How Quickly Can We Approach Channel Capacity?

Dror Baron, Mohammad Ali Khojastepour, Richard G. Baraniuk
Department of Electrical and Computer Engineering
Rice University, Houston, Texas 77005

***Abstract*— Recent progress in code design has made it crucial to understand how quickly communication systems can approach their limits. To address this issue for the channel capacity C, we define the *non-asymptotic capacity* $C_{\rm NA}(n,\epsilon)$ as the maximal rate of codebooks that achieve a probability ϵ of codeword error while using codewords of length n. We prove for the binary symmetric channel that $C_{\rm NA}(n,\epsilon) = C - K(\epsilon)/\sqrt{n} + o(1/\sqrt{n})$, where $K(\epsilon)$ is available in closed form. We also describe similar results for the Gaussian channel. These results may lead to more efficient resource usage in practical communication systems.**

I. Introduction

A. Motivation

Shannon theory [1–3] has provided a variety of bounds on the efficiency of communication systems. Unfortunately, most of these results rely on asymptotics; the bounds reflect the limiting performance using infinitely long codewords. In contrast, practical communication systems live in a non-asymptotic world; they must use finite computational resources, incur finite delay, and so on. Therefore, in order to use these resources most efficiently, it has become crucial to understand how quickly practical communication systems can approach their ultimate performance limits in the non-asymptotic regime.

B. Non-asymptotic capacity

Practical codes use codewords of a finite length n (n may be large [4]), and the probability of error (either the bit error rate or codeword error probability ϵ) must be reasonably small. With these conditions on n and ϵ, it is not clear what rates are actually achievable, because Shannon theory has not addressed this point. Despite being an imprecise measure for practical channel codes [4], the capacity is the standard benchmark for verifying the efficiency of channel coding techniques.

Consider the question: "If the codeword length n is given, then what is the maximum possible rate of transmission R for which the probability of codeword error is bounded by a given ϵ?" To address this question, we denote a code with codeword length n, codebook size M, and probability of error no larger than ϵ by (n, M, ϵ). We define the *rate* of an (n, M, ϵ) code as $R \triangleq \log(M)/n$, and the *non-asymptotic capacity* as

$$C_{\rm NA}(n,\epsilon) \triangleq \max_{\left\{\widetilde{M}\colon \exists \text{ code } (n,\widetilde{M},\epsilon)\right\}} \frac{\log(\widetilde{M})}{n}.$$

This definition provides a more informative benchmark for verifying the efficiency of practical channel codes.

C. Related results

Although the notion of non-asymptotic capacity has not been clearly stated in the literature, there exist several related results. For the Gaussian channel, Shannon [5] derived the error performance of optimal codes. However, Shannon did not dwell on the notion of non-asymptotic capacity; in a world in which capacity-achieving codes seemed far off, the rate of convergence to capacity was unimportant.

For the binary symmetric channel (BSC), Wolfowitz provided the converse capacity [2, Theorem 3.4.1]. Wolfowitz also provided a more general result [2, Theorem 3.3.1] for discrete memoryless channels, which is used to prove the strong convese of the direct coding theorem [2, Theorem 3.2.1]. However, the strong converse hides the notion of non-asymptotic capacity. Furthermore, Wolfowitz's bounds are loose. In contrast, our bounds are precise to within $o(1/\sqrt{n})$ rate.[1]

Error exponent bounds [6, 7], which provide lower and upper bounds on the probability ϵ of codeword error, can be viewed as dual to our results. However, for rates near capacity, the lower and upper bounds provided by error exponents are often relatively loose (details in Baron et al. [8]). Our results, which are based on the central limit theorem (CLT), are more effective for rates near (within $O(1/\sqrt{n})$ of the) capacity. Our results are less effective for rates substantially below capacity. In fact, for such rates ϵ becomes very small, and the CLT no longer applies.

