see Section VI for details. The constant δ in (19) could, in principle, vary with l and n, but a fixed value will suffice for our purposes.

In the case that $L = 0$, it should be understood that $\mathcal{D}_n$ contains all $\boldsymbol{x}$ sequences. In this case, (18) reduces to the i.i.d. ensemble, which is characterized by

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} Q(x_i). \quad (20)$$

A less obvious special case of (18) is the constant-composition ensemble, which is characterized by

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{|T^n(Q_n)|}\mathbb{1}\big\{\boldsymbol{x} \in T^n(Q_n)\big\}, \quad (21)$$

where Q_n is a type such that $\max_x |Q_n(x) - Q(x)| \le \frac{1}{n}$. That is, each codeword is generated uniformly over the type class $T^n(Q_n)$, and hence each codeword has the same composition. To recover this ensemble from (18), we replace Q by Q_n and choose the parameters $L = |\mathcal{X}|$, $\delta < 1$ and

$$a_l(x) = \mathbb{1}\{x = l\}, \quad l = 1, \dots, |\mathcal{X}|, \tag{22}$$

where we assume without loss of generality that $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$.

The following proposition shows that the normalizing constant μ_n in (18) decays at most polynomially in n. When $|\mathcal{X}|$ is finite, this can easily be shown using the method of types. In particular, choosing the functions given in the previous paragraph to recover the constant-composition ensemble, we have $\mu_n \ge (n+1)^{-(|\mathcal{X}|-1)}$ [14, p. 17]. For the sake of generality, we present a proof which applies to more general alphabets, subject to minor technical conditions. The case $L = 1$ was handled in [9, Ch. 7.3].

Proposition 1. *Fix an input alphabet $\mathcal{X}$ (possibly infinite or continuous), an input distribution $Q \in \mathcal{P}(\mathcal{X})$ and the auxiliary cost functions $a_1(\cdot), \dots, a_L(\cdot)$. If $\mathbb{E}_Q[a_l(X)^2] < \infty$ for $l = 1, \dots, L$, then there exists a choice of $\delta > 0$ such that the normalizing constant in* (18) *satisfies $\mu_n = \Omega(n^{-L/2})$.*

Proof: This result follows from the multivariate local limit theorem in [25, Cor. 1], which gives asymptotic expressions for probabilities of i.i.d. random vectors taking values in sets of the form (19). Let Σ denote the covariance matrix of the vector $[a_1(X), \dots, a_L(X)]^T$. We have by assumption that the entries of Σ are finite. Under the additional assumption $\det(\Sigma) > 0$, [25, Cor. 1] states that $\mu_n = \Theta(n^{-L/2})$ provided that δ is at least as high as the largest span of the $a_l(X)$ $(X \sim Q)$ which are lattice variables.[1] If all such variables are non-lattice, then δ can take any positive value.

It only remains to handle the case $\det(\Sigma) = 0$. Suppose that Σ has rank $L' < L$, and assume without loss of generality that $a_1(\cdot), \dots, a_{L'}(\cdot)$ are linearly independent. Up to sets whose probability with respect to Q is zero, the remaining costs $a_{L'+1}(\cdot), \dots, a_L(\cdot)$ can be written as linear combinations of the first L' costs. Letting α denote the largest magnitude of the scalar coefficients in these linear combinations, we conclude that $\boldsymbol{x} \in \mathcal{D}_n$ provided that

$$\left| \frac{1}{n} \sum_{i=1}^{n} a_l(x_i) - \phi_l \right| \le \frac{\delta}{\alpha L' n} \tag{23}$$

for $l = 1, \dots, L'$. The proposition follows by choosing δ to be at least as high as $\alpha L'$ times the largest span of the $a_l(X)$ which are lattice variables, and analyzing the first L' costs analogously to the case that $\det(\Sigma) > 0$. ■

In accordance with Proposition 1, we henceforth assume that the choice of δ for the cost-constrained ensemble is such that $\mu_n = \Omega(n^{-L/2})$.

III. Random-Coding Error Exponents

Error exponents characterize the asymptotic exponential behavior of the error probability in coded communication systems, and can thus provide additional insight beyond capacity results. In the matched setting, error exponents were studied by Fano [26, Ch. 9], and later by Gallager [9, Ch. 5] and Csiszár-Körner [14, Ch. 10]. The ensemble tightness of the exponent (cf. (4)) under ML decoding was studied by Gallager [27] and D'yachkov [28] for the i.i.d. and constant-composition ensembles respectively.

In this section, we present the ensemble-tight error exponent for cost-constrained random coding, yielding results for the i.i.d. and constant-composition ensembles as special cases.

A. Cost-Constrained Ensemble

We define the sets

$$\mathcal{S}(\{a_l\}) \triangleq \big\{ P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \mathbb{E}_P[a_l(X)] = \phi_l \, (l = 1, \dots, L) \big\} \tag{24}$$

$$\mathcal{T}(P_{XY}, \{a_l\}) \triangleq \Big\{ \widetilde{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \mathbb{E}_{\widetilde{P}}[a_l(X)] = \phi_l \ (l = 1, \dots, L), \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)] \Big\}, \tag{25}$$

where the notation $\{a_l\}$ is used to denote dependence on $a_1(\cdot), \dots, a_L(\cdot)$. The dependence of these sets on Q (via $\phi_l = \mathbb{E}_Q[a_l(X)]$) is kept implicit.

Theorem 2. *The random-coding error probability for the cost-constrained ensemble in* (18) *satisfies*

$$\lim_{n \to \infty} -\frac{1}{n} \log \bar{p}_e(n, e^{nR}) = E_r^{\text{cost}}(Q, R, \{a_l\}), \tag{26}$$

where

$$E_r^{\text{cost}}(Q, R, \{a_l\}) \triangleq \min_{P_{XY} \in \mathcal{S}(\{a_l\})} \min_{\widetilde{P}_{XY} \in \mathcal{T}(P_{XY}, \{a_l\})} D(P_{XY} \| Q \times W) + \big[D(\widetilde{P}_{XY} \| Q \times P_Y) - R \big]^+. \tag{27}$$

Proof: See Appendix A. ■

The optimization problem in (27) is convex when the input distribution and auxiliary cost functions are fixed. The following theorem gives an alternative expression based on Lagrange duality [17].

Theorem 3. *The error exponent in* (27) *can be expressed as*

$$E_r^{\text{cost}}(Q, R, \{a_l\}) = \max_{\rho \in [0,1]} E_0^{\text{cost}}(Q, \rho, \{a_l\}) - \rho R, \tag{28}$$

where

$$E_0^{\text{cost}}(Q, \rho, \{a_l\}) \triangleq \sup_{s \ge 0, \{r_l\}, \{\overline{r}_l\}} -\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X}, Y)^s e^{\sum_{l=1}^{L} \overline{r}_l (a_l(\overline{X}) - \phi_l)} \,|\, Y \big]}{q(X,Y)^s e^{\sum_{l=1}^{L} r_l (a_l(X) - \phi_l)}} \right)^{\rho} \right] \tag{29}$$

and $(X, Y, \overline{X}) \sim Q(x)W(y|x)Q(\bar{x})$.

Proof: See Appendix B. ■

The derivation of (28)–(29) via Theorem 2 is useful for proving ensemble tightness, but has the disadvantage of being applicable only in the case of finite alphabets. We proceed

[1] We say that X is a lattice random variable with offset γ and span h if its support is a subset of $\{\gamma + ih : i \in \mathbb{Z}\}$, and the same cannot remain true by increasing h.

by giving a direct derivation which does not prove ensemble tightness, but which extends immediately to more general alphabets provided that the second moments associated with the cost functions are finite (see Proposition 1). The extension to channels with input constraints is straightforward; see Section VI for details.

Using Theorem 1 and applying $\min\{1, \alpha\} \le \alpha^{\rho}$ ($\rho \in [0, 1]$) to rcu_s in (14), we obtain[2]

$$\bar{p}_e(n, M) \le \frac{1}{\mu_n^{1+\rho}} M^{\rho} \sum_{\boldsymbol{x} \in \mathcal{D}_n, \boldsymbol{y}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \times \left(\frac{\sum_{\bar{\boldsymbol{x}} \in \mathcal{D}_n} Q^n(\bar{\boldsymbol{x}}) q^n(\bar{\boldsymbol{x}}, \boldsymbol{y})^s}{q^n(\boldsymbol{x}, \boldsymbol{y})^s} \right)^{\rho}, \tag{30}$$

where $Q^n(\boldsymbol{x}) \triangleq \prod_{i=1}^n Q(x_i)$. From (19), each codeword $\boldsymbol{x} \in \mathcal{D}_n$ satisfies

$$e^{r(a_l^n(\boldsymbol{x}) - n\phi_l)} e^{|r|\delta} \ge 1 \tag{31}$$

for any real number r, where $a_l^n(\boldsymbol{x}) \triangleq \sum_{i=1}^n a_l(x_i)$. Weakening (30) by applying (31) multiple times, we obtain

$$\bar{p}_e(n, M) \le \frac{e^{\rho \sum_l (|r_l| + |\overline{r}_l|)\delta}}{\mu_n^{1+\rho}} M^{\rho} \sum_{\boldsymbol{x} \in \mathcal{D}_n, \boldsymbol{y}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \times \left(\frac{\sum_{\bar{\boldsymbol{x}} \in \mathcal{D}_n} Q^n(\bar{\boldsymbol{x}}) q^n(\bar{\boldsymbol{x}}, \boldsymbol{y})^s e^{\sum_l \overline{r}_l (a_l^n(\bar{\boldsymbol{x}}) - n\phi_l)}}{q^n(\boldsymbol{x}, \boldsymbol{y})^s e^{\sum_l r_l (a_l^n(\boldsymbol{x}) - n\phi_l)}} \right)^{\rho}, \tag{32}$$

where $\{r_l\}$ and $\{\overline{r}_l\}$ are arbitrary. Further weakening (32) by replacing the summations over $\mathcal{D}_n$ with summations over all sequences, and expanding each term in the outer summation as product from $i = 1$ to n, we obtain

$$\bar{p}_e(n, M) \le \frac{e^{\rho \sum_l (|r_l| + |\overline{r}_l|)\delta}}{\mu_n^{1+\rho}} M^{\rho} \left(\sum_{x,y} Q(x) W(y|x) \times \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\sum_l \overline{r}_l (a_l(\bar{x}) - \phi_l)}}{q(x, y)^s e^{\sum_l r_l (a_l(x) - \phi_l)}} \right)^{\rho} \right)^n. \tag{33}$$

Since μ_n decays to zero subexponentially in n (cf. Proposition 1), we conclude that the prefactor in (33) does not affect the exponent. Hence, and setting $M = e^{nR}$, we obtain (28).

The preceding analysis can be considered a refinement of that of Shamai and Sason [16], who showed that an achievable error exponent in the case that $L = 1$ is given by

$$E_r^{\mathrm{cost}'}(Q, R, a_1) \triangleq \max_{\rho \in [0,1]} E_0^{\mathrm{cost}'}(Q, \rho, a_1) - \rho R, \tag{34}$$

where

$$E_0^{\mathrm{cost}'}(Q, \rho, a_1) \triangleq \sup_{s \ge 0} -\log \mathbb{E}\left[\left(\frac{\mathbb{E}[q(\overline{X}, Y)^s e^{a_1(\overline{X})} \mid Y]}{q(X, Y)^s e^{a_1(X)}} \right)^{\rho} \right]. \tag{35}$$

By setting $r_1 = \overline{r}_1 = 1$ in (29), we see that E_r^{cost} with $L = 1$ is at least as high as $E_r^{\mathrm{cost}'}$. In Section III-C, we show that the former can be strictly higher.

B. i.i.d. and Constant-Composition Ensembles

Setting $L = 0$ in (29), we recover the exponent of Kaplan and Shamai [3], namely

$$E_r^{\mathrm{iid}}(Q, R) \triangleq \max_{\rho \in [0,1]} E_0^{\mathrm{iid}}(Q, \rho) - \rho R, \tag{36}$$

where

$$E_0^{\mathrm{iid}}(Q, \rho) \triangleq \sup_{s \ge 0} -\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X}, Y)^s \mid Y\big]}{q(X, Y)^s} \right)^{\rho} \right]. \tag{37}$$

In the special case of constant-composition random coding (see (21)–(22)), the constraints $\mathbb{E}_{\widetilde{P}}[a_l(X)] = \phi_l$ for $l = 1, \dots, |\mathcal{X}|$ yield $P_X = Q$ and $\widetilde{P}_X = Q$ in (24) and (25) respectively, and thus (27) recovers Csiszár's exponent for constant-composition coding [1]. Hence, the exponents of [1], [3] are tight with respect to the ensemble average.

We henceforth denote the exponent for the constant-composition ensemble by $E_r^{\mathrm{cc}}(Q, R)$. We claim that

$$E_r^{\mathrm{cc}}(Q, R) = \max_{\rho \in [0,1]} E_0^{\mathrm{cc}}(Q, \rho) - \rho R, \tag{38}$$

where

$$E_0^{\mathrm{cc}}(Q, \rho) = \sup_{s \ge 0, a(\cdot)} \mathbb{E}\left[-\log \mathbb{E}\left[\left(\frac{\mathbb{E}\big[q(\overline{X}, Y)^s e^{a(\overline{X})} \mid Y\big]}{q(X, Y)^s e^{a(X)}} \right)^{\rho} \,\middle|\, X \right] \right]. \tag{39}$$

To prove this, we first note from (22) that

$$\begin{aligned} \sum_l r_l (a_l(x) - \phi_l) &= \sum_{\widetilde{x}} r_{\widetilde{x}} (\mathbb{1}\{x = \widetilde{x}\} - Q(\widetilde{x})) && (40) \\ &= r(x) - \phi_r, && (41) \end{aligned}$$

where (40) follows since $\phi_l = \mathbb{E}_Q[\mathbb{1}\{x = l\}] = Q(l)$, and (41) follows by defining $r(x) \triangleq r_x$ and $\phi_r \triangleq \mathbb{E}_Q[r(X)]$. Defining $\overline{r}(x)$ and $\phi_{\overline{r}}$ similarly, we obtain the following E_0 function from (29):

$$\begin{aligned} E_0^{\mathrm{cc}}(Q, \rho) &\triangleq \sup_{s \ge 0, r(\cdot), \overline{r}(\cdot)} -\log \sum_{x,y} Q(x) W(y|x) \times \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\overline{r}(\bar{x}) - \phi_{\overline{r}}}}{q(x, y)^s e^{r(x) - \phi_r}} \right)^{\rho} && (42) \\ &\le \sup_{s \ge 0, r(\cdot), \overline{r}(\cdot)} -\sum_x Q(x) \log \sum_y W(y|x) \times \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\overline{r}(\bar{x}) - \phi_{\overline{r}}}}{q(x, y)^s e^{r(x) - \phi_r}} \right)^{\rho} && (43) \\ &= \sup_{s \ge 0, \overline{r}(\cdot)} -\sum_x Q(x) \log \sum_y W(y|x) \times \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\overline{r}(\bar{x})}}{q(x, y)^s e^{\overline{r}(x)}} \right)^{\rho}, && (44) \end{aligned}$$

where (43) follows from Jensen's inequality, and (44) follows by using the definitions of ϕ_r and $\phi_{\overline{r}}$ to write

$$-\sum_x Q(x) \log \left(\frac{e^{-\phi_{\overline{r}}}}{e^{r(x) - \phi_r}} \right)^{\rho} = -\sum_x Q(x) \log \left(\frac{1}{e^{\overline{r}(x)}} \right)^{\rho}. \tag{45}$$

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

Renaming $\overline{r}(\cdot)$ as $a(\cdot)$, we see that (44) coincides with (39). It remains to show that equality holds in (43). This is easily seen by noting that the choice

$$r(x) = \frac{1}{\rho}\log\sum_{y} W(y|x)\bigg(\frac{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^{s}e^{\overline{r}(\bar{x})}}{q(x,y)^{s}}\bigg)^{\rho} \quad (46)$$

makes the logarithm in (43) independent of x, thus ensuring that Jensen's inequality holds with equality.

