New Non-asymptotic Random Channel Coding Theorems

En-hui Yang and Jin Meng

***Abstract*—New non-asymptotic random coding theorems (with error probability ϵ and finite block length n) based on Gallager parity check ensemble are established for binary input arbitrary output channels. The resulting non-asymptotic achievability bounds, when combined with non-asymptotic equipartition properties, can be easily computed. Analytically, these non-asymptotic achievability bounds are shown to be asymptotically tight up to the second order of the coding rate as n goes to infinity with either constant or sub-exponentially decreasing ϵ. Numerically, they are also compared favourably, for finite n and ϵ of practical interest, with existing non-asymptotic achievability bounds in the literature in general.**

***Index Terms*—Channel capacity, non-asymptotic coding theorems, non-asymptotic equipartition properties, random linear codes, Gallager parity check ensemble.**

I. Introduction

Recently, there have been great research interests in non-asymptotic channel coding theorems in information theory. By non-asymptotic coding theorems, we mean tight lower and upper bounds on the rate of certain codes or code ensembles in the regime of finite block length n (typically ranging from hundreds to thousands) and (word) error probability ϵ (typically ranging from 10^{-1} to 10^{-9}), which is loosely referred to hereafter as the non-asymptotic regime. For example, several non-asymptotic achievability bounds on Shannon random code ensemble have been reported in [1], which, coupled with non-asymptotic converse theorems therein, were shown to be very tight by numeric calculation in the non-asymptotic regime for some special channels such as a binary symmetric channel (BSC), a binary erasure channel (BEC), and an additive white gaussian noise (AWGN) channel.

Following [1], we are motivated in this paper to investigate if similar tight bounds are still valid for some structured ensembles and general memoryless channels with finite input alphabet and arbitrary output alphabet. Of particular interest is Gallager parity check ensemble [2], in which each element of the parity check matrix of a (linear) code is independently and uniformly generated from the finite field input alphabet. Note that for Gallager parity check ensemble, codewords are not pairwise independent, and therefore, bounding techniques on Shannon random code ensemble can not be applied in general.

Let $P = \{p(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ be a channel with binary input alphabet $\mathcal{X}$. The channel P is said to be memoryless binary-input output-symmetric (MBIOS) if the transition probability distribution of the channel satisfies $p(y|0) = p(-y|1)$ for any $y \in \mathcal{Y}$. In the literature, several non-asymptotic achievability bounds of linear codes have been developed for MBIOS channels. They more or less followed the approach invented by Gallager in [2]. Specifically, given a linear code $\mathcal{C}_n$ and a transmitted codeword c^n, the channel output space $\mathcal{Y}^n$ is divided into two parts $\mathcal{Y}_b^n$ (a bad region) and $\mathcal{Y}_g^n$ (a good region); the error probability (conditioned on the codeword c^n) then is bounded as follows

$$\begin{aligned} P_e(\mathcal{C}_n|c^n) \quad \leq \quad & \Pr\{Y^n \in \mathcal{Y}_b^n | X^n = c^n\} \\ & + \Pr\left\{\text{error}, Y^n \in \mathcal{Y}_g^n | X^n = c^n\right\}; \end{aligned} \tag{1.1}$$

and the union bound with respect to all codewords other than c^n is then applied to the second probability term. Using chernoff bounds [3], Gallager [2] then derived an achievability bound for any deterministic code of block length n with respect to its Hamming weight profile $\{N(l)\}_{i=1}^n$, where $N(l)$ is the number of codewords with Hamming weight l, and further showed that substituting $\{N(l)\}_{i=1}^n$ in this achievability bound with the average Hamming weight profile of Gallager parity check ensemble yields a bound equal to the Error Exponent bound for Shannon random code ensemble in [4], multiplied by a non-exponential term*. For some special MBIOS channels, analysis of those two probabilities in (1.1) can be further refined. Particularly, $\mathcal{Y}_b^n$ can be properly selected such that the exact calculation of the first probability is feasible for any finite block length, while for the second probability, the union bound can be applied conditioned on channel noise. Well known results along this line include those of Poltyrev [6] for a BSC and binary input additive Gaussian channel (BIAGC). For BSCs, it was shown in [1] that Poltyrev's bound on Gallager parity check ensemble turns out to be the tightest achievability bound in the non-asymptotic regime among all non-asymptotic achievabilities on BSCs in the literature. For BIAGCs, however, it was shown [6] that the corresponding bound (i.e., Tangential Sphere Bound (TSB)), applied to Gallager parity check ensemble, does not yield the same error exponent as that of Shannon random code ensemble (especially when the coding rate is close to Shannon capacity of the channel), and therefore would be expected to be worse than Error Exponent bound in the non-asymptotic regime. To the best of our knowledge, for general MBIOS channels, Error

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN203035-11, and by the Canada Research Chairs Program.

En-hui Yang and Jin Meng are with the Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: ehyang@uwaterloo.ca, j4meng@uwaterloo.ca

*This result on Gallager parity check ensemble was later enhanced by Shulman and Feder [5], who showed that the non-exponential term could be further eliminated.

Exponent bound remains the tightest achievability on Gallager parity check ensemble; it is also efficiently computable.

In this paper, a new non-asymptotic achievability bound is proved based on Gallager parity check ensemble, which are applicable to any binary input memoryless channel[†] (BIMC). For some special channels such as BSC and BEC, exact calculation of this bound is easy, which is shown (analytically and numerically) to be almost the same as Dependence Testing bound in [1]. For general BIMCs, especially those with continuous output such as BIAGC, the difficulty of calculation of this bound is mitigated by applying non-asymptotic equipartition property developed in [7] [8], and the resulted achievability bound can be efficiently evaluated for any BIMC. Asymptotic analysis then shows that this bound is tight up to the second order on any BIMC with certain symmetry, and numerical calculation on BIAGC shows that this bound is tighter than TSB and Error Exponent bound in the non-asymptotic regime.

