New Channel Coding Achievability Bounds

Yury Polyanskiy
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: ypolyans@princeton.edu

H. Vincent Poor
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: poor@princeton.edu

Sergio Verdú
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*—Three essentially different approaches to the constructive part of the channel coding theorem have been proposed by Shannon, Feinstein and Gallager, respectively, leading to upper bounds on the minimal error probability achievable with a given rate and blocklength. Here, new upper bounds are given on both average and maximal error probability, which are tighter than existing bounds for many ranges of blocklength and channel parameters of interest. Along with converse bounds, the new achievability bounds allow to approximate tightly the maximum rate achievable for a given blocklength and error probability for blocklengths as short as $n = 200$ for both the BSC and the BEC.**

I. Introduction

The proof of the channel coding theorem involves three stages:

- *Converse:* an upper bound on the size of any code with given arbitrary blocklength and error probability.
- *Achievability:* a lower bound on the size of a code that can be guaranteed to exist with given arbitrary blocklength and error probability.
- *Asymptotics:* the bounds on the log size of the code normalized by blocklength asymptotically coincide as a result of the law of large numbers or the ergodic theorem (for channels with memory).

As propounded in [1], it is pedagogically sound to separate clearly the third stage from the derivation of the upper and lower bounds. The bounds need not impose assumptions on the channel such as memorylessness, stationarity, and ergodicity and they can be extremely useful in assessing the highest rate that can be achieved when operating with a given blocklength and error probability.

Three fundamentally different achievability approaches have been proposed:

- Shannon [2]: Random coding with suboptimal decoding.
- Feinstein [3]: Greedy coding with suboptimal decoding. For blocklength n and maximal error probability ϵ, a code is shown to exist with M codewords such that

$$M \geq \sup_{\rho>0} \left\{ \rho\,\epsilon - \rho \inf_{P_{X^n}} \mathbb{P}\left[\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \rho \right] \right\} \quad (1)$$

  A formalization of the approach in [2] leads to the same bound under the weaker average error probability criterion.

The research was supported by the National Science Foundation under Grants ANI-03-38807, CCF-06-35154 and CNS-06-25637.

- Gallager [4]: Random coding with maximum likelihood decoding. A code is guaranteed to exist with blocklength n, average error probability ϵ, and M codewords such that [4], [5, Section 5.6, Example 1]

$$\epsilon \leq \inf_{\rho>0} M^{\rho} \inf_{P_{X^n}} \sum_{b^n \in \mathcal{B}^n} \left(\sum_{a^n \in \mathcal{A}^n} P_{X^n}(a^n) P_{Y^n|X^n}^{1/(1+\rho)}(b^n|a^n) \right)^{1+\rho} \quad (2)$$

In this paper we find a new upper bound on the average error probability of a code with rate R and blocklength n, which in a weakened succinct version becomes

$$\epsilon \leq \inf_{P_{X^n}} \mathbb{E} \exp\left\{ -\left(\frac{1}{n} i(X^n, Y^n) - R \right)^+ \right\}, \quad (3)$$

where the information density is defined as

$$i(x, y) = \log \frac{P_{Y|X}(y|x)}{P_Y(y)}\,. \quad (4)$$

We also show that (3) is also satisfied by the maximal error probability for several classes of channels. In addition, for general classes of channels we obtain a different upper bound on maximal error probability by following Feinstein's approach of greedy coding and suboptimal decoding, but imposing additional constraints on the freedom with which the sequential build up of the codebook proceeds. As we illustrate with the binary symmetric channel and the binary erasure channel, the new bounds are tighter than the previous ones for large ranges of blocklength, rate and channels parameters.

The tightness of the bounds is illustrated by the binary symmetric channel with crossover probability equal to 0.11 (capacity = 0.5): the maximum rate that can be achieved with codes with blocklength $n = 500$, and maximal error probability not exceeding 10^{-3} is shown to be between 0.36 and 0.39. For the binary erasure channel with erasure probability equal to 0.5 (capacity = 0.5): the maximum rate that can be achieved with codes with blocklength $n = 200$, and maximal error probability not exceeding 10^{-3} is shown to be between 0.38 and 0.4.