D. Contributions

In Section II we provide converse and achievable results for the BSC such that

$$C_{\rm NA}(n,\epsilon) = C - K(\epsilon)/\sqrt{n} + o(1/\sqrt{n}),$$

where $K(\epsilon) = \Phi^{-1}(\epsilon)\sqrt{p(1-p)}\log\left(\frac{1-p}{p}\right)$. For $\epsilon < 0.5$ we have $K(\epsilon) > 0$, which indicates a *gap to capacity* in the non-asymptotic regime. This gap can be significant in practice, because $K(\epsilon)$ is large, especially for small ϵ and small n. Similar results for the Gaussian channel are discussed in Section III.

Our results do not contradict the traditional capacity analysis. Rather, they refine and strengthen the classical results

This research was supported by AFOSR, ONR, NSF, and the Texas Instruments Leadership University Program. Email: {drorb,amir,richb}@rice.edu. Web: dsp.rice.edu.

[1] For two functions $f(n)$ and $g(n)$, $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$. Similarly, $f(n) = O(g(n))$ if $\exists C > 0$ such that $\lim_{n\to\infty} |f(n)/g(n)| \le C$.

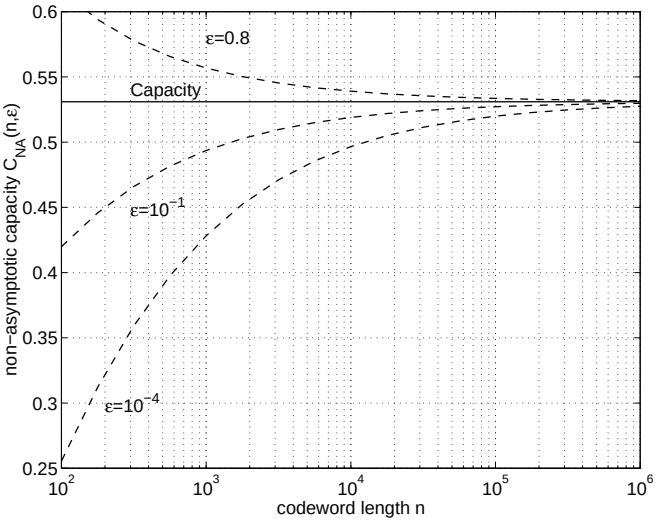

Fig. 1. **Non-asymptotic capacity as a refinement of the Shannon capacity.** We provide results for the BSC with cross-over probability $p = 0.1$. Our converse and achievable regions of Theorems 1 and 2 indicate that $C_{\mathrm{NA}}(n,\epsilon) = C - K(\epsilon)/\sqrt{n}$. As the codeword length n increases, these bounds approach the classical Shannon capacity C for any $\epsilon \in (0,1)$. Analogous results are available for the Gaussian channel.

because for any $\epsilon \in (0,1)$ we have $\lim_{n\to\infty} C_{\mathrm{NA}}(n,\epsilon) = C$. This convergence of $C_{\mathrm{NA}}(n,\epsilon)$ to C for fixed ϵ appears in Figure 1. Whereas error exponents [6, 7] characterize the codeword error probability well for rates substantially below capacity, our bounds are more precise for rates near capacity.

In the sequel, we first consider the BSC in Section II. We then describe similar results for other types of channels in Section III, and provide a discussion in Section IV. To make the paper more digestible, we include our proofs for the BSC in two appendices.

II. The Binary Symmetric Channel

We begin the discussion of the binary symmetric channel (BSC) by defining our notation and providing an overview of the classical converse proof. We then present tight converse and achievable regions, which also yield a precise expression for the capacity gap.

A. Notation and overview

Consider the BSC with cross-over probability p [3]. The channel input is a length-n sequence $x = x_1 \dots x_n$ over the binary alphabet $\{0,1\}$. The channel adds a binary noise sequence $z = z_1 \dots z_n$ to the input, where the symbols of z are independent and identically distributed (i.i.d.) such that $\Pr(z_i = 1) = p$ and $\Pr(z_i = 0) = 1 - p$. Therefore, the channel output $y = y_1 \dots y_n$ satisfies $y = x \oplus z$.