The rest of the paper is organized as follows. Non-asymptotic coding theorems for Gallager parity check ensemble on BIMC and their asymptotic results are presented in Section II. Section III is devoted to comparison between our non-asymptotic achievabilities and existing results in the literature, and the conclusion is drawn in Section IV.

II. Non-asymptotic Coding Theorem for Gallager Parity Check Ensemble

In this section, we present non-asymptotic coding results for random linear codes of block length n based on Gallager parity check ensemble for any BIMC.

For an arbitrary BIMC $\{p(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ with $\mathcal{X} = \{0, 1\}$, denote its channel capacity as C_{BIMC} and define its linear capacity as

$$C_{\mathrm{BIMC-L}} = \ln 2 - H(X|Y)$$

where X is the uniform input random variable, and Y is the corresponding output of the BIMC. (Here and throughout the rest of the paper, information quantities such as entropy, conditional entropy, mutual information, and divergence (or relative entropy) are measured in nats, and $\ln$ stands for the logarithm with base e.) Let $p(y)$ be the pmf or pdf (as the case may be) of Y, and $p(x|y)$ the conditional pmf of X given by Y. It is easy to see that

$$p(y) = \frac{1}{2}[p(y|0) + p(y|1)]$$

and

$$p(x|y) = \frac{p(y|x)}{p(y|0) + p(y|1)} .$$

Let $\mathfrak{C}_{n,k}$ be a linear code with block length n and parity check matrix $\mathbf{H}_{(n-k)\times n}$. Assuming codewords are ordered in some manner, we shall refer to the q-th codeword in $\mathfrak{C}_{n,k}$ as $x^n(q)$. We say $\mathbf{H}_{(n-k)\times n}$ is randomly picked from Gallager parity check ensemble $\mathcal{H}_{n,k}$ if entries of $\mathbf{H}_{(n-k)\times n}$ are independently and uniformly generated from $\mathcal{X} = \{0, 1\}$. Denote the ensemble of linear codes with their parity check matrices from $\mathcal{H}_{n,k}$ by $\mathcal{C}_{n,k}^{(Gal)}$. To facilitate our subsequent discussion, we also specify the encoding procedure (i.e. the mapping from messages to codewords) of $\mathcal{C}_{n,k}^{(Gal)}$: given $\mathbf{H}_{(n-k)\times n}$, $x^n(q)$ is the q-th vector in the null space of $\mathbf{H}_{(n-k)\times n}$ by lexicographical order for $0 \le q \le 2^{n-rank(\mathbf{H}_{(n-k)\times n})} - 1$. By convention, we assume that all messages are equally likely. With slight abuse of notation, we shall use q to represent both the uniformly distributed random message and its specific realization; its exact meaning, however, will be clear from the context. Note that all codes in $\mathcal{C}_{n,k}^{(Gal)}$ have the channel coding rate greater than or equal to $\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}) \triangleq \frac{k}{n}\ln 2$ (in nats). The decoding procedure (named as *jar decoding*) is then specified as follows: given the channel output y^n, the decoder forms the set $J(y^n)$ (also called BIMC-L jar for convenience) as

$$\left\{ x^n \in \mathcal{X}^n : -\frac{\sum_{i=1}^n \ln \frac{p(y_i|x_i)}{p(y_i|0)+p(y_i|1)}}{n} \le H(X|Y) + \delta \right\}, \tag{2.1}$$

declares an error if no codeword is inside $J(y^n)$, and pick an arbitrary codeword in $J(y^n)$ to be the estimate of the transmitted codeword otherwise. (Note that the case when more than one codeword is inside $J(y^n)$ is considered a tie by the decoder, which is broken in an arbitrary way[‡].) It is easy to verify that

$$|J(y^n)| \le e^{n(H(X|Y)+\delta)} \tag{2.2}$$

for any y^n.

Further define

$$P_\delta \triangleq \Pr\left\{ -\frac{1}{n}\sum_{i=1}^n \ln p(X_i|Z_i) > H(X|Y) + \delta \right\} \tag{2.3}$$

where $X_1X_2\cdots X_n$ is an independently, identically and uniformly distributed sequence and $Z_1Z_2\cdots Z_n$ is the corresponding BIMC output.

Puncture 0 from the message space and ignore the insignificant effect on the rate, we have the following non-asymptotic coding theorem.

Theorem 1. *Given a BIMC with linear capacity $C_{\mathrm{BIMC-L}}$, let $P_e(\mathcal{C}_{n,k}^{(Gal)})$ denote the average word error probability (under jar decoding) of $\mathcal{C}_{n,k}^{(Gal)}$ with respect to the random message*

[†]Our new non-asymptotic achievability bound is also applicable to any memoryless channel with a finite field input alphabet. To facilitate our discussion, however, we choose to focus on the case of binary input alphabet when Gallager parity check ensemble is considered.

