

Random-Coding Joint Source-Channel Bounds

Adrià Tauste Campo
Department of Engineering
University of Cambridge
atauste@ieee.org

Gonzalo Vazquez-Vilar
Universidad de Vigo
gvazquez@ieee.org

Albert Guillén i Fàbregas
Department of Engineering
University of Cambridge
guillen@ieee.org

Alfonso Martinez
Department of Engineering
University of Cambridge
alfonso.martinez@ieee.org

Abstract—Random-coding exact characterizations and bounds to the error probability of joint source-channel coding are presented. In particular, upper bounds using maximum-a-posteriori and threshold decoding are derived as well as a lower bound motivated by Verdú-Han’s lemma.

I. INTRODUCTION

The joint source-channel coding (JSCC) theorem consists of two parts. The direct part states that reliable transmission of a source over a channel is possible if the minimum achievable coding rate of a given source is strictly below the channel capacity. The converse part states that reliable transmission is not possible for source coding rates strictly larger than the channel capacity. The theorem was first proved by Shannon in [2] for the case of stationary memoryless sources and channels. Since then, the result has been generalized in various ways. For memoryless sources and channels, the approach based on the *reliability function* (or *error exponent*), pioneered by Gallager [3], was elaborated by Csiszár [4], [5] and more recently by Zhong *et al.* [6]. In parallel, the JSCC theorem has been extended to general classes of sources and channels [1], [7]–[10] refining the sufficient and necessary conditions for a source to be transmissible over a channel. These conditions are based on the tail distribution of specific random variables called information spectrum measures (see [11] for a review on information spectrum methods).

Most of the above results rely on the assumption that the source and channel blocklength can increase without bound. Instead, in this paper, we study the error probability of joint source-channel codes for arbitrary, i.e., finite, blocklengths. More specifically, we derive new tight achievability bounds based on the average error probability under maximum a posteriori (MAP) and threshold decoding. The new bounds are inspired by [12], strengthen classical results [1], [3], [13], [14], and can be applied to prove the direct part of the general JSCC theorem [8].

II. JOINT SOURCE-CHANNEL CODING

We define a *discrete source* [11] over a finite alphabet \mathcal{V} as a sequence of k -dimensional random variables $\mathbf{V} \triangleq \{V^k\}_{k=1}^{\infty}$, where each V^k takes values in \mathcal{V}^k . Similarly, a *discrete channel* [11] is defined as a sequence $\mathbf{W} \triangleq \{W^n : \mathcal{X}^n \rightarrow$

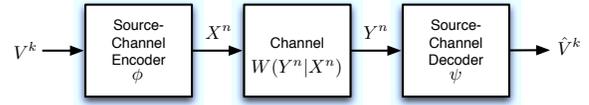


Fig. 1. Block diagram of joint source-channel coding.

$\mathcal{Y}^n\}_{n=1}^{\infty}$ of n -dimensional probability transition matrices W^n , where $W^n(y|x)$ ($x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$) is the conditional probability of y^n given x^n .

A joint source-channel code with transmission rate t is defined as follows. Given a source \mathbf{V} , and a channel \mathbf{W} , we fix k and n such that $t = \frac{k}{n}$, and we consider an encoder mapping, $\phi : \mathcal{V}^k \rightarrow \mathcal{X}^n$, and a decoder mapping, $\psi : \mathcal{Y}^n \rightarrow \mathcal{V}^k$. If $X^n = \phi(V^k)$ is the encoded n -dimensional vector (associated to the k -dimensional source vector V^k), Y^n the output vector corresponding to the input X^n , and $\hat{V}^k = \psi(Y^n)$ the decoded vector, the joint source-channel code is characterized by the following Markov-chain relation (see Fig. 1),

$$V^k \xrightarrow{\phi} X^n \rightarrow Y^n \xrightarrow{\psi} \hat{V}^k. \quad (1)$$

We wish to evaluate the average error probability of code (ϕ, ψ) defined as

$$\epsilon_{k,n} \triangleq \sum_{v \in \mathcal{V}^k} P_{V^k}(v) \Pr\{\psi(Y^n) \neq v\}. \quad (2)$$

In the following, a joint source-channel code (ϕ, ψ) with average error probability equal to $\epsilon_{k,n} > 0$ will be called a $\epsilon_{k,n}$ -code.

