The Performance of Block Codes

Elwyn Berlekamp

In his classic 1948 paper [10], Claude Shannon introduced the notions of communication channels and codes to communicate over them. During the following two decades, he remained active in refining and extending the theory. One of Shannon's favorite research topics was the fundamental performance capabilities of long block codes. In the 1950s and 1960s this topic also attracted the active involvement of several of Shannon's distinguished colleagues both at Bell Telephone Laboratories and at MIT, including P. Elias, R. M. Fano, R. G. Gallager, E. N. Gilbert, R. W. Hamming, I. M. Jacobs, B. Reiffen, D. Slepian, and J. M. Wozencraft, and several graduate students, including G. D. Forney and me. The work culminated in a book by Gallager [6] and in a sequence of two "Information and Control" papers by Shannon, Gallager, and Berlekamp [11]. In this article I present an overview of some of the more salient results and their impact.

Error Exponent Theory

A discrete memoryless *channel* has finite input and output alphabets related by any given matrix of transition probabilities. A *block code* of length N consists of M codewords, each of which is a sequence of N symbols selected from the channel's input alphabet. The *rate* R of the code in natural units is defined as $R = (\ln M)/N$. A message source selects one of the M equiprobable codewords and transmits it over the channel. The received word is determined, symbol by symbol, according to the row of the transmission matrix specified by the corresponding input symbol. A *decoder* examines the received word, computes the a posteriori probabilities of each of the M possible input words, and rank orders them. The decoder is considered correct if the transmitted codeword is the most likely. In a revealing generalization first studied by Elias [3], a "list of L" decoder is considered correct if the transmitted codeword appears anywhere among the L most likely choices.

For any fixed channel, one seeks bounds on $P_e(N, M, L)$, the *error probability* of the best code of the specified codebook size M and list size L. (In the most general case, in which the probability of decoding error may depend on which codeword is selected, one evaluates the code according to one of its worst codewords.) Even today, the best codes are known for relatively few channels, and even then only for very high or very low coderates and/or relatively short block lengths. Many of those that are known have interesting connections with other areas of mathematics.

Shannon was interested in getting bounds on the behavior of

$$P_e(N, \lfloor \exp(NR) \rfloor, L)$$

for fixed L (often $L = 1$) and fixed R as N goes to ∞. For all R less than the channel capacity C, it was shown that P_e is an exponentially decreasing function of N. In particular, there is an *error exponent* $E_L(R)$ such that $P_e < \exp(-NE_L(R))$ for all sufficiently large N.

Elwyn Berlekamp is professor of mathematics at the University of California at Berkeley. His e-mail address is `berlek@math.berkeley.edu`.

In the 1950s and 1960s, significant efforts were made to determine the best possible error exponent, which one might hope to define as

$$E_L(R) = \lim_{N\to\infty} (-1/N) \ln P_e(N, \lfloor \exp(NR) \rfloor, L).$$

This definition measures information in natural units, called "nats". If the natural logarithm is replaced by the base 2 logarithm, then the units are converted from nats to bits. In either case, $E_L(R)$ is also called the channel's "reliability function". Since it is conceivable that there might be circumstances under which this limit might not exist, one interprets upper and lower bounds on $E_L(R)$ as lim sup and lim inf.

Upper bounds on $E_L(R)$ correspond to lower bounds on the probability of decoding error. To obtain such a bound for a given list size, a given code, and a corresponding decoding algorithm, one defines M sets of received words, each set consisting of all words that are list-decoded into a particular codeword. One introduces an appropriate weighting function on the received words and computes the "volume" (total weight) of each such set and the volume V of the entire space of all possible received words. For sufficiently symmetric channels, the weighting function is uniform, and the volume is simply a count of the number of points in the relevant set. Since each received word can belong to at most L sets, it follows that there must be at least one particular codeword whose corresponding decoding region has volume no greater than LV/M. The probability of the received word lying within an arbitrary region of this given volume is maximal when the region is a sphere. That probability can be computed and used to determine the probability of decoding error of an idealized "perfect" code, whose decoding regions partition the set of all possible received sequences into perfect spheres.