[‡]This decoding rule is closely related to Feinstein's threshold decoding. The difference lies in that when more than one codeword is inside jar or passes the threshold, the jar decoder treats the case as a tie, which is arbitrarily broken, while the threshold decoder will select the codeword with the lowest index. The reason for us to call this decoding rule jar decoding instead of modified threshold decoding is three fold: (1) it leads us to a philosophically different way to handle the second probability in (1.1), as discussed in Remark 1 and illustrated in the proof of Theorem 1; (2) it allows us to easily identify which probability in (1.1) is dominating, as discussed in Remark 4; and (3) by treating all codewords inside the jar equally, the decoder is not confined to solve any specific optimization problem, which, along with the flexibility of the formation of jar itself, we hope may lead one to look at practical decoding in a different way.

q, the BIMC, and the random linear code $\mathcal{C}_{n,k}^{(Gal)}$ itself. Then for any block length n and $\delta > 0$

$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \frac{1}{1-2^{-n}} P_\delta + e^{-n(C_{\mathrm{BIMC-L}} - \delta - \mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}))}. \tag{2.4}$$

Remark 1. The key idea of the proof of Theorem 1 is to bound the error probability (under jar decoding) in two parts

$$\begin{aligned} P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq & \Pr\left\{X^n(q) \notin J(Y^n)\right\} \\ & + \Pr\left\{\exists z^n \in J(Y^n) \cap \mathcal{C}_{n,k}^{(Gal)} / \{X^n(q)\}\right\}. \end{aligned}$$

Although this approach shares certain similarities with Gallager's proof technique illustrated in Section I, the key difference lies in that since all codewords inside the jar are treated equally, the second probability is handled by the union bound applied to all sequences inside $J(Y^n)$, instead of all codewords other than $X^n(q)$. Therefore, no symmetry of channel is required in our proof.

Remark 2. The purpose of puncturing $q = 0$ from the message space is to make the proof a little bit simpler. In fact, if we add $q = 0$ back, it only increases the error probability upper bound by $2^{-n\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)})}$. Moreover, when the channel has certain symmetry, i.e. $-\ln p(0|Y)$ given $X = 0$ and $-\ln p(1|Y)$ given $X = 1$ share the same distribution (we call such a channel a binary input memoryless symmetric channel (BIMSC)), punctuation of zero message is not necessary and the term $\frac{1}{1-2^{-n}}$ in (2.4) can be dropped. Note that the set of BIMSCs includes both MBIOS channels and weakly symmetric channels defined in [9] as a special case, and in the case of BIMSC, $C_{\mathrm{BIMSC}} = C_{\mathrm{BIMSC-L}}$ always holds.

Remark 3. The proof technique of Theorem 1 can be also applied to Shannon random code ensemble (with uniform input distribution) and Elias generator ensemble [10], in which the generator matrices of linear codes are generated in the same way as that for parity check matices in Gallager ensemble. In fact, the proof for those ensembles will be simpler, and the term $\frac{1}{1-2^{-n}}$ in (2.4) can be dropped.

As can be seen, the error probability bound in (2.4) is in a parametric form with respect to δ. In other words, given the block length n and the channel coding rate $\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)})$ (or equivalently k), (2.4) holds for any value of δ. And it is not hard to see that P_δ and $e^{-n(C_{\mathrm{BIMC-L}} - \delta - \mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}))}$ are respectively decreasing and increasing functions of δ. Consequently, there is an optimal δ which minimizes (2.4). For some special channels such as BSC and BEC, P_δ can be efficiently calculated for any δ, and therefore the optimization of (2.4) with respect to δ can be exactly solved. However, for other channels, especially those with continuous output (like BIAGC), it is extremely difficult to directly evaluate P_δ. To overcome this problem, tight upper and lower bounds on P_δ are established in [7] and [8]. By combining these bounds on P_δ with Theorem 1, we then derive an achievability bound of an analytic form. Towards this, some definitions are needed.

Let us temporarily drop the assumption that $\mathcal{X}$ is discrete and adopt the convention that $\int dx$ is interpreted as $\sum_{x\in\mathcal{X}}$ when $\mathcal{X}$ is discrete. Now given a random variable pair (X, Y) with distribution $p(x, y)$, let

$$\lambda^*(X|Y) \triangleq \sup\left\{\lambda \geq 0 : \iint p(y)p^{-\lambda+1}(x|y)dxdy < \infty\right\}.$$

Suppose that

$$\lambda^*(X|Y) > 0. \tag{2.5}$$

Define $r_{X|Y}(\delta)$ for any $\delta \geq 0$ as

$$\sup_{\lambda \geq 0}\left[\lambda\left(H(X|Y) + \delta\right) - \ln \iint p(y)p^{-\lambda+1}(x|y)dxdy\right]$$

and for $\lambda \in [0, \lambda^*(X|Y))$

$$f_\lambda(x, y) \triangleq \frac{p^{-\lambda}(x|y)}{\iint p(v)p^{-\lambda+1}(u|v)dudv}$$

$$\delta(\lambda) \triangleq \mathbb{E}\left[-\ln p(X_\lambda|Y_\lambda)\right] - H(X|Y)$$

$$\sigma_H^2(X|Y, \lambda) \triangleq \mathbb{VAR}\left[-\ln p(X_\lambda|Y_\lambda)\right]$$

$$M_H(X|Y, \lambda) \triangleq \mathbb{E}_3\left[-\ln p(X_\lambda|Y_\lambda)\right],$$

where (X_λ, Y_λ) follows the distribution $p(x, y)f_\lambda(x, y)$, and $\mathbb{E}$, $\mathbb{VAR}$ and $\mathbb{E}_3$ are expectation, variance and absolute third central moment operators on random variables. Denote $\sigma_H^2(X|Y, 0)$ by $\sigma_H^2(X|Y)$ and $M_H(X|Y, 0)$ by $M_H(X|Y)$. Let

$$\Delta^*(X|Y) \triangleq \lim_{\lambda \uparrow \lambda^*(X|Y)} \delta(\lambda)$$

and

$$\begin{aligned} \bar{\xi}_H(X|Y, \lambda, n) = & \frac{2C_{BE}M_H(X|Y, \lambda)}{\sqrt{n}\sigma_H^3(X|Y, \lambda)} \\ & + e^{\frac{n\lambda^2\sigma_H^2(X|Y,\lambda)}{2}}\left[Q(\sqrt{n}\lambda\sigma_H(X|Y, \lambda))\right. \\ & \left. - \; Q(\rho^* + \sqrt{n}\lambda\sigma_H(X|Y, \lambda))\right] \end{aligned} \tag{2.6}$$

where

$$Q(s) = \frac{1}{\sqrt{2\pi}} \int_s^\infty e^{-x^2/2} dx$$

$Q(\rho^*) = \frac{C_{BE}M_H(X|Y,\lambda)}{\sqrt{n}\sigma_H^3(X|Y,\lambda)}$, and $0 < C_{BE} < 0.4784$ is the universal constant in the Berry-Esseen central limit theorem [11].