For ease of exposition of our results, we shall drop the dimension of the vector random variables V^k, X^n, Y^n and transition probabilities $W^n(\cdot, \cdot)$, as it will be clear from the context that the source outputs are k -dimensional and the channel inputs and outputs are n -dimensional. We shall use $\Pr\{\cdot\}$ to denote the probability of an event. Similarly $\mathbb{E}[\cdot]$ will be used to denote the expectation of a random variable.

III. MAP DECODING UPPER BOUNDS

In this section, we present new random-coding upper bounds to the error probability with MAP decoding. As we shall see, each bound follows from subsequent relaxations stemming from an exact characterization. For a particular y let $\mathcal{D}_\phi(y)$ be the set of source outputs whose MAP decoding metric is maximum and $|\mathcal{D}_\phi(y)| = \ell$ its cardinality. For a given v , let

¹This work has been supported in part by the European Research Council under ERC grant agreement 259663. University of Vigo contribution supported in part by the Spanish Government under project COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010).

the ensemble of subsets of $\mathcal{T}(v) \triangleq \mathcal{V}^k \setminus \{v\}$ with cardinality ℓ , be defined as $\mathcal{P}(\mathcal{T}(v), \ell) \triangleq \{\mathcal{A} \subseteq \mathcal{T}(v), |\mathcal{A}| = \ell\}$, and be denoted by $\mathcal{P}(\mathcal{T}(V), \ell)$ when v is not specified. Hence, for a given transmitted message v , $\mathcal{P}(\mathcal{T}(v), \ell)$ denotes the ensemble of all possible combinations of decoding ties with v . Then, simple combinatorial arguments yield the following result.

Theorem 1: The random-coding average error probability of JSCC under MAP decoding and arbitrary $P_{X|V}$ is given by

$$\bar{\epsilon}_{k,n}^{\text{MAP}} \triangleq 1 - \sum_{\ell \geq 0} \frac{1}{\ell + 1} \mathbb{E} [\Pr\{\ell | VXY\}], \quad (3)$$

where

$$\begin{aligned} & \Pr\{\ell | VXY\} \\ &= \sum_{\mathcal{A} \in \mathcal{P}(\mathcal{T}(V), \ell)} \prod_{v' \in \mathcal{A}} \Pr \left\{ \frac{P_V(v')W(Y|\bar{X})}{P_V(V)W(Y|X)} = 1 \middle| VXY \right\} \times \\ & \quad \times \prod_{v'' \in \mathcal{T}(V) \setminus \{\mathcal{A}\}} \Pr \left\{ \frac{P_V(v'')W(Y|\bar{X})}{P_V(V)W(Y|X)} < 1 \middle| VXY \right\}, \end{aligned} \quad (4)$$

and \bar{X} (the codeword corresponding to v' , $v'' \neq v$) is assumed to be independent of X .

The computation of the exact random coding error probability is challenging even for small (k, n) . In the following, we provide a looser version of Eq. (3) by assuming that the decoder always resolves all ties in error.

Theorem 2 (Max-bound): For every (k, n) , there exists an $\epsilon_{k,n}$ -code such that

$$\epsilon_{k,n} \leq \mathbb{E} \left[\Pr \left\{ \frac{\max_{v' \neq V} P_V(v')W(Y|\bar{X})}{W(Y|X)P_V(V)} \geq 1 \middle| VXY \right\} \right], \quad (6)$$

where $v' \neq V$ is an abuse of notation.

In practice, the computation of the bound (6) typically involves an exponential number of products, which requires a high degree of numerical precision as each factor can be fairly small. In order to obtain simpler bounds we can use the union bound

$$\Pr \left\{ \max_i A_i \geq 1 \right\} \leq \min \left\{ 1, \sum_i \Pr\{A_i \geq 1\} \right\}, \quad (7)$$

where A_i is a real-valued random variable.

Based on Eq. (7), we can extend the RCU bound for channel coding [12, Th. 16] to JSCC.