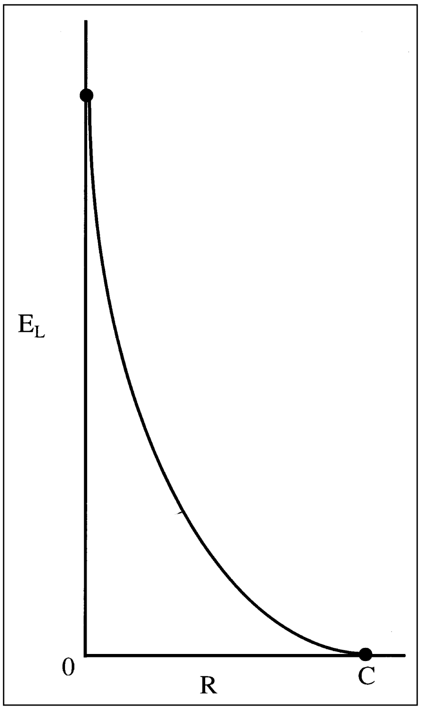

Figure 1.

This probability of decoding error for perfect codes can be used to calculate an upper bound on $E_L(R)$. For any fixed positive integer L and any fixed rate R between 0 and C, this technique yields as N goes to ∞ an upper bound on $E_L(R)$ that is independent of L. This bound is called either the "volume bound" or the "sphere-packing bound". As shown in Figure 1, this function is typically analytic. It is tangent to the R axis at $R = C$, and tangent to the E axis at $R = 0$. As we shall soon see, it happens to give the correct values of $E_\infty(R)$ for all values of R between 0 and C, assuming that we interpret $E_\infty(R)$ as the limit of $E_L(R)$ as L goes to ∞. Since each $E_L(R)$ is itself a limit as N goes to ∞, it is important that the two limits be taken in the appropriate order. If L and N were allowed both to go to ∞ in such a way that L were an exponentially growing function of N, we would get a different answer.

The volume bound remains valid even if the problem is modified to include a noiseless, delayless feedback link, which allows the encoder and the decoder to revise their code depending on previously received symbols.

Lower bounds on $E_L(R)$ correspond to upper bounds on the probability of decoding error. An important technique to get such bounds is the random coding argument introduced in Shannon's 1948 paper, in which all of the MN letters in the codebook are selected independently at random from the same input distribution, which is determined by a careful optimization that depends critically on the statistics of the noisy channel. As L goes to ∞, the random bounds approach a limit, which happens to coincide with the volume bound shown in Figure 1. The conclusion is that the volume bound is also the correct value of the function $E_\infty(R)$ for all rates R between 0 and C. This is a very remarkable result: when the list size is sufficiently large, random codes are virtually as good as perfect codes. Although the proof of this result for sufficiently symmetric channels is reasonably straightforward, it is much more difficult for the general asymmetric case. One problem is that the reliability function itself is not very tractable. The most successful approach is Gallager's formulation, which expresses the function in terms of a quantity ρ, which turns out to be the negative slope, $-dE/dR$.

For each finite value of L, there is a corresponding critical rate R_L such that the random coding bound on $E_L(R)$ agrees with $E_\infty(R)$ for coderates in the interval $R_L \le R < C$. For coderates below R_L, however, the random coding bound is a straight line of slope $-L$ that joins the function $E_\infty(R)$ tangentially at the point $R = R_L$. The random coding bounds for $L = 1, 2, 3, 4, 5$ are shown in Figure 2.

In most applications, a list of size $L = 1$ is required. The coding theorems show that at sufficiently high coderates ($R_{\text{crit}} = R_1 \le R < C$), random codes are exponentially as good as perfect codes. At lower rates, however, the error exponent for random codes becomes a straight line of slope -1 that veers off tangentially from the volume bound. With a list of size L, the error exponent for random codes remains equal to the volume bound for another interval of rates, $R_L \le R$, but then it veers off tangentially from the volume bound with slope $-L$.

The limit of $E_1(R)$ as R approaches 0 is known precisely, and for most channels it is strictly between the random bound and the volume bound. An upper bound on $E_1(0)$ is provided by

$$\lim_{M\to\infty}\lim_{N\to\infty} -(1/N)\ln P_e(N,M,1).$$

For sufficiently symmetric channels, this double limit is relatively easy to compute, because for any fixed M and very large N, the best codes can be constructed explicitly. Their key property is that they are equidistant, in the sense that the frequency with which any particular ordered pair of input symbols occurs in one pair of codewords is the same as the frequency with which it occurs in any other pair of codewords. Equidistant codes also happen to optimize the above double limit for arbitrary asymmetric channels, although that result was rather more difficult to prove.

