

The Information-Outage Probability of Finite-Length Codes over AWGN Channels

David Buckingham and Matthew C. Valenti
Lane Dept. of Comp. Sci. and Elect. Eng.
West Virginia University
Morgantown, WV 26506-6109
Email: {david.buckingham,valenti}@ieee.org

Abstract—The performance of random error control codes approaches the Shannon capacity limit as the code length goes to infinity. When the code length is finite, then the code will be unable to achieve arbitrarily low error probability and a nonzero codeword error rate is inevitable. Information-theoretic bounds on codeword error rate may be found as a function of length through traditional methods such as sphere packing. Alternatively, the behavior of finite-length codes can be characterized in terms of an information-outage probability. The information-outage probability is the probability that the mutual information, which is a random variable, is less than the rate. In this paper, a Gaussian approximation is proposed that accurately models the information-outage probability for moderately small codes. The information-outage probability is related to several previously derived bounds, including Shannon’s sphere-packing and random coding bounds, as well as a bound on maximal error probability known as Feinstein’s lemma. It is shown that the information-outage probability is a useful predictor of achievable error rate.

I. INTRODUCTION

The Shannon capacity does not exist for nonergodic channels, such as slow fading channels or channels with finite-length inputs. For such systems, the mutual information between channel input and output is a random variable whose value will change from one channel realization to the next. The probability that the mutual information is less than the code rate R is called the *information-outage probability* (IOP) [1], [2]. The IOP is the cumulative distribution function (CDF) of the mutual information evaluated at R . Information outage probability provides some insight into the block-error rate (BLER) that may be achieved by practical codes operating in nonergodic settings.

Past work [2], [3] on information-outage probability has concentrated on the block-fading channel. The mutual information per block is taken to be $\log(1 + SNR)$ while the SNR is allowed to vary from block to block. The overall mutual information is then found by averaging across the blocks. In effect, it is assumed that the channel is able to achieve the Shannon capacity for the duration of each block and the randomness in the mutual information is due to the randomness of the SNR. However, there is a paradox in that a block must be sufficiently small to span just one coherence interval, yet sufficiently long that the Shannon capacity is achieved.

Even in the additive white Gaussian noise (AWGN) channel, the mutual information between channel input and output is a random variable if the blocklength is finite. An information-outage probability can be defined for the AWGN channel with finite blocklengths just as it can be defined for the block-fading channel. This outage probability has been called the *information spectrum* [4], [5] in the recent literature. In [6], [7], the mutual information is shown to be a Bessel-K distribution which occurs when n independent and identically distributed (i.i.d.) Laplacian random variables are averaged, where n is the number of symbols per codeword. In this paper, we revisit this distribution and derive a Gaussian approximation for it. The approximation is accurate for surprisingly small values of n , and is much easier to manipulate than the true Bessel-K distributions.

Having found an easily computable method for approximating the information-outage probability, we compare the IOP with the block-error rate (BLER) that may be achieved with actual codes (in this case, the turbo code from the UMTS-LTE standard). A relationship is observed between the IOP and the simulated BLER; the two curves have a similar slope with a fixed horizontal offset. The relationship is made more formal by recalling Feinstein’s lemma [8] which states that a code exists with maximal codeword error probability that is incrementally higher than the information-outage probability. For moderate values of n the maximal error rate due to Feinstein’s lemma approaches the information-outage probability, suggesting that the IOP could perhaps be used directly to approximate the codeword error probability that may be achieved in practice. The utility of the information-outage probability is that it is easily generated using the Gaussian Q-function, in contrast with simulating the actual code. It also serves as an alternative to other methods for bounding blocklength limited performance such as sphere-packing and random coding bounds [9], [10], [11].

This paper continues by reviewing mutual information from the perspective of it being a random variable in Section II. Next, the mutual information of channels with finite-length vector inputs is considered in Section III. The pdf and CDF of the mutual information between the channel’s input and output vectors are discussed in Sections IV and V, respectively, with a new Gaussian approximation proposed. The usefulness of the information-outage probability is discussed in Section VI,

where it is compared against the sphere-packing and random coding bounds, as well as the bound due to Feinstein's lemma. Finally, the paper concludes in Section VII.