The converse coding theorem for the BSC [1–3] can be easily obtained using sphere packing arguments. Given a codeword, there are roughly $2^{nH(p)}$ typical channel output sequences, where

$$H(p) = -p\log(p) - (1-p)\log(1-p).$$

However, the entire output space consists of 2^n sequences. Therefore, for a codebook of size 2^{nR}, an ideal partitioning of the output space allocates $\frac{2^n}{2^{nR}} = 2^{n(1-R)}$ outputs to each codeword. If $R > C = 1 - H(p)$, then asymptotically the spheres are packed too closely, the noise drives us into another sphere, and the input codeword cannot be decoded correctly.

B. Converse region

The insights from the converse proof apply to the asymptotic regime. However, for any finite n there is a positive probability that the channel output is atypical, i.e., that it falls outside the set of typical channel outputs. Therefore we have $\epsilon > 0$. The following theorem provides an upper bound on $C_{\mathrm{NA}}(n,\epsilon)$, and is very similar to a result of Wolfowitz [2, Theorem 3.4.1]; we provide the proof in Appendix A.

Theorem 1: The non-asymptotic capacity of the BSC with cross-over probability p satisfies

$$C_{\mathrm{NA}}(n,\epsilon) < 1 - H\left(p + \Phi^{-1}(\epsilon)\sqrt{\frac{p(1-p)}{n}}\right) + o(1/\sqrt{n}). \quad (1)$$

The proof considers the number of channel errors $e(z) = \sum_{i=1}^{n} z_i$. Because $E[e(z)] = np$ and $\mathrm{var}(e(z)) = np(1-p)$, the CLT indicates that

$$\Pr\left(e(z) > np + \tau\sqrt{np(1-p)}\right) = \Phi(\tau) + o(1).$$

Therefore, ϵ can be attained by defining a noise sphere that allows $e(z)$ to exceed its expected value by $\Phi^{-1}(\epsilon)$ variance terms. This is the smallest volume noise sphere that yields ϵ. The bound (1) is obtained by dividing 2^n by the number of codewords in such a noise sphere.

C. Achievable region

We now provide an achievable region for the BSC, which lower bounds the non-asymptotic capacity. This achievable region, which refines a similar proof by Wolfowitz for the discrete memoryless channel [2, Theorem 3.2.1], is tight with respect to the converse region of Theorem 1.

Theorem 2: The non-asymptotic capacity of the BSC with cross-over probability p satisfies

$$C_{\mathrm{NA}}(n,\epsilon) > 1 - H\left(p + \Phi^{-1}(\epsilon)\sqrt{\frac{p(1-p)}{n}}\right) + o(1/\sqrt{n}). \quad (2)$$

We outline the proof here and include the details in Appendix B. The key idea is to use a codebook construction wherein we sequentially add codewords $x(1), x(2), \dots$ to the codebook until there are no "good" codewords left. For each codeword $x(i)$ that is added to the codebook, there is a set

$$T(x(i)) = \left\{y:\ y = x(i) + z,\ e(z) < np + \Phi^{-1}(\epsilon)\sqrt{np(1-p)}\right\}$$

of typical channel outputs and a set $A(i) \triangleq T(x(i)) - \cup_{j=1}^{i-1} A(j)$ of channel outputs that will be decoded as $x(i)$. A new codeword $x(i)$ is added to the codebook only if $\Pr\left(y \in \cup_{j=1}^{i-1} A(j) | x(i)\right) < \epsilon$. The construction halts with a codebook of size M only after $\cup_{i=1}^{M} A(i) > \epsilon 2^n$. On the other hand, $\cup_{i=1}^{M} A(i)$ is a union of M sets, each of cardinality

$2^{nH(p+O(1/\sqrt{n}))}$ at most. A lower bound for M is obtained by dividing $\epsilon 2^n$ by the maximal cardinality of $A(i)$.

Remark 1: Our achievable region refines a similar result by Wolfowitz for the discrete memoryless channel [2, Theorem 3.2.1]. Wolfowitz proved the existence of a constant K and codebooks of size M that attain probability of codeword error ϵ, where $\log(M)/n = C - K/\sqrt{n}$. Unfortunately, K was not specified precisely. Novelties in our proof include a refined definition of $T(x(i))$, a modified condition for adding codewords to the codebook, and a more precise characterization of the maximal cardinality of $A(i)$.

