Delay-Constrained Capacity and Optimal Block Codes

Xiao-Yu Hu
IBM Research, Zurich Research Laboratory, Säumerstrasse 4, 8803, Rüschlikon, Switzerland
Email: xhu@zurich.ibm.com

Abstract— **The Noisy Channel Coding Theorem discovered by C. E. Shannon assumes infinite coding latency. The objective of this work is to identify the maximal achievable (transmit) rates over noisy, delay-constrained channels, referred to as (ϵ, n)-capacity C_ϵ^n with ϵ denoting target error probability and n coding latency (viz. block length). We investigate a family of block codes based on a probabilistic construction that approaches delay-constrained capacity closely and provably achieves the Shannon limit over an additive white Gaussian noise (AWGN) channel. In the full version of this paper we present an improved construction of a probabilistic code with *correlated* codewords, enhancing its asymptotic distance by introducing a specific amount of correlation between codewords. Analytical results show that, if the correlation coefficients are chosen uniformly to be $-1/(M-1)$, where M denotes the number of codewords, the corresponding probabilistic code is asymptotically (in the sense of block length) the "best-$d_{\min}$" code.**

Keywords— **AWGN, Shannon limit, probabilistic codes, coding latency, (ϵ, n)-capacity**

I. Introduction

Ideas presented in the special issue [1] on codes and graphs and iterative algorithms have allowed us to approach the Shannon limit of an additive white Gaussian noise (AWGN) channel to within hundredths of a decibel at the expense of very long block lengths. However, in most applications where the system delay is strictly limited, approaching the Shannon limit becomes problematic. To construct good block codes, the major parameters of interest are the probability of block (word) decoding error p_w, the code block length n, and the rate R. The Shannon Noisy Channel Coding Theorem [2] states that, if R is less than the Shannon limit C, no matter how close they are, surely there exist codes for which the word error probability p_w becomes small exponentially with increasing n. There are, of course, prices to be paid for increasing the block length, one of which is *coding latency*. At the transmitter side, the first information bit in a block of incoming data stream must generally be delayed by n samples (or symbols) before a codeword can be formed, and at the receiver it is the same case that decoding also requires a complete codeword. The block length n is thus referred to as coding latency. Note that the coding latency differs from the processing delay in that it is inherent in a coding scheme and can not be reduced by increasing the processing capability.

II. (ϵ, n)-Capacity

The definition of channel capacity involving coding latency n and a target block error probability ϵ is the following.

Definition 1: Consider an (n, M) code with a block length n, M codewords, and a block error probability not greater than ϵ. $R \geq 0$ is an (ϵ, n)-achievable rate if, for every $\delta \geq 0$, there exists at least such a (n, M) code with rate

$$\frac{\log_2 M}{n} \geq R - \delta.$$

The maximum (ϵ, n)-achievable rate is called the (ϵ, n)-capacity C_ϵ^n. The Shannon limit C is defined as the maximal rate that is (ϵ, n)-achievable for all $0 < \epsilon < 1$ and for all positive integer n. It follows immediately from the definition that

$$C = \lim_{n \to \infty} \lim_{\epsilon \to 0} C_\epsilon^n.$$

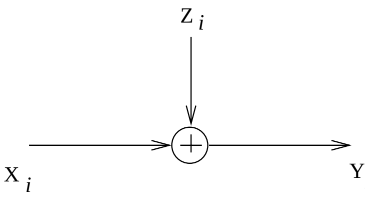

Fig. 1. The discrete-time additive white Gaussian noise channel.

For the sake of simplicity, we concentrate on a discrete-time memoryless AWGN channel (as shown in Fig. 1), which usually serves as a basic analysing tool for all other kinds of non-ideal channels. Before going on, we cite the well-known Shannon channel capacity formula — the supremum of all rates R for which there exists at least one code with vanishing error probability, that is

$$C = \max_{p_X} I(X; Y), \tag{1}$$

where p_X denotes the probability distribution of the real-valued channel input X and Y is the real-valued channel output. This formula holds for any ergodic memoryless channels [3]. Specifically, the Shannon limit of a Gaussian channel with power constraint P and noise variance N is [1]

$$\begin{aligned} C &= \max_{EX^2 \leq P} I(X; Y) \\ &= \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right) \end{aligned} \tag{2}$$

and the maximum is attained only when $X \sim \mathcal{N}(0, P)$, where $\mathcal{N}(0, P)$ is a zero-mean Gaussian distribution of variance P.

