



HAL
open science

On the saddlepoint approximation of the dependence testing bound in memoryless channels

Dadja Anade, Jean-Marie Gorce, Philippe Mary, Samir Perlaza

► To cite this version:

Dadja Anade, Jean-Marie Gorce, Philippe Mary, Samir Perlaza. On the saddlepoint approximation of the dependence testing bound in memoryless channels. ICC 2020 - IEEE International Conference on Communications, Jun 2020, Dublin, Ireland. pp.1-5. hal-02457361v1

HAL Id: hal-02457361

<https://inria.hal.science/hal-02457361v1>

Submitted on 28 Jan 2020 (v1), last revised 3 Mar 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the tightness of dependence testing bound approximations in point-to-point memoryless channels

Dadja T. Anade A.^{*}, Jean-Marie Gorce^{*}, Philippe Mary[†], Samir Perlaza^{*}

^{*}Univ Lyon, INSA Lyon, Inria, CITI, F-69621 Villeurbanne, France

[†]Univ. Rennes, INSA Rennes, CNRS, IETR, Rennes, France

Email: dadja-toussaint.anade-apko@insa-lyon.fr

Abstract—In this paper, upper and lower bounds on the dependence testing (DT) bound on the average decoding error probability in channel coding are derived. These bounds are obtained when the saddlepoint approximation is used to approach the DT bound, known for its computational complexity. The formulation is valid for any point-to-point memoryless channel configuration in a wide range of system parameters. These bounds allow to estimate the accuracy of the known normal and the new saddlepoint approximations, even when the fundamental limits are not known. The proposed bounds are simple to compute and provide an interesting addendum to the classical normal or saddlepoint approximations. We illustrate the interest of this approach for the binary symmetric channel and for the binary Gaussian channel. In all scenarios we studied, the bounds obtained from the saddlepoint approximation outperform the bounds obtained from the normal approximation.

Index Terms—Finite block length, Berry-Esseen, saddlepoint approximation, binary symmetric channel.

I. INTRODUCTION

The advent of the Internet of Things and massive machine type communications (mMTCs) raised up new technical and scientific issues. Indeed, mMTCs imply a paradigm shift in the design of wireless systems to cope with the high number of devices transmitting simultaneously and with the short length of packets used to transmit small information quantities [1]. From this perspective, classical results of information theory in the asymptotic regime, i.e. when the number of channel uses tends to infinity are not relevant.

Exploring the fundamental relationship between the information rate and the error probability when the number of channel uses is finite has been studied in [2], [3], [4] and references therein. In particular, achievable bounds on the decoding error probability for general point-to-point (P2P) channels have been derived using the random coding union (RCU) bound [2, Th. 16] and dependence testing (DT) bound [2, Th. 17 and 18]. The DT bound is easier to evaluate than the RCU bound and sometimes outperforms the RCU bound, as it is the case in the binary erasure channel. The converse rely on the *meta-converse* approach and is the solution to a minimax problem on the input and output distributions.

Other recent works dealt with the multi-user case, extending the results in [2], and rely on the normal approximation. In [5] and [6] converse and achievable bounds are provided for

discrete memoryless multiple access channel (DM-MAC). In [7] the asymptotic expansion on DM-MAC is addressed. In [8], the authors dealt with an achievable bound for the second-order rate region in Gaussian MAC. The achievable second-order rate region is derived thanks to the multi-dimensional Berry-Esseen theorem in [7] and with the central limit theorem for functions of random vectors, in [8]. In both works, the tractable expression of the finite block length rate region is obtained while considering n sufficiently large. Therefore, the accuracy of the region derived for small n is not guaranteed and cannot be evaluated since the theoretical rate region is not known. Broadcast channel (BC) has been dealt with in [7], [9], [10] but [7] suffers from the same limitation in small n and [9], [10] did not provide an asymptotic expansion of the rate region.

In [11], a new approach to compute the asymptotic expansion of [2] via Laplace integrals is introduced. Nonetheless, very little can be said about the approximation error. In [4], a saddlepoint approximation of the meta-converse and the RCU bounds for channel coding is proposed. This approach outperforms the normal approximation but this is proved only by comparison to the true minimal error probability assessed with Monte Carlo simulations.

In all the above mentioned contributions, the normal and saddlepoint approximations appeared to be good even for moderate n but these results come without a formal proof of their accuracy.

In this paper, we aim at bounding the error induced when approaching the DT bound by saddlepoint approximation. This approximation rely on the substitution of an unknown distribution by another one. This paper provides a bound on the error of such an approximation, whatever the value of n . The resulting bounds when sufficiently tight, can justify the use of the saddlepoint technique, even in channels where the exact computation of the error probability is unavailable for moderate n . The approach is illustrated on the binary symmetric channel (BSC) and the binary input Additive White Gaussian noise (BI-AWGN).

In Section II, a system model of point-to-point memoryless channels is presented and the computation problem of DT bound is introduced. In Section III, the preliminaries of probability approximation are introduced and previous results re-

called. In Section IV, saddlepoint approximation is introduced and a new upper bound on the absolute approximation error is presented. In Section V, existing and new tractable bounds on the DT bound for point-to-point memoryless channels with i.i.d. channel inputs are presented using the results of Section V. Then, a numerical analysis for the BSC and BI-AWGN channels is conducted. Section VII concludes the paper.

Notation: A set is represented by a calligraphic upper letter, e.g. \mathcal{S} . A scalar (resp. vector) random variable is denoted by an upper case letter (resp. bold upper case), e.g. X (\mathbf{X}), while its realization is represented with a lower case letter (resp. bold lower case), e.g. x (\mathbf{x}). A probability distribution of the random variable X is denoted by P_X and $P_{Y|X}$ is the conditional distribution of Y given X . Moreover, $\mathbb{1}\{c\}$ is the indicator function that is 1 if the condition c is true and \mathbb{E}_{P_X} denotes the expectation over the probability distribution P_X .