II. A REVIEW OF MUTUAL INFORMATION

Before delving into the details, we must first comment on our notation. We use capital letters to indicate random variables and lowercase letters to indicate sample values. Let $Z \sim \mathcal{N}(m, \sigma^2)$ indicate that Z is a real Gaussian random variable with mean m and variance σ^2 , and let $Z \sim \mathcal{N}_c(m, \sigma^2)$ indicate that Z is a circular complex Gaussian random variable with mean m and variance σ^2 . The probability density function (pdf) of a complex Gaussian Z is given by

$$f_Z(z) = \frac{1}{\pi\sigma^2} \exp\left\{-\frac{|z-m|^2}{\sigma^2}\right\}, \quad (1)$$

and its cumulative distribution function (CDF) is indicated by $F_Z(z)$. The probability of an event \mathcal{A} is indicated by $P[\mathcal{A}]$.

The discrete-time output of a capacity-achieving additive white Gaussian noise (AWGN) channel is

$$Y = X + N, \quad (2)$$

where $X \sim \mathcal{N}_c(0, \mathcal{E}_s)$ is the channel input and $N \sim \mathcal{N}_c(0, N_0)$ is the noise. Note that the input has been chosen to be Gaussian in the interest of achieving the AWGN capacity, and the output $Y \sim \mathcal{N}_c(0, \mathcal{E}_s + N_0)$.

The *mutual information* between a channel's input sample x and output sample y is the function [11], [12]

$$i(x; y) = \log \frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)}. \quad (3)$$

The *average* mutual information is found by taking the expectation over the joint pdf $f_{X,Y}(x, y)$

$$I(X; Y) = E[i(X; Y)]. \quad (4)$$

While some authors [13] use the term *mutual information* to refer to (4), we follow the convention of [11], [12] and reserve the term for (3).

Using the definition of conditional probability, (3) can be expressed as

$$i(x; y) = \log \frac{f_{Y|X}(y|x)}{f_Y(y)}, \quad (5)$$

where $f_{Y|X}(y|x)$ is the conditional pdf of Y given X which is $\mathcal{N}_c(X, N_0)$. Substituting the Gaussian pdf (1) for $Y|X$ and Y into (3) yields

$$i(x; y) = \log \left(1 + \frac{\mathcal{E}_s}{N_0}\right) + \frac{|y|^2}{\mathcal{E}_s + N_0} - \frac{|y-x|^2}{N_0}. \quad (6)$$

The average mutual information is then found by substituting (6) into (4), resulting in

$$I(X; Y) = \log \left(1 + \frac{\mathcal{E}_s}{N_0}\right) + \frac{E[|Y|^2]}{\mathcal{E}_s + N_0} - \frac{E[|Y-X|^2]}{N_0}. \quad (7)$$

Because $E[|Y|^2] = \mathcal{E}_s + N_0$ and $E[|Y-X|^2] = E[|N|^2] = N_0$, the second line of (7) is zero and thus

$$I(X; Y) = \log \left(1 + \frac{\mathcal{E}_s}{N_0}\right), \quad (8)$$

which is the ergodic capacity of the AWGN channel.

III. MUTUAL INFORMATION RATE

Capacity is related to the average mutual information only in the limit of infinite-length codewords. In this paper, we are concerned with information-theoretic measures of performance for codewords of finite length. Let n be the length of the codeword and let the output $\mathbf{Y} \in \mathbb{C}^n$ be

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \quad (9)$$

where \mathbf{X} is a vector of independent and identically distributed (i.i.d.) variables $X_i \sim \mathcal{N}_c(0, \mathcal{E}_s)$, $1 \leq i, \leq n$, and \mathbf{N} is white noise with i.i.d. elements $N_i \sim \mathcal{N}_c(0, N_0)$, $1 \leq i, \leq n$.

For the vector channel, the mutual information between an input sample vector \mathbf{x} and output sample vector \mathbf{y} is defined as the function

$$i(\mathbf{x}; \mathbf{y}) = \frac{1}{n} \log \frac{f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y})}. \quad (10)$$

This function is called the (normalized) *mutual-information rate* in [5], [7]. Because the inputs to the AWGN channel are i.i.d. and there is no correlation between noise samples, the mutual-information rate can be rewritten as

$$i(\mathbf{x}; \mathbf{y}) = \frac{1}{n} \sum_{k=1}^n i(x_k; y_k), \quad (11)$$

with $i(x, y)$ as defined in (3) and given for the AWGN channel by (6).