Evaluating the (ϵ, n)-capacity is not as simple as calculating the Shannon limit, but we can nevertheless employ known analytical results to obtain an upper-bound as well

[1] In this paper we deal with only real-valued channels; however the results can easily be extended to complex channels, i.e., bandpass signals represented in the equivalent complex baseband. In this case, the factor of 1/2 in (2) does not appear.

as a lower-bound on the (ϵ, n)-capacity. A classic lower-bound on the error probability for codes of a specific block size is the sphere-packing bound developed by Shannon [4]. This bound has been recently employed as a useful tool to evaluate the "imperfectness" of turbo codes [5]. The problem posed by Shannon is to estimate, as well as possible, the probability of error for a "best" code of length n containing M codewords, each of power P and perturbed by Gaussian noise of variance N. We denote this minimal or optimal probability of error by $p_w^{\text{opt}}(M, n, \sqrt{P/N})$. The sphere-packing bound is equal to the probability that the output sequence Y of the AWGN channel will not be confined to the cone with a solid half-angel θ centralized with respect to the transmitted codeword, which can be expressed in the form

$$p_w^{\text{opt}}(M, n, \sqrt{P/N}) \geq Q_{sp}(\theta) = \int_\theta^\pi \frac{(n-1)(\sin\phi)^{n-2}}{2^{n/2}\sqrt{\pi}\Gamma(\frac{n+1}{2})} \cdot \int_0^\infty r^{n-1} e^{-(r^2 + nA^2 - 2r\sqrt{n}A\cos\phi)/2} dr d\phi, \tag{3}$$

where A is the squared root of the signal-to-noise ratio (SNR), i.e., $\sqrt{P/N}$, $\Gamma(p)$ is the Gamma function $\int_0^\infty t^{p-1}e^{-t}dt$, and θ is the root of the following equation:

$$\int_0^\theta \frac{n-1}{n} \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n+1}{2})\sqrt{\pi}} (\sin\phi)^{n-2} d\phi = 2^{-nR}. \tag{4}$$

For moderate to large n, (3) can be approximated with great accuracy by

$$Q_{sp}(\theta) \approx \frac{[G(\theta)\sin\theta e^{-(A^2 - AG(\theta)\cos\theta)/2}]^n}{\sqrt{n\pi}\sqrt{1 + G^2(\theta)\sin\theta}[AG(\theta)\sin^2\theta - \cos\theta]}, \tag{5}$$

where $G(\theta) = (1/2)[A\cos\theta + \sqrt{A^2\cos^2\theta + 4}]$, and (4) becomes, asymptotically,

$$\frac{\Gamma(\frac{n}{2}+1)(\sin\theta)^{n-1}}{n\Gamma(\frac{n+1}{2})\sqrt{\pi}\cos\theta} = 2^{-nR}. \tag{6}$$

The sphere-packing lower bound on word error probability would be reached with equality only if the code were a *perfect* code for the channel, i.e., if equal-size non-intersecting cones could be drawn around every codeword to completely fill the n-dimensional space. Such a partitioning is clearly possible only for n=1 or 2, if $M > 2$ [4]. It is very plausible intuitively that any actual code would have a higher probability of error than a sphere-packing code. Recognizing the monotonically increasing error probability with more codewords, the rates specified by the sphere-packing bound can naturally be utilized to *upper-bound* the (ϵ, n)-capacity.

Shannon also computed an upper bound on word error probability by a spherical "random coding" method [4]. The random coding bound gives an expression for the ensemble average word error probability, averaged over the ensemble of all possible spherical codes, where each codeword is selected independently and completely at random, subject to an equal energy constraint. As n grows large enough, an asymptotic formula of the random coding bound turns out to be the sphere-packing bound multiplied by a factor essentially independent of n, that is

$$\begin{aligned} & p_w^{\text{opt}}(M, n, \sqrt{P/N}) \leq Q_{rc}(\theta) \\ = \quad & Q_{sp}(\theta) + 2^{nR}\int_0^\theta \frac{\Gamma(\frac{n}{2}+1)(\sin\phi)^{n-1}}{n\Gamma(\frac{n+1}{2})\pi^{1/2}\cos\phi}\sqrt{\frac{n}{\pi}} \\ & \cdot \frac{[G(\theta)\sin\phi\exp(-\frac{P}{2N} + \frac{1}{2}\sqrt{\frac{P}{N}}G(\theta)\cos\phi)]^n}{\sqrt{1 + G^2(\theta)}\sin^2\phi} d\phi \\ \approx \quad & Q_{sp}(\theta)\left(1 + \frac{AG(\theta)\sin^2\theta - \cos\theta}{2\cos\theta - AG(\theta)\sin^2\theta}\right). \end{aligned} \tag{7}$$

Since the average error probability over the ensemble of spherical random codes satisfies (7), it is clear that at least one code in the ensemble must have a sufficiently small error probability, i.e. at least one code of block length n meets the target error probability ϵ with a certain rate, which in turn, gives rise to a *lower bound* on the (ϵ, n)-capacity. It is worth emphasizing that, in the case of moderate to large n, the multiplying factor in (7) is just a little over unity; the sphere-packing and the random-coding bounds are close together, thereby yielding a sharp estimate of the (ϵ, n)-capacity.