The exponent $E_r^{\rm iid}(Q,R)$ is positive for all rates below $I_{\rm GMI}(Q)$ [3], whereas $E_r^{\rm cc}$ recovers the stronger rate $I_{\rm LM}(Q)$. Similarly, both $E_r^{\rm cost}$ ($L=1$) and $E_r^{\rm cost'}$ recover the LM rate provided that the auxiliary cost is optimized [16].

C. Number of Auxiliary Costs Required

We claim that

$$E_r^{\rm iid}(Q,R) \le E_r^{\rm cost}(Q,R,\{a_l\}) \le E_r^{\rm cc}(Q,R). \quad (47)$$

The first inequality follows by setting $r_l = \overline{r}_l = 0$ in (29), and the second inequality follows by setting $r(x) = \sum_l r_l a_l(x)$ and $\overline{r}(x) = \sum_l \overline{r}_l a_l(x)$ in (29), and upper bounding the objective by taking the supremum over all $r(\cdot)$ and $\overline{r}(\cdot)$ to recover $E_0^{\rm cc}$ in the form given in (42). Thus, the constant-composition ensemble yields the best error exponent of the three ensembles.

In this subsection, we study the number of auxiliary costs required for cost-constrained random coding to achieve $E_r^{\rm cc}$. Such an investigation is of interest in gaining insight into the codebook structure, and since the subexponential prefactor in (33) grows at a slower rate when L is reduced (see Proposition 1). Our results are summarized in the following theorem.

Theorem 4. *Consider a DMC W and input distribution Q.*

1) *For any decoding metric, we have*

$$\sup_{a_1(\cdot),a_2(\cdot)} E_r^{\rm cost}(Q,R,\{a_1,a_2\}) = E_r^{\rm cc}(Q,R) \quad (48)$$

$$\max_{Q}\sup_{a_1(\cdot)} E_r^{\rm cost'}(Q,R,a_1) = \max_{Q} E_r^{\rm cc}(Q,R). \quad (49)$$

2) *If $q(x,y) = W(y|x)$ (ML decoding), then*

$$\sup_{a_1(\cdot)} E_r^{\rm cost}(Q,R,a_1) = E_r^{\rm cc}(Q,R) \quad (50)$$

$$\sup_{a_1(\cdot)} E_r^{\rm cost'}(Q,R,a_1) = E_r^{\rm iid}(Q,R) \quad (51)$$

$$\max_{Q} E_r^{\rm iid}(Q,R) = \max_{Q} E_r^{\rm cc}(Q,R). \quad (52)$$

Proof: We have from (47) that $E_r^{\rm cost} \le E_r^{\rm cc}$. To obtain the reverse inequality corresponding to (48), we set $L=2$, $r_1 = \overline{r}_2 = 1$ and $r_2 = \overline{r}_1 = 0$ in (29). The resulting objective coincides with (42) upon setting $a_1(\cdot) = r(\cdot)$ and $a_2(\cdot) = \overline{r}(\cdot)$.

To prove (49), we note the following observation from Appendix C: Given $s > 0$ and $\rho > 0$, any pair (Q,a) maximizing the objective in (39) must satisfy the property that the logarithm in (39) has the same value for all x such that $Q(x) > 0$. It follows that the objective in (39) is unchanged when the expectation with respect to X is moved inside the logarithm, thus yielding the objective in (35).

We now turn to the proofs of (50)–(52). We claim that, under ML decoding, we can write $E_0^{\rm cc}$ as

$$E_0^{\rm cc}(Q,\rho) = \sup_{a(\cdot)} -\log\sum_{y}\bigg(\sum_{x} Q(x)W(y|x)^{\frac{1}{1+\rho}}e^{a(x)-\phi_a}\bigg)^{1+\rho}, \quad (53)$$

where $\phi_a \triangleq \mathbb{E}_Q[a(X)]$. To show this, we make use of the form of $E_0^{\rm cc}$ given in (42), and write the summation inside the logarithm as

$$\sum_{y}\bigg(\sum_{x} Q(x)W(y|x)^{1-s\rho}e^{-\rho(r(x)-\phi_r)}\bigg) \times\bigg(\sum_{\bar{x}} Q(\bar{x})W(y|\bar{x})^{s}e^{\overline{r}(\bar{x})-\phi_{\overline{r}}}\bigg)^{\rho}. \quad (54)$$

Using Hölder's inequality in an identical fashion to [9, Ex. 5.6], this summation is lower bounded by

$$\sum_{y}\bigg(\sum_{x} Q(x)W(y|x)^{\frac{1}{1+\rho}}e^{\overline{r}(x)-\phi_{\overline{r}}}\bigg)^{1+\rho} \quad (55)$$

with equality if and only if $s = \frac{1}{1+\rho}$ and $\overline{r}(\cdot) = -\rho r(\cdot)$. Renaming $\overline{r}(\cdot)$ as $a(\cdot)$, we obtain (53). We can clearly achieve $E_r^{\rm cc}$ using $L=2$ with the cost functions $r(\cdot)$ and $\overline{r}(\cdot)$. However, since we have shown that one is a scalar multiple of the other, we conclude that $L=1$ suffices.

A similar argument using Hölder's inequality reveals that the objective in (35) is maximized by $s = \frac{1}{1+\rho}$ and $a_1(\cdot) = 0$, and the objective in (37) is maximized by $s = \frac{1}{1+\rho}$, thus yielding (51). Finally, combining (49) and (51), we obtain (52). ■

Theorem 4 shows that the cost-constrained ensemble recovers $E_r^{\rm cc}$ using at most two auxiliary costs. If either the input distribution or decoding rule is optimized, then $L=1$ suffices (see (49) and (50)), and if both are optimized then $L=0$ suffices (see (52)). The latter result is well-known [15] and is stated for completeness. While (49) shows that $E_r^{\rm cost}$ and $E_r^{\rm cost'}$ coincide when Q is optimized, (50)–(51) show that the former can be strictly higher for a given Q even when $L=1$, since $E_r^{\rm cc}$ can exceed $E_r^{\rm iid}$ even under ML decoding [15].

D. Numerical Example

We consider the channel defined by the entries of the $|\mathcal{X}|\times|\mathcal{Y}|$ matrix

$$\begin{bmatrix} 1-2\delta_0 & \delta_0 & \delta_0 \\ \delta_1 & 1-2\delta_1 & \delta_1 \\ \delta_2 & \delta_2 & 1-2\delta_2 \end{bmatrix} \quad (56)$$

with $\mathcal{X} = \mathcal{Y} = \{0,1,2\}$. The mismatched decoder chooses the codeword which is closest to $\boldsymbol{y}$ in terms of Hamming distance. For example, the decoding metric can be taken to be the entries of (56) with δ_i replaced by $\delta \in (0,\frac{1}{3})$ for $i = 1,2,3$. We let $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (0.1,0.3,0.6)$. Under these parameters, we have $I_{\rm GMI}(Q) = 0.387$, $I_{\rm LM}(Q) = 0.449$ and $I(X;Y) = 0.471$ bits/use.

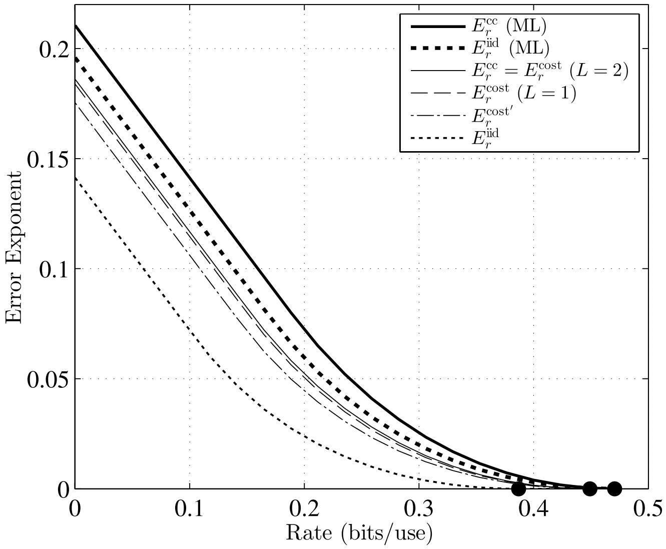

Fig. 1. Error exponents for the channel defined in (56) with $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (0.1, 0.3, 0.6)$. The mismatched decoder uses the minimum Hamming distance metric. The corresponding achievable rates $I_{\mathrm{GMI}}(Q)$, $I_{\mathrm{LM}}(Q)$ and $I(X;Y)$ are respectively marked on the horizontal axis.

We evaluate the exponents using the optimization software YALMIP [29]. For the cost-constrained ensemble with $L = 1$, we optimize the auxiliary cost. As expected, Fig. 1 shows that the highest exponent is E_r^{cc}. The exponent E_r^{cost} $(L = 1)$ is only marginally lower than E_r^{cc}, whereas the gap to $E_r^{\mathrm{cost}'}$ is larger. The exponent E_r^{iid} is not only lower than each of the other exponents, but also yields a worse achievable rate. In the case of ML decoding, E_r^{cc} exceeds E_r^{iid} for all $R < I(X;Y)$.

IV. Second-Order Coding Rates

In the matched setting, the finite-length performance limits of a channel are characterized by $M^*(n, \epsilon)$, defined to be the maximum number of codewords of length n yielding an error probability not exceeding ϵ for some encoder and decoder. The problem of finding the second-order asymptotics of $M^*(n, \epsilon)$ for a given ϵ was studied by Strassen [10], and later revisited by Polyanskiy *et al.* [11] and Hayashi [12], among others. For DMCs, we have under mild technical conditions that

$$\log M^*(n, \epsilon) = nC - \sqrt{nV}\mathsf{Q}^{-1}(\epsilon) + O(\log n), \qquad (57)$$

where C is the channel capacity, and V is known as the channel dispersion. Results of the form (57) provide a quantification of the speed of convergence to the channel capacity as the block length increases.

In this section, we present achievable second-order coding rates for the ensembles given in Section I, i.e. expansions of the form (57) with the equality replaced by $\geq$. To distinguish between the ensembles, we define $M^{\mathrm{iid}}(Q, n, \epsilon)$, $M^{\mathrm{cc}}(Q, n, \epsilon)$ and $M^{\mathrm{cost}}(Q, n, \epsilon)$ to be the maximum number of codewords of length n such that the random-coding error probability does not exceed ϵ for the i.i.d., constant-composition and cost-constrained ensembles respectively, using the input distribution Q. We first consider the discrete memoryless setting, and then discuss more general memoryless channels.

A. Cost-Constrained Ensemble

A key quantity in the second-order analysis for ML decoding is the information density, given by

$$i(x, y) \triangleq \log \frac{W(y|x)}{\sum_x Q(x) W(y|x)}, \qquad (58)$$

where Q is a given input distribution. In the mismatched setting, the relevant generalization of $i(x, y)$ is

$$i_{s,a}(x, y) \triangleq \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}, \qquad (59)$$

where $s \geq 0$ and $a(\cdot)$ are fixed parameters. We write $i_{s,a}^n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \sum_{i=1}^n i_{s,a}(x_i, y_i)$ and similarly $Q^n(\boldsymbol{x}) \triangleq \prod_{i=1}^n Q(x_i)$ and $a^n(\boldsymbol{x}) \triangleq \sum_{i=1}^n a(x_i)$. We define

$$I_{s,a}(Q) \triangleq \mathbb{E}[i_{s,a}(X, Y)] \qquad (60)$$

$$U_{s,a}(Q) \triangleq \mathrm{Var}[i_{s,a}(X, Y)] \qquad (61)$$

$$V_{s,a}(Q) \triangleq \mathbb{E}\big[\mathrm{Var}[i_{s,a}(X, Y) \,|\, X]\big], \qquad (62)$$

where $(X, Y) \sim Q \times W$. From (9), we see that the LM rate is equal to $I_{s,a}(Q)$ after optimizing s and $a(\cdot)$.

We can relate (60)–(62) with the E_0 functions defined in (35) and (39). Letting $E_0^{\mathrm{cost}'}(Q, \rho, s, a)$ and $E_0^{\mathrm{cc}}(Q, \rho, s, a)$ denote the corresponding objectives with fixed (s, a) in place of the supremum, we have $I_{s,a} = \frac{\partial E_0^{\mathrm{cost}'}}{\partial \rho}\Big|_{\rho=0} = \frac{\partial E_0^{\mathrm{cc}}}{\partial \rho}\Big|_{\rho=0}$, $U_{s,a} = -\frac{\partial^2 E_0^{\mathrm{cost}'}}{\partial \rho^2}\Big|_{\rho=0}$, and $V_{s,a} = -\frac{\partial^2 E_0^{\mathrm{cc}}}{\partial \rho^2}\Big|_{\rho=0}$. The latter two identities generalize a well-known connection between the exponent and dispersion in the matched case [11, p. 2337].

The main result of this subsection is the following theorem, which considers the cost-constrained ensemble. Our proof differs from the usual proof using threshold-based random-coding bounds [10], [11], but the latter approach can also be used in the present setting [30]. Our analysis can be interpreted as performing a normal approximation of rcu_s in (14).

Theorem 5. *Fix the input distribution Q and the parameters $s \geq 0$ and $a(\cdot)$. Using the cost-constrained ensemble in* (18) *with $L = 2$ and*

$$a_1(x) = a(x) \qquad (63)$$

$$a_2(x) = \mathbb{E}_{W(\cdot|x)}[i_{s,a}(x, Y)], \qquad (64)$$

the following expansion holds:

$$\log M^{\mathrm{cost}}(Q, n, \epsilon) \geq n I_{s,a}(Q) - \sqrt{n V_{s,a}(Q)}\mathsf{Q}^{-1}(\epsilon) + O(\log n). \qquad (65)$$

Proof: Throughout the proof, we make use of the random variables $(X, Y, \overline{X}) \sim Q(x)W(y|x)Q(\bar{x})$ and $(\boldsymbol{X}, \boldsymbol{Y}, \overline{\boldsymbol{X}}) \sim P_{\boldsymbol{X}}(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) P_{\boldsymbol{X}}(\bar{\boldsymbol{x}})$. Probabilities, expectations, etc. containing a realization $\boldsymbol{x}$ of $\boldsymbol{X}$ are implicitly defined to be conditioned on the event $\boldsymbol{X} = \boldsymbol{x}$.