Because the random processes $\{x_k\}$ and $\{y_k\}$ are ergodic in the mean, the average mutual information (4) may be found by taking the time average of any pair of infinite-length sample processes. This implies that the capacity can be found as

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n i(x_k; y_k). \quad (12)$$

The above expression is the key to computing the AWGN capacity via Monte Carlo methods. All one needs to do is simulate a large number of channel input/output pairs (x_i, y_i) and average the corresponding realizations of the mutual-information function computed using (6). While there is no benefit for using Monte Carlo methods to determine the unconstrained AWGN capacity, the basic procedure could be easily adapted to handle more complicated channels or constrained modulations.

IV. THE PDF OF MUTUAL INFORMATION

When n is finite, the quantity $i(\mathbf{x}; \mathbf{y})$ given by (11) is a random variable. Let Z_n be the mutual information between a length- n \mathbf{x} and length- n \mathbf{y} , which has form

$$Z_n = \log \left(1 + \frac{\mathcal{E}_s}{N_0}\right) + W_n, \quad (13)$$

where

$$\begin{aligned}
W_n &= \frac{1}{n} \sum_{k=1}^n \left(\frac{|Y_k|^2}{\mathcal{E}_s + N_0} - \frac{|Y_k - X_k|^2}{N_0} \right) \\
&= \frac{1}{n} \sum_{k=1}^n \left(\frac{|Y_k|^2}{\mathcal{E}_s + N_0} - \frac{|N_k|^2}{N_0} \right) \\
&= \frac{1}{n} \sum_{k=1}^n \left(|Y'_k|^2 - |N'_k|^2 \right), \tag{14}
\end{aligned}$$

where $Y'_k \sim \mathcal{N}_c(0, 1)$ and $N'_k \sim \mathcal{N}_c(0, 1)$ are a pair of correlated complex Gaussian variables with correlation coefficient that depends on the ratio \mathcal{E}_s/N_0 .

When $n = 1$, W_n is a Laplacian random variable with zero mean and variance

$$\sigma_W^2 = \frac{2\mathcal{E}_s}{\mathcal{E}_s + N_0}. \tag{15}$$

See [6] and [7] for details of this derivation.

For $n > 1$, W_n is the average of n i.i.d. Laplacian random variables and is a Bessel-K random variable with pdf given by

$$f_{W_n}(w) = \frac{2^{1-n}}{\sqrt{\pi}\Gamma(n)\sigma_W} \left(\frac{\sqrt{2}|w|}{\sigma_W} \right)^{n-\frac{1}{2}} K_{n-\frac{1}{2}} \left(\frac{\sqrt{2}|w|}{\sigma_W} \right) \tag{16}$$

where

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

is the gamma function and $K_\nu(z)$ is the modified Bessel function of the second kind with index ν .

For large n , it is inconvenient to deal directly with the pdf of the Bessel-K random variable. However for large n , the Central Limit Theorem applies, and thus W_n may be approximated by a Gaussian variable. Let \tilde{W}_n represent the Gaussian approximation to W_n . Since \tilde{W}_n is the average of n i.i.d. variables with zero mean and variance σ_W^2 , \tilde{W}_n will be a scalar Gaussian with zero mean and variance σ_W^2/n . Since Z_n is merely W_n with shifted mean, it may be approximated with \tilde{Z}_n , where

$$\tilde{Z}_n \sim \mathcal{N} \left(\log \left(1 + \frac{\mathcal{E}_s}{N_0} \right), \frac{2\mathcal{E}_s}{n(\mathcal{E}_s + N_0)} \right). \tag{17}$$

A comparison of the exact and approximate densities is shown in Fig. 1 for $n = \{1, 5, 10\}$ at $\mathcal{E}_s/N_0 = 3/4$. The Laplacian density is apparent from the $n = 1$ case. For n as small as 10, the exact and approximate densities are already taking on the same appearance. Further illustration of the approximation is shown in Fig. 2, which shows a normalized histogram produced by drawing 10^6 realizations of the mutual-information random variable using (6), $n = 10$, and $\mathcal{E}_s/N_0 = 3/4$. The exact and approximate densities are superimposed on the plot to verify their correctness.