The bounds validate the use of Strassen's normal approximation to the maximal achievable rate [6] of discrete memoryless channels without cost constraints.

II. Average Probability of Error

Let us introduce measurable spaces of inputs A and outputs B and a conditional probability measure $P_{Y|X} : \mathsf{A} \mapsto \mathsf{B}$. We denote a codebook with M codewords by $\{c_1, \ldots, c_M\} \subset \mathsf{A}$. A (possibly randomized) decoder is a random transformation $P_{Z|Y} : \mathsf{B} \mapsto \{0, 1, \ldots M\}$ (where '0' indicates that the decoder chooses "error"). The average error probability is

$$1 - \frac{1}{M} \sum_{m=1}^{M} P_{Z|X}(m|c_m).$$

In the application of our results, we will take A and B as n-fold cartesian products of alphabets $\mathcal{A}$ and $\mathcal{B}$, and a channel is a sequence of conditional probabilities $\{P_{Y^n|X^n} : \mathcal{A}^n \to \mathcal{B}^n\}$ [7]. Thus, to focus ideas the random variables X and Y throughout Sections II-IV can be viewed as vectors of fixed dimension equal to the blocklength.

Theorem 1: For any distribution P_X on A, there exists a code with M codewords and average probability of error not exceeding

$$\mathbb{P}\left[i(X,Y) \le \log \frac{M-1}{2}\right] + \frac{M-1}{2}\mathbb{P}\left[i(X,\bar{Y}) > \log \frac{M-1}{2}\right] \quad (5)$$

$$= \mathbb{E} \exp\left\{-\left(i(X,Y) - \log \frac{M-1}{2}\right)^+\right\} \quad (6)$$

where $P_{XY\bar{Y}}(a,b,c) = P_X(a)P_{Y|X}(b|a)P_Y(c)$.

The proof[1] of Theorem 1 uses Shannon's random coding along with Feinstein's decoder. Note that unlike the existing bounds (1), (2), the bound in Theorem 1 requires no optimization or selection of auxiliary constants.

It can be shown that the code size compatible with a given average error probability given by Theorem 1 is larger than Feinstein's (1), obtained for a given maximal error probability. It can be easily seen from (6) that Theorem 1 can be used to prove the achievability part of the most general known channel capacity formula [7].

Note that (5) is $\frac{M+1}{2}$ times the Bayesian minimal error probability of a binary hypothesis test of dependence:

$$\begin{array}{lll} \mathrm{H}_1: & P_{XY} & \text{with probability } \frac{2}{M+1} \\ \mathrm{H}_0: & P_X P_Y & \text{with probability } \frac{M-1}{M+1} \end{array}$$

III. Maximal Probability of Error

A. Bounds fixing input distribution

The details of the proof of Theorem 1 reveal that we could have generated the random codebook with only pairwise independent codewords. Thus, for some channels (e.g., discrete channels with additive noise) we can generate the codebook by imposing a distribution on the generating matrix of a linear code. Then Theorem 1 implies the existence of a linear code with average probability of error upper-bounded by (5). But the maximal and average probability of error coincide for a linear code and hence for additive-noise discrete channels the bound in Theorem 1 is also in the sense of maximal probability of error. The following bound on maximal error probability holds in general.

Theorem 2: For any input distribution P_X, and measurable $\gamma : \mathsf{A} \to [0, \infty]$ there exists a code with M codewords such that the j-th codeword's probability of error satisfies

$$\epsilon_j \le \mathbb{P}[i(X,Y) \le \log \gamma(X)] + (j-1) \sup_x \mathbb{P}[i(x,Y) > \log \gamma(x)]\,. \quad (7)$$

where the first probability is with respect to P_{XY} and the second is with respect to P_Y. In particular, the maximal probability of error satisfies

$$\epsilon \le \mathbb{P}[i(X,Y) \le \log \gamma(X)] + (M-1) \sup_x \mathbb{P}[i(x,Y) > \log \gamma(x)]\,, \quad (8)$$

Some symmetric channels and choices of P_X (most notably the binary erasure channel (BEC) and the binary symmetric channel (BSC) under equiprobable P_X satisfy the sufficient condition in the next result