II. PROBLEM STATEMENT

A. System Model

Consider a point-to-point communication system in which a transmitter aims at sending information to one receiver through a noisy memoryless channel. Such a channel can be modeled by a random transformation

$$(\mathcal{X}^n, \mathcal{Y}^n, P_{Y|X}), \quad (1)$$

where $n \in \mathbb{N}$ is the blocklength. Given the input $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$, the output $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ is observed at the receiver with probability

$$P_{Y|X}(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n P_{Y|X}(y_t|x_t). \quad (2)$$

The message index to be sent by the transmitter to the receiver is a realization of a random variable W that is uniformly distributed over the set $\mathcal{W} = \{1, 2, \dots, M\}$ with $1 < M < \infty$. To send the message index within n channel uses, the transmitter uses an (n, M, λ) -code.

Definition 1 ((n, M, λ) -code). *Given $M \in \mathbb{N}$, a blocklength $n \in \mathbb{N}$, and $\lambda \in [0, 1]$, an (n, M, λ) -code for the random transformation in (1) is a system*

$$\left\{ \left(\mathbf{x}(1), \mathcal{D}(1) \right), \left(\mathbf{x}(2), \mathcal{D}(2) \right), \dots, \left(\mathbf{x}(M), \mathcal{D}(M) \right) \right\}, \quad (3)$$

where $\forall (j, l) \in \mathcal{W}^2$, with $j \neq l$:

$$\mathbf{x}(j) = (x_1(j), x_2(j), \dots, x_n(j)) \in \mathcal{X}^n, \quad (4a)$$

$$\mathcal{D}(j) \cap \mathcal{D}(l) = \emptyset, \quad (4b)$$

$$\bigcup_{j \in \mathcal{W}} \mathcal{D}(j) \subseteq \mathcal{Y}^n, \quad (4c)$$

$$\text{and } 1 - \frac{1}{M} \sum_{l=1}^M \Pr[\mathbf{Y} \in \mathcal{D}(j) | \mathbf{X} = \mathbf{x}(j)] \leq \lambda. \quad (4d)$$

The transmitter uses the codeword $\mathbf{x}(j)$ to transmit the message index $j \in \mathcal{W}$. At channel use t , $t \in \{1, 2, \dots, n\}$,

the transmitter inputs the symbol $x_t(j)$ to the channel. After n channel uses, the receiver observes the output $\mathbf{y} = (y_1, y_2, \dots, y_n)$ and determines that the symbol j was transmitted if $\mathbf{y} \in \mathcal{D}(j)$. The probability in (4d) is with respect to the conditional distribution $P_{Y|X}$ in (1) and λ is the upper bound on the average decoding error probability of the code in (3).

B. Dependence Testing Bound Expression

Using random coding with input distribution P_X , Polyanskiy *et al.* [2, Theo. 18] show the existence of an (n, M, λ) -code for the random transformation in (1) such that

$$\lambda \leq \lambda_{\text{DT}} = \lambda_{\text{DT},o} + \frac{M-1}{2} \lambda_{\text{DT},c} \quad (5)$$

where

$$\lambda_{\text{DT},o} \triangleq \mathbb{E}_{P_X P_{Y|X}} \left[\mathbb{1} \left\{ i(\mathbf{X}; \mathbf{Y}) \leq \ln \left(\frac{M-1}{2} \right) \right\} \right], \quad (6)$$

$$\lambda_{\text{DT},c} \triangleq \mathbb{E}_{P_X P_Y} \left[\mathbb{1} \left\{ i(\mathbf{X}; \mathbf{Y}) > \ln \left(\frac{M-1}{2} \right) \right\} \right], \quad (7)$$

$$\text{and } i(\mathbf{x}; \mathbf{y}) \triangleq \ln \left(\frac{dP_{Y|X=\mathbf{x}}}{dP_Y}(\mathbf{y}) \right) \quad (8)$$

where P_Y is the marginal probability distribution induced by P_X and $P_{Y|X}$. The function $\frac{dP_{Y|X=\mathbf{x}}}{dP_Y}(\cdot)$ is the Radon-Nikodym derivative of $P_{Y|X=\mathbf{x}}$ with respect to P_Y . In [2, Sec. III.C] λ_{DT} is known as the DT bound on the error probability.

The computation of $\lambda_{\text{DT},o}$, in (6), and $\lambda_{\text{DT},c}$, in (7), is often difficult to compute especially when the distribution of $i(\mathbf{X}; \mathbf{Y})$ is not known in closed form. This paper aims at deriving tractable bounds on λ_{DT} in (5) by bounding the error done when approaching the exact DT bound with the saddlepoint approximations for any blocklength n and a fixed number of message indices for i.i.d. channel inputs. That is, for all $\mathbf{x} \in \mathcal{X}^n$

$$P_X(\mathbf{x}) = \prod_{i=1}^n P_X(x_i). \quad (9)$$

This problem is nothing but a problem of probability approximation, that is formalized in the next section and for which a new bound is derived.