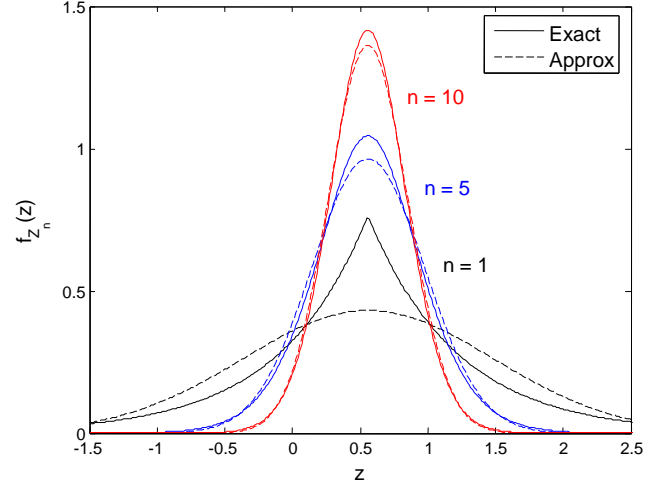


Fig. 1. The exact Bessel-K pdf and its Gaussian approximation for $n = \{1, 5, 10\}$ and $\mathcal{E}_s/N_0 = 3/4$.

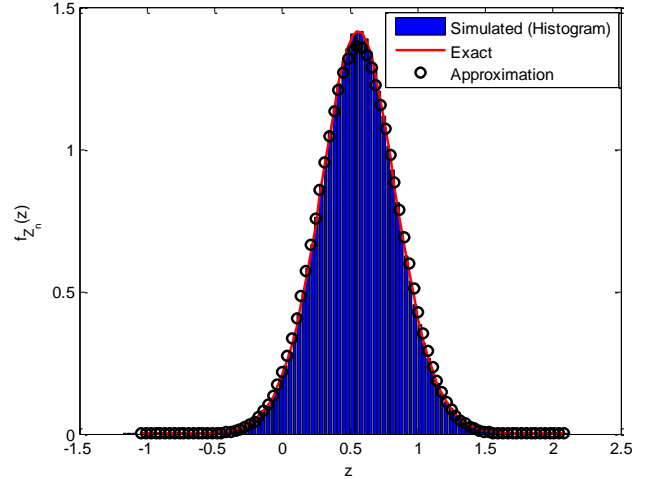


Fig. 2. A histogram of 10^6 realizations of Z_{10} for $\mathcal{E}_s/N_0 = 3/4$ along with the exact and approximate pdfs.

V. INFORMATION OUTAGE PROBABILITY

Let $R_2 = k/n$ represent the code rate in bits per symbol, where k is the number of information bits per message and n is the number of modulated symbols per codeword. When natural logarithms are used, the rate should be converted to units of nats per symbol, i.e. $R_e = \log(2)R_2$. The information-outage probability P_o is the probability that the mutual-information random variable is less than the code rate

$$\begin{aligned}
P_o &= P[Z_n \leq R_e] \\
&= F_{Z_n}(R_e), \tag{18}
\end{aligned}$$

which, as indicated by the second line, is the CDF of Z_n evaluated at the rate. This statistic is also known as the *distribution of mutual information* or the *mutual-information spectrum* [5].

Using (13), the information-outage probability can be expressed in terms of the CDF of W_n

$$\begin{aligned} P_o &= P \left[\log \left(1 + \frac{\mathcal{E}_s}{N_0} \right) + W_n \leq R_e \right] \\ &= F_{W_n} \left(R_e - \log \left(1 + \frac{\mathcal{E}_s}{N_0} \right) \right). \end{aligned} \quad (19)$$

As derived in [7], the Bessel-K variable W_n has CDF

$$F_{W_n}(w) = 1 - \sum_{l=0}^{n-1} \frac{\Gamma(n+l)\Gamma(n-l, \sqrt{2}w/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)} 2^{-n-l} \quad (20)$$

for $w \geq 0$ and

$$F_{W_n}(w) = \sum_{l=0}^{n-1} \frac{\Gamma(n+l)\Gamma(n-l, \sqrt{2}w/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)} 2^{-n-l} \quad (21)$$

for $w < 0$, where

$$\Gamma(a, z) = \int_z^{\infty} t^{a-1} e^{-t} dt$$

is the upper incomplete Gamma function [14].