Further assume that

$$\sigma_H^2(X|Y) > 0 \text{ and } M_H(X|Y) < \infty. \tag{2.7}$$

Let X be the uniform input random variable to the BIMC, and Y the corresponding output random variable of the BIMC. Combining Theorem 1 with non-asymptotic bounds on P_δ developed in [7] [8], we then get the following result.

Theorem 2. *Given a BIMC with $\sigma_H^2(X|Y) > 0$, $\lambda^*(X|Y) > 0$, and $M_H(X|Y) < \infty$ and any block length n, the following hold:*

1) *For any* $\delta \in (0, \Delta^*(X|Y))$

$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \left(\frac{1}{1-2^{-n}} + \lambda\right) \bar{\xi}_H(X|Y,\lambda,n) e^{-n r_{X|Y}(\delta)} \tag{2.8}$$

whenever

$$\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}) \leq C_{\text{BIMC-L}} - \delta - r_{X|Y}(\delta) + \frac{\ln \lambda \bar{\xi}_H(X|Y,\lambda,n)}{n} \tag{2.9}$$

where $\lambda = r'_{X|Y}(\delta)$.

2) *For any real number* c

$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \frac{1}{1-2^{-n}} Q\left(\frac{c}{\sigma_H(X|Y)}\right) + \frac{1}{\sqrt{n}}\left(\frac{C_{BE} M_H(X|Y)}{\sigma_H^3(X|Y)} + \frac{e^{-\frac{c^2}{2\sigma_H^2(X|Y)}}}{\sqrt{2\pi}\sigma_H(X|Y)}\right) \tag{2.10}$$

whenever

$$\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}) \leq C_{\text{BIMC-L}} - \frac{c}{\sqrt{n}} - \frac{\ln n}{2n} - \frac{\frac{c^2}{2\sigma_H^2(X|Y)} + \left[\ln \sqrt{2\pi}\sigma_H(X|Y)\right]}{n}. \tag{2.11}$$

Remark 4. Given the coding rate $\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)})$, the optimal δ is yielded by making

$$e^{-n(C_{\text{BIMC-L}} - \delta + \mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}))} \approx \lambda P_\delta$$

and

$$e^{-n(C_{\text{BIMC-L}} - \delta + \mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}))} \approx \frac{1}{\sqrt{n}} P_\delta$$

in part 1) and 2) of Theorem 2 respectively. In both cases,

$$P_\delta \gg e^{-n(C_{\text{BIMC-L}} - \delta + \mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}))}$$

for the optimal δ when $\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)})$ is close to $C_{\text{BIMC-L}}$. On the contrary, in Gallager's error exponent analysis illustrated in the introduction section, $\mathcal{Y}_b^n$ was chosen such that the first and second probabilities share the same exponent, for the sake of the tightness of error exponent. This difference, coupled with the fact that non-asymptotic bounds on P_δ in [7] [8] are tighter than chernoff bound, explains why our achievability can be tighter than Error Exponent bound in the non-asymptotic regime. Another advantage of applying non-asymptotic bounds on P_δ is that we do not have to choose $J(Y^n)$ for the sake of easy computation of P_δ, which explains why our achievability can be tighter than TSB on BIAGC.

Remark 5. The inequalities (2.10) and (2.11) show that if the word error probability is kept slightly above 0.5, the code rate can be even slightly above the capacity of the BIMC with $C_{\text{BIMC}} = C_{\text{BIMC-L}}$! Figure 1 shows the tradeoff between the word error probability and block length when the code rate is 0.21% above the capacity for the BSC with cross-over probability $p = 0.12$, where in Figure 1, both the capacity and code rate are expressed in terms of bits. As can be seen from Figure 1, at the block length 1000, the word error probability is around 0.65, and the code rate is 0.21% above the capacity! Although this phenomenon has been implied by the second order analysis of the coding rate as n goes to ∞ [1], [12]–[15] , the inequalities (2.10) and (2.11) allow us to demonstrate this for specific values of n and for random linear codes based on Gallager parity check ensemble.

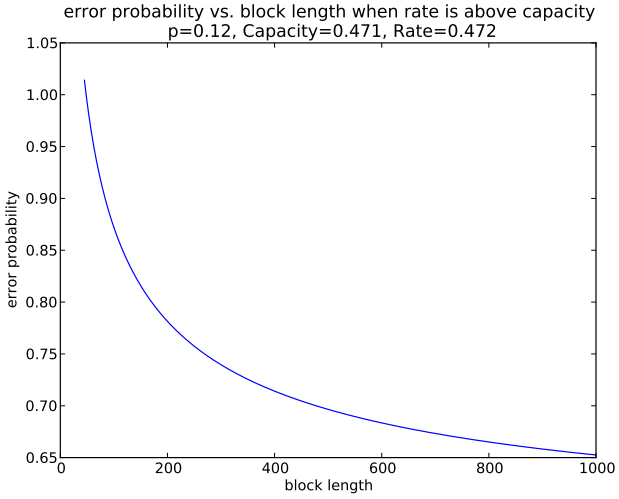


Fig. 1. Tradeoff between the word error probability and block length when the code rate is above the capacity with $p = 0.12$.