A lower bound on $E_1(R)$ that coincides with the correct value at $R = 0$ is obtained by a technique called "expurgation". One first selects a code at random. Then one computes the probability of confusing each pair of codewords, as if $M = 2$ and these two words were the only words in the code. A threshold is carefully selected, and any pair of codewords whose confusion probability exceeds this threshold is expurgated, meaning that both of those codewords are removed from the code. The threshold is chosen so that on the average, at least half of the codewords survive the expurgation. What remains is an expurgated random code that contains no pair of too-easily confused codewords. As shown in Figure 3, this yields an error exponent $E_x(R)$, which improves the random coding bound in the region of rates $0 \le R < R_x$. The $E_x(R)$ curve joins the random coding bound tangentially at $R = R_x$.

Finally, there is another general result that allows any upper bound on $E_1(R)$ to be joined to any point on the volume bound by a straight line bound (also known as the SGB bound). As shown in Figure 3, the best such bound is attained by selecting points such that the straight line bound is tangent to the volume bound at its high-rate endpoint and to a low-rate bound at its low-rate endpoint. In Figure 3, the low-rate bound is the single-point double-limit upper bound on $E_1(R)$ described above.

One communications channel of considerable practical importance is the binary-input Gaussian noise channel. This channel has only two inputs, which are taken as the real numbers +1 and −1. The input number is called the "signal". The noise is taken as an additive real number having zero mean and given variance. The output, also a real number, is the sum of the signal and the noise. In order to make the channel discrete, the output range is commonly partitioned into a finite number of intervals. Usually this is done by uniformly quantizing the output over some finite range, allowing the extreme intervals at either end to run to ∞. For example, in the 2-input 8-output version of this model, the seven breakpoints between different output intervals are commonly set at −1.5, −1, −.5, 0, +.5, +1, and +1.5.

More intricate models of this channel assume that the noise is white and that an appropriate figure of throughput for the entire system is the number of nats/sec or bits/sec. This turns out to be maximized by making the time interval for each transmitted bit very small, even though this yields a discrete memoryless channel that is very noisy and that has a very small capacity per transmitted bit.

This motivation led to the study of "very noisy channels", whose probability transition matrices could be written as $P_{i|j} = P_i(1 + \epsilon_{i,j})$, where $\epsilon_{i,j}$ is small for all i and j. Reiffen [9] showed that the error exponent for any such channel depends on only one parameter, the capacity C. As shown in Figure 4, the graph of $E_\infty(R)$ is a parabola that attains a slope of −1 at the critical rate $R_1 = C/4$. At rate $R = 0$, the average distance upper bound coincides with the random coding bound. So for such channels, expurgation gains nothing, and the random coding bound is optimal at all rates between 0 and the capacity C. It coincides with the straight-line SGB bound at rates

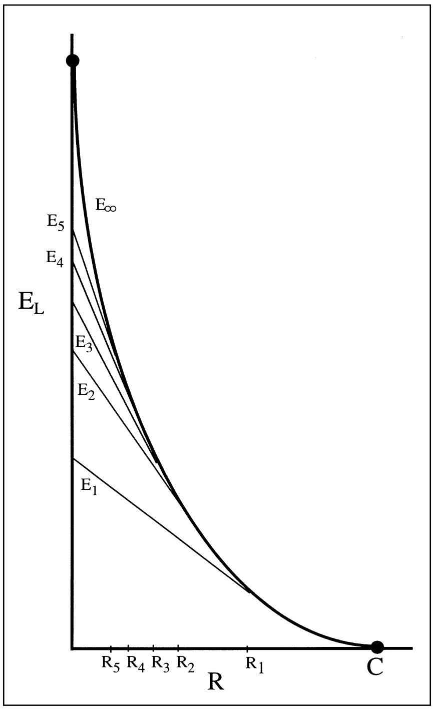

Figure 2.

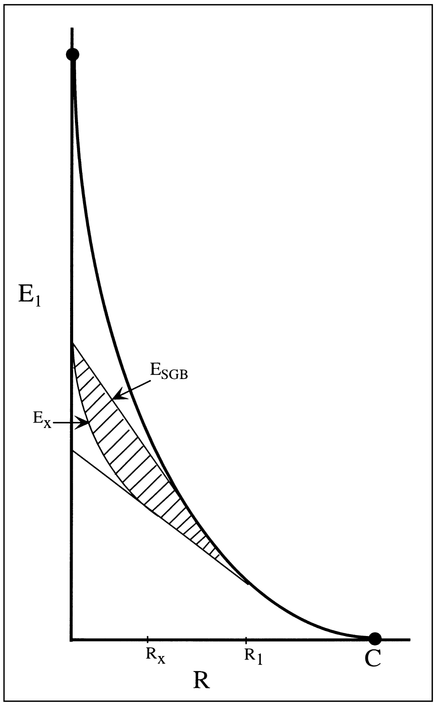

Figure 3.