D. Capacity gap and non-asymptotic capacity

We define the non-asymptotic *capacity gap* as

$$G_{\rm NA}(n,\epsilon) \triangleq C - C_{\rm NA}(n,\epsilon),$$

and derive this value for the BSC. To derive expressions for $G_{\rm NA}(n,\epsilon)$ and $C_{\rm NA}(n,\epsilon)$, we first approximate the entropy function around $H(p)$. By taking the derivative

$$\frac{d}{dp}H(p) = \log\left(\frac{1-p}{p}\right),$$

the first-order Taylor approximation of the entropy around p is

$$H(p+\Delta) = H(p) + \log\left(\frac{1-p}{p}\right)\Delta + O(\Delta^2).$$

Therefore,

$$H\left(p + \Phi^{-1}(\epsilon)\sqrt{\frac{p(1-p)}{n}}\right) + o(1/\sqrt{n}) =$$
$$H(p) + \Phi^{-1}(\epsilon)\log\left(\frac{1-p}{p}\right)\sqrt{\frac{p(1-p)}{n}} + o(1/\sqrt{n}).$$

Taking into account Theorems 1 and 2, the approximation of the entropy, and the capacity $C = 1 - H(p)$ of the BSC, the non-asymptotic capacity gap $G_{\rm NA}(n,\epsilon)$ is given by

$$G_{\rm NA}(n,\epsilon) \;=\; \frac{\Phi^{-1}(\epsilon)\sqrt{p(1-p)}\log\left(\frac{1-p}{p}\right)}{\sqrt{n}} + o(1/\sqrt{n}).$$

Therefore, for $\epsilon < 0.5$ the non-asymptotic capacity is smaller than the Shannon capacity, and the capacity gap is strictly proportional to $1/\sqrt{n}$. Finally, the constant $K(\epsilon)$ is given by

$$K(\epsilon) = \Phi^{-1}(\epsilon)\sqrt{p(1-p)}\log\left(\frac{1-p}{p}\right),$$

and the non-asymptotic capacity is

$$C_{\rm NA}(n,\epsilon) = C - \frac{\Phi^{-1}(\epsilon)\sqrt{p(1-p)}\log\left(\frac{1-p}{p}\right)}{\sqrt{n}} + o(1/\sqrt{n}).$$

III. Additional types of channels

Consider the Gaussian channel. The channel input is a length-n sequence $x \in \mathbb{R}^n$ with peak power constraint S, i.e., $\sum_{i=1}^n (x_i)^2 \le nS$. The channel output is a length-n sequence y such that $y_i = x_i + z_i$, where z is i.i.d. zero-mean Gaussian with variance N.

Shannon [5] considered the structure and error performance of optimal codes for the Gaussian channel. In particular, he provided the following result.

Theorem 3: [5] Using a peak power constraint S, the non-asymptotic capacity of the Gaussian channel with noise variance N satisfies

$$C_{\rm NA}(n,\epsilon) = C - \frac{\Phi^{-1}(\epsilon)}{\ln(2)}\sqrt{\frac{S(2N+S)}{2n(N+S)^2}} + o(1/\sqrt{n}),$$

where

$$C = \frac{1}{2}\log\left(1 + \frac{S}{N}\right).$$

Shannon's converse proof [5] relies on codewords that lie on the signal sphere, i.e., $\sum_{i=1}^n (x_i)^2 = nS$. Therefore, the decoding regions are multi-dimensional pyramids whose apexes are at the origin. This enables Shannon to use cone packing arguments [5] to provide a lower bound on the attainable probability ϵ of codeword error. In our recent work [8], we have simplified Shannon's tedious derivations using arguments based on the CLT.

Shannon's achievable proof is also quite involved. Instead, we provide achievable results using information spectrum methods [9], which specify precise bounds on the probability of codeword error. In our recent work [8], we used information spectrum methods to analyze the non-asymptotic performance of Gaussian codebooks, where each codeword x is i.i.d. zero-mean Gaussian with variance S. Such Gaussian codebooks have often been used [3, 7] to prove direct coding theorems for the Gaussian channel. The non-asymptotic performance of Gaussian codebooks is strictly sub-optimal in the sense that the non-asymptotic rate $R_{\rm NA}(n,\epsilon)$ that yields probability ϵ of codeword error is smaller than $C_{\rm NA}(n,\epsilon)$. This surprising result was originally provided by Rice [10].