The CDF of the Gaussian approximation \tilde{Z}_n to Z_n can be found using the Q-function

$$F_{\tilde{Z}_n}(z) = Q \left(\frac{\log \left(1 + \frac{\mathcal{E}_s}{N_0} \right) - z}{\sqrt{\frac{2\mathcal{E}_s}{n(\mathcal{E}_s + N_0)}}} \right). \quad (22)$$

For sufficiently large n , the information-outage probability can be approximated by $P_o \approx F_{\tilde{Z}_n}(R_e)$.

A comparison of the CDF using the exact expression and the Gaussian approximation is shown in Fig. 3. In this curve, $n = 20$ and $\mathcal{E}_s/N_0 = 3/4$. As can be seen, the two curves coincide quite closely.

Rather than fixing the SNR and varying the rate, it is often more insightful to fix the rate and plot information-outage probability as a function of SNR. In Fig. 4 the exact and approximate outage probabilities are shown as a function of $\mathcal{E}_b/N_0 = (\mathcal{E}_s/N_0)/R_2$ for codewords of binary rate $R_2 = 2/3$ over several different blocklengths. At low SNR, the Gaussian approximation seems to be a good fit, even for smaller blocklengths. However at higher SNR, the Gaussian approximation tends to diverge from the exact outage curve when n is small. As the blocklength increases, the approximation becomes a tighter fit. Based on this figure, it can be concluded that the Gaussian approximation is accurate for $P_o \geq 10^{-3}$ when $n \geq 60$.

Further insight into the tightness of the Gaussian approximation is provided by Table I, which shows the minimum SNR required to achieve outage probabilities $P_o = \{10^{-3}, 10^{-4}, 10^{-5}\}$ and $n = \{6, 15, 30, 60, 120\}$. In agreement with Fig. 4, we see that the approximation diverges from the exact curve with decreasing P_o and n . However, even for blocklengths as small as $n = 120$ and outage probabilities as small as $P_o = 10^{-5}$, the difference in the two curves is

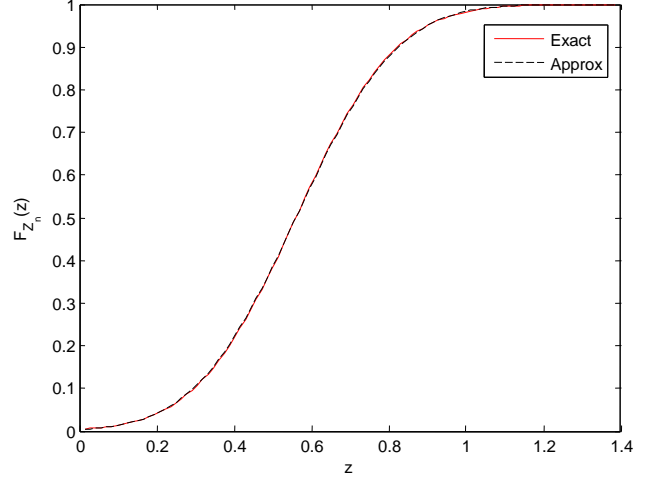


Fig. 3. CDF of the exact and approximate mutual-information variables for $n = 20$ and $\mathcal{E}_s/N_0 = 3/4$.

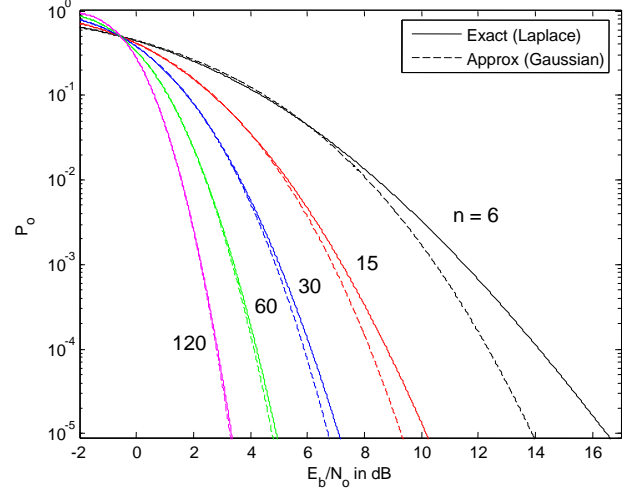


Fig. 4. Information outage probability vs. SNR for various blocklengths at binary rate $R_2 = 2/3$.

less than 0.1 dB. Due to the Central Limit Theorem it is safe to assume that this gap will decrease as the blocklength n increases until it becomes infinitesimally small as $n \rightarrow \infty$, when the AWGN channel capacity is reached. The SNR required to achieve $P_o = 10^{-4}$ is also shown in Fig. 5 for a range of blocklengths, again illustrating the tightness of the Gaussian approximation.