The significance of the definition of (ϵ, n)-capacity can be seen from the following numerical example. Consider an AWGN channel on which we wish to transmit information with rate 1 bit per sample, for which the minimum SNR specified by the Shannon limit is 3.0. By applying the sphere-packing bound, Fig. 2 shows that, for the same code rate, the minimum threshold for reliable communication (in the sense of achieving a target error probability) is significantly higher than the corresponding Shannon limit, provided that the code block length is constrained to a relatively small size. For some real-time applications where large delay is not tolerable, the Shannon limit does not convey much useful information, but the (ϵ, n)-capacity reveals the ultimate limit in such cases instead. It is suggested that, even if a code operates far from the Shannon limit it might perform nearly as well as the best code possible of the same length.

A quantitative overview of (ϵ, n)-capacity (upper bound) versus the Shannon limit is exhibited in Fig. 3. It should not be surprising that, if the code block length is less than 10^4, only the rates significantly lower than the Shannon limit are achievable. For example, codes of block size 100 have a penalty of 0.3 bits per sample with $p_w = 10^{-3}$, and even a penalty of over 0.5 bits per sample with $p_w = 10^{-10}$, as compared with the Shannon limit. The Shannon limit can be approached within 0.05 bits per sample only for block sizes 100,000 and greater. Again, it is evident that, not the Shannon limit, but the (ϵ, n)-capacity should be employed in evaluating a practical coding scheme with finite block lengths.

Plotted in Figs. 4 and 5 are comparisons of the sphere-packing bound and the random-coding bound with varying block sizes. In particular, information on the upper and

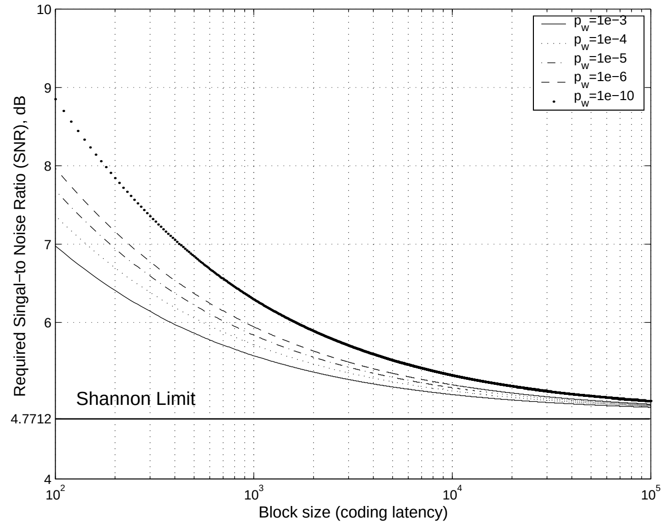

Fig. 2. The minimum required signal-to-noise ratio (SNR) by the Shannon sphere-packing bound for codes with varying block size n and rate 1 bit per sample, operating over a continuous-input AWGN channel at $p_w = 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-10}$, respectively.

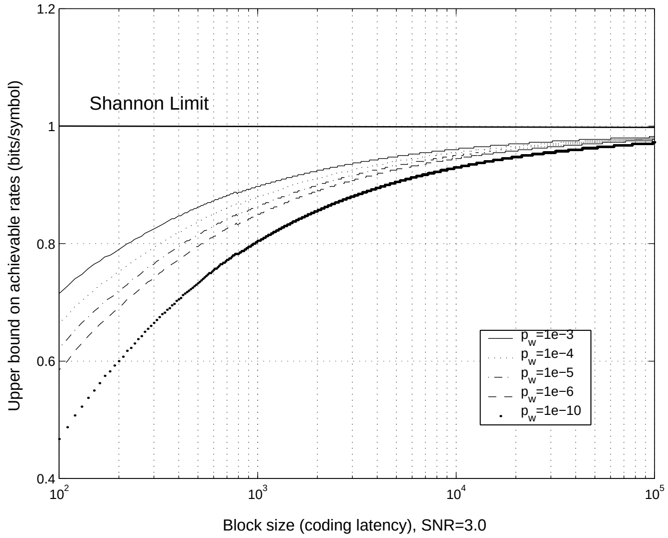

Fig. 3. Upper bound on (ϵ, n)-capacity by the sphere-packing bound for codes with varying block size n, operating over a continuous-input AWGN channel at a SNR of 3.0 (4.7712 dB), and $p_w = 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-10}$, respectively.