Corollary 1 (RCU bound): There exists an $\epsilon_{k,n}$ -code such that

$$\begin{aligned} \epsilon_{k,n} &\leq \text{RCU}(k, n) \\ &\triangleq \mathbb{E} \left[\min \left\{ 1, \sum_{v' \neq V} \Pr \left\{ \frac{P_V(v')W(Y|\bar{X})}{P_V(V)W(Y|X)} \geq 1 \middle| VXY \right\} \right\} \right]. \end{aligned} \quad (8)$$

The bound in Eq. (8) turns out to be a good bound, especially when the error event is sufficiently small, as most of the pairwise error probabilities are very small. Furthermore,

we can relax Eq. (8) by applying Markov's inequality to each term in the sum, which yields

$$\bar{\epsilon}_{k,n} \leq \mathbb{E} \left[\min \left\{ 1, \sum_{v' \neq V} \frac{\mathbb{E} [P_V(v')^s W(Y|\bar{X})^s | Y]}{P_V(V)^s W(Y|X)^s} \right\} \right]. \quad (10)$$

By using the identity $\min\{1, x\} = \exp(-|\log x^{-1}|^+)$, $\forall x \in \mathbb{R}^+$, where $|x|^+ = \max(x, 0)$, a new family of upper bounds parameterized by $s > 0$ follow.

Theorem 3 (Tilted RCU bounds): For $s > 0$, there exists a $\epsilon_{k,n}$ -code such that

$$\epsilon_{k,n} \leq \text{RCU}_s(k, n) = \mathbb{E} \left[e^{-n|j_s(V, X, Y)|^+} \right], \quad (11)$$

where

$$j_s(V, X, Y) \triangleq \frac{1}{n} \log \frac{P_V(V)^s W(Y|X)^s}{\sum_{v'} P_V(v')^s \sum_{x'} P_{X|V}(x'|v') W(Y|x')^s}. \quad (12)$$

In particular, when $s = 1$ equation (12) takes the form $j_1(V, X, Y) = i(X, Y) - th(V)$, where $i(X, Y)$ and $h(V)$ are the information and entropy density rates, respectively [11]. Hence, Theorem 3 can be used to prove the direct part of the general JSCC theorem [8]. For ease of notation, we shall use $j(V, X, Y)$ to denote $j_1(V, X, Y)$ in the rest of the paper. Moreover, when X is independent of V , $j_s(V, X, Y)$ decomposes into a *source* and a *channel* term

$$j_s(V, X, Y) = i_s(X, Y) - th_s(V), \quad (13)$$

where $i_s(X, Y) \triangleq \frac{1}{n} \log \frac{W(Y|X)^s}{\mathbb{E}[W(Y|X)^s | Y]}$ and $h_s(V) \triangleq \frac{1}{n} \log \frac{P_V(V)^s}{\sum_{v'} P_V(v')^s}$.

Gallager's upper bound [3, Prob. 5.16] can be obtained from Eq. (10) by assuming that V and X are independent, by using $\min(1, x) \leq x^\rho$ for $0 \leq \rho \leq 1$ and by letting $s = \frac{1}{1+\rho}$:

$$\bar{\epsilon}_{k,n} \leq e^{-\tilde{E}_0(\rho, \frac{1}{1+\rho}, W) + E_s(\rho, \frac{1}{1+\rho}, P_V)}, \quad (14)$$

where

$$\tilde{E}_0(\rho, s, W) \triangleq \max_{P_X} E_0(\rho, s, P_X, W) \quad (15)$$

$$E_0(\rho, s, P_X, W) \triangleq -\log \mathbb{E} \left[\left(\frac{\mathbb{E} [W(Y|X)^s | Y]}{W(Y|X)^s} \right)^\rho \right] \quad (16)$$

$$E_s(\rho, s, P_V) \triangleq \log \mathbb{E} \left[\left(\frac{\mathbb{E} [P_V(V)^{s-1}]}{P_V(V)^s} \right)^\rho \right]. \quad (17)$$

Equation (14) gives a lower bound to the JSCC exponent that was strengthened for memoryless sources and channels by Csiszár [4]. Csiszár's original bound has been formulated as [6]

$$\bar{\epsilon}_{k,n} \leq e^{-n(\bar{E}_0(\rho, W) - tE_s(\rho, \frac{1}{1+\rho}, P_V)) + o(n)}, \quad (18)$$

where $\bar{E}_0(\rho, W)$ is the convex hull of $\tilde{E}_0(\rho, \frac{1}{1+\rho}, W)$ over $\rho \in [0, 1]$. However, the original proof is based on the method of types [15] and uses a maximum mutual information decoder at the receiver. In this context, we have recently rederived Csiszár's lower bound as a random-coding exponent with MAP decoding [16].