We start with Theorem 1 and weaken rcu_s in (14) as follows:

$$\mathrm{rcu}_s(n, M)$$

$$= \mathbb{E}\left[\min\left\{1, (M-1)\frac{\sum_{\bar{\boldsymbol{x}} \in \mathcal{D}_n} P_{\boldsymbol{X}}(\bar{\boldsymbol{x}}) q^n(\bar{\boldsymbol{x}}, \boldsymbol{Y})^s}{q^n(\boldsymbol{X}, \boldsymbol{Y})^s}\right\}\right] \qquad (66)$$

$$\leq \mathbb{E}\left[\min\left\{1, M e^{2\delta} \frac{\sum_{\bar{\boldsymbol{x}} \in \mathcal{D}_n} P_{\boldsymbol{X}}(\bar{\boldsymbol{x}}) q^n(\bar{\boldsymbol{x}}, \boldsymbol{Y})^s e^{a^n(\bar{\boldsymbol{x}})}}{q^n(\boldsymbol{X}, \boldsymbol{Y})^s e^{a^n(\boldsymbol{X})}}\right\}\right] \qquad (67)$$

$$\leq \mathbb{E}\left[\min\left\{1, \frac{Me^{2\delta}}{\mu_n}\frac{\sum_{\bar{\boldsymbol{x}}} Q^n(\bar{\boldsymbol{x}})q^n(\bar{\boldsymbol{x}},\boldsymbol{Y})^s e^{a^n(\bar{\boldsymbol{x}})}}{q^n(\boldsymbol{X},\boldsymbol{Y})^s e^{a^n(\boldsymbol{X})}}\right\}\right] \tag{68}$$

$$= \mathbb{P}\left[i_{s,a}^n(\boldsymbol{X},\boldsymbol{Y}) + \log U \leq \log\frac{Me^{2\delta}}{\mu_n}\right] \tag{69}$$

$$\leq \mathbb{P}\left[i_{s,a}^n(\boldsymbol{X},\boldsymbol{Y}) + \log U \leq \log\frac{Me^{2\delta}}{\mu_n} \cap \boldsymbol{X} \in \mathcal{A}_n\right] + \mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big] \tag{70}$$

$$\leq \max_{\boldsymbol{x}\in\mathcal{A}_n} \mathbb{P}\left[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y}) + \log U \leq \log\frac{Me^{2\delta}}{\mu_n}\right] + \mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big], \tag{71}$$

where (67) follows from (31), (68) follows by substituting the random-coding distribution in (18) and summing over all $\bar{\boldsymbol{x}}$ instead of $\bar{\boldsymbol{x}} \in \mathcal{D}_n$, and (69) follows from the definition of $i_{s,a}^n$ and the identity

$$\mathbb{E}[\min\{1, A\}] = \mathbb{P}[A \geq U], \tag{72}$$

where A is an arbitrary non-negative random variable, and U is uniform on $(0, 1)$ and independent of A. Finally, (70) holds for any set $\mathcal{A}_n$ by the law of total probability.

We treat the cases $V_{s,a}(Q) > 0$ and $V_{s,a}(Q) = 0$ separately. In the former case, we choose

$$\mathcal{A}_n = \left\{\boldsymbol{x} \in \mathcal{D}_n \,:\, \left|\frac{1}{n}v_{s,a}^n(\boldsymbol{x}) - V_{s,a}(Q)\right| \leq \zeta\sqrt{\frac{\log n}{n}}\right\}, \tag{73}$$

where ζ is a constant, and $v_{s,a}^n(\boldsymbol{x}) \triangleq \sum_{i=1}^n v_{s,a}(x_i)$ with

$$v_{s,a}(x) \triangleq \mathrm{Var}_{W(\cdot|x)}[i_{s,a}(x, Y)]. \tag{74}$$

Using this definition along with that of $\mathcal{D}_n$ in (19) and the cost function in (64), we have for any $\boldsymbol{x} \in \mathcal{A}_n$ that

$$\left|\mathbb{E}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y})] - nI_{s,a}(Q)\right| \leq \delta \tag{75}$$

$$\left|\mathrm{Var}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y})] - nV_{s,a}(Q)\right| \leq \zeta\sqrt{n\log n}. \tag{76}$$

Since $\log U$ has finite moments, this implies

$$\left|\mathbb{E}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y}) + \log U] - nI_{s,a}(Q)\right| = O(1) \tag{77}$$

$$\left|\mathrm{Var}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y}) + \log U] - nV_{s,a}(Q)\right| = O\big(\sqrt{n\log n}\big). \tag{78}$$

Using (18) and defining $\boldsymbol{X}' \sim Q^n(\boldsymbol{x}')$, we have

$$\mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big] \leq \frac{1}{\mu_n}\mathbb{P}\big[\boldsymbol{X}' \notin \mathcal{A}_n\big]. \tag{79}$$

We claim that there exists a choice of ζ such that the right-hand side of (79) behaves as $O\big(\frac{1}{\sqrt{n}}\big)$, thus yielding

$$\mathbb{P}\big[\boldsymbol{X} \notin \mathcal{A}_n\big] = O\left(\frac{1}{\sqrt{n}}\right). \tag{80}$$

Since Proposition 1 states that $\mu_n = \Omega(n^{-L/2})$, it suffices to show that $\mathbb{P}[\boldsymbol{X}' \notin \mathcal{A}_n]$ can be made to behave as $O(n^{-(L+1)/2})$. This follows from the following moderate deviations result of [31, Thm. 2]: Given an i.i.d. sequence $\{Z_i\}_{i=1}^n$ with $\mathbb{E}[Z_i] = \mu$ and $\mathrm{Var}[Z_i] = \sigma^2 > 0$, we have $\mathbb{P}\big[\big|\frac{1}{n}\sum_{i=1}^n Z_i - \mu\big| > \eta\sigma\sqrt{\frac{\log n}{n}}\big] \asymp \frac{2}{\eta\sqrt{2\pi\log n}}n^{-\eta^2/2}$ provided that $\mathbb{E}[Z_i^{\eta^2+2+\delta}] < \infty$ for some $\delta > 0$. The latter condition is always satisfied in the present setting, since we are considering finite alphabets.

We are now in a position to apply the Berry-Esseen theorem for independent and non-identically distributed random variables [32, Sec. XVI.5]. The relevant first and second moments are bounded in (77)–(78), and the relevant third moment is bounded since we are considering finite alphabets. Choosing

$$\log M = nI_{s,a}(Q) - \log\mu_n - 2\delta - \xi_n \tag{81}$$

for some ξ_n, and also using (71) and (80), we obtain from the Berry-Esseen theorem that

$$\bar{p}_e \leq \mathsf{Q}\left(\frac{\xi_n + O(1)}{\sqrt{nV_{s,a}(Q) + O(\sqrt{n\log n})}}\right) + O\left(\frac{1}{\sqrt{n}}\right). \tag{82}$$

By straightforward rearrangements and a first-order Taylor expansion of the square root function and the Q^{-1} function, we obtain

$$\xi_n \leq \sqrt{nV_{s,a}(Q)}\,\mathsf{Q}^{-1}(\bar{p}_e) + O\big(\sqrt{\log n}\big). \tag{83}$$

The proof for the case $V_{s,a}(Q) > 0$ is concluded by combining (81) and (83), and noting from Proposition 1 that $\log\mu_n = O(\log n)$.

In the case that $V_{s,a}(Q) = 0$, we can still make use of (77), but the variance is handled differently. From the definition in (62), we in fact have $\mathrm{Var}[i_{s,a}(x, Y)] = 0$ for all x such that $Q(x) > 0$. Thus, for all $\boldsymbol{x} \in \mathcal{D}_n$ we have $\mathrm{Var}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y})] = 0$ and hence $\mathrm{Var}[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y}) + \log U] = O(1)$. Choosing M as in (81) and setting $\mathcal{A}_n = \mathcal{D}_n$, we can write (71) as

$$\mathrm{rcu}_s(n, M) \leq \max_{\boldsymbol{x}\in\mathcal{D}_n} \mathbb{P}\big[i_{s,a}^n(\boldsymbol{x},\boldsymbol{Y}) + \log U - nI_{s,a}(Q) \leq -\xi_n\big] \tag{84}$$

$$\leq \frac{O(1)}{(\xi_n - O(1))^2}, \tag{85}$$

where (85) holds due to (77) and Chebyshev's inequality provided that ξ_n is sufficiently large so that the $\xi_n - O(1)$ term is positive. Rearranging, we see that we can achieve any target value $\bar{p}_e = \epsilon$ with $\xi_n = O(1)$. The proof is concluded using (81). ■

Theorem 5 can easily be extended to channels with more general alphabets. However, some care is needed, since the moderate deviations result [31, Thm. 2] used in the proof requires finite moments up to a certain order depending on ζ in (73). In the case that *all* moments of $i_{s,a}(X, Y)$ are finite, the preceding analysis is nearly unchanged, except that the third moment should be bounded in the set $\mathcal{A}_n$ in (73) in the same way as the second moment. An alternative approach is to introduce two further auxiliary costs into the ensemble:

$$a_3(x) = v_{s,a}(x) \tag{86}$$

$$a_4(x) = \mathbb{E}\big[|i_{s,a}(x, Y) - I_{s,a}(Q)|^3\big], \tag{87}$$

where $v_{s,a}$ is defined in (74). Under these choices, the relevant second and third moments for the Berry-Esseen theorem are bounded within $\mathcal{D}_n$ similarly to (77). The only further requirement is that the sixth moment of $i_{s,a}(X, Y)$ is finite under $Q \times W$, in accordance with Proposition 1.

We can easily deal with additive input constraints by handling them similarly to the auxiliary costs (see Section VI for details). With these modifications, our techniques provide, to our knowledge, the most general known second-order achievability proof for memoryless input-constrained channels with infinite or continuous alphabets.[3] In particular, for the additive white Gaussian noise (AWGN) channel with a maximal power constraint and ML decoding, setting $s = 1$ and $a(\cdot) = 0$ yields the achievability part of the dispersion given by Polyanskiy *et al.* [11], thus providing a simple alternative to the proof therein based on the $\kappa\beta$ bound.

B. i.i.d. and Constant-Composition Ensembles

The properties of the cost-constrained ensemble used in the proof of Theorem 5 are also satisfied by the constant-composition ensemble, so we conclude that (65) remains true when M^{cost} is replaced by M^{cc}. However, using standard bounds on μ_n in (71) (e.g. [14, p. 17]), we obtain a third-order $O(\log n)$ term which grows linearly in $|\mathcal{X}|$. In contrast, by Proposition 1 and (81), the cost-constrained ensemble yields a third-order term of the form $-\frac{L}{2}\log n + O(1)$, where L is independent of $|\mathcal{X}|$.

The second-order asymptotic result for i.i.d. coding does not follow directly from Theorem 5, since the proof requires the cost function in (64) to be present. However, using similar arguments along with the identities $\mathbb{E}[i_s^n(\boldsymbol{X},\boldsymbol{Y})] = nI_s(Q)$ and $\mathrm{Var}[i_s^n(\boldsymbol{X},\boldsymbol{Y})] = nU_s(Q)$ (where $\boldsymbol{X} \sim Q^n$), we obtain

$$\log M^{\text{iid}}(Q,n,\epsilon) \ge nI_s(Q) - \sqrt{nU_s(Q)}\,\mathsf{Q}^{-1}(\epsilon) + O(1) \tag{88}$$

for $s \ge 0$, where $I_s(Q)$ and $U_s(Q)$ are defined as in (60)–(61) with $a(\cdot) = 0$. Under some technical conditions, the $O(1)$ term in (88) can be improved to $\frac{1}{2}\log n + O(1)$ using the techniques of [33, Sec. 3.4.5]; see Section V-C for further discussion.

C. Number of Auxiliary Costs Required

For ML decoding ($q(x,y) = W(y|x)$), we immediately see that $a_1(\cdot)$ in (63) is not needed, since the parameters maximizing $I_{s,a}(Q)$ in (60) are $s = 1$ and $a(\cdot) = 0$, thus yielding the mutual information.

We claim that, for any decoding metric, the auxiliary cost $a_2(\cdot)$ in (64) is not needed in the case that Q and $a(\cdot)$ are optimized in (65). This follows from the following observation proved in Appendix C: Given $s > 0$, any pair (Q,a) which maximizes $I_{s,a}(Q)$ must be such that $\mathbb{E}_{W(\cdot|x)}[i_{s,a}(x,Y)]$ has the same value for all x such that $Q(x) > 0$. Stated differently, the conditional variance $V_{s,a}(Q)$ coincides with the unconditional variance $U_{s,a}(Q)$ after the optimization of the parameters, thus generalizing the analogous result for ML decoding [11].

We observe that the number of auxiliary costs in each case coincides with that of the random-coding exponent (see Section III-C): $L = 2$ suffices in general, $L = 1$ suffices if the metric or input distribution is optimized, and $L = 0$ suffices is both are optimized.

V. Saddlepoint Approximations

Random-coding error exponents can be thought of as providing an estimate of the error probability of the form $p_e \approx e^{-nE_r(R)}$. More refined estimates can be obtained having the form $p_e \approx \alpha_n(R)e^{-nE_r(R)}$, where $\alpha_n(R)$ is a subexponential prefactor. Early works on characterizing the subexponential prefactor for a given rate under ML decoding include those of Elias [23] and Dobrushin [34], who studied specific channels exhibiting a high degree of symmetry. More recently, Altuğ and Wagner [22], [35] obtained asymptotic prefactors for arbitrary DMCs.

In this section, we take an alternative approach based on the saddlepoint approximation [13]. Our goal is to provide approximations for rcu and rcu_s (see Theorem 1) which are not only tight in the limit of large n for a fixed rate, but also when the rate varies. In particular, our analysis will cover the regime of a fixed target error probability, which was studied in Section IV, as well as the moderate deviations regime, which was studied in [36] and [37]. We focus on i.i.d. random coding, which is particularly amenable to a precise asymptotic analysis.

A. Preliminary Definitions and Results

Analogously to Section IV, we fix Q and $s > 0$ and define the quantities

$$i_s(x,y) \triangleq \log\frac{q(x,y)^s}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s} \tag{89}$$

$$i_s^n(\boldsymbol{x},\boldsymbol{y}) \triangleq \sum_{i=1}^n i_s(x_i,y_i) \tag{90}$$

$$I_s(Q) \triangleq \mathbb{E}[i_s(X,Y)] \tag{91}$$

$$U_s(Q) \triangleq \mathrm{Var}[i_s(X,Y)], \tag{92}$$

where $(X,Y) \sim Q \times W$. We write rcu_s in (14) (with $P_{\boldsymbol{X}} = Q^n$) as

$$\text{rcu}_s(n,M) = \mathbb{E}\Big[\min\big\{1, (M-1)e^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})}\big\}\Big]. \tag{93}$$

We let

$$E_0^{\text{iid}}(Q,\rho,s) \triangleq -\log\mathbb{E}\big[e^{-\rho i_s(X,Y)}\big] \tag{94}$$

denote the objective in (37) with a fixed value of s in place of the supremum. The optimal value of ρ is given by

$$\hat{\rho}(Q,R,s) \triangleq \operatorname*{arg\,max}_{\rho\in[0,1]} E_0^{\text{iid}}(Q,\rho,s) - \rho R. \tag{95}$$

and the critical rate is defined as

$$R_s^{\text{cr}}(Q) \triangleq \sup\big\{R : \hat{\rho}(Q,R,s) = 1\big\}. \tag{96}$$

Furthermore, we define the following derivatives associated with (95):

$$c_1(Q,R,s) \triangleq R - \frac{\partial E_0^{\text{iid}}(Q,\rho,s)}{\partial\rho}\bigg|_{\rho=\hat{\rho}(Q,R,s)} \tag{97}$$

$$c_2(Q,R,s) \triangleq -\frac{\partial^2 E_0^{\text{iid}}(Q,\rho,s)}{\partial\rho^2}\bigg|_{\rho=\hat{\rho}(Q,R,s)}. \tag{98}$$

[3] Analogous results were stated in [12], but the generality of the proof techniques is unclear. In particular, the quantization arguments on page 4963 therein require that the rate of convergence from $I(X_m;Y)$ to $I(X;Y)$ is sufficiently fast, where X_m is the quantized input variable with a support of cardinality m.

The following properties of the above quantities are analogous to those of Gallager for ML decoding [9, pp. 141–143], and can be proved in a similar fashion:

1) For all $R \geq 0$, we have $c_2(Q,R,s) > 0$ if $U_s(Q) > 0$, and $c_2(Q,R,s) = 0$ if $U_s(Q) = 0$. Furthermore, we have $c_2(Q, I_s(Q), s) = U_s(Q)$.
2) If $U_s(Q) = 0$, then $R_s^{\mathrm{cr}}(Q) = I_s(Q)$.
3) For $R \in \big[0, R_s^{\mathrm{cr}}(Q)\big)$, we have $\hat{\rho}(Q,R,s) = 1$ and $c_1(Q,R,s) < 0$.
4) For $R \in \big[R_s^{\mathrm{cr}}(Q), I_s(Q)\big]$, $\hat{\rho}(Q,R,s)$ is strictly decreasing in R, and $c_1(Q,R,s) = 0$.
5) For $R > I_s(Q)$, we have $\hat{\rho}(Q,R,s) = 0$ and $c_1(Q,R,s) > 0$.