Theorem 3: Fix an arbitrary input distribution P_X. If the cdf $\mathbb{P}[i(x,Y) \le \alpha]$ does not depend on x for any α when Y is distributed according to P_Y, then there exists an (M, ϵ) code with maximal probability of error satisfying (for any $x \in \mathsf{A}$)

$$\epsilon \quad \le \quad \mathbb{E} \exp\left\{-\left(i(X,Y) - \log(M-1)\right)^+\right\} \quad (9)$$

B. Bounds fixing output distribution

All the previous achievability bounds fixed some input distribution P_X and then proved that a certain codebook exists. However, in some cases (most notably, the additive white Gaussian noise (AWGN) channel which is outside the no-cost-constraint scope of this paper) it is desirable to consider auxiliary distributions on the output alphabet that are not necessarily induced by an input distribution.

Binary hypothesis tests between the conditional distribution $P_{Y|X=x}$ and an auxiliary unconditional distribution Q_Y on B play an important role in this subsection. A randomized test between those two distributions is defined by a random transformation $P_{Z|Y} : \mathsf{B} \mapsto \{0, 1\}$ where 0 indicates that the test chooses Q_Y. The best performance achievable among those randomized tests is given by

$$\beta_\alpha(x, Q_Y) = \min_{\substack{P_{Z|Y}: \\ P_{Z|X}(1|x) \ge \alpha}} \sum_{y \in \mathsf{B}} Q_Y(y) P_{Z|Y}(1|y)\,, \quad (10)$$

where the minimum is guaranteed to be achieved by the Neyman-Pearson lemma. Note that the conditional distribution that achieves the minimum depends on x.

For an arbitrary $\mathsf{F} \subset \mathsf{A}$, we define a related measure of performance for the composite hypothesis test between Q_Y

[1]Space limitations prevent us from including proofs of the results, which can be found in [8].

and the collection $\{P_{Y|X=x}\}_{x\in\mathsf{F}}$:

$$\kappa_\tau(\mathsf{F},Q_Y)=\inf_{\substack{P_{Z|Y}:\\ \inf_{x\in\mathsf{F}}P_{Z|X}(1|x)\geq\tau}}\sum_{y\in\mathsf{B}}Q_Y(y)P_{Z|Y}(1|y)\,. \tag{11}$$

As long as Q_Y is the output distribution induced by an input distribution Q_X, the quantity (11) satisfies the bound

$$\tau Q_X(\mathsf{F})\leq\kappa_\tau(\mathsf{F},Q_Y)\leq\tau\,. \tag{12}$$

Theorem 4: For any $0<\epsilon<1$, there exists a code with maximal error probability not exceeding ϵ, and M codewords chosen from $\mathsf{F}\subset\mathsf{A}$, satisfying

$$M\geq\sup_{0<\tau<\epsilon}\sup_{Q_Y}\frac{\kappa_\tau(\mathsf{F},Q_Y)}{\sup_{x\in\mathsf{F}}\beta_{1-\epsilon+\tau}(x,Q_Y)}\,. \tag{13}$$

Using (12) in Theorem 4 we obtain a weakened but useful bound:

$$M\geq\sup_{0<\tau<\epsilon}\sup_{Q_X}\frac{\tau Q_X(\mathsf{F})}{\sup_{x\in\mathsf{F}}\beta_{1-\epsilon+\tau}(x,Q_Y)}. \tag{14}$$

where the supremum is over all input distributions, and Q_Y denotes the distribution induced by Q_X on the output.

IV. Converse Bound

The approach for the converse comes from classical sphere-packing for the binary symmetric channel (BSC) with the exception that instead of measuring decoding sets by their cardinality, they are measured with an arbitrary probability distribution Q_Y [6], [9], [10].

Theorem 5: The size of a code with maximal error probability ϵ and with codewords belonging to F is upper bounded by

$$M\leq\inf_{Q_Y}\sup_{x\in\mathsf{F}}\frac{1}{\beta_{1-\epsilon}(x,Q_Y)}\,, \tag{15}$$

where the infimum is over all distributions Q_Y on B.

V. Normal approximation for discrete memoryless channels

We now consider the particularization of the abstract setup we have used so far to the case $\mathsf{A}=\mathcal{A}^n$ and $\mathsf{B}=\mathcal{B}^n$.