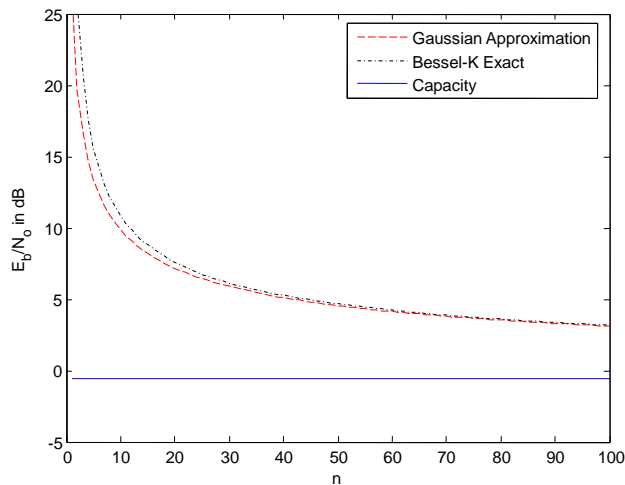
VI. BOUNDING THE ACHIEVABLE ERROR PROBABILITY

Having established the information-outage probability as a function of blocklength, it is worth discussing the meaning of the quantity. As seen in Fig. 4, the information-outage probability becomes steeper with increasing n . In the limit, as $n \rightarrow \infty$, the information-outage probability when plotted as a function of SNR for a fixed rate will converge to a vertical line

TABLE I

MINIMUM SNR (E_b/N_o) IN DB REQUIRED FOR VARIOUS OUTAGE PROBABILITIES AND BLOCKLENGTHS (BINARY RATE = 2/3)

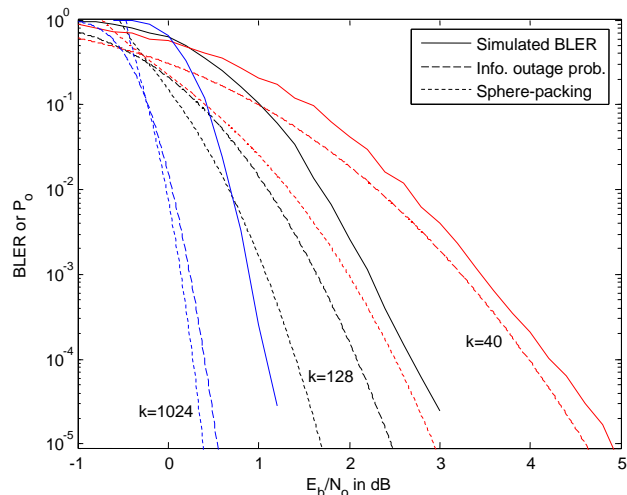
Blocklength	N = 6			N = 15			N = 30			N = 60			N = 120		
P_o	10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}
Exact	11.53	14.14	16.51	7.22	8.8	10.18	5.02	6.14	7.11	3.43	4.23	4.92	2.28	2.85	3.35
Approximation	10.51	12.35	13.9	6.9	8.21	9.3	4.89	5.9	6.74	3.38	4.13	4.77	2.26	2.82	3.29
dB Difference	1.02	1.79	2.61	0.32	0.59	0.88	0.13	0.24	0.37	0.05	0.1	0.15	0.02	0.03	0.06

Fig. 5. Minimum SNR (E_b/N_o) in dB required to achieve $P_o = 10^{-4}$ at binary rate $R_2 = 2/3$ as a function of blocklength, n .

at the location of the ergodic capacity limit for that rate. From the Shannon capacity theorem, it is well known that infinitely long codes exist that achieve arbitrarily low error rates for SNRs that are greater than the capacity limit. However, what if n is finite? Do codes exist that can achieve a block-error rate that approaches the information-outage probability?