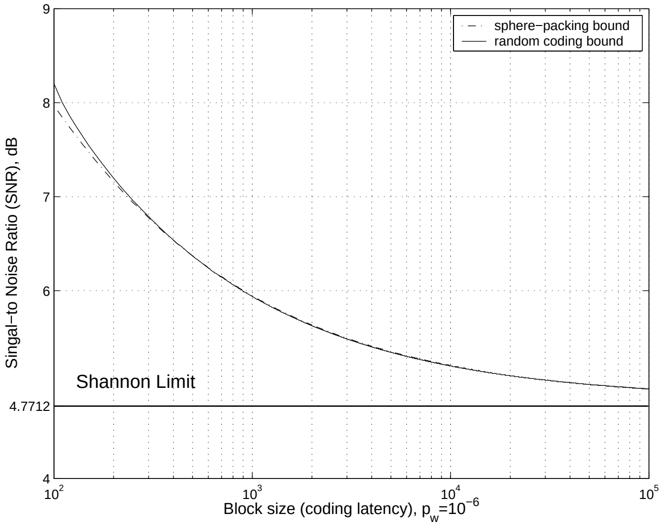

Fig. 4. Signal-to-noise ratio by the spherical random-coding bound (as compared with the sphere-packing bound) for codes with varying block size n and rate 1 bit per sample, operating over a continuous-input AWGN channel at $p_w = 10^{-6}$.

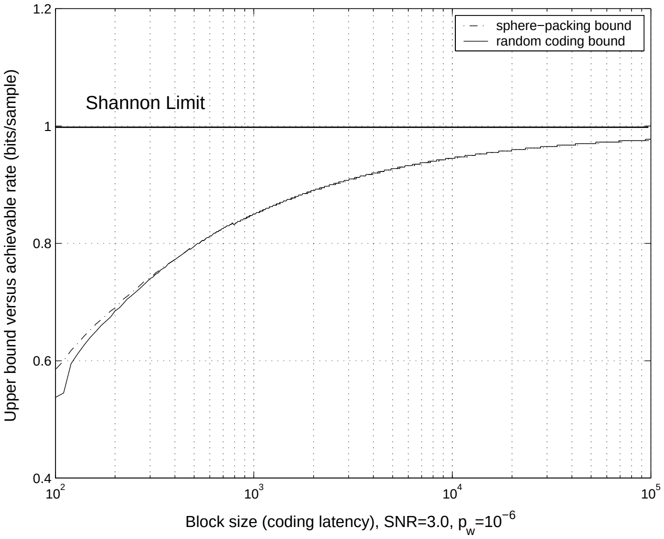

Fig. 5. Achievable rates by the spherical random-coding bound (as compared with the sphere-packing bound) for codes with varying block size n, operating over a continuous-input AWGN channel at a SNR of 3.0 and $p_w = 10^{-6}$. Note that the discontinuity in the short blocklength region is caused by numerical difficulty.

lower bounds on the (ϵ, n)-capacity is shown. In this specific setting and for $n \geq 100$, the upper and lower bounds on the (ϵ, n)-capacity are close together enough, thereby delivering precise information concerning the (ϵ, n)-capacity, whereas when $n < 100$, the upper bound and the lower bound are apart and thus the question of determining (ϵ, n)-capacity for this blocklength region still remains open.

Additional insight into the implications of Figs. 4 and 5 may be obtained by re-examining the definitions of the sphere-packing bound and the spherical random-coding bound. As we know, the performance limit corresponding to the sphere-packing bound would be reached with equality only if the code were a *perfect* spherical code for the continuous-input AWGN, i.e., if equal-sized cones could be drawn around every codeword so as to completely fill the n-dimensional space without intersecting. Actually this is impossible for all $n > 2$. On the other hand, we demonstrate that the spherical random-coding bound is virtually indistinguishable from the sphere-packing bound for block sizes greater than a few hundred. Therefore it is tempting to construct probabilistic codes to "approximate" the ensemble of spherical random codes, rather than search for a deterministic "best" code which does not exist at all.