Theorem 6: (Strassen [6]) Let $M^*(n,\epsilon)$ be the largest size of a code with blocklength n and maximal error probability upper bounded by ϵ.

Then, for any discrete memoryless channel with capacity C and $0<\epsilon\leq 1/2$, we have

$$\log M^*(n,\epsilon)=nC-\sqrt{nV}Q^{-1}(\epsilon)+O(\log n)\,, \tag{16}$$

where $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\,dt$ and[2]

$$V=\min_{\substack{X:\\ C=I(X;Y)}}\operatorname{Var}\left[\log\frac{P_{Y|X}(Y|X)}{P_Y(Y)}\right]. \tag{17}$$

[2]Unlike Sections II-IV, in (17), X and Y refer to an individual input and output, respectively.

Theorem 6 formalizes the notion that $\log M^*(n,\epsilon)$ behaves asymptotically as the ϵ-quantile of the information density of the capacity achieving distribution [7]; for memoryless channels that information density is a sum of independent random variables whose small-deviation behavior is governed by the central limit theorem.

To prove Theorem 6, Strassen [6] used Theorem 5 for the converse and Feinstein's lemma for achievability. However, Feinstein's bound is too loose to yield the optimal $\sqrt{n}$ term for channels with cost constraints, most notably for the additive white Gaussian noise channel. Instead, for that purpose, Theorem 4 succeeds [8].

VI. Typical applications: BSC and BEC

A. Binary Symmetric Channel (BSC)

This section illustrates the application of developed theory to the BSC with crossover probability $\delta<1/2$. The input and output alphabets are binary, $\mathsf{A}=\mathsf{B}=\{0,1\}^n$, and the channel is defined as

$$P_{Y^n|X^n}(y^n|x^n)=\delta^{|y^n-x^n|}(1-\delta)^{n-|y^n-x^n|}\,, \tag{18}$$

where $|z^n|$ denotes the Hamming weight of the binary vector z^n.

Taking P_{X^n} equiprobable on $\{0,1\}^n$, the bound of Theorem 1 is equal to $\frac{M+1}{2}$ times the minimal probability of error of an optimal binary hypothesis test between n fair coin flips (with prior probability $\frac{M-1}{M+1}$) and n bias-δ coin flips (with prior probability $\frac{2}{M+1}$). In the form (6) the upper bound on average error probability becomes

$$f(M)\stackrel{\triangle}{=}\mathbb{E}\left[2^{-\left(na-bZ-\log\frac{M-1}{2}\right)^+}\right], \tag{19}$$

where

$$a=1+\log_2(1-\delta),\ b=\log_2\frac{1-\delta}{\delta}\,,$$

and $Z\sim B(n,\delta)$ is a binomial random variable with parameters n and δ. We want to have an achievability result in the form of a lower-bound on $M^*(n,\epsilon)$. Going from average probability of error to maximal in (19) can be done using the random linear code trick, which is applicable as long as $M=2^k$. Thus the algorithm for finding a lower-bound on $M^*(n,\epsilon)$ is to find the maximum $M=2^k$ such that the right-hand side of (19), $f(M)$, is still below prescribed ϵ:

$$M^*(n,\epsilon)\geq\max\left\{2^k: f(2^k)\leq\epsilon\right\}. \tag{20}$$

For comparison, Feinstein's lemma, with equiprobable P_X, yields the following bound:

$$M^*(n,\epsilon)\geq\sup_{t>0}2^{nt}\left(\epsilon-P\left[Z\geq n(a-t)/b\right]\right), \tag{21}$$

where $Z\sim B(n,\delta)$.