Theorem 4: [10] Using an average power S, the non-asymptotic rate achieved by Gaussian codebooks over a Gaussian channel with noise variance N satisfies

$$R_{\rm NA}(n,\epsilon) = C - \frac{\Phi^{-1}(\epsilon)}{\ln(2)}\sqrt{\frac{S}{n(N+S)}} + o(1/\sqrt{n}).$$

We have also used information spectrum bounds to analyze the performance of Shannon's cone construction. The outcome of our derivation [8] is a somewhat less tedious achievable proof. The problem with Gaussian codebooks, as pointed out by Shannon [5], is that codewords in the interior of the signal sphere incur a relatively high probability of error. In contrast, the constraint that codewords lie on the signal sphere eliminates this problem. However, such a constraint introduces dependencies between the codeword symbols. Therefore, optimal codebooks for the Gaussian channel must use a non-i.i.d. codeword distribution.

We hope to extend these results on the non-asymptotic capacity of the BSC and Gaussian channel to an expression for general memoryless channels. To do so, we must establish tight converse and achievable regions for arbitrary memoryless channels. Our results suggest that information spectrum bounds [9] are precise enough to derive achievable bounds on the non-asymptotic capacity. Unfortunately, because optimal codebooks for the Gaussian channel must use a non-i.i.d. codeword distribution, we must consider all possible codeword distributions in order to establish similar converse bounds. Therefore, it seems that the general characterization of the non-asymptotic capacity may be difficult.

IV. Discussion

The non-asymptotic capacity provides a more informative benchmark for verifying the efficiency of practical channel codes. The prior art has concentrated on error exponents, which are informative when dealing with very small probabilities of error. However, for rates near capacity, the lower and upper bounds provided by error exponents are often relatively loose [8]. Our results, which are based on the CLT, are more effective for rates near capacity.

One extension of our work is the derivation of the non-asymptotic capacity of general memoryless channels. As discussed in Section III, the general characterization of the non-asymptotic capacity may be difficult, because non-i.i.d. codeword distributions must be considered.

Another extension would investigate the non-asymptotic performance of additional communication systems. We have studied this problem for distributed source coding. Our results [11] considered a specific Bernoulli setup where side information is passed through a correlation channel. In this setup, which is similar to the BSC investigated in this paper, the redundancy rates of non-asymptotic Slepian-Wolf coding are equivalent to our gap to capacity $G_{\mathrm{NA}}(n,\epsilon)$.

Yet another potential extension is the case where the channel and source statistics are unknown. For Slepian-Wolf coding, we proposed a scheme [11] that relies on a trickle of information between linked encoders. In some cases this resulted in an $O(1/\sqrt{n})$ redundancy rate, similar to the gaps to capacity of this paper. Unfortunately, in some cases, the redundancy rate increases to $O(n^{-1/3})$. This result also suggests that the redundancy of universal channel coding systems with feedback could turn out to be exceedingly high [11].

Finally, we emphasize that the importance of non-asymptotic characterization of communication systems extends to a wide range of additional problems in Shannon theory. We have seen that the penalties for using finite n and lack of knowledge of source and channel statistics may be quite large. Therefore, an understanding of these issues is crucial in order to improve the performance of communication systems.

Appendix A. Proof of Theorem 1

Define $e(z) \triangleq \sum_{i=1}^{n} z_i$ as the number of errors over the channel. Because z is i.i.d., $e(z)$ has a binomial distribution, which can be approximated by a Gaussian distribution with mean np and variance $np(1-p)$. Let t be the number of errors $\Phi^{-1}(\epsilon)+o(1)$ standard deviations above the mean, i.e.,

$$t \triangleq np + [\Phi^{-1}(\epsilon)+o(1)]\sqrt{np(1-p)}, \tag{A.1}$$

where the importance of the $o(1)$ term will be explained shortly. Define the *typical set* T as

$$T \triangleq \{z:\ e(z) \le t\};$$

then the atypical set is $T^C \triangleq \{z:\ e(z) > t\}$. It is well known from the CLT that

$$\Pr\left(z \in T^C\right) = \Phi\left(\Phi^{-1}(\epsilon)+o(1)\right) + o(1) > \epsilon,$$

where the second inequality comes from the first $o(1)$ term, which cancels out the second $o(1)$ term.