Remark 6. Parts 1) and 2) of Theorem 2 both provide non-asymptotic achievability bounds on the error probability and coding rate of Gallager's ensemble, which begs a comparison between them. It turns out that given block length, either of

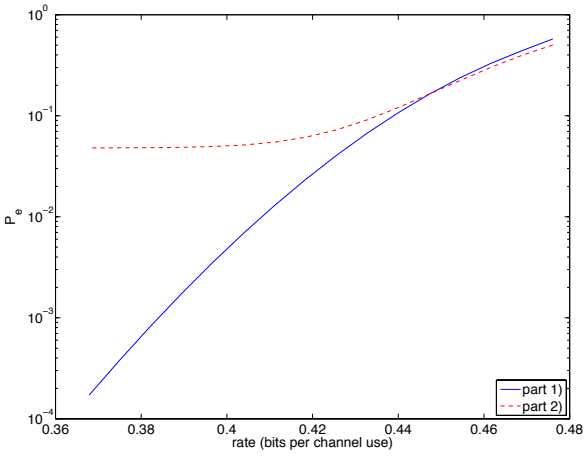


Fig. 2. Part 1) vs Part 2) of Theorem 2 on BIAGC with block length $n = 1000$ and snr=0dB

those achievability bounds can be tighter than the other for different coding rate regions. When the coding rate is above capacity, part 1) is not applicable, while part 2) can still bound the error probability strictly lower than 1, shown in the above

discussion. However, when the coding rate is below capacity, part 1) will be tighter than part 2) as long as the coding rate is not too close to the channel capacity. A numeric comparison between part 1) and part 2) is shown in Figure 2 for BIAGC with block length 1000 and snr 0dB, where the coding rate is kept less than the channel capacity ≈ 0.4847 (bits per channel use). As can be seen, when the coding rate is moving away from the channel capacity, part 1) becomes much tighter.

Although our focus in this paper is on non-asymptotic coding theorems, it is instructive to see how tight our achievability bounds in Theorem 2 are asymptotically as n goes to ∞. Then we get the following asymptotic result.

Corollary 1. *Given a BIMC with $\sigma_H^2(X|Y) > 0$, $\lambda^*(X|Y) > 0$, and $M_H(X|Y) < \infty$, let $\delta_n = \frac{\sigma_H(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon_n)$ for $0 < \epsilon_n < 1$. Suppose $\frac{-\ln \epsilon_n}{n} = o(1)$ as $n \to +\infty$. Then we have*

$$\mathcal{R}(\mathcal{C}_{n,k}^{(Gal)}) \geq C_{\mathrm{BIMC-L}} - \delta_n - o(\delta_n) \tag{2.12}$$

while $P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \epsilon_n$.

Remark 7. Given a BIMSC, results in [1], [12]–[17] imply that C_{BIMSC} and $-\delta_n$ are the first and second order of the best coding rate that can be achieved by any code when the error probability is a constant or sub-exponentially decreasing with respect to n. Corollary 1 shows that the optimal first and second order coding performance can be achieved by Gallager ensemble under jar decoding as well. This in turn implies that the achievability bounds in Theorem 2 are asymptotically tight as n goes to ∞ with either a constant or sub-exponentially decreasing error probability with respect to n.

III. Comparison with Existing Non-Asymptotic Achievability

Although there are tremendous achievable bounds [18], [19] (and references therein) on channel coding rate in the prosperous literature of information theory, where various code ensembles and bounding techniques are used, it does not seem that any of our random coding theorems (Theorems 1 and 2) could be implied by existing achievability bounds in the literature because of either the generality of our channel models or the special structure of our random code ensembles in our random coding theorems. For example, Theorems 1 and 2 are concerned with Gallager parity check ensemble, wherein codewords are not necessarily pairwise independent, and applicable to any binary input memoryless channel without any symmetry constraint whatsoever. On the other hand, most achievability bounds on linear block codes are for binary input memoryless channels with symmetry [18]. Nonetheless, it is instructive to compare our achievability bounds in Theorems 1 and 2, with existing bounds in the literature whenever possible.

Random linear code ensembles include Elias generator ensemble and Gallager parity check ensemble. While codewords generated in Elias ensemble are pairwise independent, it is not true for Gallager ensemble. Consequently, non-asymptotic coding theorems on Shannon random code ensemble in the literature, whose proof relies on pairwise independence of codewords, apply only to Elias ensemble, but not to Gallager ensemble. Here we focus on those achievabilities applicable to random linear code ensembles, with the emphasis on Gallager ensemble. Furthermore, as some achievability bounds are only applicable to special channels, we divide our discussion into four parts: 1) bounds for BSC; 2) bounds for BEC; 3) bounds for BIAGC; and 4) bounds for memoryless binary-input output-symmetric (MBIOS) channels.