III. PRELIMINARIES

Let X be a random variable. The problem is to find a good approximation of the probability that $X \in \mathcal{A}$, denoted by $P_X(\mathcal{A})$, where $\mathcal{A} \subseteq \mathbb{R}$. The accuracy of the approximation of $P_X(\mathcal{A})$ is measured by the absolute approximation error given by

$$\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})} \triangleq \left| P_X(\mathcal{A}) - \hat{P}_X(\mathcal{A}) \right|. \quad (10)$$

Without loss of generality, let us consider the case where $P_X(\mathcal{A}) < 0.5^1$. The approximation is said to be accurate at order k , $\forall k \in \mathbb{R}$, when the following condition holds:

$$\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})} \leq 10^{-k} P_X(\mathcal{A}). \quad (11)$$

However, when $P_X(\mathcal{A})$ is not available (or at least hard to compute), it can be shown that if the following condition

$$\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})} \leq \frac{10^{-k}}{1 + 10^{-k}} \hat{P}_X(\mathcal{A}) \quad (12)$$

holds, then (11) is verified. In this paper, we propose an approximate $\hat{P}_X(\mathcal{A})$ and an upper bound on $\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})}$ relying on the saddlepoint technique.

The following definitions will be useful for the rest of the paper.

Definition 2. Given a probability distribution P_X , the moment generating function, denoted by φ_X , is defined as

$$\begin{aligned} \varphi_X : \mathbb{R} &\rightarrow \mathbb{R}_+ \\ \theta &\rightarrow \mathbb{E}_{P_X}[\exp(\theta X)]. \end{aligned} \quad (13)$$

The domain of convergence of φ_X , denoted by Θ_X , is defined by

$$\Theta_X = \{\theta \in \mathbb{R} : \varphi_X(\theta) < \infty\}. \quad (14)$$

Definition 3. Given a probability distribution P_X , for all $\theta \in \Theta_X$, the probability distribution $G_{X;\theta}$ is defined such that for all $x \in \mathbb{R}$

$$\frac{dG_{X;\theta}}{dP_X}(x) \triangleq \frac{\exp(\theta x)}{\varphi_X(\theta)}. \quad (15)$$

A. Existing results

Theorem 1 (Berry-Esseen [12]). Let Y_1, Y_2, \dots, Y_n be i.i.d random variables, $Y_1 \sim P_Y$ with

$$\begin{aligned} \mu &\triangleq \mathbb{E}_{P_Y}[Y_1], \quad \sigma^2 \triangleq \mathbb{E}_{P_Y}[|Y_1 - \mu|^2], \\ \gamma &\triangleq \mathbb{E}_{P_Y}[|Y_1 - \mu|^3]. \end{aligned} \quad (16)$$

Let X be the random variable

$$X = \sum_{j=1}^n Y_j \quad (17)$$

and Z a Gaussian random variable with mean $n\mu$ and variance $n\sigma^2$. Then,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{E}_{P_X}[\mathbb{1}_{\{X \leq x\}}] - \mathbb{E}_{P_Z}[\mathbb{1}_{\{Z \leq x\}}] \right| \leq \frac{c\gamma}{\sqrt{n}\sigma^3}, \quad (18)$$

with $c = 0.476$ [13].

Theorem 1 gives an upper bound on $\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})}$

$$\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})} \leq \frac{c\gamma}{\sqrt{n}\sigma^3} \quad (19)$$

for $\mathcal{A} = (-\infty, x]$, $x \in \mathbb{R}$ when the true distribution of $X = \sum_{j=1}^n Y_j$ with Y_j i.i.d. random variables, is approached by a Gaussian random variable. The drawback of the upper bound

¹This consideration does not remove any generality as the complement $(1 - P_X(\mathcal{A}))$ is considered when $P_X(\mathcal{A}) \geq 0.5$

in (19) is that it does not depend on \mathcal{A} . If $\hat{P}_X(\mathcal{A})$ is lower than the right hand side of (19), the accuracy of the bound (19) can not be assessed unless $P_X(\mathcal{A})$ can be computed.

IV. NEW RESULTS: BOUNDING ERROR OF SADDLEPOINT APPROXIMATIONS

A. Saddlepoint Approximations

Saddlepoint approximation is used to improve the approximation of the probability $P_X(\mathcal{A})$ when $\mathcal{A} = (a, \infty)$, $a \in \mathbb{R}$, $a \geq \mathbb{E}_{P_X}[X]$. For such \mathcal{A} and for all $x \in \mathbb{R}$, the following holds [14, Chapter 1, Section 1.4]

$$P_X(\mathcal{A}) = \mathbb{E}_{P_X}[\mathbb{1}_{\{X \in \mathcal{A}\}}] \quad (20)$$

$$= \mathbb{E}_{G_{X;\theta}} \left[\left(\frac{dG_{X;\theta}}{dP_X}(X) \right)^{-1} \mathbb{1}_{\{X \in \mathcal{A}\}} \right] \quad (21)$$

$$= \left(\frac{dG_{X;\theta}}{dP_X}(x) \right)^{-1} \mathbb{E}_{G_{X;\theta}} \left[\frac{dG_{X;\theta}}{dP_X}(x) \cdot \left(\frac{dG_{X;\theta}}{dP_X}(X) \right)^{-1} \mathbb{1}_{\{X \in \mathcal{A}\}} \right] \quad (22)$$

$$= \varphi_X(\theta) \exp(-\theta x) \mathbb{E}_{G_{X;\theta}} \left[\exp(\theta(x - X)) \cdot \mathbb{1}_{\{X \in \mathcal{A}\}} \right] \quad (23)$$

The saddlepoint approximation is about choosing $(\theta, x) \in \Theta_X \cdot \mathbb{R}$ such that

$$(\theta, x) = \underset{(t,w) \in \Theta_X \cdot \mathbb{R} : \forall v \in \mathcal{A}, \exp(t(w-v)) \leq 1}{\operatorname{argmin}} \varphi_X(t) \exp(-tw) \quad (24)$$

and approximating $G_{X;\theta}$ in (23).