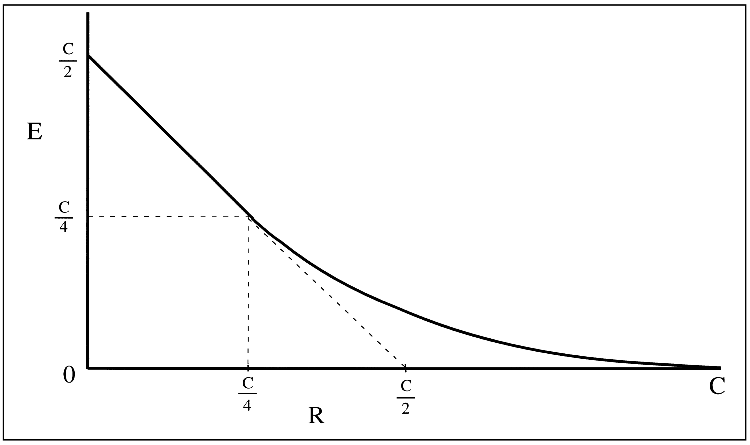

Figure 4.

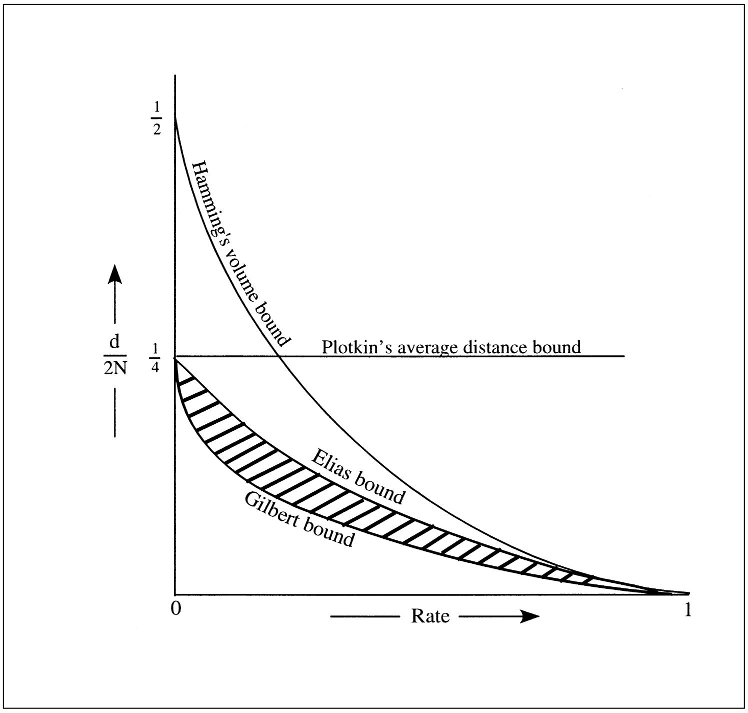

Figure 5.

$0 < R < C/4$ and with the volume bound at rates $C/4 < R < C$.

In general, sophisticated codes attain reliability at the cost of code redundancy and digital electronics required to implement the coding and decoding algorithms. Often the greatest potential system gains are attained by designing a channel that is too noisy to use without coding. Such a channel allows such significant benefits as lower power, faster speeds, and perhaps greater storage density. The local cost is a much higher raw error rate, which is eliminated from the system by the coding.

Many early communication and memory systems were designed and deployed before sophisticated coding was practical. Others were designed and deployed before the benefits of sophisticated coding were understood and appreciated by the system designers. In either case, such systems worked by placing heavy investments in power and in analog equipment, and by constraining the users to relatively low throughput and often only marginally acceptable error-performance as well. When users of such systems expressed willingness to make further sacrifices in throughput in order to gain better error-performance, the coding engineers and theorists were presented with "very quiet channels". Most such channels turned out to be highly symmetric. This is fortunate because, to this day, relatively little is known about very quiet asymmetric channels.