A. BSC

To make comparison transparent, we rewrite Theorem 1. Let $M = 2^k$ be the number of codewords, and $p \in (0, 0.5)$ be the crossover probability. By (2.4) in Theorem 1 and Remark 2, it is not hard to verify that

$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \underbrace{\sum_{n\left(p+\frac{\delta}{\ln\frac{1-p}{p}}\right)<w\leq n} \binom{n}{w} p^w (1-p)^{n-w}}_{\Pr\{X^n \notin J(Y^n)\}} + \sum_{0\leq w\leq n\left(p+\frac{\delta}{\ln\frac{1-p}{p}}\right)} \binom{n}{w} 2^{-n} M. \tag{3.1}$$

Further optimizing δ implies that

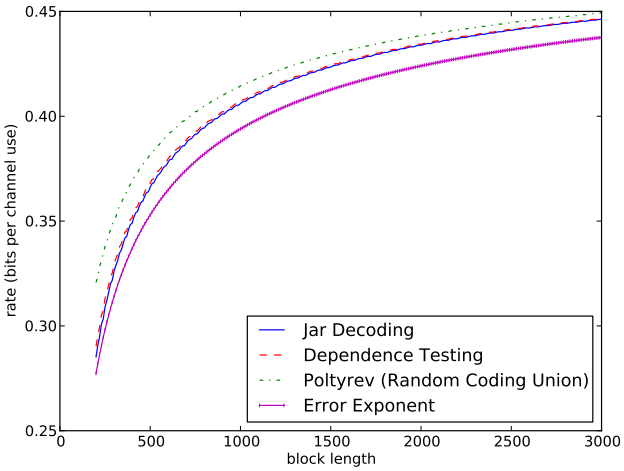
$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \sum_{w=0}^{n} \binom{n}{w} \min\left\{p^w(1-p)^{n-w}, 2^{-n}M\right\} \tag{3.2}$$

and (3.2) is essentially the same (except for a minor difference[§]) as the Dependence Testing Bound recently established in [1, Theorem 34] for Shannon random code ensemble and Elias ensemble over the BSC.

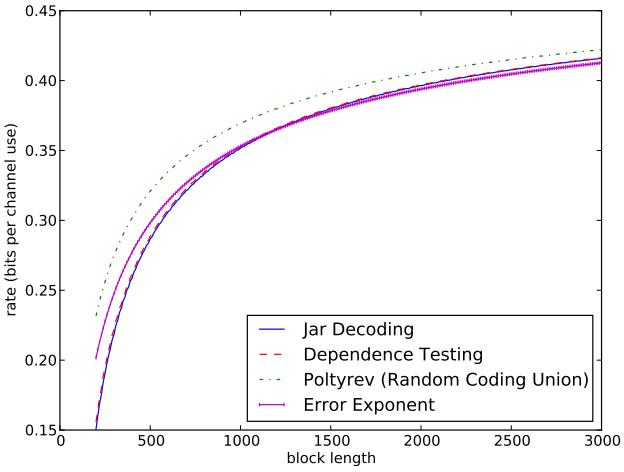
As discussed in the introduction section, Poltyrev derived an achievability bound for any deterministic code in terms of its hamming weight profile $\{N(l)\}_{l=1}^n$ on BSC, and by replacing $N(l)$ with $2^{-(n-k)}\binom{n}{l}$, the resulted bound holds for Gallager ensemble $\mathcal{C}_{n,k}^{(Gal)}$, as well as Elias ensemble. In addition, it was shown that Random Coding Union Bound [1, Theorem 33] derived for Shannon random code ensemble and Elias ensemble is the same as Poltyrev's bound.

Figure 3 shows the numeric comparison (with block length range $[200, 3000]$ and fixed word error probability 10^{-3} and 10^{-6}) among Theorem 1, Poltyrev's Bound [6, Lemma 1] (Random Coding Union Bound [1, Theorem 33]) and Error Exponent Bound on a BSC with cross-over probability $p = 0.11$, where Dependence Testing Bound [1, Theorem 34] is also included for a benchmark. As can be seen, the numeric result confirms that Theorem 1 is essentially the same as Dependence Testing Bound and further shows that Poltyrev's Bound (Random Coding Union Bound) is better than Dependence Testing Bound and Theorem 1 by a small margin, while Dependence Testing Bound and Theorem 1 outperform Error Exponent Bound when word error probability is relatively

[§]Replacing M in (3.2) by $(M-1)/2$ yields exactly the Dependence Testing Bound [1, Theorem 34].



(a) $P_e = 10^{-3}$



(b) $P_e = 10^{-6}$

Fig. 3. Comparison of Achievability for BSC with cross-over probability $p = 0.11$

large with respect to block length, which is consistent with the observation in [1].

B. BEC

Now let us focus on BEC. In this case, Theorem 1 on BEC can be further improved as follows. Let $M = 2^k$ be the number of codewords and p be the erasure probability. It is then easy to verify that

$$H(X|Y) = p \ln 2$$

and in this case, the BIMC-L jar reduces to

$$J(y^n) = \{x^n : x_i = y_i \text{ if } y_i \neq e\}$$

if $|\{i : y_i = e\}| \leq n\left(p + \frac{\delta}{\ln 2}\right)$, and an empty set otherwise. Following the argument in the proof of Theorem 1, it is not hard to show that

$$\begin{aligned} & P_e(\mathcal{C}_{n,k}^{(Gal)}) \\ & \leq \underbrace{\sum_{n(p+\frac{\delta}{\ln 2})<t\leq n} \binom{n}{t} p^t(1-p)^{n-t}}_{\Pr\{X^n(q)\notin J(Y^n)\}} \\ & \quad + \Pr\left\{\exists z^n \in J(Y^n) \cap \mathcal{C}_{n,k}^{(Gal)}/\{X^n(q)\}\right\} \\ & \leq \sum_{n(p+\frac{\delta}{\ln 2})<t\leq n} \binom{n}{t} p^t(1-p)^{n-t} \\ & \quad + \sum_{1\leq t\leq n(p+\frac{\delta}{\ln 2})} \binom{n}{t} p^t(1-p)^{n-t}2^{t-n}M \end{aligned} \tag{3.3}$$

and optimizing δ yields

$$P_e(\mathcal{C}_{n,k}^{(Gal)}) \leq \sum_{t=1}^{n} \binom{n}{t} p^t(1-p)^{n-t}2^{-[n-t-\log_2 M]^+} \tag{3.4}$$

which is again essentially the same (except for a minor difference[¶]) as the Dependence Testing Bound [1, Theorem 37] for Shannon random code ensemble and Elias generator ensemble. Note that $\frac{1}{1-2^{-n}}$ in Theorem 1 is dropped here according to Remark 2.