We now derive the cardinality of the typical set

$$|T| = \sum_{j=0}^{\lfloor t \rfloor} |\{z:\ e(z)=j\}| = \sum_{j=0}^{\lfloor t \rfloor} \binom{n}{j},$$

where $\lfloor \cdot \rfloor$ denotes the floor operator. Note that

$$\frac{|\{z:\ e(z)=j+1\}|}{|\{z:\ e(z)=j\}|} = \frac{\frac{n!}{(j+1)!(n-j-1)!}}{\frac{n!}{j!(n-j)!)}} = \frac{n-j}{j+1}.$$

Furthermore, this ratio decreases as j increases, so that for $j \le t$ the ratio is lower bounded by $\frac{1-p}{p} + O(1/\sqrt{n}) > 1$. Therefore,

$$\binom{n}{j} < \binom{n}{\lfloor t \rfloor}\left(\frac{p}{1-p}+O(1/\sqrt{n})\right)^{\lfloor t \rfloor - j},$$

and so we have

$$\begin{aligned} |T| &< \binom{n}{\lfloor t \rfloor}\sum_{j=0}^{\lfloor t \rfloor}\left(\frac{p}{1-p}+O(1/\sqrt{n})\right)^{\lfloor t \rfloor - j} \\ &< \binom{n}{\lfloor t \rfloor}\frac{1}{1-\frac{p}{1-p}+O(1/\sqrt{n})} \\ &= \binom{n}{\lfloor t \rfloor}O(1). \end{aligned} \tag{A.2}$$

We now invoke Stirling's formula $\log(k!) \approx (k+0.5)\log(k/e) + O(1)$ and obtain

$$\begin{aligned} \log\left(\binom{n}{\lfloor t \rfloor}\right) &= \log(n!) - \log(t!) - \log((n-t)!) \\ &\approx (n+0.5)\log(n/e) - (t+0.5)\log(t/e) \\ &\quad -(n-t+0.5)\log((n-t)/e) + O(1) \\ &= (n+0.5)\log(n) - (t+0.5)\log(t) \\ &\quad -(n-t+0.5)\log(n-t) + O(1). \end{aligned}$$

Using (A.1), (A.2) and substituting into the entropy function,

$$\begin{aligned} \log(|T|) &= nH\left(p + [\Phi^{-1}(\epsilon)+o(1)]\sqrt{\frac{p(1-p)}{n}}\right) \\ &\quad -0.5\log(n) + O(1). \end{aligned} \tag{A.3}$$

Because there are 2^n total possible outputs, it is impossible to construct a codebook with more than $2^n(1-\log(|T|))$

codewords whose typical sets are disjoint. The CLT indicates that the probability of error of such a codebook is at least ϵ, which completes the proof. □

Appendix B. Proof of Theorem 2

In our codebook construction, we sequentially add codewords $x(1), x(2), \ldots$ to the codebook until there are no "good" codewords left. Fix $\eta \in (0, \epsilon)$, and let T_z be the typical set of noise sequences z such that

$$T_z \triangleq \left\{ z : \ e(z) < np + \Phi^{-1}(\epsilon - \eta)\sqrt{np(1-p)} \right\}.$$

For a candidate codeword x, let

$$T(x) \triangleq \{ y : \ y = x + z, \ z \in T_z \}$$

be the set of typical channel outputs. For each $x(i)$ that we add to the codebook, we define $A(i) \triangleq T(x(i)) - \cup_{j=1}^{i-1} A(j)$ as the set of channel outputs that will be decoded as $x(i)$. Therefore, $\cup_i A(i) = \cup_i T(x(i))$ and $\{A(i)\}_i$ are disjoint. We add $x(i)$ to the codebook only if

$$\Pr\left(y \in T(x) \cap [\cup_{j=1}^{i-1} A(j)] \ | x(i) \right) < \eta. \tag{B.1}$$

The construction continues until there are no potential codewords left that satisfy (B.1). Denote the codebook size when the construction halts by M.