IV. THRESHOLD DECODING UPPER BOUNDS

We now study upper bounds to the error probability using (suboptimal) threshold decoders. We assume that the source vectors v_1, v_2, \dots, v_{M_s} are indexed according to their corresponding probabilities such that $P_V(v_1) \geq P_V(v_2) \geq \dots \geq P_V(v_{M_s})$. In case of equality between source probabilities the indexes are set arbitrarily. We define the decoding function $T_\gamma : \mathcal{V} \times \mathcal{Y} \rightarrow \{0, 1\}$ by

$$T_\gamma(v_m, y) = \mathbb{1} \{j(v_m, x(v_m), y) > \gamma(v_m)\}, \quad (19)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Given a transmitted codeword x and a received sequence y the threshold decoder ψ_γ returns the lowest index \hat{m} (associated to the highest probable source vector) for which $T_\gamma(v_{\hat{m}}, y) = 1$. For this decoder, we have the following exact expression.

Theorem 4: The average random-coding error probability of JSCC under threshold decoding is given by

$$\bar{\epsilon}_{k,n}^{\text{TD}} = 1 - \sum_{m=1}^{M_s} P_V(v_m) \Pr \{\psi(Y) = v_m\}, \quad (20)$$

where

$$\Pr \{\hat{v}_m = v_m\} = \mathbb{E} \left[\Pr \{j(v_m, X, Y) > \gamma(v_m) | Y\} \times \prod_{l=1}^{m-1} \Pr \{j(v_l, \bar{X}, Y) \leq \gamma(v_l) | Y\} \right], \quad (21)$$

where the probabilities in the second term assume that \bar{X} and Y are independent.

Optimizing the above exact expression over all possible source-dependent thresholds is computationally challenging. In order to overcome this limitation, we follow the arguments in [12] and apply the union bound to Eq. (21). This results in the extension of the dependence-testing (DT) bound [12, Th. 17] to JSCC:

$$\bar{\epsilon}_{k,n} \leq \Pr \{j(V, X, Y) \leq \gamma(V)\} + \sum_v \Pr \{\mathcal{S}(v)\} \Pr \{j(v, X, \bar{Y}) > \gamma(v)\}, \quad (22)$$

where

$$\mathcal{S}(v) \triangleq \{v' \in \mathcal{V}^k \mid P_V(v') \leq P_V(v)\}, \quad (23)$$

and \bar{Y} is independent of X . Following the arguments of [12] we find that the optimal threshold is given by $\gamma(v) = \frac{1}{n} \log \Pr \{\mathcal{S}(v)\}$, and that Eq. (22) can be rewritten as follows.

Theorem 5 (Optimized DT Bound): There exists an $\epsilon_{k,n}$ -code that satisfies

$$\epsilon_{k,n} \leq \text{DT}_*(k, n) = \mathbb{E} \left[e^{-n|j(V, X, Y) - \gamma(V)|^+} \right]. \quad (24)$$

We can apply the Chernoff bound to (24) to lower-bound the exponent achieved by threshold decoding as [16]

$$\text{DT}_*(k, n) \leq e^{-\tilde{E}_0(\rho, 1, W) + E_s(\rho, \frac{1}{1+\rho}, P_V)}. \quad (25)$$

The particularization of Theorem 5 to channel coding leads to a marginally tighter result than the original DT bound as it

is observed in [17]. Also $\text{DT}_*(k, n)$ is tighter than $\text{RCU}_s(k, n)$ with $s = 1$ as $\gamma(v) \leq 0, \forall v \in \mathcal{V}^k$.

We now use equation (22) to recover Feinstein's bound [8, Lemma 3.1]. If we fix $\gamma(V) = \gamma$ to be a constant for each source sequence and upper-bound the second term in (22) by using Markov's inequality we obtain

$$\bar{\epsilon}_{k,n} \leq \Pr \{j(V, X, Y) \leq \gamma\} + e^{-n\gamma} \sum_v P_V(v) \Pr \{\mathcal{S}(v)\} \quad (26)$$

$$= \Pr \{j(V, X, Y) \leq \gamma\} + e^{-n\gamma}, \quad (27)$$

which corresponds to Feinstein's bound.