Perhaps the most-studied channel is the binary symmetric channel. It is a special case (with $q = 2$) of the q-ary symmetric channel, which has q inputs and q outputs. Each input goes to its corresponding output with probability $1 - p$. The parameter p, which is the probability of symbol error, is the same for all inputs. All kinds of symbol errors are equally likely, each occurring with probability $p/(q-1)$. As p approaches $(q-1)/q$, the channel becomes very noisy, and its capacity approaches 0. But as for all very noisy channels, when properly normalized, the reliability approaches the function shown in Figure 4. On the other hand, when p approaches 0, the capacity approaches $\ln q$ bits/symbol. Since R_{crit} and R_x both approach the capacity C, the high-rate region above R_x disappears. Random codes become uniformly bad, because the reliability is dominated by minimum distance considerations rather than by volume considerations. Here minimum distance is the smallest number of input symbols in which any pair of codewords differ. The reliability approaches ∞ at all coderates between 0 and the capacity C, and the rate at which it approaches ∞ is proportional to $-\ln p$. So the interesting function to study is the "normalized reliability", defined as

$$e(R) = E_1(R)/(-\ln p).$$

For the binary symmetric channel, $N \cdot e(R)$ is the best possible minimum distance of any block code of the specified (long) length and rate. From the mid-1950s through the mid-1970s, the classic bounds on $e(R)$ shown for the binary symmetric channel in Figure 5 were the best results known. Results for the more general symmetric q-ary channels were similar. But eventually, both the upper and lower bounds on $e(R)$ were partially improved. The first improvement on the Elias bound, at sufficiently small positive rates, was by Welch, McEliece, and Rumsey [15]. The first improvement on the Gilbert bound, for $q \geq 49$, was due to

Tsfasman, Vlăduț, and Zink [13]. The latter result was a novel application of algebraic geometry. Despite further improvements, a gap still remains between the improved bounds.

Algebraic Coding

Even before Shannon's retirement, many engineers, computer scientists, and mathematicians were creating innovative algorithms to specify and implement feasible coding strategies for particular noisy channels. The first such were the single-bit-error-correcting binary Hamming codes. Shannon mentioned the binary Hamming code with 2^4 codewords of length $n = 7$ as an example in his 1948 paper. Binary Hamming codes with 2^{32} or 2^{64} codewords later became very widely used in computer memories.

In 1960, Bose-Chaudhuri-Hocquenghem and Reed-Solomon applied the theory of finite fields to construct error-correcting codes. Although subsequently seen to be special cases of a common generalization, the initial BCH codes were binary, while the initial RS codes used a large prime-power alphabet of size $q = N + 1$. The RS codes proved directly suitable for the original q-ary symmetric channel and for several of its close relatives with q symmetric inputs but with more outputs. One such channel has an additional $(q + 1)$st "erasure" output, accessible from all q inputs with the same transition probability. Another such channel has $2q$ output symbols, each pair of which corresponds to a "more reliable" and a "less reliable" estimate of each input signal.

The most widespread application of RS codes has been to a bursty (rather than a memoryless) version of the binary symmetric channel, the bit error probability varying with time. Any sequence of bits can be partitioned into m-bit characters, each of which can be viewed as a symbol in a finite field of order $q = 2^m$. This is particularly appropriate for certain binary channels in which errors tend to come in short bursts of lengths comparable to m. For such channels, doubly-interleaved RS codes are superior to multiply-interleaved binary BCH codes in numerous respects. They have superior performance. Surprisingly, they are easier to encode. They are also easier to decode. Partly because of this, and partly because of unusually fast and efficient decoding algorithms now associated with them, RS codes with high coderates have become very widely used in many magnetic and optical storage systems.

The emphasis on high coderates arises from a coincidence of several factors. On the system side, considerations of factors such as bit synchronization, latency, and delay make high coderates attractive. On the coding side, all of the costs associated with RS codes grow linearly or as the square of the redundancy, $(\ln q - R)N$, and these costs become very small when R is large. There is also a performance factor, because traditional RS decoding algorithms perform only as block codes with list size $L = 1$.

Sudan [12] introduced an innovative method to decode RS codes with low coderates using lists of size greater than 1. He also applied these methods to random proof-checking problems in theoretical computer science. When it is feasible to trade an increase in decoding complexity for improved performance, Sudan's algorithm can be used to attain much higher reliability for low-rate RS codes even when $L = 1$. One first uses Sudan's algorithm like a search committee, to reduce a large pool of candidates down to a short list. In many applications, there is auxiliary information available that allows another part of the decoder to scrutinize each of these candidates more carefully before making the final selection.