To provide some insight into these questions, it is worth comparing the block-error rate of some known good codes against the information-outage probability. To do this, we have simulated the performance of the turbo code from the UMTS-LTE (long term evolution) standard [15] with QPSK modulation. The binary rate of the code with QPSK modulation is $R_2 = 2k/(3k + 12) \approx 2/3$ bits per symbol, where k is the number of data bits per block. The LTE standard supports 188 distinct values of k ranging from 40 to 6144. In Fig. 6, block-error rates of the LTE code with several values of k are compared against the corresponding information-outage probability (produced using the Gaussian approximation) as well as the sphere-packing bound which will be discussed later. For the two longer blocks, P_0 is approximately 0.7 dB better than the BLER achieved with the actual turbo code, while for the shorter block, P_0 is only 0.3 dB better. It is interesting to note that the slope of each outage probability curve is approximately the same as the corresponding turbo code BLER.

While Fig. 6 suggests that the information-outage probability is a useful quantity for predicting the performance of good

Fig. 6. Block error rate (BLER) of the turbo code from the LTE standard and the information-outage probability of the same blocklength. k is the number of bits per block and the number of symbols per block is $(3k + 12)/2$.

finite-length codes, it is not clear if a code exists that can achieve a BLER equal to the information-outage probability. To test the usefulness of the information-outage probability further, we compare it with previously derived bounds on error performance, namely Shannon's sphere-packing and random coding bounds [9], [10], [16], and Feinstein's lemma [8].

Shannon's sphere-packing bound is a lower bound on code performance which is based on n -dimensional Euclidian space (for a length- n codeword). A sphere in n -dimensional Euclidian space can be packed with $M = 2^k$ n -dimensional cones, one for each possible codeword. The codeword-error probability is then found to be greater than or equal to the probability that an n -dimensional Gaussian random vector falls outside a specific cone. The sphere-packing bound is shown along with the information-outage probability and the BLER of the LTE turbo code in Fig. 6.

Feinstein's lemma [8] (see also [5], [7]) states that a length- n , rate- R_e code exists that can achieve a codeword error probability ϵ that satisfies

$$\epsilon \leq P \left[\frac{1}{n} \log \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})f_{\mathbf{X}}(\mathbf{x})} \leq R + \gamma \right] + e^{-n\gamma}, \quad (23)$$

for any real valued $\gamma : 0 < \gamma < 1$. For the AWGN channel, this is related to the CDF of Z_n by

$$\epsilon \leq P_{Z_n} [R + \gamma] + e^{-n\gamma}. \quad (24)$$

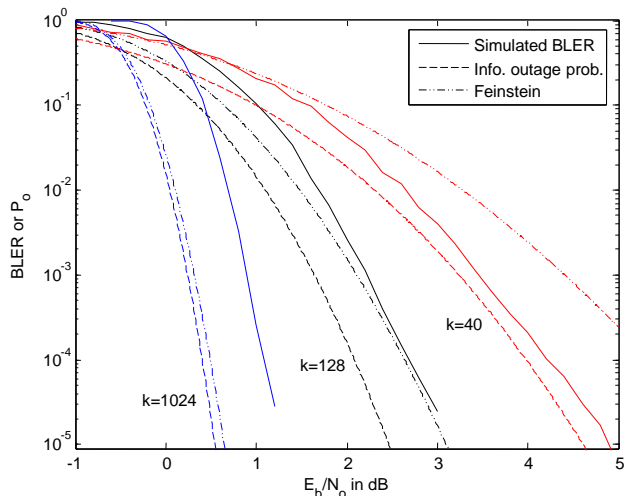


Fig. 7. The same curves from Fig. 6 (without sphere-packing bound) with the maximal error rate determined from Feinstein’s lemma superimposed.

Since $P_{Z_n}[R + \gamma] \geq P_{Z_n}(R)$ for $\gamma > 0$, this bound is incrementally larger than the outage probability. Note that ϵ is an upper bound on achievable codeword error rate performance, so codes may exist that can perform better than this bound. The bound is made sharpest by minimizing the function over γ .