III. Probabilistic Codes

Definition 2: *Probabilistic code with independent codewords*—An (n, M) probabilistic code for a certain channel with power constraint P consists of the following:

- For each encoding block, generate M codewords $X_1^n, X_2^n, \ldots, X_M^n$, that satisfy the power constraint P, i.e., for every codeword

$$\sum_{i=1}^{n} x_{i,w}^2 \leq nP, \qquad w = 1, 2, \ldots, M, \tag{8}$$

where codewords X_w^n are created by independent identically distributed random variables X_w, subject to a common distribution P_X [3] maximizing input-output mutual information $I(X;Y)$, e.g., for an AWGN channel, $P_X \sim \mathcal{N}(0, P)$.

- An encoding function $X : \{1, 2, \ldots, M\} \rightarrow \mathcal{X}^n$, selecting one codeword from the codebook and passing it through the channel.
- A synchronization scheme between the encoder and decoder that guarantees that the decoder will generate an exact copy of the codebook of encoder for each block. In other words, the receiver has perfect knowledge of the random sources X_w.
- A decoding function

$$g : \mathcal{Y}^n \rightarrow \{1, 2, \ldots, M\}, \tag{9}$$

which is a deterministic rule that assigns an estimate to each possible received vector.

The definition of probabilistic code has the effect of "combined coding and modulation" (baseband), i.e. the encoder feeds its output directly to the AWGN channel. In our notation, the transmission rate is measured as $R = \frac{\log_2 M}{n}$, which can be made readily larger than 1 bit per sample simply by generating a large number of codewords such that $M = 2^{nR}$.

The probabilistic code is inherently time-varying, i.e., the codebook varies from block to block and the codeword with the same index w does not remain the same for different blocks. This scheme differs from the conventional concept wherein a *deterministic* codebook is selected once and used repetitively. The time-varying nature ensures that the channel input resembles a stochastic process with an appropriate distribution, which maximizes the mutual information of channel input and output.

The probabilistic code should not be confused with the standard method of proof of coding theorems based on a *random-coding argument.* Whereas a probabilistic code constitutes a communication technique, a random-coding argument is a proof technique often used to establish the existence of a (single) deterministic code which yields good performance on a specific channel without actually constructing the code. This is done by introducing a probability mass function (pmf) on an ensemble of codes, computing the corresponding average performance over such an ensemble, and then invoking the argument to show that if this average performance is good, then there must exist at least one code in the ensemble with good performance. In contrast, a probabilistic code constitutes a communication technique, the implementation of which requires the availability of a common source of randomness at the transmitter and receiver.

It can be proved that if block length n tends to infinity, the probabilistic code is capable of achieving the Shannon limit by means of a suboptimal decoding procedure—*typical set decoding* [6]. Intuitively, the performance of a probabilistic code with independent codewords approaches closely the average performance of the ensemble of spherical random codes for moderate to large block lengths. This suggests that the probabilistic code is capable of approaching delay-constrained capacity closely except for very small block lengths.

Due to the space limitations we cannot give a detailed account of the probabilistic code with correlated codewords, but we summarize the main results instead. Using a linear transformation, we derive a new construction of a probabilistic code with correlated codewords, thus improving its asymptotic distance by introducing a controlled amount of correlation between codewords [7]. Analytical results show that, if the correlation coefficients are chosen uniformly to be $-1/(M-1)$, the corresponding probabilistic code is asymptotically (in the sense of block length) the "best-$d_{\min}$" code.

IV. Remarks

The decoding complexity of probabilistic codes with independent or correlated codewords grows exponentially with the block lengths, while they are the right codes capable of approaching the (ϵ, n)-capacity closely except for very small block lengths. The theoretical characterization of the (ϵ, n)-capacity and its corresponding optimal codes offers insights into how the optimal block codes look like and what is the maximum achievable rate under a critical coding latency constraint. The (ϵ, n)-capacity can be used as a natural criterion against how good a practical coding scheme is with a finite block length.

References

[1] "Special issue on codes and graphs and iterative algorithms," *IEEE Trans. Inform. Theory,* vol. 47, pp. 493-849, Feb. 2001.

[2] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 and pp. 623-656, July and Oct. 1948.

[3] S. Shamai (Shitz), and S. Verdu, "The empirical distribution of good codes," *IEEE Trans. Inform. Theory,* vol. 43, pp. 836-846, May 1997.

[4] C.E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.,* vol. 38, pp. 611-656, May 1959.

[5] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," *TMO Progress Report 42-133,* Jet Propulsion Laboratory, Pasadena, California, pp. 1-23, May 1998.

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

[7] X.-Y Hu, "Delay–constrained capacity and probabilistic codes," *IBM Research Report,* 2001.