Convolutional Codes

Originally, Shannon used the block length N as a rough measure of the delay and complexity of a coding/decoding system. Many coding theorists have long sought to refine this measure. In the 1950s, Elias introduced binary "convolutional codes" in which message and check bits were intermingled in the transmitted data stream, and in which each check bit was constrained to depend on only N prior message bits. This approach replaced the block length with a constraint length and led Wozencraft, Fano, Jacobs-Berlekamp, and others to study sequential decoding algorithms. Viterbi [14] presented a very innovative approach showing that maximum-likelihood sequential decoding was feasible on sufficiently short convolutional codes and that the performance was adequate for many purposes, particularly on the binary-input Gaussian noise channel. This led to widespread use of Viterbi decoding on many communications channels. In the mid-1980s, NASA adopted standards for deep-space communications that required a concatenation of two codes: an inner Viterbi code of coderate 1/2, and an outer RS code of coderate 223/255 or 239/255. The Viterbi code converts a rather noisy memoryless Gaussian channel into a rather bursty binary symmetric channel, on which the RS code then attains a very low probability of decoding error.

Low Density Parity Check Codes

In 1962, Gallager [5] introduced "low density parity check codes" as a way to get some of the advantages of long block codes while constraining the cost of complexity. Although the topic then remained relatively dormant for nearly twenty years, the phenomenal decrease in the cost of memory led a number of computer scientists and engineers to revisit or rediscover and improve this approach in the early 1980s. Gallager's methods had been random, but the modern resurgence of interest in this area has focussed more attention on specific constructions using expander graphs such as those constructed by Margulis [7] and by Lubotzky, Phillips, and Sarnak. This construction is a novel application of modular forms.

Conclusion

Most mathematicians are probably aware of names such as Conway, Margulis, and Sarnak, but they may not be aware of the enormous impact the coding work inspired by Shannon has had in other areas. Forney, Ungerboeck, and others used increasingly sophisticated coding techniques to improve speeds over local phone lines by almost two orders of magnitude, well beyond crude early estimates of capacity. Jacobs and Viterbi founded a company called Qualcomm, which dominates the intellectual property in the cell phone industry. Today, Reed-Solomon codes are used in most magnetic disks and optical CDs. The large part of Bell Telephone Labs that stayed with AT&T at the time of the Lucent spinoff was renamed "Shannon Labs". Statues of Shannon have been dedicated in his hometown of Gaylord, Michigan, and at both of his alma maters: University of Michigan and MIT, and at AT&T's Shannon Labs and at Lucent's Bell Labs, and at the Center for Magnetic Recording Research of the University of California at San Diego.

References

[1] Elwyn R. Berlekamp, *Algebraic coding theory*, McGraw-Hill, 1968, revised edition, Aegean Park Press, 1984.

[2] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, third edition, Springer, New York, 1999.

[3] Peter Elias, List decoding for noisy channels, Research Laboratory of Electronics Report Number 335, Massachusetts Institute of Technology, 1957.

[4] G. David Forney Jr., *Concatenated codes*, M.I.T. Press, Cambridge, Mass., 1966.

[5] Robert G. Gallager, *Low-density parity-check codes*, M.I.T. Press, Cambridge, Mass., 1963.

[6] _____ , *Information theory and reliable communication*, Wiley, New York, 1968.

[7] G. A. Margulis, Explicit constructions of expanders, *Problemy Peredachi Informatsii* **9** (1973) 71–80; English translation in *Problems of Information Transmission* **9** (1973), no. 4, 325–332 (1975).

[8] I. S. Reed and G. Solomon, Polynomial codes over certain finite fields, *J. Soc. Indust. Appl. Math.* **8** (1960) 300–304.

[9] Barney Reiffen, A note on "very noisy" channels, *Information and Control* **6** (1963) 126–130.

[10] C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27** (1948) 379–423, 623–656.

[11] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channels, I and II, *Information and Control* **10** (1967) 65–103, 522–552.

[12] Madhu Sudan, Decoding of Reed-Solomon codes beyond the error-correction bound, *J. Complexity* **13** (1997), no. 1, 180–193.

[13] M. A. Tsfasman, S. G. Vlăduţ, and Th. Zink, Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound, *Math. Nachr.* **109** (1982) 21–28.

[14] Andrew J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Information Theory* **IT-13** (1967) 260–269.

[15] Lloyd R. Welch, Robert J. McEliece, and Howard Rumsey Jr., A low-rate improvement on the Elias bound, *IEEE Trans. Information Theory* **IT-20** (1974), 676–678.