Throughout this section, the arguments to $\hat{\rho}$, c_1, etc. will be omitted, since their values will be clear from the context.

The density function of a $N(\mu,\sigma^2)$ random variable is denoted by

$$\phi(z;\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{99}$$

When studying lattice random variables (see Footnote 1 on Page 2650) with span h, it will be useful to define

$$\phi_h(z;\mu,\sigma^2) \triangleq \frac{h}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, \tag{100}$$

which can be interpreted as an approximation of the integral of $\phi(\,\cdot\,;\mu,\sigma^2)$ from z to $z+h$ when h is small.

B. Approximation for $\mathrm{rcu}_s(n,M)$

In the proof of Theorem 6 below, we derive an approximation $\widehat{\mathrm{rcu}}_s$ of rcu_s taking the form

$$\widehat{\mathrm{rcu}}_s(n,M) \triangleq \alpha_n(Q,R,s) e^{-n(E_0^{\mathrm{iid}}(Q,\hat{\rho},s) - \hat{\rho}R)}, \tag{101}$$

where $R = \frac{1}{n}\log M$, and the prefactor α_n varies depending on whether $i_s(X,Y)$ is a lattice variable. In the non-lattice case, the prefactor is given by

$$\alpha_n^{\mathrm{nl}}(Q,R,s) \triangleq \int_0^{\infty} e^{-\hat{\rho}z}\phi(z; nc_1, nc_2)dz + \int_{-\infty}^{0} e^{(1-\hat{\rho})z}\phi(z; nc_1, nc_2)dz. \tag{102}$$

In the lattice case, it will prove convenient to deal with $R - i_s(X,Y)$ rather than $i_s(X,Y)$. Denoting the offset and span of $R - i_s(X,Y)$ by γ and h respectively, we see that $nR - i_s^n(\boldsymbol{X},\boldsymbol{Y})$ has span h, and its offset can be chosen as

$$\gamma_n \triangleq \min\Big\{ n\gamma + ih \,:\, i \in \mathbb{Z}, n\gamma + ih \geq 0 \Big\}. \tag{103}$$

The prefactor for the lattice case is given by

$$\alpha_n^{\mathrm{l}}(Q,R,s) \triangleq \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)}\phi_h(\gamma_n + ih; nc_1, nc_2) + \sum_{i=-\infty}^{-1} e^{(1-\hat{\rho})(\gamma_n + ih)}\phi_h(\gamma_n + ih; nc_1, nc_2), \tag{104}$$

and the overall prefactor in (101) is defined as

$$\alpha_n \triangleq \begin{cases} \alpha_n^{\mathrm{nl}} & i_s(X,Y) \text{ is non-lattice} \\ \alpha_n^{\mathrm{l}} & R - i_s(X,Y) \text{ has offset } \gamma \text{ and span } h. \end{cases} \tag{105}$$

While (102) and (104) are written in terms of integrals and summations, both prefactors can be computed efficiently to a high degree of accuracy. In the non-lattice case, this is easily done using the identity

$$\int_a^{\infty} e^{bz}\phi(z;\mu,\sigma^2)dz = e^{\mu b + \frac{1}{2}\sigma^2 b^2}\mathsf{Q}\Big(\frac{a - \mu - b\sigma^2}{\sigma}\Big). \tag{106}$$

In the lattice case, we can write each of the summations in (104) in the form

$$\sum_i e^{b_0 + b_1 i + b_2 i^2} = e^{-\frac{b_1^2}{4b_2} + b_0}\sum_i e^{b_2(i + \frac{b_1}{2b_2})^2}, \tag{107}$$

where $b_2 < 0$. We can thus obtain an accurate approximation by keeping only the terms in the sum such that i is sufficiently close to $-\frac{b_1}{2b_2}$. Overall, the computational complexity of the saddlepoint approximation is similar to that of the exponent alone.

Theorem 6. *Fix an input distribution* Q *and parameter* $s > 0$ *such that* $U_s(Q) > 0$*. For any sequence* $\{M_n\}$ *such that* $M_n \to \infty$*, we have*

$$\lim_{n\to\infty} \frac{\widehat{\mathrm{rcu}}_s(n,M_n)}{\mathrm{rcu}_s(n,M_n)} = 1. \tag{108}$$

Proof: See Appendix E. ■

A heuristic derivation of the non-lattice version of $\widehat{\mathrm{rcu}}_s$ was provided in [38]; Theorem 6 provides a formal derivation, along with a treatment of the lattice case. It should be noted that the assumption $U_s(Q) > 0$ is not restrictive, since in the case that $U_s(Q) = 0$ the argument to the expectation in (93) is deterministic, and hence rcu_s can easily be computed exactly.

In the case that the rate R is fixed, simpler asymptotic expressions can be obtained. In Appendix D, we prove the following (here $f_n \asymp g_n$ denotes the relation $\lim_{n\to\infty}\frac{f_n}{g_n} = 1$):

- If $R \in [0, R_s^{\mathrm{cr}}(Q))$ or $R > I_s(Q)$, then

$$\alpha_n(Q,R,s) \asymp 1. \tag{109}$$

- If $R = R_s^{\mathrm{cr}}(Q)$ or $R = I_s(Q)$, then

$$\alpha_n(Q,R,s) \asymp \frac{1}{2}. \tag{110}$$

- If $R \in (R_s^{\mathrm{cr}}(Q), I_s(Q))$, then

$$\alpha_n^{\mathrm{nl}}(Q,R,s) \asymp \frac{1}{\sqrt{2\pi nc_2}\hat{\rho}(1-\hat{\rho})} \tag{111}$$

$$\alpha_n^{\mathrm{l}}(Q,R,s) \asymp \frac{h}{\sqrt{2\pi nc_2}} \times \bigg(e^{-\hat{\rho}\gamma_n}\Big(\frac{1}{1 - e^{-\hat{\rho}h}}\Big) + e^{(1-\hat{\rho})\gamma_n}\Big(\frac{e^{-(1-\hat{\rho})h}}{1 - e^{-(1-\hat{\rho})h}}\Big)\bigg). \tag{112}$$

The asymptotic prefactors in (109)–(112) are related to the problem of *exact asymptotics* in the statistics literature, which seeks to characterize the subexponential prefactor for probabilities that decay at an exponential rate (see [39]). These prefactors are useful in gaining further insight into the behavior of the error probability compared to the error exponent alone. However, there is a notable limitation which is best demonstrated here using (111). The right-hand side of (111) characterizes the prefactor to within a multiplicative $1 + o(1)$ term for a given rate, but it diverges as $\hat{\rho} \to 0$ or $\hat{\rho} \to 1$. Thus, unless n is large, the estimate obtained by omitting the higher-order terms is inaccurate for rates slightly above $R_s^{\mathrm{cr}}(Q)$ or slightly below $I_s(Q)$.

In contrast, the right-hand side of (102) (and similarly (104)) remains bounded for all $\hat{\rho} \in [0, 1]$. Furthermore, as Theorem 6 shows, this expression characterizes the true behavior of rcu_s to within a multiplicative $1 + o(1)$ term not only for fixed rates, but also when the rate varies with the block length. Thus, it remains suitable for characterizing the behavior of rcu_s even when the rate approaches $R_s^{\mathrm{cr}}(Q)$ or $I_s(Q)$. In particular, this implies that $\widehat{\mathrm{rcu}}_s$ gives the correct second-order asymptotics of the rate for a given target error probability (see (88)). More precisely, the proof of Theorem 6 reveals that $\widehat{\mathrm{rcu}}_s = \mathrm{rcu}_s + O\big(\frac{1}{\sqrt{n}}\big)$, which implies (via a Taylor expansion of Q^{-1} in (88)) that the two yield the same asymptotics for a given error probability up to the $O(1)$ term.

C. Approximation for rcu(n, M)

In the proof of Theorem 1, we obtained rcu_s from rcu using Markov's inequality. In this subsection we will see that, under some technical assumptions, a more refined analysis yields a bound which is tighter than rcu_s, but still amenable to the techniques of the previous subsection.

1) Technical Assumptions: Defining the set

$$\mathcal{Y}_1(Q) \triangleq \Big\{y \,:\, q(x, y) \neq q(\bar{x}, y) \text{ for some } \\ x, \bar{x} \text{ such that } Q(x)Q(\bar{x})W(y|x)W(y|\bar{x}) > 0\Big\}, \tag{113}$$

the technical assumptions on (W, q, Q) are as follows:

$$q(x, y) > 0 \iff W(y|x) > 0 \tag{114}$$

$$\mathcal{Y}_1(Q) \neq \emptyset. \tag{115}$$

When $q(x, y) = W(y|x)$, (114) is trivial, and (115) is the *non-singularity* condition of [22]. A notable example where this condition fails is the binary erasure channel (BEC) with $Q = \big(\frac{1}{2}, \frac{1}{2}\big)$. It should be noted that if (114) holds but (115) fails then we in fact have $\mathrm{rcu} = \mathrm{rcu}_s$ for any $s > 0$, and hence $\widehat{\mathrm{rcu}}_s$ also approximates rcu. This can be seen by noting that rcu_s is obtained from rcu using the inequality $\mathbb{1}\{\overline{q} \geq q\} \leq \big(\frac{\overline{q}}{q}\big)^s$, which holds with equality when $\frac{\overline{q}}{q} \in \{0, 1\}$.

2) Definitions: Along with the definitions in Section V-A, we will make use of the reverse conditional distribution

$$\widetilde{P}_s(x|y) \triangleq \frac{Q(x)q(x, y)^s}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x}, y)^s}, \tag{116}$$

the joint tilted distribution

$$P^*_{\hat{\rho},s}(x, y) = \frac{Q(x)W(y|x)e^{-\hat{\rho}i_s(x,y)}}{\sum_{x',y'} Q(x')W(y'|x')e^{-\hat{\rho}i_s(x',y')}}, \tag{117}$$

and its Y-marginal $P^*_{\hat{\rho},s}(y)$, and the conditional variance

$$c_3(Q, R, s) \triangleq \mathbb{E}\Big[\mathrm{Var}\big[i_s(X^*_s, Y^*_s)\big|Y^*_s\big]\Big], \tag{118}$$

where $(X^*_s, Y^*_s) \sim P^*_{\hat{\rho},s}(y)\widetilde{P}_s(x|y)$. Furthermore, we define

$$\mathcal{I}_s \triangleq \Big\{i_s(x, y) \,:\, Q(x)W(y|x) > 0, y \in \mathcal{Y}_1(Q)\Big\} \tag{119}$$

and let

$$\psi_s \triangleq \begin{cases} 1 & \mathcal{I}_s \text{ does not lie on a lattice} \\ \frac{\overline{h}}{1-e^{-\overline{h}}} & \mathcal{I}_s \text{ lies on a lattice with span } \overline{h}. \end{cases} \tag{120}$$

The set $\mathcal{I}_s$ is the support of a random variable which will appear in the analysis of the inner probability in (13). While $\overline{h}$ in (120) can differ from h (the span of $i_s(X, Y)$) in general, the two coincide whenever $\mathcal{Y}_1(Q) = \mathcal{Y}$.

We claim that the assumptions in (114)–(115) imply that $c_3 > 0$ for any R and $s > 0$. To see this, we write

$$\mathrm{Var}_{\widetilde{P}_s(\cdot|y)}[i_s(X, y)] = 0$$

$$\iff \log \frac{\widetilde{P}_s(x|y)}{Q(x)} \text{ is independent of } x \text{ where } \widetilde{P}_s(x|y) > 0 \tag{121}$$

$$\iff q(x, y) \text{ is independent of } x \text{ where } Q(x)q(x, y) > 0 \tag{122}$$

$$\iff y \notin \mathcal{Y}_1(Q), \tag{123}$$

where (121) and (122) follow from the definition of $\widetilde{P}_s$ in (116) and the assumption $s > 0$, and (123) follows from (114) and the definition of $\mathcal{Y}_1(Q)$ in (113). Using (89), (114) and (117), we have

$$P^*_{\hat{\rho},s}(y) > 0 \iff \sum_x Q(x)W(y|x) > 0. \tag{124}$$

Thus, from (115), we have $P^*_{\hat{\rho},s}(y) > 0$ for some $y \in \mathcal{Y}_1(Q)$, which (along with (123)) proves that $c_3 > 0$.

3) Main Result: The main result of this subsection is written in terms of an approximation of the form

$$\widehat{\mathrm{rcu}}^*_s(n, M) \triangleq \beta_n(Q, R, s)e^{-n(E_0^{\mathrm{iid}}(Q,\hat{\rho},s)-\hat{\rho}R)}. \tag{125}$$

Analogously to the previous subsection, we treat the lattice and non-lattice cases separately, writing

$$\beta_n \triangleq \begin{cases} \beta_n^{\mathrm{nl}} & i_s(X, Y) \text{ is non-lattice} \\ \beta_n^{\mathrm{l}} & R - i_s(X, Y) \text{ has offset } \gamma \text{ and span } h, \end{cases} \tag{126}$$

where

$$\beta_n^{\mathrm{nl}}(Q, R, s) \triangleq \int_{\log \frac{\sqrt{2\pi nc_3}}{\psi_s}}^{\infty} e^{-\hat{\rho}z}\phi(z; nc_1, nc_2)dz \\ + \frac{\psi_s}{\sqrt{2\pi nc_3}} \int_{-\infty}^{\log \frac{\sqrt{2\pi nc_3}}{\psi_s}} e^{(1-\hat{\rho})z}\phi(z; nc_1, nc_2)dz \tag{127}$$

$$\beta_n^{\mathrm{l}}(Q, R, s) \triangleq \sum_{i=i^*}^{\infty} e^{-\hat{\rho}(\gamma_n+ih)}\phi_h(\gamma_n + ih; nc_1, nc_2) \\ + \frac{\psi_s}{\sqrt{2\pi nc_3}} \sum_{i=-\infty}^{i^*-1} e^{(1-\hat{\rho})(\gamma_n+ih)}\phi_h(\gamma_n + ih; nc_1, nc_2), \tag{128}$$

and where in (128) we use γ_n in (103) along with

$$i^* \triangleq \min\left\{i \in \mathbb{Z} \,:\, \gamma_n + ih \geq \log\frac{\sqrt{2\pi n c_3}}{\psi_s}\right\}. \tag{129}$$

Theorem 7. *Under the setup of Theorem 6 and the assumptions in* (114)–(115)*, we have for any $s > 0$ that*

$$\mathrm{rcu}(n, M_n) \leq \mathrm{rcu}_s^*(n, M_n)(1 + o(1)), \tag{130}$$

where

$$\mathrm{rcu}_s^*(n, M) \triangleq \mathbb{E}\left[\min\left\{1, \frac{M\psi_s}{\sqrt{2\pi n c_3}} e^{-i_s^n(\boldsymbol{X},\boldsymbol{Y})}\right\}\right]. \tag{131}$$

Furthermore, we have

$$\lim_{n\to\infty} \frac{\widehat{\mathrm{rcu}}_s^*(n, M_n)}{\mathrm{rcu}_s^*(n, M_n)} = 1. \tag{132}$$

Proof: See Appendix F. ■

When the rate does not vary with n, we can apply the same arguments as those given in Appendix D to obtain the following analogues of (109)–(112):

- If $R \in [0, R_s^{\mathrm{cr}}(Q))$, then

$$\beta_n(Q, R, s) \asymp \frac{\psi_s}{\sqrt{2\pi n c_3}}, \tag{133}$$

and similarly for $R = R_s^{\mathrm{cr}}(Q)$ after multiplying the right-hand side by $\frac{1}{2}$.