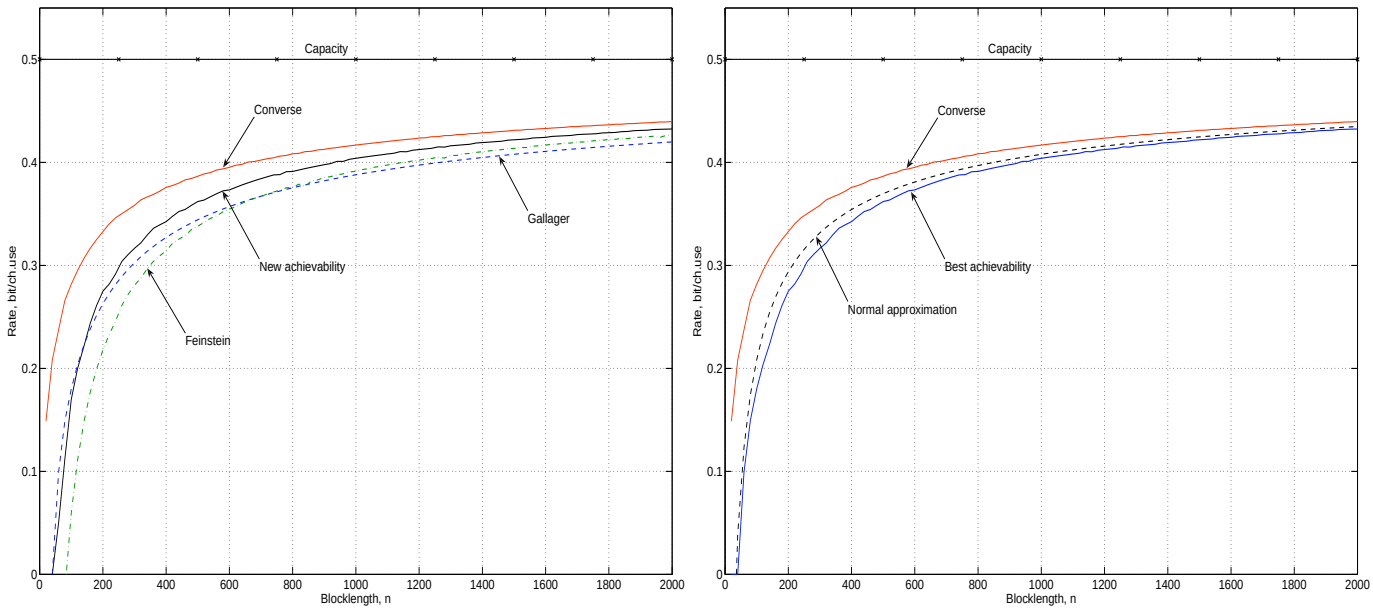

Fig. 1. Rate-blocklength tradeoff for the BSC with crossover probability $\delta = 0.11$ and maximal block error rate $\epsilon = 10^{-3}$.