The idea behind saddlepoint is to use another parametric probability distribution, i.e. $G_{X;\theta}$ (see Definition 3), such that $P_X(\mathcal{A})$ can be factorized in two terms of different magnitudes such that the smallest can be easily and exactly computed and the biggest fairly approximated. By factorizing $P_X(\mathcal{A})$ in such a manner, $\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})}$ can be made dependent on \mathcal{A} and then, avoiding the drawback of Berry-Esseen Theorem. The choice in (24) allows to observe that

$$\epsilon_{P_X(\mathcal{A}), \hat{P}_X(\mathcal{A})} \leq \min_{\substack{(t,w) \in \Theta_X \cdot \mathbb{R} : \\ \forall v \in \mathcal{A}, \exp(t(w-v)) \leq 1}} \varphi_X(t) \exp(-tw) \quad (25)$$

which is related to \mathcal{A} .

B. Main Result

The main contribution of this paper consists of giving explicit upper bound on the error of saddlepoint approximation when dealing with tail probability of the sum of i.i.d. random variables. The reason is that the average decoding error probability of any point-to-point memoryless channel can be upper bounded by a linear function of tail probabilities of a sum of i.i.d. random variables. The following Theorem gives our main result.

Theorem 2. Let Y_1, Y_2, \dots, Y_n be i.i.d. random variables in \mathbb{R} , $Y_1 \sim P_Y$ such that φ_Y in (13) exists and $a \in \mathbb{R}$ such that there exists $\theta \in \Theta_Y$ (defined in (14)), verifying

$$\frac{d}{d\theta} (\ln(\varphi_Y(\theta))) = \frac{a}{n}, \quad (26)$$

and μ_θ , V_θ , and ξ_θ defined as

$$\begin{aligned} \mu_\theta &\triangleq \mathbb{E}_{P_Y} \left[\frac{Y \exp(\theta Y)}{\varphi_Y(\theta)} \right], \\ V_\theta &\triangleq \mathbb{E}_{P_Y} \left[\frac{(Y - \mu_\theta)^2 \exp(\theta Y)}{\varphi_Y(\theta)} \right], \\ \xi_\theta &\triangleq \mathbb{E}_{P_Y} \left[\frac{|Y - \mu_\theta|^3 \exp(\theta Y)}{\varphi_Y(\theta)} \right] \end{aligned} \quad (27)$$

be finite. Let $X = \sum_{j=1}^n Y_j$ and $\mathcal{A} = (a, \infty)$. Then,

$$\begin{aligned} \widehat{P}_X(\mathcal{A}) &= \mathbb{1}\{\theta \leq 0\} + (-1)^{\mathbb{1}\{\theta \leq 0\}} \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \\ &\quad \exp\left(\frac{1}{2}n\theta^2 V_\theta\right) Q\left(|\theta|\sqrt{nV_\theta}\right), \end{aligned} \quad (28)$$

$$\epsilon_{P_X(\mathcal{A}), \widehat{P}_X(\mathcal{A})} \leq c \frac{2(\varphi_Y(\theta))^n}{\exp(\theta a)} \frac{\xi_\theta}{V_\theta^{3/2} \sqrt{n}} \quad (29)$$

where c is the same constant than in Theorem 1.

Sketch of proof. To obtain the result of Theorem 2

- choose x and θ in (23) such that

$$x = a, \quad \frac{d \ln(\varphi_Y(\theta))}{d\theta} = \frac{a}{n}. \quad (30)$$

- Approximate the probability distribution $G_{X;\theta}$ by the probability distribution P_Z of a Gaussian random variable Z with mean μ and variance σ^2 given by

$$\begin{aligned} \mu &\triangleq n \mathbb{E}_{P_Y} \left[\frac{Y \exp(\theta Y)}{\varphi_Y(\theta)} \right] \quad \text{and} \\ \sigma^2 &\triangleq n \mathbb{E}_{P_Y} \left[\frac{(Y - \mu_\theta)^2 \exp(\theta Y)}{\varphi_Y(\theta)} \right] \end{aligned} \quad (31)$$

- $X = \sum_{j=1}^n Y_j$, Y_j i.i.d random variables, observe that $\varphi_X(\theta) = (\varphi_Y(\theta))^n$.
- For $\theta > 0$ by solving (30), take

$$\begin{aligned} \widehat{P}_X(\mathcal{A}) &= (\varphi_Y(\theta))^n \exp(-\theta a) \\ &\quad \mathbb{E}_{P_Z} \left[\exp(\theta(a - Z)) \mathbb{1}_{\{Z \in \mathcal{A}\}} \right] \end{aligned} \quad (32)$$

and for $\theta \leq 0$, take

$$\begin{aligned} 1 - \widehat{P}_X(\mathcal{A}) &= (\varphi_Y(\theta))^n \exp(-\theta a) \\ &\quad \mathbb{E}_{P_Z} \left[\exp(\theta(a - Z)) \mathbb{1}_{\{Z \in \mathcal{A}^c\}} \right] \end{aligned} \quad (33)$$

where \mathcal{A}^c is the complement of \mathcal{A} in \mathbb{R} .