- If $R \in (R_s^{\mathrm{cr}}(Q), I_s(Q))$, then

$$\beta_n^{\mathrm{nl}}(Q, R, s) \asymp \left(\frac{\psi_s}{\sqrt{2\pi n c_3}}\right)^{\hat{\rho}} \frac{1}{\sqrt{2\pi n c_2 \hat{\rho}(1-\hat{\rho})}} \tag{134}$$

$$\beta_n^{\mathrm{l}}(Q, R, s) \asymp \left(\frac{\psi_s}{\sqrt{2\pi n c_3}}\right)^{\hat{\rho}} \frac{h}{\sqrt{2\pi n c_2}} \times\left(e^{-\hat{\rho}\gamma_n'}\left(\frac{1}{1-e^{-\hat{\rho}h}}\right) + e^{(1-\hat{\rho})\gamma_n'}\left(\frac{e^{-(1-\hat{\rho})h}}{1-e^{-(1-\hat{\rho})h}}\right)\right), \tag{135}$$

where $\gamma_n' \triangleq \gamma_n + i^*h - \log\frac{\sqrt{2\pi n c_3}}{\psi_s} \in [0, h)$ (see (129)).

- For $R \geq I_s(Q)$, the asymptotics of β_n coincide with those of α_n (see (109)–(110)).

When combined with Theorem 7, these expansions provide an alternative proof of the main result of [22], along with a characterization of the multiplicative $\Theta(1)$ terms which were left unspecified in [22]. A simpler version of the analysis in this paper can also be used to obtain the prefactors with unspecified constants; see [40] for details.

Analogously to the previous section, in the regime of fixed error probability we can write (132) more precisely as $\widehat{\mathrm{rcu}}_s^* = \mathrm{rcu}_s^* + O\left(\frac{1}{\sqrt{n}}\right)$, implying that the asymptotic expansions of the rates corresponding to rcu_s^* and $\widehat{\mathrm{rcu}}_s^*$ coincide up to the $O(1)$ term. From the analysis given in [33, Sec. 3.4.5], rcu_s^* yields an expansion of the form (88) with the $O(1)$ term replaced by $\frac{1}{2}\log n + O(1)$. It follows that the same is true of $\widehat{\mathrm{rcu}}_s^*$.

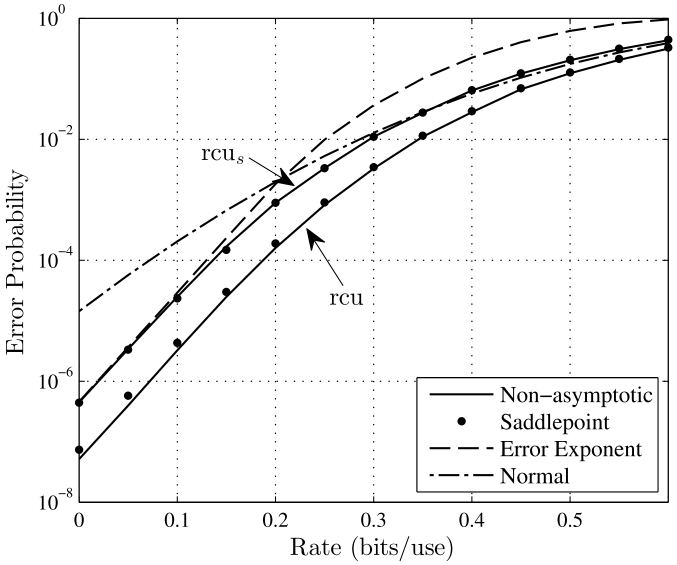

Fig. 2. i.i.d. random-coding bounds for the channel defined in (56) with minimum Hamming distance decoding. The parameters are $n = 60$, $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

D. Numerical Examples

Here we provide numerical examples to demonstrate the utility of the saddlepoint approximations given in this section. Along with $\widehat{\mathrm{rcu}}_s$ and $\widehat{\mathrm{rcu}}_s^*$, we consider (i) the normal approximation, obtained by omitting the remainder term in (88), (ii) the error exponent approximation $p_e \approx e^{-nE_r^{\mathrm{iid}}(Q,R)}$, and (iii) exact asymptotics approximations, obtained by ignoring the implicit $1 + o(1)$ terms in (112) and (135). We use the lattice-type versions of the approximations, since we consider examples in which $i_s(X, Y)$ is a lattice variable. We observed no significant difference in the accuracy of each approximation in similar non-lattice examples.

We consider the example given in Section III-D, using the parameters $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$, and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. For the saddlepoint approximations, we approximate the summations of the form (107) by keeping the 1000 terms[4] whose indices are closest to $-\frac{b_1}{2b_2}$. We choose the free parameter s to be the value which maximizes the error exponent at each rate. For the normal approximation, we choose to s achieve the GMI in (11). Defining $R^{\mathrm{cr}}(Q)$ to be the supremum of all rates such that $\hat{\rho} = 1$ when s is optimized, we have $I_{\mathrm{GMI}}(Q) = 0.643$ and $R^{\mathrm{cr}}(Q) = 0.185$ bits/use.

In Fig. 2, we plot the error probability as a function of the rate with $n = 60$. Despite the fact that the block length is small, we observe that rcu_s and $\widehat{\mathrm{rcu}}_s^*$ are indistinguishable at all rates. Similarly, the gap from rcu to $\widehat{\mathrm{rcu}}_s^*$ is small. Consistent with the fact that Theorem 7 gives an asymptotic upper bound on rcu rather than an asymptotic equality, $\widehat{\mathrm{rcu}}_s^*$ lies slightly above rcu at low rates. The error exponent approximation is close to rcu_s at low rates, but it is pessimistic at high rates. The normal approximation behaves somewhat similarly to rcu_s, but it is less precise than the saddlepoint approximation, particularly at low rates.

[4] The plots remained the same when this value was increased or decreased by an order of magnitude.

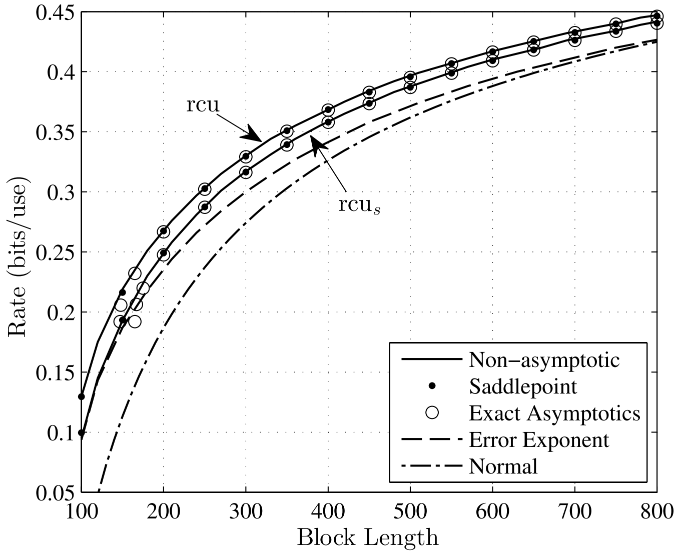

Fig. 3. Rate required to achieve a target error probability ϵ for the channel defined in (56) with ML decoding. The parameters are $\epsilon = 10^{-8}$, $\delta_0 = \delta_1 = \delta_2 = \delta = 0.1$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

To facilitate the computation of rcu and rcu_s at larger block lengths, we consider the symmetric setup of $\delta_0 = \delta_1 = \delta_2 = \delta = 0.1$ and $Q = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Under these parameters, we have $I(X;Y) = 0.633$ and $R^{\mathrm{cr}}(Q) = 0.192$ bits/use. In Fig. 3, we plot the rate required for each random-coding bound and approximation to achieve a given error probability $\epsilon = 10^{-8}$, as a function of n. Once again, $\widehat{\mathrm{rcu}}_s$ is indistinguishable from rcu_s, and similarly for $\widehat{\mathrm{rcu}}_s^*$ and rcu. The error exponent approximation yields similar behavior to rcu_s at small block lengths, but the gap widens at larger block lengths. The exact asymptotics approximations are accurate other than a divergence near the critical rate, which is to be expected from the discussion in Section V-B. In contrast to similar plots with larger target error probabilities (e.g. [11, Fig. 8]), the normal approximation is inaccurate over a wide range of rates.

VI. Discussion and Conclusion

We have introduced a cost-constrained ensemble with multiple auxiliary costs which yields similar performance gains to constant-composition coding, while remaining applicable in the case of infinite or continuous alphabets. We have studied the number of auxiliary costs required to match the performance of the constant-composition ensemble, and shown that the number can be reduced when the input distribution or decoding metric is optimized. Using the saddlepoint approximation, refined asymptotic estimates have been given for the i.i.d. ensemble which unify the regimes of error exponents, second-order rates and moderate deviations, and provide accurate approximations of the random-coding bounds.

Extension to Channels With Input Constraints

Suppose that each codeword $\boldsymbol{x}$ is constrained to satisfy $\frac{1}{n}\sum_{i=1}^{n} c(x_i) \le \Gamma$ for some (system) cost function $c(\cdot)$. The i.i.d. ensemble is no longer suitable, since in all non-trivial cases it has a positive probability of producing codewords which violate the constraint. On the other hand, the results for the constant-composition ensemble remain unchanged provided that Q itself satisfies the cost constraint, i.e. $\sum_x Q(x)c(x) \le \Gamma$.

For the cost-constrained ensemble, the extension is less trivial but still straightforward. The main change required is a modification of the definition of $\mathcal{D}_n$ in (19) to include a constraint on the quantity $\frac{1}{n}\sum_{i=1}^{n} c(x_i)$. Unlike the auxiliary costs in (19), where the sample mean can be above or below the true mean, the system cost of each codeword is constrained to be less than or equal to its mean. That is, the additional constraint is given by

$$\frac{1}{n}\sum_{i=1}^{n} c(x_i) \le \phi_c \triangleq \sum_x Q(x)c(x), \tag{136}$$

or similarly with both upper and lower bounds (e.g. $-\frac{\delta}{n} \le \frac{1}{n}\sum_{i=1}^{n} c(x_i) - \phi_c \le 0$). Using this modified definition of $\mathcal{D}_n$, one can prove the subexponential behavior of μ_n in Proposition 1 provided that Q is such that $\phi_c \le \Gamma$, and the exponents and second-order rates for the cost-constrained ensemble remain valid under any such Q.

Appendix

A. Proof of Theorem 2

The proof is similar to that of Gallager for the constant-composition ensemble [15], so we omit some details. The codeword distribution in (18) can be written as

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbb{1}\left\{\hat{P}_{\boldsymbol{x}} \in \mathcal{G}_n\right\}, \tag{137}$$

where $\hat{P}_{\boldsymbol{x}}$ is the empirical distribution (type) of $\boldsymbol{x}$, and $\mathcal{G}_n$ is the set of types corresponding to sequences $\boldsymbol{x} \in \mathcal{D}_n$ (see (19)). We define the sets

$$\mathcal{S}_n(\mathcal{G}_n) \triangleq \left\{P_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}) \,:\, P_X \in \mathcal{G}_n\right\} \tag{138}$$

$$\begin{aligned}\mathcal{T}_n(P_{XY}, \mathcal{G}_n) \triangleq \big\{\widetilde{P}_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y}) \,:\, \widetilde{P}_X \in \mathcal{G}_n,\\ \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\big\}.\end{aligned} \tag{139}$$

We have from Theorem 1 that $\bar{p}_e \doteq \mathrm{rcu}$. Expanding rcu in terms of types, we obtain

$$\begin{aligned}\bar{p}_e \doteq \sum_{P_{XY}\in\mathcal{S}_n(\mathcal{G}_n)} &\mathbb{P}\big[(\boldsymbol{X},\boldsymbol{Y}) \in T^n(P_{XY})\big]\\ &\times \min\left\{1, (M-1)\sum_{\widetilde{P}_{XY}\in\mathcal{T}_n(P_{XY},\mathcal{G}_n)} \mathbb{P}\big[(\overline{\boldsymbol{X}},\boldsymbol{y}) \in T^n(\widetilde{P}_{XY})\big]\right\},\end{aligned} \tag{140}$$

where $\boldsymbol{y}$ denotes an arbitrary sequence with type P_Y.

From Proposition 1, the normalizing constant in (137) satisfies $\mu_n \doteq 1$, and thus we can safely proceed from (140) as if the codeword distribution were $P_{\boldsymbol{X}} = Q^n$. Using the property of types in [15, Eq. (18)], it follows that the two probabilities in (140) behave as $e^{-nD(P_{XY}\|Q\times W)}$ and

$e^{-nD(\widetilde{P}_{XY}\|Q\times P_Y)}$ respectively. Combining this with the fact that the number of joint types is polynomial in n, we obtain $\bar{p}_e \doteq e^{-nE_{r,n}(Q,R,\mathcal{G}_n)}$, where

$$E_{r,n}(Q,R,\mathcal{G}_n) \triangleq \min_{P_{XY}\in\mathcal{S}_n(\mathcal{G}_n)} \min_{\widetilde{P}_{XY}\in\mathcal{T}_n(P_{XY},\mathcal{G}_n)} D(P_{XY}\|Q\times W) + \big[D(\widetilde{P}_{XY}\|Q\times P_Y) - R\big]^+. \quad (141)$$

Using a simple continuity argument (e.g. see [28, Eq. (30)]), we can replace the minimizations over types by minimizations over joint distributions, and the constraints of the form $|\mathbb{E}_P[a_l(X)] - \phi_l| \le \frac{\delta}{n}$ can be replaced by $\mathbb{E}_P[a_l(X)] = \phi_l$. This concludes the proof.

B. Proof of Theorem 3

Throughout the proof, we make use of Fan's minimax theorem [41], which states that $\min_a \sup_b f(a,b) = \sup_b \min_a f(a,b)$ provided that the minimum is over a compact set, $f(\cdot,b)$ is convex in a for all b, and $f(a,\cdot)$ is concave in b for all a. We make use of Lagrange duality [17] in a similar fashion to [5, Appendix A]; some details are omitted to avoid repetition with [5].