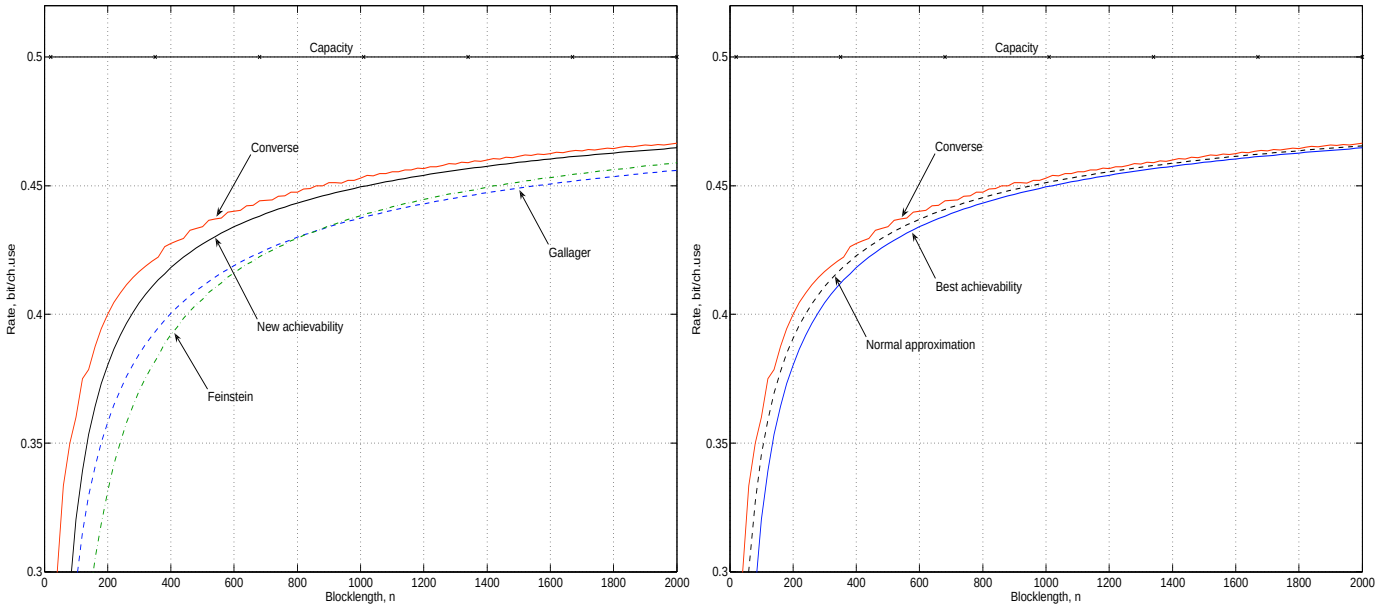

Fig. 2. Rate-blocklength tradeoff for the BEC with erasure probability $\delta = 0.5$ and maximal block error rate $\epsilon = 10^{-3}$.

Gallager's random coding bound (2) also with equiprobable P_X, assures that[3]

$$\log_2 M^*(n,\epsilon) \geq nE_r^{-1}\left(\frac{1}{n}\log_2\frac{1}{\epsilon}\right), \qquad (22)$$

where [5, Theorem 5.6.2, Corollary 2 and Example 1 in Section 5.6.]

$$E_r(1-h(s)) = \begin{cases} d(s||\delta), & s \in (\delta, s^*], \\ h(s) - 2\log s_1, & s > s^*, \end{cases}$$

[3]This bound holds for average probability of error. Fig. 1 shows the corresponding bound on maximal error probability where we drop half of the codewords with worse error probability. This results in an additional term of -1 appended to the right side of (22), while $\frac{1}{\epsilon}$ becomes $\frac{2}{\epsilon}$ therein.

and $s^* = \frac{\sqrt{\delta}}{\sqrt{\delta}+\sqrt{1-\delta}}$, $s_1 = \sqrt{\delta} + \sqrt{1-\delta}$.

We now turn our attention to the computation of the converse bound of Theorem 5 choosing Q_{Y^n} equiprobable on $\{0,1\}^n$. To streamline notation, we denote $\beta_\alpha^n = \beta_\alpha(x^n, Q_{Y^n})$ since it does not depend on x^n, and Q_{Y^n} is fixed.

Taking the log-likelihood ratio of (18) and $Q_{Y^n}(y^n) = 2^{-n}$ we observe that the Hamming weight $|Y^n|$ is a sufficient statistic for discriminating between $P_{Y^n|X^n=\mathbf{0}}$ and Q_{Y^n}. Thus, the optimal randomized test is

$$P_{Z_0|Y^n}(1|y^n) = \begin{cases} 0, & |y^n| > K_\alpha^n, \\ L_\alpha^n, & |y^n| = K_\alpha^n, \\ 1, & |y^n| < K_\alpha^n. \end{cases}$$

where $K_\alpha^n \in \mathbb{Z}_+$ and $L_\alpha^n \in [0,1)$ are uniquely determined by the condition

$$\sum_{y^n \in \mathsf{A}} P_{Y^n|X^n}(y^n|\mathbf{0}) P_{Z_0|Y^n}(1|y^n) = \alpha\,.$$

Then we find that

$$\beta_\alpha^n = L_\alpha^n \binom{n}{K_\alpha^n} 2^{-n} + \sum_{k=0}^{K_\alpha^n - 1} \binom{n}{k} 2^{-n}\,. \tag{23}$$

Thus, by Theorem 5

$$M^*(n,\epsilon) \le \frac{1}{\beta_{1-\epsilon}^n}\,. \tag{24}$$

Note that (24) is exactly the classical sphere-packing bound.