Consider the uniform prior $\Pr(x) = 2^{-n}$ for any possible input. Owing to symmetry we also have $\Pr(y) = 2^{-n}$ for any possible output. Using this uniform prior, when the construction halts,

$$\begin{aligned}
&\Pr\left(y \in \cup_{i=1}^{M} A(i) \right) = \\
&\qquad \sum_{y \in \{0,1\}^n} \Pr(y) 1_{\left\{ y \in \cup_{i=1}^{M} A(i) \right\}} && \text{(B.2)} \\
&= \sum_{y \in \{0,1\}^n} \sum_{x \in \{0,1\}^n} \Pr(x)\Pr(y|x) 1_{\left\{ y \in \cup_{i=1}^{M} A(i) \right\}} && \text{(B.3)} \\
&= \sum_{x \in \{0,1\}^n} \Pr(x) \Pr\left(y \in \cup_{i=1}^{M} A(i) | x \right) && \text{(B.4)} \\
&> 2^{-n} \sum_{x \in \{0,1\}^n} \Pr\left(y \in T(x) \cap [\cup_{i=1}^{M} A(i)] | x \right) && \text{(B.5)} \\
&\geq 2^{-n} \sum_{x \in \{0,1\}^n} \eta = \eta. && \text{(B.6)}
\end{aligned}$$

where in (B.2) we sum over all possible outputs, and $1_{\{\cdot\}}$ denotes an indicator function, (B.3) performs a double summation over all possible inputs and outputs, (B.4) changes the order of summation, intersection may decrease the cardinality of a set in (B.5), and (B.6) relies on (B.1). We conclude that $\cup_{i=1}^{M} A(i)$ covers at least $\eta 2^n$ possible outputs. Using (A.3), we also have

$$\begin{aligned}
\left| \cup_{i=1}^{M} A(i) \right| &= \sum_{i=1}^{M} |A(i)| \\
&\leq \sum_{i=1}^{M} |T(x(i))| = M|T_z| \\
&= M 2^{nH\left(p + [\Phi^{-1}(\epsilon-\eta) + o(1)]\sqrt{\frac{p(1-p)}{n}} \right) - 0.5\log(n) + O(1)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\log(M) \ \geq \ & n\left[1 - H\left(p + [\Phi^{-1}(\epsilon-\eta) + o(1)]\sqrt{\frac{p(1-p)}{n}} \right) \right] \\
& + \log(\eta) + 0.5\log(n) + O(1).
\end{aligned}$$

The choice $\eta = 1/\sqrt{n}$ provides the required bound (2). □

Acknowledgments

We thank Behnaam Aazhang, Robert Gallager, Shlomo Shamai, Sergio Verdú, Pramod Viswanath, and Tsachy Weissman for insightful comments.

References

[1] 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.

[2] J. Wolfowitz, *Coding theorems of information theory*, 3d ed. Springer-Verlag, Berlin ; New York, 1978.

[3] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, New York, 1991.

[4] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Information Theory*, vol. IT-47, no. 2, pp. 619-637, February 2001.

[5] C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," *Bell System Tech. Journal*, vol. 38, pp. 611-656, 1959.

[6] C. E. Shannon, R. G. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I," *Information and Control*, vol. 10, no. 1, pp. 65-103, January 1967.

[7] R. G. Gallager, *Information theory and reliable communication*. John Wiley and Sons, New York, 1968.

[8] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Refined bounds on non-asymptotic channel capacity," *in preparation*, 2004.

[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Information Theory*, vol. IT-40, no. 4, pp. 1147-1157, July 1994.

[10] S. O. Rice, "Communication in the presence of noise – Probability of error for two encoding schemes," *Bell System Tech. Journal*, vol. 29, pp. 60-93, 1950.

[11] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Redundancy rates of Slepian-Wolf coding," *Proc. 42nd Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, September 2004.