- Let $\delta_\theta(\mathcal{A})$ be defined by

$$\begin{aligned} \delta_\theta(\mathcal{A}) &= \left| \mathbb{E}_{G_{X;\theta}} \left[\exp(\theta(a - X)) \mathbb{1}_{\{X \in \mathcal{B}_\theta\}} \right] \right. \\ &\quad \left. - \mathbb{E}_{P_Z} \left[\exp(\theta(a - Z)) \mathbb{1}_{\{Z \in \mathcal{B}_\theta\}} \right] \right| \end{aligned} \quad (34)$$

where $\mathcal{B}_\theta = \mathcal{A}$ for $\theta > 0$ and $\mathcal{B}_\theta = \mathcal{A}^c$ for $\theta \leq 0$. Then,

$$\epsilon_{P_X(\mathcal{A}), \widehat{P}_X(\mathcal{A})} \leq \frac{(\varphi_Y(\theta))^n}{\exp(\theta a)} \delta_\theta(\mathcal{A}) \quad (35)$$

- using the following inequality

$$\delta_\theta(\mathcal{A}) \leq 2 \cdot \sup_{x \in \mathbb{R}} |G_{X;\theta}(X \leq x) - P_Z(Z \leq x)| \quad (36)$$

- Observe that for $X = \sum_{j=1}^n Y_j$, Y_j i.i.d random variables, $G_{X;\theta}$ is the distribution of the sum of i.i.d. random variables with distribution $G_{Y;\theta}$. Then, Berry-Esseen Theorem can be used to estimate the upper bound in (36).

This concludes the proof. \square

The result of (28) is not new since it appears in [4] in another form. However, the upper bound in (29) and the tools used to derive it constitute original results and our main contribution. Thanks to this result, bounds on $P_X(\mathcal{A})$ can be obtained which would have remained out of reach otherwise. Indeed, (28) does not give any further insights on its accuracy, except if $P_X(\mathcal{A})$ is available.

V. APPLICATION: BOUNDS ON THE DEPENDENCE TESTING BOUND

Consider λ_{DT} , $\lambda_{DT,o}$ and $\lambda_{DT,c}$ as in (5), (6) and (7) respectively. $\forall (\widehat{\lambda}_{DT}, \widehat{\lambda}_{DT,o}, \widehat{\lambda}_{DT,c}) \in [0, 1]^3$ and $(\epsilon_o, \epsilon_c) \in \mathbb{R}_+^2$ such that

$$|\lambda_{DT,o} - \widehat{\lambda}_{DT,o}| \leq \epsilon_o, \quad (37)$$

$$|\lambda_{DT,c} - \widehat{\lambda}_{DT,c}| \leq \epsilon_c, \quad (38)$$

$$\text{and } \widehat{\lambda}_{DT} \triangleq \min \left(1, \widehat{\lambda}_{DT,o} + \frac{M-1}{2} \widehat{\lambda}_{DT,c} \right). \quad (39)$$

Then,

$$\lambda_{DT} \leq \min \left(1, \widehat{\lambda}_{DT,o} + \epsilon_o + \frac{M-1}{2} (\widehat{\lambda}_{DT,c} + \epsilon_c) \right), \quad (40)$$

$$\begin{aligned} \lambda_{DT} &\geq \max \left(0, \widehat{\lambda}_{DT,o} - \epsilon_o \right) \\ &\quad + \frac{M-1}{2} \max \left(0, \widehat{\lambda}_{DT,c} - \epsilon_c \right), \end{aligned} \quad (41)$$

and

$$|\lambda_{DT} - \widehat{\lambda}_{DT}| \leq \epsilon_o + \frac{M-1}{2} \epsilon_c. \quad (42)$$

Equations (37) and (38) give the conditions of existence for the upper (40) and lower (41) bounds on DT bound λ_{DT} in (5). In the sequel of this section, two different expressions for $\widehat{\lambda}_{DT,o}$, $\widehat{\lambda}_{DT,c}$, ϵ_o , and ϵ_c , satisfying conditions in (37) and (38), are given: (a) the existing expression using the Berry-Esseen Theorem 1; and (b) new expressions using our Theorem 2.

A. Existing Bounds: Theorem 1

Proposition 1. Consider the upper bound on the decoding error probability λ_{DT} in (5) with respect to the random transformation (1) subject to (9). Then, $\hat{\lambda}_{\text{DT},o}$, $\hat{\lambda}_{\text{DT},c}$, ϵ_o , and ϵ_c below satisfy the conditions in (37) and (38).

$$\hat{\lambda}_{\text{DT},o} = 1 - Q\left(\frac{\ln\left(\frac{M-1}{2}\right) - n\mu}{\sqrt{n\sigma^2}}\right), \quad (43)$$

$$\epsilon_o = c \frac{\xi}{\sigma\sqrt{n}}, \quad (44)$$

$$\hat{\lambda}_{\text{DT},c} = \exp\left(\frac{1}{2}n\sigma^2 - n\mu\right) \cdot Q\left(\frac{\ln\left(\frac{M-1}{2}\right) - n\mu + n\sigma^2}{\sqrt{n\sigma^2}}\right), \quad (45)$$

$$\text{and } \epsilon_c = 2c \frac{\xi}{\sigma\sqrt{n}} \frac{2}{M-1}, \quad (46)$$

where

$$\mu \triangleq \mathbb{E}_{P_X P_Y|X} [i(X; Y)], \quad (47)$$

$$\sigma^2 \triangleq \mathbb{E}_{P_X P_Y|X} \left[(i(X; Y) - \mu)^2 \right], \quad (48)$$

$$\xi \triangleq \mathbb{E}_{P_X P_Y|X} \left[|i(X; Y) - \mu|^3 \right] \quad (49)$$

$$i(X; Y) \triangleq \ln \left(\frac{dP_{Y|X}}{dP_Y}(Y|X) \right), \quad (50)$$

and $c = 0.476$ is the constant of Berry-Esseen Theorem [13].