For BECs, Ashikmin derived an expression for word error probability of full rank Elias ensemble (i.e. the generator matrix is equiprobably selected among all full rank matrices), included as Theorem 6 in [1]. Figure 4 shows the numeric comparison among (3.4), Ashikmin's Bound, Error Exponent Bound, and Dependence Testing Bound [1, Theorem 37]. Once again, our achievability is very close to Dependence Testing Bound, outperforms Error Exponent Bound, and is worse than Ashikmin's Bound (the best achievability under ML decoding known so far) by a small margin.

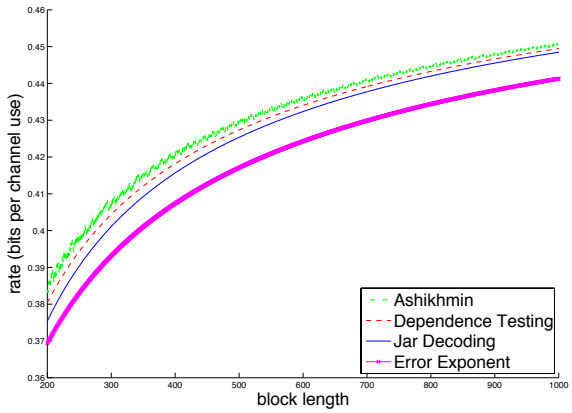
C. BIAGC

Since in this case, there is no feasible way to calculate P_δ, we apply part 1) of Theorem 2, where $\frac{1}{1-2^{-n}}$ in (2.8) can be replaced by 1 due to Remark 2.

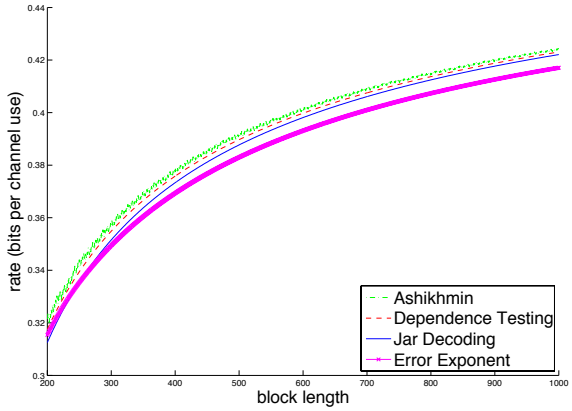
There is a rich literature about error probability bounds of linear codes for BIAGC, considering the practical impact of this research topic. One of the tightest bounds in this research area is the Tangential Sphere Bound (TSB), proved by Poltyrev in [6]. TSB was then improved by Yousefi and Khandani in [20], and Mehrabian and Yousefi in [21]. It is unclear, however, whether those two improved bounds can be efficiently evaluated for Gallager parity check ensemble.

Although TSB is one of the tightest bounds for any deterministic code in terms of its hamming weight profile on BIACG, it fails to reproduce the Gallager error exponent ([18] and references therein) for Gallager parity check ensemble.

[¶]Replacing M by $(M-1)/2$, and then starting the summation from $t = 0$ instead of $t = 1$ in (3.4) yield exactly the Dependence Testing Bound [1, Theorem 37].



(a) $P_e = 10^{-3}$



(b) $P_e = 10^{-6}$

Fig. 4. Comparison of Achievability for BEC with erasure probability $p = 0.5$

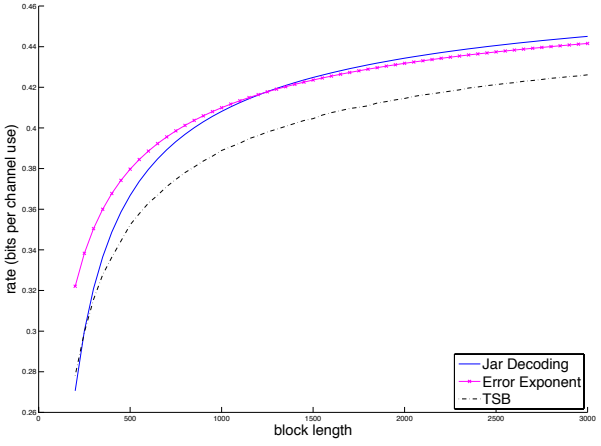


Fig. 5. Comparison of Achievability for BIAGC with snr 0dB and word error probability $P_e = 10^{-2}$

On the other hand, Error Exponent Bound holds for Gallager parity check ensemble on BIAGC, due to that BIAGC is MBIOS. Figure 5 shows numerical comparison among part 1) of Theorem 2 ((2.8) and (2.9)), TSB, and Error Exponent Bound, where the signal-to-noise ratio (snr) is 0dB and the word error probability is kept to be 10^{-2}. As can be seen, TSB is worse than Error Exponent Bound, while our achievability is better than Error Exponent Bound in certain block length region. To the best knowledge of authors, this is the first numeric demonstration in the literature that Error Exponent Bound can be beaten in the non-asymptotic regime on BIAGC as well.