Using the identity $[\alpha]^+ = \max_{\rho\in[0,1]}\rho\alpha$ and Fan's minimax theorem, the expression in (27) can be written as

$$E_r^{\mathrm{cost}}(Q,R,\{a_l\}) = \max_{\rho\in[0,1]} \hat{E}_0^{\mathrm{cost}}(Q,\rho,\{a_l\}) - \rho R, \quad (142)$$

where

$$\hat{E}_0^{\mathrm{cost}}(Q,\rho,\{a_l\}) \triangleq \min_{P_{XY}\in\mathcal{S}(\{a_l\})} \min_{\widetilde{P}_{XY}\in\mathcal{T}(P_{XY},\{a_l\})} D(P_{XY}\|Q\times W) + \rho D(\widetilde{P}_{XY}\|Q\times P_Y). \quad (143)$$

It remains to show that $\hat{E}_0^{\mathrm{cost}} = E_0^{\mathrm{cost}}$. We will show this by considering the minimizations in (143) one at a time. We can follow the steps of [5, Appendix A] to conclude that

$$\min_{\widetilde{P}_{XY}\in\mathcal{T}(P_{XY},\{a_l\})} D(\widetilde{P}_{XY}\|Q\times P_Y) = \sup_{s\ge0,\{\overline{r}_l\}} \sum_{x,y} P_{XY}(x,y)\log\frac{q(x,y)^s}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{\sum_l \overline{r}_l(a_l(\bar{x})-\phi_l)}}, \quad (144)$$

where s and $\{\overline{r}_l\}$ are Lagrange multipliers. It follows that the right-hand side of (143) equals

$$\min_{P_{XY}\in\mathcal{S}(\{a_l\})} \sup_{s\ge0,\{\overline{r}_l\}} \sum_{x,y} P_{XY}(x,y)\bigg(\log\frac{P_{XY}(x,y)}{Q(x)W(y|x)} + \rho\log\frac{q(x,y)^s}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{\sum_l \overline{r}_l(a_l(\bar{x})-\phi_l)}}\bigg). \quad (145)$$

Since the objective is convex in P_{XY} and jointly concave in $(s,\{\overline{r}_l\})$, we can apply Fan's minimax theorem. Hence, we consider the minimization of the objective in (145) over $P_{XY}\in\mathcal{S}(\{a_l\})$ with s and $\{\overline{r}_l\}$ fixed. Applying the techniques of [5, Appendix A] a second time, we conclude that this minimization has a dual form given by

$$\sup_{\{r_l\}} -\log\sum_{x,y} Q(x)W(y|x)\bigg(\frac{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{\sum_l \overline{r}_l(a_l(\bar{x})-\phi_l)}}{q(x,y)^s e^{\sum_l r_l(a_l(x)-\phi_l)}}\bigg)^{\rho}, \quad (146)$$

where $\{r_l\}$ are Lagrange multipliers. The proof is concluded by taking the supremum over s and $\{\overline{r}_l\}$.

C. Necessary Conditions for the Optimal Parameters

1) Optimization of $E_0^{\mathrm{cc}}(Q,\rho)$: We write the objective in (39) as

$$E_0^{\mathrm{cc}}(Q,\rho,s,a) \triangleq \rho\sum_x Q(x)a(x) - \sum_x Q(x)\log f(x), \quad (147)$$

where

$$f(x) \triangleq \sum_y W(y|x)q(x,y)^{-\rho s}\bigg(\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}\bigg)^{\rho}. \quad (148)$$

We have the partial derivatives

$$\frac{\partial f(x)}{\partial Q(x')} = \rho g(x,x') \quad (149)$$

$$\frac{\partial f(x)}{\partial a(x')} = \rho Q(x')g(x,x'), \quad (150)$$

where

$$g(x,x') \triangleq \sum_y W(y|x)q(x,y)^{-\rho s} \times \rho\bigg(\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}\bigg)^{\rho-1} q(x',y)^s e^{a(x')} \quad (151)$$

We proceed by analyzing the necessary Karush-Kuhn-Tucker (KKT) conditions [17] for (Q,a) to maximize $E_0^{\mathrm{cc}}(Q,\rho,s,a)$. The KKT condition corresponding to the partial derivative with respect to $a(x')$ is

$$\rho Q(x') - \sum_x Q(x)\frac{\rho Q(x')g(x,x')}{f(x)} = 0, \quad (152)$$

or equivalently

$$\sum_x Q(x)\frac{g(x,x')}{f(x)} = 1. \quad (153)$$

Similarly, the KKT condition corresponding to $Q(x')$ gives

$$\rho a(x') - \log f(x') - \rho\sum_x Q(x)\frac{g(x,x')}{f(x)} - \lambda = 0 \quad (154)$$

for all x' such that $Q(x') > 0$, where λ is the Lagrange multiplier associated with the constraint $\sum_x Q(x) = 1$. Substituting (153) into (154) gives

$$-\log\big(f(x')e^{-\rho a(x')}\big) = \lambda + \rho. \quad (155)$$

Using the definition of $f(\cdot)$ in (148), we see that (155) implies that the logarithm in (39) is independent of x.

2) Optimization of $I_{s,a}(Q)$: We write $I_{s,a}(Q)$ in (60) as

$$I_{s,a}(Q) = s\sum_{x,y} Q(x)W(y|x)\log q(x,y) + \sum_x Q(x)a(x) - \sum_{x,y} Q(x)W(y|x)\log\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})} \quad (156)$$

and analyze the KKT conditions associated with the maximization over (Q, a). We omit some details, since the steps are similar to those above. The KKT condition for $a(x')$ is

$$\sum_{x,y} Q(x)W(y|x)\frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} = 1, \qquad (157)$$

and the KKT condition for $Q(x')$ gives

$$\begin{aligned} s\sum_{y} W(y|x')\log q(x',y) + a(x') \\ - \sum_{y} W(y|x')\log\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})} \\ - \sum_{x,y} Q(x)W(y|x)\frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} - \lambda = 0 \qquad (158) \end{aligned}$$

for all x' such that $Q(x') > 0$, where λ is a Lagrange multiplier. Substituting (157) into (158) and performing some simple rearrangements, we obtain

$$\sum_{y} W(y|x')\log\frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} = \lambda + 1. \qquad (159)$$

D. Asymptotic Behavior of the Saddlepoint Approximation

Here we prove the asymptotic relations given in (109)–(112). We will make frequent use of the properties of $\hat{\rho}$, c_1 and c_2 given in Section V-A.

We first prove (109)–(110) in the non-lattice case. Suppose that $R < R_s^{\mathrm{cr}}(Q)$, and hence $\hat{\rho} = 1$, $c_1 < 0$ and $c_2 > 0$. Using (106) and the identity $\mathsf{Q}(z) \le \frac{1}{2}\exp\left(\frac{-z^2}{2}\right)$ for $z > 0$, it is easily verified that the first term in (102) decays to zero exponentially fast. The second term is given by $\mathsf{Q}\big(-c_1\sqrt{\frac{n}{c_2}}\big)$, which tends to one since $\lim_{z\to\infty}\mathsf{Q}(z) = 1$. We thus obtain (109). When $R = R_s^{\mathrm{cr}}(Q)$, the argument is similar except that $c_1 = 0$, yielding the following: (i) From (106), the first term in (102) equals $\mathsf{Q}(\sqrt{nc_2})e^{nc_2/2} \asymp \frac{1}{\sqrt{2\pi nc_2}}$, rather than decaying exponentially fast, (ii) The second term in (102) equals $\mathsf{Q}(0) = \frac{1}{2}$, rather than one. For $R > I_s(Q)$ (respectively, $R = I_s(Q)$) the argument is similar with the roles of the two terms in (102) reversed, and with $\hat{\rho} = 0$ and $c_1 > 0$ (respectively, $c_1 = 0$).

In the lattice case, the arguments in proving (109)–(110) are similar to the non-lattice case, so we focus on (109) with $R < R_s^{\mathrm{cr}}(Q)$. Similarly to the non-lattice case, it is easily shown that the first summation in (104) decays to zero exponentially fast, so we focus on the second. Since $\hat{\rho} = 1$, the second summation is given by

$$\begin{aligned} &\sum_{i=-\infty}^{-1} \phi_h(\gamma_n + ih; nc_1, nc_2) \\ &\qquad = (1 + o(1))\sum_{i=-\infty}^{\infty} \phi_h(\gamma_n + ih; nc_1, nc_2) \qquad (160) \\ &\qquad = 1 + o(1), \qquad (161) \end{aligned}$$

where (160) follows since the added terms from $i = 0$ to ∞ contribute an exponentially small amount to the sum since $c_1 < 0$, and (161) is easily understood by interpreting the right-hand side of (160) as approximating the integral over the real line of a Gaussian density function via discrete sampling. Since the sampling is done using intervals of a fixed size h but the variance nc_2 increases, the approximation improves with n and approaches one.

Finally, we consider the case that $R \in (R_s^{\mathrm{cr}}(Q), I_s(Q))$, and hence $\hat{\rho} \in (0,1)$, $c_1 = 0$ and $c_2 > 0$. In the non-lattice case, we can substitute $c_1 = 0$ and (106) into (104) to obtain

$$\alpha_n = e^{\frac{1}{2}nc_2\hat{\rho}^2}\mathsf{Q}\big(\hat{\rho}\sqrt{nc_2}\big) + e^{\frac{1}{2}nc_2(1-\hat{\rho})^2}\mathsf{Q}\big((1-\hat{\rho})\sqrt{nc_2}\big). \quad (162)$$

Using the fact that $\mathsf{Q}(z)e^{z^2/2} \asymp \frac{1}{z\sqrt{2\pi}}$ as $z \to \infty$, along with the identity $\frac{1}{\rho} + \frac{1}{1-\rho} = \frac{1}{\rho(1-\rho)}$, we obtain (111).

We now turn to the lattice case. Setting $c_1 = 0$ in (104) yields

$$\begin{aligned} \alpha_n = \frac{h}{\sqrt{2\pi nc_2}}\Bigg(&\sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih) - \frac{(\gamma_n+ih)^2}{2nc_2}} \\ &+ \sum_{i=-\infty}^{-1} e^{(1-\hat{\rho})(\gamma_n+ih) - \frac{(\gamma_n+ih)^2}{2nc_2}}\Bigg). \quad (163) \end{aligned}$$

The two summations are handled in a nearly identical fashion, so we focus on the first. Using the identity $1 - x \le e^{-x} \le 1$, we can write

$$\begin{aligned} &\left|\sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n+ih) - \frac{(\gamma_n+ih)^2}{2nc_2}} - \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n+ih)}\right| \\ &\qquad \le \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n+ih)}\frac{(\gamma_n+ih)^2}{2nc_2} \qquad (164) \\ &\qquad = O\Big(\frac{1}{n}\Big), \qquad (165) \end{aligned}$$

where (165) follows since the summation $\sum_{i=0}^{\infty} e^{-\zeta i}p(i)$ is convergent for any polynomial $p(i)$ and $\zeta > 0$. Furthermore, we have from the geometric series that

$$\sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n+ih)} = e^{-\hat{\rho}\gamma_n}\left(\frac{1}{1 - e^{-\hat{\rho}h}}\right), \qquad (166)$$

We have thus weakened the first summation in (163) to the first term in the sum in (112) (up to an $O\left(\frac{1}{n}\right)$ remainder term). The second term is obtained in a similar fashion.

E. Proof of Theorem 6

Since $M_n \to \infty$ by assumption, we can safely replace M_n by $M_n + 1$ without affecting the theorem statement. We begin by considering fixed values of n, M and $R = \frac{1}{n}\log M$.

Using (93) and the identity in (72), we can can write

$$\mathrm{rcu}_s(n, M+1) = \mathbb{P}\left[nR - \sum_{i=1}^{n} i_s(X_i, Y_i) \ge \log U\right]. \quad (167)$$

This expression resembles the tail probability of an i.i.d. sum of random variables, for which asymptotic estimates were given by Bahadur and Rao [39] (see also [9, Appendix 5A]). There are two notable differences in our setting which mean that the results of [9], [39] cannot be applied directly. First, the right-hand side of the event in (167) is random rather than deterministic. Second, since we are allowing for rates below $R_s^{\mathrm{cr}}(Q)$ or above $I_s(Q)$, we cannot assume that the derivative of the moment generating function of $R - i_s(X,Y)$ at zero (which we will shortly see equals c_1 in (97)) is equal to zero.

1) Alternative Expressions for rcu_s*:* Let $F(t)$ denote the cumulative distribution function (CDF) of $R - i_s(X,Y)$, and let $Z_1, \dots, Z_n$ be i.i.d. according to the tilted CDF

$$F_Z(z) = e^{E_0^{\mathrm{iid}}(Q,\hat{\rho},s)-\hat{\rho}R} \int_{-\infty}^{z} e^{\hat{\rho}t} dF(t). \tag{168}$$

It is easily seen that this is indeed a CDF by writing

$$\int_{-\infty}^{\infty} e^{\hat{\rho}t} dF(t) = \mathbb{E}\big[e^{\hat{\rho}(R-i_s(X,Y))}\big] = e^{-(E_0^{\mathrm{iid}}(Q,\hat{\rho},s)-\hat{\rho}R)}, \tag{169}$$

where we have used (94). The moment generating function (MGF) of Z is given by

$$\begin{aligned} M_Z(\tau) &\triangleq \mathbb{E}\big[e^{\tau Z}\big] & (170)\\ &= e^{E_0^{\mathrm{iid}}(Q,\hat{\rho},s)-\hat{\rho}R}\mathbb{E}\big[e^{(\hat{\rho}+\tau)(R-i_s(X,Y))}\big] & (171)\\ &= e^{E_0^{\mathrm{iid}}(Q,\hat{\rho},s)} e^{-(E_0^{\mathrm{iid}}(Q,\hat{\rho}+\tau,s)-\tau R)}, & (172) \end{aligned}$$

where (171) follows from (168), and (172) follows from (94). We can now compute the mean and variance of Z in terms of the derivatives of the MGF, namely

$$\begin{aligned} \mathbb{E}[Z] &= \frac{dM_Z}{d\tau}\Big|_{\tau=0} = c_1 & (173)\\ \mathrm{Var}[Z] &= \frac{d^2 M_Z}{d\tau^2}\Big|_{\tau=0} - \mathbb{E}[Z]^2 = c_2, & (174) \end{aligned}$$

where c_1 and c_2 are defined in (97)–(98). Recall that $U_s(Q) > 0$ by assumption, which implies that $c_2 > 0$ (see Section V-A).

In the remainder of the proof, we omit the arguments $(Q, \hat{\rho}, s)$ to E_0^{iid}. Following [39, Lemma 2], we can use (168) to write (167) as follows:

$$\begin{aligned} &\mathrm{rcu}_s(n, M+1)\\ &\quad= \int\!\cdots\!\int_{\sum_i t_i \ge \log u} dF(t_1)\cdots dF(t_n)dF_U(u) & (175)\\ &\quad= e^{-n(E_0^{\mathrm{iid}}-\hat{\rho}R)}\\ &\qquad\times \int\!\cdots\!\int_{\sum_i z_i \ge \log u} e^{-\hat{\rho}\sum_i z_i} dF_Z(z_1)\cdots dF_Z(z_n)dF_U(u), & (176)\\ &\quad\triangleq I_n e^{-n(E_0^{\mathrm{iid}}-\hat{\rho}R)}, & (177) \end{aligned}$$

where $F_U(u)$ is the CDF of U. We write the prefactor I_n as

$$I_n = \int_0^1 \int_{\log u}^{\infty} e^{-\hat{\rho}z} dF_n(z) dF_U(u), \tag{178}$$

where F_n is the CDF of $\sum_{i=1}^n Z_i$. Since the integrand in (178) is non-negative, we can safely interchange the order of integration, yielding

$$\begin{aligned} I_n &= \int_{-\infty}^{\infty} \int_0^{\min\{1,e^z\}} e^{-\hat{\rho}z} dF_U(u) dF_n(z) & (179)\\ &= \int_0^{\infty} e^{-\hat{\rho}z} dF_n(z) + \int_{-\infty}^{0} e^{(1-\hat{\rho})z} dF_n(z), & (180) \end{aligned}$$

where (180) follows by splitting the integral according to which value achieves the $\min\{\cdot,\cdot\}$ in (179). Letting $\hat{F}_n$ denote the CDF of $\frac{\sum_{i=1}^n Z_i - nc_1}{\sqrt{nc_2}}$, we can write (180) as

$$\begin{aligned} I_n = &\int_{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}}^{\infty} e^{-\hat{\rho}(z\sqrt{nc_2}+nc_1)} d\hat{F}_n(z)\\ &+ \int_{-\infty}^{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}} e^{(1-\hat{\rho})(z\sqrt{nc_2}+nc_1)} d\hat{F}_n(z). \end{aligned} \tag{181}$$

2) Non-Lattice Case: Let $\Phi(z)$ denote the CDF of a zero-mean unit-variance Gaussian random variable. Using the fact that $\mathbb{E}[Z] = c_1$ and $\mathrm{Var}[Z] = c_2 > 0$ (see (173)–(174)), we have from the refined central limit theorem in [32, Sec. XVI.4, Thm. 1] that

$$\hat{F}_n(z) = \Phi(z) + G_n(z) + \tilde{F}_n(z), \tag{182}$$

where $\tilde{F}_n(z) = o(n^{-\frac{1}{2}})$ uniformly in z, and

$$G_n(z) \triangleq \frac{K}{\sqrt{n}}(1-z^2)e^{-\frac{1}{2}z^2} \tag{183}$$

for some constant K depending only on the variance and third absolute moment of Z, the latter of which is finite since we are considering finite alphabets. Substituting (182) into (181), we obtain

$$I_n = I_{1,n} + I_{2,n} + I_{3,n}, \tag{184}$$

where the three terms denote the right-hand side of (181) with Φ, G_n and $\tilde{F}_n$ respectively in place of $\hat{F}_n$. Reversing the step from (180) to (181), we see that $I_{1,n}$ is precisely α_n in (102). Furthermore, using $\frac{dG_n}{dz} = \frac{K}{\sqrt{n}}(z^3 - 3z)e^{-\frac{1}{2}z^2}$, we obtain

$$\begin{aligned} I_{2,n} = \frac{K}{\sqrt{n}}\Bigg(&\int_{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}}^{\infty} e^{-\hat{\rho}(z\sqrt{nc_2}+nc_1)}(z^3-3z)e^{-\frac{1}{2}z^2}dz\\ &+ \int_{-\infty}^{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}} e^{(1-\hat{\rho})(z\sqrt{nc_2}+nc_1)}(z^3-3z)e^{-\frac{1}{2}z^2}dz\Bigg). \end{aligned} \tag{185}$$

In accordance with the theorem statement, we must show that $I_{2,n} = o(\alpha_n)$ and $I_{3,n} = o(\alpha_n)$ even in the case that R and $\hat{\rho}$ vary with n. Let $R_n = \frac{1}{n}\log M_n$ and $\hat{\rho}_n = \hat{\rho}(Q, R_n, s)$, and let $c_{1,n}$ and $c_{2,n}$ be the corresponding values of c_1 and c_2. We assume with no real loss of generality that

$$\lim_{n\to\infty} R_n = R^* \tag{186}$$

for some $R^* \ge 0$ possibly equal to ∞. Once the theorem is proved for all such R^*, the same will follow for an arbitrary sequence $\{R_n\}$.