B. New Bounds: Theorem 2

Proposition 2. Consider the upper bound on the decoding error probability λ_{DT} in (5) with respect to the random transformation (1) subject to (9). Then, $\hat{\lambda}_{\text{DT},o}$, $\hat{\lambda}_{\text{DT},c}$, ϵ_o , and ϵ_c below satisfy the conditions in (37) and (38).

$$\hat{\lambda}_{\text{DT},o} = \mathbb{1}\{\theta > 0\} + \frac{(-1)^{\mathbb{1}\{\theta > 0\}} \varphi^n(\theta)}{\exp\left(\theta \ln\left(\frac{M-1}{2}\right)\right)}. \quad (51)$$

$$\epsilon_o = \frac{\exp(0.5\theta^2 n V_\theta) Q(\sqrt{n V_\theta} |\theta|),}{V_\theta^{3/2} \sqrt{n} \exp\left(\theta \ln\left(\frac{M-1}{2}\right)\right)} \frac{2c \cdot \xi_\theta}{\varphi^n(\theta)} \quad (52)$$

$$\hat{\lambda}_{\text{DT},c} = \mathbb{1}\{\theta + 1 \leq 0\} + \frac{(-1)^{\mathbb{1}\{\theta + 1 \leq 0\}} \varphi^n(\theta)}{\exp\left((\theta + 1) \ln\left(\frac{M-1}{2}\right)\right)}. \quad (53)$$

$$\text{and } \epsilon_c = \frac{\exp(0.5(\theta + 1)^2 n V_\theta) Q(\sqrt{n V_\theta} |\theta + 1|),}{V_\theta^{3/2} \sqrt{n} \exp\left((\theta + 1) \ln\left(\frac{M-1}{2}\right)\right)} \frac{2c \cdot \xi_\theta}{\varphi^n(\theta)} \quad (54)$$

where

$$\varphi(\theta) \triangleq \mathbb{E}_{P_X P_Y|X} [\exp(\theta \cdot i(X; Y))], \quad (55)$$

$$\mu_\theta \triangleq \mathbb{E}_{P_X P_Y|X} \left[\frac{i(X; Y) \exp(\theta \cdot i(X; Y))}{\varphi(\theta)} \right], \quad (56)$$

$$V_\theta \triangleq \mathbb{E}_{P_X P_Y|X} \left[\frac{|i(X; Y) - \mu_\theta|^2 \exp(\theta \cdot i(X; Y))}{\varphi(\theta)} \right], \quad (57)$$

$$\xi_\theta \triangleq \mathbb{E}_{P_X P_Y|X} \left[\frac{|i(X; Y) - \mu_\theta|^3 \exp(\theta \cdot i(X; Y))}{\varphi(\theta)} \right] \quad (58)$$

with θ chosen such that

$$\left. \frac{d}{dt} (\ln(\varphi(t))) \right|_{t=\theta} = \frac{\ln\left(\frac{M-1}{2}\right)}{n}. \quad (59)$$

and $c = 0.476$ is the constant of Berry-Esseen Theorem [13].

Proof: See Appendix

VI. NUMERICAL ANALYSIS

In this section, the upper and lower bounds, i.e. (40) and (41), as well as the approximation of the DT bound, i.e. (39), are compared when using Propositions 1 and 2. Throughout this section, uniform input distribution P_X with alphabet $\{1, -1\}$ is considered. Figures 1 and 2 considered a BSC with a cross probability $\delta = 0.11$ and information rate ($R = \log_2 M/n$) of 0.37 and 0.45 respectively. Curves in solid line, labeled NA, represent the upper and lower bounds and the approximation of the DT bound using Berry-Esseen Theorem and Proposition 1. Curves with the "star" marker are those obtained using our Theorem 2 and Proposition 2. One can remark that the approximation using the saddlepoint technique is always well bounded by our lower and upper bound using Proposition 2 while the normal approximation, i.e. using Berry-Esseen and Proposition 1, of the DT bound can be relatively far from its lower or upper bound, and may even diverge with the blocklength n . One may remark that some curves are missing, e.g. lower bound of NA in Fig. 1. We remind that the lower bound on the DT expression is obtained using (41) for which the best lower bound may be zero. In that case, there is no non-trivial lower bound obviously.

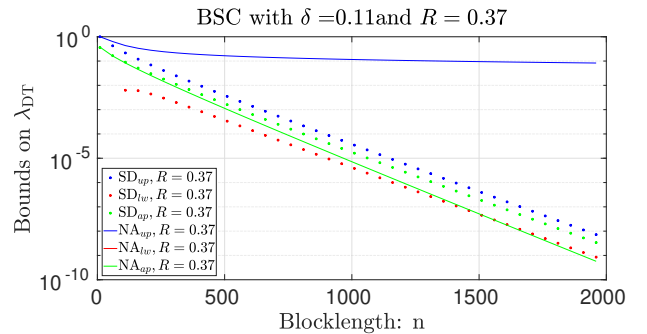


Fig. 1. Lower, upper bounds and approximations on λ_{DT} in BSC using Proposition 1, labeled NA, and Proposition 2, labeled SD, w.r.t. blocklength n for $R = 0.37$.

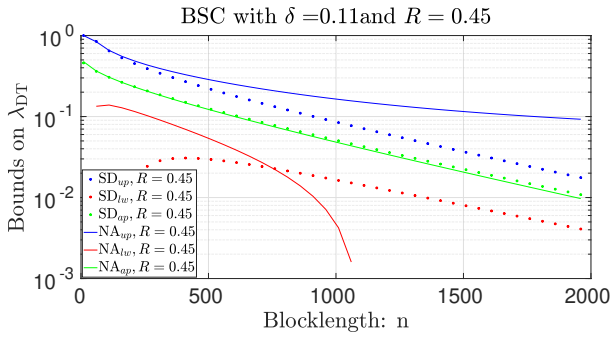


Fig. 2. Lower, upper bounds and approximations on λ_{DT} in BSC using Proposition 1, labeled NA, and Proposition 2, labeled SD, w.r.t. blocklength n for $R = 0.45$.