The numerical evaluation of (20) and (24) is shown in Fig. 1, where we also show bounds by Feinstein (21) and Gallager (22). As we anticipated analytically, the new bound is always tighter than Feinstein's bound. For $\delta = 0.11$ and $\epsilon = 0.001$, we can see in Fig. 1 (a) that for blocklengths greater than $n_* \approx 150$, Theorem 4 gives better results than Gallager's bound. In fact, for large n the gap to the converse upper bound of the new lower bound is less than half that of Gallager's bound. This tendency remains for other choices of δ and ϵ. Although, for smaller ϵ and/or δ, Gallager's bound (originally devised to analyze the regime of exponentially small ϵ) performs better (i.e., the value of n_* is greater). A similar relationship between the two bounds holds, qualitatively, in the case of the additive white Gaussian noise channel, see [8].

Fig. 1 (b) compares the upper bound (24), the maximum of the lower bounds (20) and (22), and the normal approximation (16), which becomes

$$\log M^*(n,\epsilon) \approx n - nh(\delta) - \sqrt{n} V Q^{-1}(\epsilon) \tag{25}$$

where

$$V = \delta(1-\delta)\log_2^2 \frac{\delta}{1-\delta}\,. \tag{26}$$

Fig. 1 (b) shows that the normal approximation is excellent even for rather short blocklengths. Note that for typical low-latency applications such as voice-over-IP packets of 50-80 bytes, the best achievable rate is 75-80% capacity. In fact, the asymptotic approximation

$$\log M^*(n,\epsilon) \approx n - nh(\delta)$$

is rather optimistic for the blocklengths shown in Fig. 1.

B. Binary erasure channel (BEC)

With equiprobable P_{X^n} the upper bound on error probability in (9) becomes

$$\epsilon \le \mathbb{E}\left[2^{-(Z-\log(M-1))^+}\right]\,, \tag{27}$$

where Z is binomial with parameters n and $1-\delta$, $Z \sim B(n, 1-\delta)$. We see in Fig. 2 (a) that (27) is quite a bit tighter than the Gallager and Feinstein bounds when particularized for the BEC. It is also easy to show that the bound in (27) not only leads to the achievability of capacity, but also yields Gallager's random coding exponent for the BEC.

The upper bound on code size given by Theorem 5 (with capacity achieving output distribution) is improved by the following converse result which gives an upper bound on the rate required to achieve a given average error probability (and thus, a given maximal error probability).

Theorem 7: For a binary erasure channel with erasure probability δ, the average error probability of a k-to-n code satisfies

$$\epsilon \ge \sum_{\ell=n-k+1}^{n} \binom{n}{\ell} \delta^\ell (1-\delta)^{n-\ell} \left(1 - 2^{n-\ell-k}\right) \tag{28}$$

Fig. 2 (b) compares the normal approximation for the BEC

$$\log M^*(n,\epsilon) \approx n - n\delta - \sqrt{n\delta(1-\delta)} Q^{-1}(\epsilon)\,. \tag{29}$$

with the upper and lower bounds, showing that (29) gives an excellent approximation even for very short blocklengths.

VII. Conclusion

The maximal rate achievable for a desired error probability and fixed blocklength is a fundamental limit of great importance in practice. The new achievability bounds together with classical information theoretic bounds yield tight approximations to such a fundamental limit even for very short blocklengths. We have contrasted our new bounds to the existing ones in the contexts of the BSC and BEC. These results validate the use of a simple normal approximation (16) to the best achievable rate in the finite blocklength regime. The extension of the bounds in this paper to channels with cost constraints such as the additive white Gaussian noise channel is reported in [8]. In contrast to existing bounds, our Theorem 4 succeeds in leading to a normal approximation for those cases.

References

[1] S. Verdú, "teaching it," XXVIII Shannon Lecture, *2007 IEEE International Symposium on Information Theory*, Nice, France, June 28, 2007. Also: IEEE Information Theory Society Newsletter, Dec. 2007.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623-656, 1948.

[3] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory,* vol. PGIT-4, pp. 2-22, 1954.

[4] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3-18, 1965.

[5] R. G. Gallager, *Information theory and Reliable Communication*, John Wiley & Sons, Inc. New York, 1968.

[6] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962, Czechoslovak Academy of Sciences, Prague, pp. 689-723.

[7] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147-1157, 1994.

[8] Y. Polyanskiy, H. V. Poor and S. Verdú, "Channel coding rate in the finite blocklength regime", *draft*, 2008.

[9] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405-417, 1974.

[10] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I", *Inform. Contr.*, vol. 10, pp. 65-103, 1967.