V. LOWER BOUNDS

In this section, we describe a generalization of the Verdú-Han lower bound [1] for JSCC. To do that, we define a set \mathcal{G} , and upper-bound its probability as

$$\Pr \{\mathcal{G}\} \leq \epsilon_{k,n} + \Pr \{\mathcal{G} \cap \text{no error}\}. \quad (28)$$

The choice

$$\mathcal{G} = \{(v, y) : P_V(v)W(y|x(v)) \leq \gamma(y)\} \quad (29)$$

with $\gamma(y) \geq 0$ yields the following result.

Theorem 6 (Generalized Verdú-Han lower bound): For a given source and channel, a (ϕ, ψ) -code satisfies

$$\epsilon_{k,n} \geq \Pr \{P_V(V)W(Y|\phi(V)) \leq \gamma(Y)\} - \sum_y \gamma(y). \quad (30)$$

It is easy to see that we can recover a Verdú-Han's lower bound for JSCC [8, Lemma 3,2],

$$\epsilon_{k,n} \geq \Pr \left\{ \frac{P_V(V)W(Y|\phi(V))}{P_Y(Y)} \leq -\gamma' \right\} - \gamma', \quad (31)$$

by setting $\gamma(y) = P_Y(y)\gamma'$, with $\gamma' > 0$ and P_Y the average output density for a given code. Computing (30) for the best code is in general intractable. In any event, one has the freedom to choose P_Y such that Y is independent of $X = \phi(V)$ and average (31) over the set of randomly selected codes. Indeed, by defining $\gamma(y) = \mathbb{E} [W(y|X)] e^{-n\gamma'}$ in (30) we get the following bound on the average error probability $\bar{\epsilon}_{k,n}$,

$$\bar{\epsilon}_{k,n} \geq \text{LB}(k, n, \gamma) = \Pr \{j(V, X, Y) \leq -\gamma'\} - e^{-n\gamma'}, \quad (32)$$

where $j(V, X, Y)$ is as defined in Eq. (12) for $s = 1$. While the above bound cannot be used to prove a converse, it may be used to lower-bound the random-coding error probability.

VI. APPLICATION TO BMS-BSC AND BMS-BEC

In the following, we numerically compare the upper bounds $\text{RCU}(k, n)$, $\text{RCU}_s(k, n)$, and $\text{DT}_*(k, n)$ and the lower bound $\text{LB}(k, n, \gamma)$ for the source-channel pairs BMS-BSC and BMS-BEC and establish links with classical results. BMS's are parameterized by δ denoting the probability of bit 1. For the sake of simplicity we consider $n = k$, i.e., transmission rate $t = 1$. We evaluate our bounds over the ensemble of random source-channel codes generated by the capacity achieving

distribution $P_X(x) = 2^{-n}$. In each case, the tilted RCU bound is computed with the s that optimizes Gallager's upper bound (14) and will be denoted as $\text{RCU}_{s_G}(n)$. We also denote by $\text{LB}(n)$ the lower bound obtained by optimizing $\text{LB}(n, \gamma)$ over $\gamma > 0$.

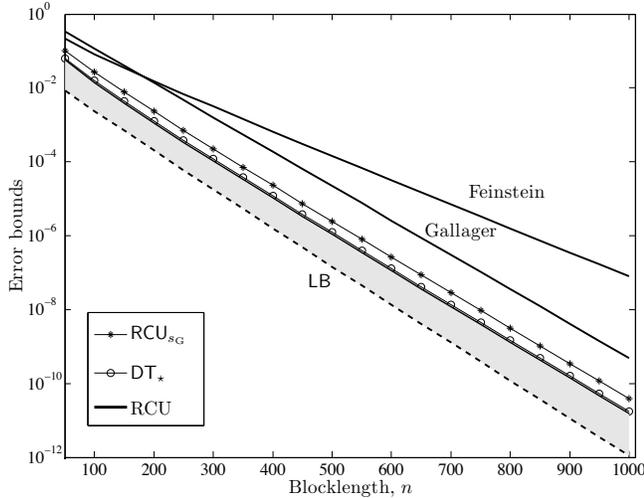


Fig. 2. JSSC error bounds for a BMS with $\delta = 0.05$ and BEC channel with erasure probability $\xi = 0.5$.