Figures 3 and 4 investigate the bounds and approximations of the DT bound in BI-AWGN channel with signal to noise ratio SNR = 1 and $R = 0.35$ and $R = 0.425$ respectively. The lower bound is not always available, e.g. Fig. 3, but our new upper bound is always better than the upper bound obtained with Berry-Esseen Theorem, making our approach based on the saddlepoint technique a valuable technique to approach the DT bound.

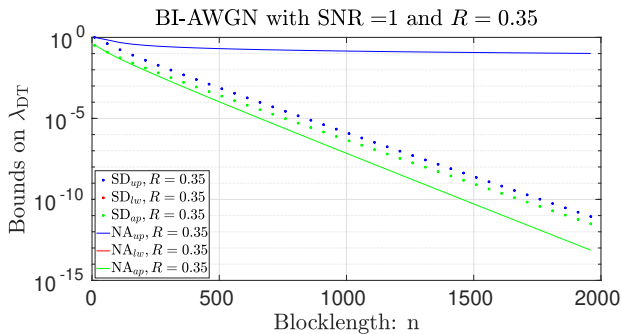


Fig. 3. Lower, upper bounds and approximations on λ_{DT} in BI-AWGN, SNR = 1, using Proposition 1, labeled NA, and Proposition 2, labeled SD, w.r.t. blocklength n for $R = 0.35$.

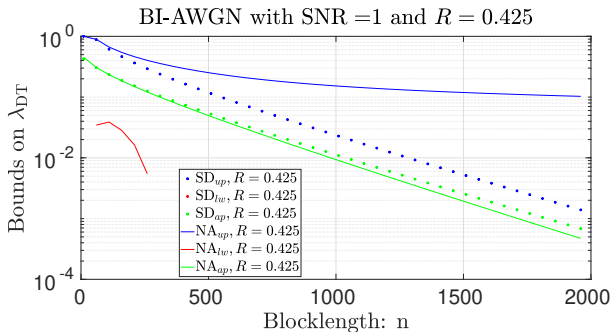


Fig. 4. Lower, upper bounds and approximations on λ_{DT} in BI-AWGN, SNR = 1, using Proposition 1, labeled NA, and Proposition 2, labeled SD, w.r.t. blocklength n for $R = 0.425$.

VII. CONCLUSION

The evaluation of the DT bound, as an upper bound of the exact message error probability in finite blocklength regime,

requires significant amount of computational resources and hence an approximation of the DT bound may be interesting. In this paper, a new bound on the absolute approximation error of saddlepoint approximation of tail probability of sum of i.i.d random variable is proposed. The bound on the error allows to bound the DT bound and it has been shown in particular that the upper bound obtained with the saddlepoint technique is always better than the upper bound obtained using the Berry-Esseen Theorem in BSC and BI-AWGN channels. These findings are interesting when investigating the achievable decoding error probability for any finite blocklength n by approaching the known exact bounds, e.g. DT bound, with an expression simpler to compute while handling the error committed. This approach will be particularly insightful in scenarios where the exact message error probability is unknown such as in the multi-user cases or non-Gaussian noise channel in finite blocklength regime.

ACKNOWLEDGMENT

This work has been (partly) funded by the French National Agency for Research (ANR) under grant ANR-16-CE25-0001 - ARBURST.

APPENDIX

Sketch of Proof of Proposition 2

Proof. The proof is based on Theorem 2. $\lambda_{DT,o}$ in (6) and $\lambda_{DT,c}$ in (7) are probability tails of the random variable $i(\mathbf{X}; \mathbf{Y})$ with respect to probability distributions $P_{\mathbf{X}}P_{\mathbf{Y}|\mathbf{X}}$ and $P_{\mathbf{X}}P_{\mathbf{Y}}$ respectively. For the random transformation in (1) subject to (9), $i(\mathbf{X}; \mathbf{Y})$ is given by

$$i(\mathbf{X}; \mathbf{Y}) = \ln \left(\frac{dP_{\mathbf{Y}|\mathbf{X}}}{dP_{\mathbf{Y}}}(\mathbf{Y}|\mathbf{X}) \right) \quad (60)$$

$$= \sum_{j=1}^n \ln \left(\frac{dP_{Y_j|\mathbf{X}}}{dP_{Y_j}}(Y_j|X_j) \right) \quad (61)$$

$$= \sum_{j=1}^n i(X_j; Y_j) \quad (62)$$

where $i(X_j; Y_j)$ are i.i.d. random variables. Then, $\hat{\lambda}_{DT,o}$ and ϵ_o are computed using Theorem 2 substituting X by $i(\mathbf{X}; \mathbf{Y})$, $P_{\mathbf{Y}}$ by $P_{\mathbf{X}\mathbf{Y}}$ and a by $\ln(\frac{M-1}{2})$. $P_{\mathbf{X}}(\mathcal{A})$ and $\hat{P}_{\mathbf{X}}(\mathcal{A})$ are respectively replaced by $(1 - \lambda_{DT,o})$ and $(1 - \hat{\lambda}_{DT,o})$ and ϵ_o is given by the right hand side of (29).

$\hat{\lambda}_{DT,c}$ and ϵ_c are computed similarly using Theorem 2 substituting $P_{\mathbf{Y}}$ by $P_{\mathbf{X}}P_{\mathbf{Y}}$. Moreover, $P_{\mathbf{X}}(\mathcal{A})$ and $\hat{P}_{\mathbf{X}}(\mathcal{A})$ are respectively replaced by $\lambda_{DT,c}$ and $\hat{\lambda}_{DT,c}$ and ϵ_c is given by the right hand side of (29).