D. General MBIOS Channel

The only existing achievability bound in the literature applicable to this general case is Error Exponent Bound for Gallager ensemble, as well as Elias ensemble. The symmetry property of MBIOS channel is essential to the proof of Error Exponent Bound for Gallager ensembles. As demonstrated already, our achievabilities in Theorems 1 and 2, applicable to any BIMC, can be tighter than Error Exponent Bound in the non-asymptotic regime.

E. Summary

Applicability (to ensembles and channels) and computational complexity of jar decoding achievability and existing achievability bounds on random linear code ensembles in the literature are summarized in Table I, where by unknown, we means that at this point we are not aware of any method which can be used to effectively compute the corresponding bound. Among all the listed results, Theorem 2 is the only achievability that can be applied to general BIMC and efficiently evaluated. Focusing on Gallager ensemble, existing achievability bounds only deal with MBIOS channels, which are a strict subset of BIMC. For some special MBIOS channels, e.g. BSC and BEC, there are bounds proved under ML decoding, which are better than our achievability in (3.2) and (3.4) by a small margin in the non-asymptotic regime. For general MBIOS channels, however, to the our best knowledge, Error Exponent Bound was the best computable achievability result in the literature before this paper. And numerical calculation shows that the achievability bound in Theorem 2 can be tighter than Error Exponent Bound in the non-asymptotic regime.

IV. Conclusion

New non-asymptotic achievabilities for random structured code ensembles, specifically Gallager parity check ensemble, are derived in this paper. It is then shown that our achievabilities are asymptotically tight up to the second order of the coding rate as the block length n goes to infinity. Compared to existing non-asymptotic achievability bounds in the literature, our achievabilities are tight and easy to compute in general. In particular, numeric evaluation demonstrates that our achievability on Gallager parity check ensemble is the tightest achievability result known so far in some non-asymptotic regime on binary input additive gaussian channel.

Achievability Bounds		Applicability		Computational Complexity
		Linear Code Ensembles	BIMC	
Jar Decoding	(3.2)	$\surd$ Elias $\surd$ Gallager	BSC	$O(n)$
	(3.4)		BEC	$O(n)$
	Theorem 2		General	$O(1)$
Poltyrev	[6, Lemma 1]	$\surd$ Elias $\surd$ Gallager	BSC	$O(n)$
Ashikmin	[1, Theorem 6]	$\surd$ Elias (full rank) $\times$ Gallager	BEC	$O(n^2)$
TSB	[6, Lemma 4]	$\surd$ Elias $\surd$ Gallager	BIAGC	$O(1)$
Error Exponent	[5]	$\surd$ Elias $\surd$ Gallager	MBIOS	$O(1)$
Random Coding Union	[1, Theorem 33]	$\surd$ Elias $\times$ Gallager	BSC	$O(n)$
	[1, Theorem 16]		General	Unknown
Dependence Testing	[1, Theorem 34]	$\surd$ Elias $\times$ Gallager	BSC	$O(n)$
	[1, Theorem 37]		BEC	$O(n)$
	[1, Theorem 17]		General	Unknown

TABLE I
ACHIEVABILITY BOUNDS OF RANDOM LINEAR CODES FOR BIMC

REFERENCES

[1] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, may 2010.

[2] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[3] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, vol. 23, pp. 493–507, 1952.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968.

[5] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *Information Theory, IEEE Transactions on*, vol. 45, no. 6, pp. 2101 –2104, sep 1999.

[6] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *Information Theory, IEEE Transactions on*, vol. 40, no. 4, pp. 1284 –1292, jul 1994.

[7] E.-H. Yang and J. Meng, "Non-asymptotic equipartition properties for independent and identically distributed sources." [Online]. Available: http://arxiv.org/abs/1204.3661

[8] ——, "New non-asymptotic random channel coding theorems," submitted to *IEEE Trans. on Inform. Theory*, 2011.

[9] T.-M. Cover and J.-A. Thomas, *Elements of Information Theory (second edition)*. Hoboken, NJ: Wiley, 2006.

[10] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[11] V. Korolev and I. Shevtsova, "An improvement of the berryesseen inequality with applications to poisson and mixed poisson random sums," *Scandinavian Actuarial Journal*, pp. 1–25, 2010. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03461238.2010.485370

[12] V. Strassen, "Asymptoticsche abschätzugen in shannon's informationstheorie," in *Proc. 3rd Conf. Inf. Theory*, Prague, Czech Republic, 1962, pp. 689–723.

[13] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 4947–4966, nov. 2009.

[14] E. H. Yang and J. Meng, "Channel capacity in the non-asymptotic regime: Taylor-type expansion and computable benchmarks," in *Proc. of Allerton'2012*, 2012.

[15] E.-H. Yang and J. Meng, "Jar decoding: Non-asymptotic converse coding theorems, taylor-type expansion, and optimality," submitted to *IEEE Trans. on Inform. Theory*, 2012. [Online]. Available: http://arxiv.org/abs/1204.3658

[16] Y. Altug and A. Wagner, "Moderate deviations in channel coding." [Online]. Available: http://arxiv.org/abs/1208.1924

[17] Y. Polyanskiy and S. Verdu, "Channel dispersion and moderate deviations limits for memoryless channels," in *Proceedings of Allerton'2010*, 2010, pp. 1334–1339.

[18] I. Sason and S. Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial." *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1/2, 2006.

[19] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton, 2010.

[20] S. Yousefi and A. Khandani, "A new upper bound on the ml decoding error probability of linear binary block codes in awgn interference," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3026 – 3036, dec. 2004.

[21] A. Mehrabian and S. Yousefi, "Improved tangential sphere bound on the ml decoding error probability of linear binary block codes in awgn and block fading channels," *Communications, IEE Proceedings-*, vol. 153, no. 6, pp. 885 –893, dec. 2006.