Table I summarizes the growth rates α_n, $I_{2,n}$ and $I_{3,n}$ for various ranges of R^*, and indicates whether the first or second integral (see (181) and (185)) dominates the behavior of each. We see that $I_{2,n} = o(\alpha_n)$ and $I_{3,n} = o(\alpha_n)$ for all values of R^*, as desired.

The derivations of the growth rates in Table I when $R^* \notin \{R_s^{\mathrm{cr}}(Q), I_s(Q)\}$ are done in a similar fashion to Appendix D. To avoid repetition, we provide details only for $R^* = R_s^{\mathrm{cr}}(Q)$; this is a less straightforward case whose analysis differs

TABLE I

GROWTH RATES OF α_n, $I_{2,n}$ AND $I_{3,n}$ WHEN THE RATE CONVERGES TO R^*

	$\hat{\rho}$	c_1	Dominant Term(s)	α_n	$I_{2,n}$	$I_{3,n}$
$R^* \in [0, R_s^{\mathrm{cr}}(Q))$	1	< 0	2	$\Theta(1)$	$\Theta\big(\frac{1}{\sqrt{n}}\big)$	$o\big(\frac{1}{\sqrt{n}}\big)$
$R^* = R_s^{\mathrm{cr}}(Q)$	$\to 1$	$\to 0$	2	$\omega\big(\frac{1}{\sqrt{n}}\big)$	$O\big(\frac{1}{\sqrt{n}}\big)$	$o\big(\frac{1}{\sqrt{n}}\big)$
$R^* \in (R_s^{\mathrm{cr}}(Q), I_s(Q))$	$\in (0,1)$	0	1,2	$\Theta\big(\frac{1}{\sqrt{n}}\big)$	$\Theta\big(\frac{1}{n}\big)$	$o\big(\frac{1}{\sqrt{n}}\big)$
$R^* = I_s(Q)$	$\to 0$	$\to 0$	1	$\omega\big(\frac{1}{\sqrt{n}}\big)$	$O\big(\frac{1}{\sqrt{n}}\big)$	$o\big(\frac{1}{\sqrt{n}}\big)$
$R^* > I_s(Q)$	0	> 0	1	$\Theta(1)$	$\Theta\big(\frac{1}{\sqrt{n}}\big)$	$o\big(\frac{1}{\sqrt{n}}\big)$

slightly from Appendix D. From Section V-A, we have $\hat{\rho}_n \to 1$ and $c_{1,n} \to 0$ from below, with $c_{1,n} < 0$ only if $\hat{\rho}_n = 1$.

For any $\hat{\rho} \in [0,1]$, the terms $e^{-\hat{\rho}(\cdot)}$ and $e^{(1-\hat{\rho})(\cdot)}$ in (185) are both upper bounded by one across their respective ranges of integration. Since the moments of a Gaussian random variable are finite, it follows that both integrals are $O(1)$, and thus $I_{2,n} = O\big(\frac{1}{\sqrt{n}}\big)$. The term $I_{3,n}$ is handled similarly, so it only remains to show that $\alpha_n = \omega\big(\frac{1}{\sqrt{n}}\big)$. In the case that $\hat{\rho} = 1$, the second integral in (102) is at least $\frac{1}{2}$, since $c_1 \le 0$. It only remains to handle the case that $c_1 = 0$ and $\hat{\rho}_n \to 1$ with $\hat{\rho}_n < 1$. For any $\delta > 0$, we have $\hat{\rho}_n \ge 1 - \delta$ for sufficiently large n. Lower bounding α_n by replacing $1 - \hat{\rho}$ by δ in the second term of (102), we have similarly to (162) that

$$\alpha_n \ge e^{\frac{1}{2}nc_2\delta^2}\mathsf{Q}\big(\delta\sqrt{nc_2}\big) \asymp \frac{1}{\sqrt{2\pi nc_2}\delta}. \tag{187}$$

Since δ is arbitrary, we obtain $\alpha_n = \omega\big(\frac{1}{\sqrt{n}}\big)$, as desired.

3) Lattice Case: The arguments following (184) are essentially identical in the lattice case, so we focus our attention on obtaining the analogous expression to (184). Letting $P_n(z)$ denote the probability mass function (PMF) of $\sum_{i=1}^n Z_i$, we can write (180) as

$$I_n = \sum_{z \ge 0} P_n(z) e^{-\hat{\rho}z} + \sum_{z<0} P_n(z) e^{(1-\hat{\rho})z}. \tag{188}$$

Using the fact that $\mathbb{E}[Z] = c_1$ and $\mathrm{Var}[Z] = c_2 > 0$, we have from the local limit theorem in [9, Eq. (5A.12)] that

$$P_n(z) = \phi_h(z; nc_1, nc_2) + \tilde{P}_n(z), \tag{189}$$

where ϕ_h is defined in (100), and $\tilde{P}_n(z) = o\big(\frac{1}{\sqrt{n}}\big)$ uniformly in z. Thus, analogously to (184), we can write

$$I_n = I_{1,n} + I_{2,n}, \tag{190}$$

where the two terms denote the right-hand side of (188) with ϕ_h and $\tilde{P}_n(z)$ respectively in place of $P_n(z)$. Using the definition of γ_n in (103) and the fact that $\sum_i Z_i$ has the same support as $nR - i_s^n(\boldsymbol{X},\boldsymbol{Y})$ (cf. (168)), we see that the first summation in (188) is over the set $\{\gamma_n + ih : i \in \mathbb{Z}, i \ge 0\}$, and the second summation is over the set $\{\gamma_n + ih : i \in \mathbb{Z}, i < 0\}$. It follows that $I_{1,n} = \alpha_n$, and similar arguments to the non-lattice case show that $I_{2,n} = o(\alpha_n)$.

F. Proof of Theorem 7

Throughout this section, we make use of the same notation as Appendix E. We first discuss the proof of (132). Using the definition of rcu_s^*, we can follow identical arguments to those following (167) to conclude that

$$\mathrm{rcu}_s^*(n, M) = I_n e^{-n(E_0^{\mathrm{iid}}(Q,\hat{\rho},s) - \hat{\rho}R)}, \tag{191}$$

where analogously to (178) and (180), we have

$$I_n = \int_0^1 \int_{\log \frac{u\sqrt{2\pi nc_3}}{\psi_s}}^{\infty} e^{-\hat{\rho}z} dF_n(z) dF_U(u) \tag{192}$$

$$= \int_{\log \frac{\sqrt{2\pi nc_3}}{\psi_s}}^{\infty} e^{-\hat{\rho}z} dF_n(z) + \frac{\psi_s}{\sqrt{2\pi nc_3}} \int_{-\infty}^{\log \frac{\sqrt{2\pi nc_3}}{\psi_s}} e^{(1-\hat{\rho})z} dF_n(z). \tag{193}$$

The remaining arguments in proving (132) follow those given in Appendix E, and are omitted.

To prove (130), we make use of two technical lemmas, whose proofs are postponed until the end of the section. The following lemma can be considered a refinement of [11, Lemma 47].

Lemma 1. *Fix $K > 0$, and for each n, let $(n_1, \dots, n_K)$ be integers such that $\sum_k n_k = n$. Fix the PMFs $Q_1, \dots, Q_K$ on a finite subset of $\mathbb{R}$, and let $\sigma_1^2, \dots, \sigma_K^2$ be the corresponding variances. Let $Z_1, \dots, Z_n$ be independent random variables, n_k of which are distributed according to Q_k for each k. Suppose that* $\min_k \sigma_k > 0$ *and* $\min_k n_k = \Theta(n)$. *Defining*

$$\mathcal{I}_0 \triangleq \bigcup_{k : \sigma_k > 0} \big\{z : Q_k(z) > 0\big\} \tag{194}$$

$$\psi_0 \triangleq \begin{cases} 1 & \mathcal{I}_0 \text{ does not lie on a lattice} \\ \frac{h_0}{1-e^{-h_0}} & \mathcal{I}_0 \text{ lies on a lattice with span } h_0, \end{cases} \tag{195}$$

the summation $S_n \triangleq \sum_i Z_i$ satisfies the following uniformly in t:

$$\mathbb{E}\Big[e^{-S_n}\mathbb{1}\{S_n \ge t\}\Big] \le e^{-t}\bigg(\frac{\psi_0}{\sqrt{2\pi V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big)\bigg), \tag{196}$$

where $V_n \triangleq \mathrm{Var}[S_n]$.

Roughly speaking, the following lemma ensures the existence of a high probability set in which Lemma 1 can be applied to the inner probability in (13). We make use of the definitions in (116)–(118), and we define the random variables

$$(\boldsymbol{X}, \boldsymbol{Y}, \overline{\boldsymbol{X}}, \boldsymbol{X}_s) \sim Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) Q^n(\bar{\boldsymbol{x}}) \widetilde{P}_s^n(\boldsymbol{x}_s|\boldsymbol{y}), \tag{197}$$

where $\widetilde{P}_s^n(\boldsymbol{x}|\boldsymbol{y}) \triangleq \prod_{i=1}^n \widetilde{P}_s(x_i|y_i)$. Furthermore, we write the empirical distribution of $\boldsymbol{y}$ as $\hat{P}_{\boldsymbol{y}}$, and we let $P_{\boldsymbol{Y}}$ denote the PMF of $\boldsymbol{Y}$.

Lemma 2. *Let the parameters $s > 0$ and $\hat{\rho} \in [0, 1]$ be given. If the triplet (W, q, Q) satisfies (114)–(115), then the set*

$$\mathcal{F}^n_{\hat{\rho},s}(\delta) \triangleq \Big\{ \boldsymbol{y} \,:\, P_{\boldsymbol{Y}}(\boldsymbol{y}) > 0,\ \max_{y} \big|\hat{P}_{\boldsymbol{y}}(y) - P^*_{\hat{\rho},s}(y)\big| \le \delta \Big\} \tag{198}$$

satisfies the following properties:

1) *For any $\boldsymbol{y} \in \mathcal{F}^n_{\hat{\rho},s}(\delta)$, we have*

$$\mathrm{Var}\big[i^n_s(\boldsymbol{X}_s, \boldsymbol{Y}) \,|\, \boldsymbol{Y} = \boldsymbol{y}\big] \ge n(c_3 - r(\delta)), \tag{199}$$

where $r(\delta) \to 0$ as $\delta \to 0$.

2) *For any $\delta > 0$, we have*

$$\liminf_{n\to\infty} -\frac{1}{n} \log \frac{\sum_{\boldsymbol{x},\boldsymbol{y} \notin \mathcal{F}^n_{\hat{\rho},s}(\delta)} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{-\hat{\rho} i^n_s(\boldsymbol{x},\boldsymbol{y})}}{\sum_{\boldsymbol{x},\boldsymbol{y}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{-\hat{\rho} i^n_s(\boldsymbol{x},\boldsymbol{y})}} > 0. \tag{200}$$

It should be noted that since the two statements of Lemma 2 hold true for any $\hat{\rho} \in [0, 1]$, they also hold true when $\hat{\rho}$ varies within this range, thus allowing us to handle rates which vary with n. Before proving the lemmas, we show how they are used to obtain the desired result.

Proof of (130) *Based on Lemmas 1–2:* By upper bounding $M - 1$ by M and splitting rcu (see (13)) according to whether or not $\boldsymbol{y} \in \mathcal{F}^n_{\hat{\rho},s}(\delta)$, we obtain

$$\begin{aligned} \mathrm{rcu}(n, M) \le & \sum_{\boldsymbol{x},\boldsymbol{y} \in \mathcal{F}^n_{\hat{\rho},s}(\delta)} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \\ & \times \min\Big\{1, M\mathbb{P}[i^n_s(\overline{\boldsymbol{X}}, \boldsymbol{y}) \ge i^n_s(\boldsymbol{x}, \boldsymbol{y})]\Big\} \\ & + M^{\hat{\rho}} \sum_{\boldsymbol{x},\boldsymbol{y} \notin \mathcal{F}^n_{\hat{\rho},s}(\delta)} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{-\hat{\rho} i^n_s(\boldsymbol{x},\boldsymbol{y})}, \end{aligned} \tag{201}$$

where we have replaced q^n by i^n_s since each is a monotonically increasing function of the other, and in the summation over $\boldsymbol{y} \notin \mathcal{F}^n_{\hat{\rho},s}(\delta)$ we further weakened the bound using Markov's inequality and $\min\{1, \cdot\} \le (\cdot)^{\hat{\rho}}$. In order to make the inner probability in (201) more amenable to an application of Lemma 1, we follow [33, Sec. 3.4.5] and note that the following holds whenever $\widetilde{P}^n_s(\bar{\boldsymbol{x}}|\boldsymbol{y}) > 0$:

$$Q^n(\bar{\boldsymbol{x}}) = Q^n(\bar{\boldsymbol{x}}) \frac{\widetilde{P}^n_s(\bar{\boldsymbol{x}}|\boldsymbol{y})}{\widetilde{P}^n_s(\bar{\boldsymbol{x}}|\boldsymbol{y})} = \widetilde{P}^n_s(\bar{\boldsymbol{x}}|\boldsymbol{y}) e^{-i^n_s(\bar{\boldsymbol{x}},\boldsymbol{y})}. \tag{202}$$

For a fixed sequence $\boldsymbol{y}$ and a constant t, summing (202) over all $\bar{\boldsymbol{x}}$ such that $i^n_s(\bar{\boldsymbol{x}}, \boldsymbol{y}) \ge t$ yields

$$\mathbb{P}[i^n_s(\overline{\boldsymbol{X}}, \boldsymbol{y}) \ge t] = \mathbb{E}\Big[e^{-i^n_s(\boldsymbol{X}_s,\boldsymbol{Y})} \mathbb{1}\{i^n_s(\boldsymbol{X}_s, \boldsymbol{Y}) \ge t\} \,\Big|\, \boldsymbol{Y} = \boldsymbol{y}\Big] \tag{203}$$

under the joint distribution in (197).