Let U and V be the random variable $i(X; Y)$ under the distributions $P_{\mathbf{X}}P_{\mathbf{Y}|\mathbf{X}}$ and $P_{\mathbf{X}}P_{\mathbf{Y}}$ respectively. With the previous observations, we are left to compute φ_U, φ_V and ξ_θ in (27), where the notations $\xi_{1,\theta}$ and $\xi_{2,\theta}$ are used to distinguish ξ_θ for U and V respectively.

Using Definition 2

$$\varphi_U(\theta) \triangleq \mathbb{E}_{P_X P_{Y|X}} [\exp(\theta \cdot i(X; Y))] \quad (63)$$

$$= \mathbb{E}_{P_X P_Y} \left[\exp(\theta \cdot i(X; Y)) \frac{dP_{Y|X}}{dP_Y}(Y|X) \right] \quad (64)$$

$$= \mathbb{E}_{P_X P_Y} [\exp((\theta + 1) \cdot i(X; Y))] \quad (65)$$

$$= \varphi_V(\theta + 1). \quad (66)$$

Using (66), then, for θ_1 and θ_2 satisfying (26) the following equation holds

$$\theta_2 = \theta_1 + 1. \quad (67)$$

Finally,

$$\xi_{1,\theta_1} \triangleq \mathbb{E}_{P_X P_{Y|X}} \left[\frac{|Y - \mu_{1,\theta_1}|^3 \exp(\theta_1 \cdot i(X; Y))}{\varphi_U(\theta_1)} \right] \quad (68)$$

$$= \mathbb{E}_{P_X P_Y} \left[\frac{|Y - \mu_{1,\theta_1}|^3 \exp(\theta_1 \cdot i(X; Y))}{\varphi_U(\theta_1)} \cdot \frac{dP_{Y|X}}{dP_Y}(Y|X) \right] \quad (69)$$

$$= \mathbb{E}_{P_X P_Y} \left[\frac{|Y - \mu_{1,\theta_1}|^3 \exp((1 + \theta_1) \cdot i(X; Y))}{\varphi_U(\theta_1)} \right] \quad (70)$$

$$= \mathbb{E}_{P_X P_Y} \left[\frac{|Y - \mu_{2,\theta_2}|^3 \exp(\theta_2 \cdot i(X; Y))}{\varphi_V(\theta_2)} \right] \quad (71)$$

$$= \xi_{2,\theta_2} \quad (72)$$

with

$$\mu_{1,\theta} \triangleq \mathbb{E}_{P_X P_{Y|X}} \left[\frac{i(X; Y) \exp(\theta \cdot i(X; Y))}{\varphi_U(\theta)} \right] \quad (73)$$

$$= \mathbb{E}_{P_X P_Y} \left[\frac{i(X; Y) \exp(\theta \cdot i(X; Y))}{\varphi_U(\theta)} \frac{dP_{Y|X}}{dP_Y}(Y|X) \right] \quad (74)$$

$$= \mathbb{E}_{P_X P_Y} \left[\frac{i(X; Y) \exp((\theta + 1) \cdot i(X; Y))}{\varphi_V(\theta + 1)} \right] \text{ and} \quad (75)$$

$$= \mu_{2,\theta+1}. \quad (76)$$

Using (66), (67), and (72) in the relationship between $\hat{\lambda}_{DT,o}$, ϵ_o , $\hat{\lambda}_{DT,c}$, ϵ_c and $P_X(\mathcal{A})$, $\hat{P}_X(\mathcal{A})$ of Theorem 2, the proof of Proposition 2 is complete. \square

REFERENCES

- [1] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmcc: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.
- [2] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, pp. 2307–2359, December 2010.
- [3] T. Erseghe, "On the Evaluation of the Polyanskiy-Poor-Verdú Converse Bound for Finite Block-Length Coding in AWGN," *IEEE Transactions on Information Theory*, vol. 61, no. 12, 2015.
- [4] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint Approximations of Lower and Upper Bounds to the Error Probability in Channel Coding," in *52nd Annual Conference on Information Sciences and Systems (CISS)*, May 2018.
- [5] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *IEEE International Symposium on Information Theory (ISIT)*, 2012.
- [6] P. Moulin, "A New Metaconverse and Outer Region for Finite Blocklength MACs," in *Information Theory and Applications Workshop (ITA)*, February 2013.
- [7] V. Y. F. Tan and O. Kosut, "On the Dispersion of Three Network Information Theory Problems," *IEEE Transaction on Information Theory*, vol. 60, no. 2, pp. 1–184, 2014.
- [8] E. MolavianJazi and J. Laneman, "A Second-Order Achievable Rate Region for Gaussian Multi-Access Channels via a Central Limit Theorem for Functions," *IEEE Transactions on Information Theory*, vol. 61, pp. 6719–6733, December 2015.
- [9] A. Unsal and J.-M. Gorce, "The Dispersion of Superposition Coding for Gaussian Broadcast Channels," in *IEEE Information Theory Workshop 2017*, Kaohsiung, Taiwan, Nov. 2017. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01643260>
- [10] J. Liu, R. van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 943–947.
- [11] T. Erseghe, "Coding in the Finite-Blocklength Regime: Bounds Based on Laplace Integrals and Their Asymptotic Approximations," *IEEE Transactions on Information Theory*, vol. 62, no. 12, 2016.
- [12] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York, NY, USA: John Wiley and Sons, 1971, vol. 2.
- [13] I. Shevtsova, "On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands," *ArXiv e-prints*, Nov. 2011.
- [14] L. J. Jens, *Saddlepoint Approximations*. Clarendon press - Oxford, 1995.