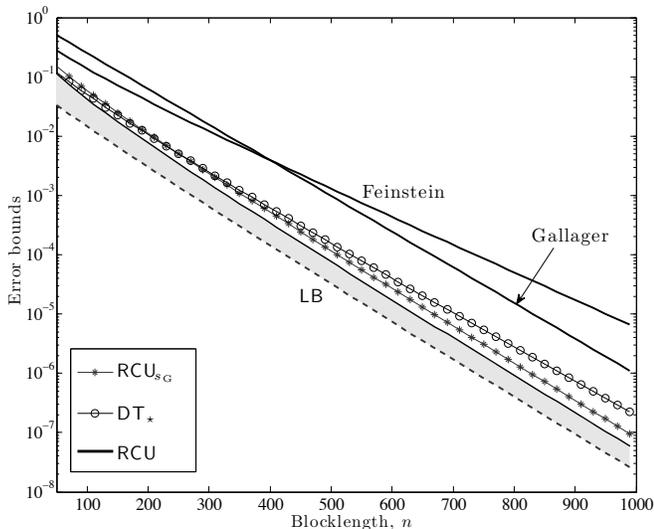


Fig. 3. JSSC error bounds for a BMS with $\delta = 0.05$ and a BSC with crossover probability $\xi = 0.11$.

Fig. 2 shows upper and lower bounds for a BMS with $\delta = 0.05$ that is transmitted over a BEC with erasure probability $\xi = 0.5$. The three new upper bounds

showed in Fig. 2 significantly outperform Gallager's (14) and Feinstein's (27) bounds in the whole range of blocklengths. In particular, the threshold decoding bound $\text{DT}_*(n)$ (24) has the same exponent as the MAP decoding bounds $\text{RCU}(n)$ (8) and $\text{RCU}_{s_G}(n)$ (11). For the BEC, the \tilde{E}_0 function (15) is independent of s and thus, equation (25) coincides with (14) and the exponents are equal. Besides, Gallager's error exponent is equal to Csiszár's (18) as the channel is symmetric [4], [6].

Fig. 3 shows the error bounds for the same BMS transmitted over a BSC with crossover probability $\xi = 0.11$. The three new bounds are also tighter than Gallager's and Feinstein's bounds over the blocklength range in Fig. 3. However, the threshold decoding bound $\text{DT}_*(n)$ does not have the optimal exponent and it is expected to perform poorer than Gallager's bound for sufficiently large blocklength. Indeed, in this example $s_G < 1$ and Gallager's exponent is strictly larger than $\text{DT}_*(n)$ on account of (24).

REFERENCES

- [1] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
- [2] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York, 1968.
- [4] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inf. Theory*, vol. 9, pp. 315–328, 1980.
- [5] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 823–828, Nov. 1982.
- [6] Y. Zhong, F. Alajaji, and L. Lorne Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1450–1468, April 2006.
- [7] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, 1995.
- [8] T. S. Han, "Joint source-channel coding revisited: Information-spectrum approach," *Arxiv preprint arXiv:0712.2959v1*.
- [9] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Math., Statist., Probabil.*, 1961, vol. 1, pp. 211–252.
- [10] S. Verdú and T. S. Han, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
- [11] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer-Verlag, Berlin, Germany, 2003.
- [12] Y. Polyanskiy, H.V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [13] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inf. Theory*, vol. IT-4, pp. 2–22, 1954.
- [14] C. E. Shannon, "Certain results on the error probability of channel coding," *Inf. Contr.*, vol. 10, no. 1, pp. 65–103.
- [15] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.
- [16] A. Tauste Campo, *Topics in joint source-channel coding and multiuser detection*, Ph.D. thesis, University of Cambridge, Cambridge, UK, June 2011.
- [17] A. Martínez and A. Guillén i Fàbregas, "Random-Coding Bounds for Threshold Decoders: Error Exponent and Saddlepoint Approximation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, July-August 2011.