We now observe that (203) is of the same form as the left-hand side of (196). We apply Lemma 1 with Q_k given by the PMF of $i_s(X_s, y_k)$ under $X_s \sim \widetilde{P}_s(\cdot\,|y_k)$, where y_k is the k-th output symbol for which $\sum_x Q(x)W(y|x) > 0$. The conditions of the lemma are easily seen to be satisfied for sufficiently small δ due to the definition of $\mathcal{F}^n_{\hat{\rho},s}(\delta)$ in (198), the assumption in (115), and (123). We have from (196), (199) and (203) that

$$\mathbb{P}\big[i^n_s(\overline{\boldsymbol{X}}, \boldsymbol{y}) \ge t\big] \le \frac{\psi_s}{\sqrt{2\pi n(c_3 - r(\delta))}} e^{-t}(1 + o(1)) \tag{204}$$

for all $\boldsymbol{y} \in \mathcal{F}^n_{\hat{\rho},s}(\delta)$ and sufficiently small δ. Here we have used the fact that ψ_0 in (195) coincides with ψ_s in (120), which follows from (123) and the fact that $\widetilde{P}_s(x|y) > 0$ if and only if $Q(x)W(y|x) > 0$ (see (114) and (116)).

Using the uniformity of the $o(1)$ term in t in (204) (see Lemma 1), taking $\delta \to 0$ (and hence $r(\delta) \to 0$), and writing

$$\min\{1, f_n(1 + \zeta_n)\} \le (1 + |\zeta_n|) \min\{1, f_n\}, \tag{205}$$

we see that the first term in (201) is upper bounded by $\mathrm{rcu}^*_s(n, M)(1 + o(1))$. To complete the proof of (130), we must show that the second term in (201) can be incorporated into the multiplicative $1 + o(1)$ term. To see this, we note from (125) and (132) that the exponent of rcu^*_s is given by $E^{\mathrm{iid}}_0(Q, \hat{\rho}, s) - \hat{\rho}R$. From (94), the denominator in the logarithm in (200) equals $e^{-nE^{\mathrm{iid}}_0(Q,\hat{\rho},s)}$. Combining these observations, the second part of Lemma 2 shows that the second term in (201) decays at a faster exponential rate than rcu^*_s, thus yielding the desired result.

Proof of Lemma 1: The proof makes use of the local limit theorems given in [42, Thm. 1] and [43, Sec. VII.1, Thm. 2] for the non-lattice and lattice cases respectively. We first consider the summation $S'_n \triangleq \sum_{i=1}^{n'} Z_i$, where we assume without loss of generality that the first $n' = \Theta(n)$ indices correspond to positive variances, and the remaining $n - n'$ correspond to zero variances. We similarly assume that $\sigma_k > 0$ for $k = 1, \dots, K'$, and $\sigma_k = 0$ for $k = K' + 1, \dots, K$. We clearly have $\mathrm{Var}[S'_n] = \mathrm{Var}[S_n] = V_n$.

We first consider the non-lattice case. We claim that the conditions of the lemma imply the following local limit theorem given in [42, Thm. 1]:

$$\mathbb{P}\big[S'_n \in [z, z + \eta)\big] = \frac{\eta}{\sqrt{2\pi V_n}} e^{-\frac{(z-\mu'_n)^2}{2V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{206}$$

uniformly in z, where $\mu'_n \triangleq \mathbb{E}[S'_n]$, and $\eta > 0$ is arbitrary. To show this, we must verify the technical assumptions of [42, p. 593]. First, [42, Cond. (α)] states that there exists $Z_{\max} < \infty$ and $c > 0$ such that

$$\frac{1}{\mathrm{Var}[Z]} \mathbb{E}\big[(Z - \mathbb{E}[Z])^2 \mathbb{1}\{|Z - \mathbb{E}[Z]| \le Z_{\max}\}\big] > c \tag{207}$$

under $Z \sim Q_k$ and each $k = 1, \dots, K'$. This is trivially satisfied since we are considering finite alphabets, which implies that the support of each Q_k is bounded. The Lindeberg condition is stated in [42, Cond. (γ)], and is trivially satisfied due to the assumption that $n_k = \Theta(n)$ for all k. The only non-trivial condition is [42, Cond. (β)], which can be written as follows in the case of finite alphabets: For any given lattice, there exists $\delta > 0$ such that

$$\frac{1}{\log V_n} \sum_{i=1}^{n'} \mathbb{P}[Z_i \text{ is not } \delta\text{-close to a lattice point}] \to \infty. \tag{208}$$

Since we are considering the case that $\mathcal{I}_0$ does not lie on a lattice, we have for sufficiently small δ that the summation grows linearly in n, whereas $\log V_n$ only grows as $\log n$. We have thus shown that the technical conditions of [42] are satisfied, and hence (206) holds.

Upper bounding the exponential term in (206) by one, and noting that $S_n - S'_n$ has zero variance, we obtain

$$\mathbb{P}\big[S_n \in [z, z+\eta)\big] \le \frac{\eta}{\sqrt{2\pi V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{209}$$

uniformly in z. We can now prove the lemma similarly to [11, Lemma 47] by writing

$$\begin{aligned}
&\mathbb{E}\Big[e^{-S_n}\mathbb{1}\{S_n \ge t\}\Big] \\
&\quad \le \sum_{l=0}^{\infty} e^{-t-l\eta}\mathbb{P}\Big[t + l\eta \le S_n \le t + (l+1)\eta\Big] && (210) \\
&\quad \le \sum_{l=0}^{\infty} e^{-t-l\eta}\bigg(\frac{\eta}{\sqrt{2\pi V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big)\bigg) && (211) \\
&\quad = e^{-t}\bigg(\frac{\eta}{(1-e^{-\eta})\sqrt{2\pi V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big)\bigg), && (212)
\end{aligned}$$

where (212) follows by evaluating the summation using the geometric series. The proof is concluded by taking $\eta \to 0$ and using the identity $\lim_{\eta\to 0}\frac{\eta}{1-e^{-\eta}} = 1$. The uniformity of (196) in t follows from the uniformity of (209) in z.

In the lattice case, the argument is essentially unchanged, but we instead use the local limit theorem given in [43, Sec. VII.1, Thm. 2], which yields

$$\mathbb{P}[S'_n = z] = \frac{h_0}{\sqrt{2\pi V_n}}e^{-\frac{(z-\mu'_n)^2}{2V_n}} + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{213}$$

uniformly in z on the lattice corresponding to S'_n (with span h_0). The remaining arguments are identical to the non-lattice case, with $\eta = h_0$ instead of $\eta \to 0$.

Proof of Lemma 2: We obtain (199) by using the definitions of c_3 and $\mathcal{F}^n_{\hat{\rho},s}(\delta)$ (see (118) and (198)) to write

$$\begin{aligned}
&\mathrm{Var}[i^n_s(\boldsymbol{X}_s, \boldsymbol{Y}) \,|\, \boldsymbol{Y} = \boldsymbol{y}] \\
&\quad = \sum_y n\hat{P}_{\boldsymbol{y}}(y)\mathrm{Var}[i_s(X_s, Y) \,|\, Y = y] && (214) \\
&\quad \ge \sum_y n(P^*_{\hat{\rho},s}(y) - \delta)\mathrm{Var}[i_s(X_s, Y) \,|\, Y = y] && (215) \\
&\quad = n\big(c_3 - o(\delta)\big), && (216)
\end{aligned}$$

where $(X_s | Y = y) \sim \widetilde{P}_s(\cdot\,|\,y)$. To prove the second property, we perform an expansion in terms of types in the same way as Appendix A to conclude that the exponent of the denominator in the logarithm in (200) is given by

$$\min_{P_{XY}} \sum_{x,y} P_{XY}(x,y)\log\bigg(\frac{P_{XY}(x,y)}{Q(x)W(y|x)}e^{\hat{\rho}i_s(x,y)}\bigg). \tag{217}$$

Similarly, using the definition of $\mathcal{F}^n_{\hat{\rho},s}(\delta)$ in (198), the exponent of the numerator in the logarithm in (200) is given by

$$\min_{P_{XY}\,:\,\max_y |P_Y(y) - P^*_{\hat{\rho},s}(y)| > \delta} \sum_{x,y} P_{XY}(x,y)\log\bigg(\frac{P_{XY}(x,y)}{Q(x)W(y|x)}e^{\hat{\rho}i_s(x,y)}\bigg). \tag{218}$$

A straightforward evaluation of the KKT conditions [17, Sec. 5.5.3] yields that (217) is uniquely minimized by $P^*_{\hat{\rho},s}$, defined in (117). On the other hand, $P^*_{\hat{\rho},s}$ does not satisfy the constraint in (218), and thus (218) is strictly greater than (217). This concludes the proof of (200).

References

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

[2] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Dept. Comput. Sci., MIT, Cambridge, MA, USA, 1983.

[3] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elek. Über.*, vol. 47, no. 4, pp. 228–239, 1993.

[4] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.

[5] 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.

[6] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.

[7] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[8] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[9] R. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[10] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Proc. Trans. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[11] 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.

[12] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[13] J. L. Jensen, *Saddlepoint Approximations*. Oxford, U.K.: Oxford Univ. Press, 1995.

[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge. U.K.: Cambridge Univ. Press, 2011.

[15] R. Gallager. (2013). *Fixed Composition Arguments and Lower Bounds to Error Probability* [Online]. Available: http://web.mit.edu/gallager/www/notes/notes5.pdf

[16] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] A. Somekh-Baruch. (2013). On achievable rates for channels with mismatched decoding. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1305.0547

[19] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas. (2013). Multiuser coding techniques for mismatched decoding. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1311.6635

[20] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas. (2013). Expurgated random-coding ensembles: Exponents, refinements and connections. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1307.6679

[21] A. Somekh-Baruch. (2013). A general formula for the mismatch capacity. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1309.7964

[22] Y. Altuğ and A. B. Wagner. (2014). Refinement of the random coding bound. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1312.6875

[23] P. Elias, "Coding for two noisy channels," in *Proc. 3rd London Symp. Inf. Theory*, 1955, pp. 1–20.

[24] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Comput. Sci., Tel Aviv Univ., Jerusalem, Israel, 2003.

[25] C. Stone, "On local and ratio limit theorems," in *Proc. 5th Berkeley Symp. Math. Stat. Prob.*, 1965, pp. 217–224.

[26] R. Fano, *Transmission of Information: A Statistical Theory of Communications*. Cambridge, MA, USA: MIT Press, 1961.

[27] R. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 244–246, Mar. 1973.

[28] A. G. D'yachkov, "Bounds on the average error probability for a code ensemble with fixed composition," *Probab. Inf. Transmiss.*, vol. 16, no. 4, pp. 3–8, 1980.

[29] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, Taiwan, 2004, pp. 284–289.

[30] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Cost-constrained random coding and applications," in *Proc. Inf. Theory Apps. Workshop*, San Diego, CA, USA, Feb. 2013, pp. 1–7.

[31] H. Rubin and J. Sethuraman, "Probabilities of moderate deviations," *Indian J. Statist.*, vol. 27, no. 2, pp. 325–346, Dec. 1965.

[32] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2, 2nd ed. New York, NY, USA: Wiley, 1971.

[33] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Dept. Electr. Eng., Princeton Univ., Princeton, NJ, USA, 2010.

[34] R. L. Dobrushin, "Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix," *Theory Probab. Appl.*, vol. 7, no. 3, pp. 270–300, 1962.

[35] Y. Altuğ and A. B. Wagner. (2012). Refinement of the sphere-packing bound: Asymmetric channels. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1211.6697

[36] Y. Altuğ and A. B. Wagner. (2012). Moderate deviations in channel coding. Submitted to *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/1208.1924

[37] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *Proc. Allerton Conf. Commun., Control Comput.*, 2010, pp. 1334–1339.

[38] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *Proc. Inf. Theory Appl. Workshop*, La Jolla, CA, USA, 2011, pp. 1–6.

[39] R. Bahadur and R. Ranga Rao, "On deviations of the sample mean," *Ann. Math. Statist.*, vol. 31, pp. 1015–1027, Dec. 1960.

[40] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "A derivation of the asymptotic random-coding prefactor," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, USA, 2013, pp. 1–6.

[41] K. Fan, "Minimax theorems," *Proc. Nat. Acad. Sci.*, vol. 39, no. 1, pp. 42–47, 1953.

[42] J. Mineka and S. Silverman, "A local limit theorem and recurrence conditions for sums of independent non-lattice random variables," *Ann. Math. Statist.*, vol. 41, no. 2, pp. 592–600, Apr. 1970.

[43] V. V. Petrov, *Sums of Independent Random Variables*. New York, NY, USA: Springer-Verlag, 1975.

Jonathan Scarlett was born in Melbourne, Australia, in 1988. He received the B.Eng. degree in electrical engineering and the B.Sci. degree in computer science from the University of Melbourne, Australia, in 2010. In 2011 he was a Research Assistant with the Department of Electrical and Electronic Engineering, University of Melbourne. He is currently pursuing the Ph.D. degree from the Signal Processing and Communications Group at the Department of Engineering, University of Cambridge, Cambridge, U.K. His current research interests include information theory and signal processing. He holds the Poynton Cambridge Australia International Scholarship.

Alfonso Martinez (SM'11) was born in Zaragoza, Spain, in October 1973. He is currently a Ramón y Cajal Research Fellow at Universitat Pompeu Fabra, Barcelona, Spain. He received the Telecommunications Engineering degree from the University of Zaragoza in 1997. From 1998 to 2003 he was a Systems Engineer at the research centre of the European Space Agency (ESA-ESTEC) in Noordwijk, The Netherlands. His work on APSK modulation was instrumental in the definition of the physical layer of DVB-S2. From 2003 to 2007 he was a Research and Teaching Assistant at Technische Universiteit Eindhoven, The Netherlands, where he conducted research on digital signal processing for MIMO optical systems and on optical communication theory. Between 2008 and 2010 he was a Post-doctoral Fellow with the Information-theoretic Learning Group at Centrum Wiskunde & Informatica (CWI), in Amsterdam, The Netherlands. In 2011 he was a Research Associate with the Signal Processing and Communications Lab at the Department of Engineering, University of Cambridge, Cambridge, U.K.

His current research interests include information theory and coding, with emphasis on digital modulation and the analysis of mismatched decoding; in this area he has coauthored a monograph on *Bit-Interleaved Coded Modulation*. He is also involved in connections between information theory, optical communications, and physics, particularly by the links between classical and quantum information theory.

Albert Guillén i Fàbregas (S'01–M'05–SM'09) was born in Barcelona, Catalunya, Spain, in 1974. In 1999 he received the Telecommunication Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL), in 2004.

Since 2011 he has been a Research Professor of the Institució Catalana de Recerca i Estudis Avançats (ICREA) hosted at the Department of Information and Communication Technologies, Universitat Pompeu Fabra. He is also an Adjunct Researcher at the Department of Engineering, University of Cambridge. He has held appoinments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, University of Cambridge where he was a Reader and a Fellow of Trinity Hall, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His current research interests include information theory, communication theory, coding theory, digital modulation and signal processing techniques.

Dr. Guillén i Fàbregas received the Starting Grant from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is a Member of the Young Academy of Europe. He is a co-author of the monograph book *Bit-Interleaved Coded Modulation*. He is also an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, an Editor of the *Foundations and Trends in Communications and Information Theory*, Now Publishers and was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).