The bound on maximal error rate computed using Feinstein’s lemma is shown in Fig. 7. As in Fig. 6, the LTE turbo code was used to produce the simulated BLER curves and the rate is $R_2 = 2k/(3k + 12) \approx 2/3$. As can be seen, the bound due to Feinstein’s lemma is rather loose for the smallest blocksize. While the lemma predicts that there is a code with an error rate no worse than ϵ , the LTE turbo code has actually beaten this bound at $n = (3 * 40 + 12)/2 = 66$. However, as n increases, the Feinstein bound approaches the information-outage probability. This reaffirms the information-outage probability’s utility as a performance limit for moderate blocklengths (e.g. $n > 1000$, which is typical for practical systems).

A comparison of each bound as a function of blocklength is shown in Fig. 8. In the figure the following bounds are shown in order of worst to best: Feinstein, information-outage, random coding [9], [10], and sphere-packing.

VII. CONCLUSIONS

The mutual information between an AWGN channel’s input and output is a random variable which converges to the ergodic capacity in the limit of large blocklength. For finite blocklengths, the mutual information follows a Bessel-K distribution, though it can be accurately modeled as a Gaussian for moderate lengths. The outage probability, which is the CDF of the mutual information, provides some insight into the achievable block-error rate of good finite-length codes. Due to Feinstein’s lemma, it is known that codes exist whose maximal error rate is incrementally larger than the outage probability.

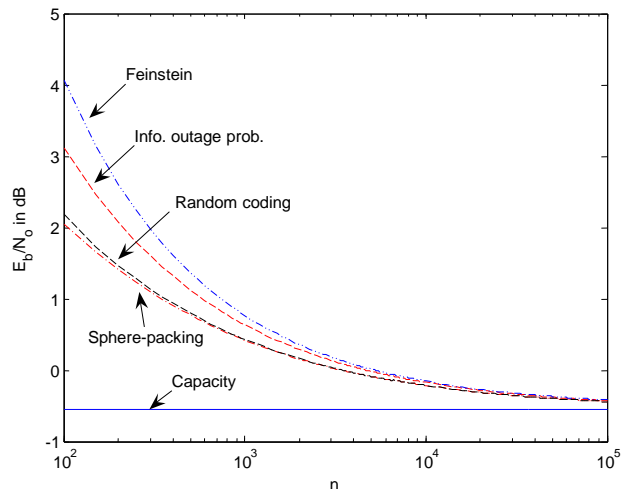


Fig. 8. Minimum SNR (E_b/N_o) in dB required to achieve $P_o = 10^{-4}$ at binary rate $R_2 = 2/3$ as a function of blocklength, n .

The gap between the outage probability and the maximal error rate due to Feinstein’s limit closes with increasing blocklength. Thus, the outage probability is a good practical metric for predicting the performance of moderate-length codes.

REFERENCES

- [1] L. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.
- [2] R. Knopp and P. A. Humblet, “On coding for block fading channels,” *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.
- [3] A. Guillén i Fàbregas and G. Caire, “Coded modulation in the block-fading channel: Coding theorems and code construction,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 262–271, Jan. 2006.
- [4] S. Verdù and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
- [5] T. S. Han, *Information Spectrum Methods in Information Theory*. Springer, 2003.
- [6] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holiday-Day, 1964.
- [7] J. N. Laneman, “On the distribution of mutual information,” in *Proc. Workshop on Information Theory and its Applications*, San Diego, CA, Feb. 2006.
- [8] A. Feinstein, “A new basic theorem of information theory,” *IRE Trans. Info. Thy.*, vol. 4, pp. 2–22, Sept. 1954.
- [9] C. E. Shannon, “Probability of error for optimal codes in a gaussian channel,” *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.
- [10] S. Dolinar, D. Divsalar, and F. Pollara, “Code performance as a function of block size,” JPL TDA Progress Report, Tech. Rep., May 1998.
- [11] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.
- [12] J. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, Inc., 2001.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [14] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. Dover, 1972.
- [15] Third Generation Partnership Project, “Evolved universal terrestrial radio access (E-UTRA): Multiplexing and channel coding,” *3GPP TS 36.212 version 8.0.0*, Sept. 12 2007.
- [16] A. Valembois and M. P. Fossorier, “Sphere-packing bounds revisited for moderate block lengths,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 2998–3014, Dec. 2004.