Since the inequality

$$\left|\ln x - (x-1) + \frac{(x-1)^2}{2}\right| < \delta \mid x-1 \mid^2 \qquad \text{(A.43)}$$

is valid for $\mid \ln x \mid < \frac{\delta}{2}$, we obtain, for $y \in \bar{B}_{K\epsilon,\delta}(\underline{x}_0)$ and $\underline{X} \in \mathcal{S}$ that

$$\left| f(\underline{X},y) \ln \sqrt{\frac{Ef(\underline{X},y)}{f(\underline{X},y)}} - \sqrt{f(\underline{X},y)}\left(\sqrt{Ef(\underline{X},y)} - \sqrt{f(\underline{X},y)}\right) + \frac{1}{2}\left(\sqrt{Ef(\underline{X},y)} - \sqrt{f(\underline{X},y)}\right)^2 \right| < \delta \mid \sqrt{Ef(\underline{X},y)} - \sqrt{f(\underline{X},y)} \mid^2 \quad a.s. \qquad \text{(A.44)}$$

Hence, since

$$2\int_{\bar{B}_{K\epsilon,\delta}(\underline{x}_0)} E\left\{\sqrt{f(\underline{X},y)}\left(\sqrt{f(\underline{X},y)} - \sqrt{Ef(\underline{X},y)}\right)\right\}\nu(dy) = \int_{\bar{B}_{K\epsilon,\delta}(\underline{x}_0)} E\left\{\sqrt{f(\underline{X},y)} - \sqrt{Ef(\underline{X},y)}\right\}^2 \nu(dy) = G \qquad \text{(A.45)}$$

it follows from (A.44) that

$$\mid J_1 - 2G \mid \leq 2\delta G. \qquad \text{(A.46)}$$

Therefore, applying Corollary A.1, we have the following estimate for sufficiently small values of $D(\underline{X})$:

$$\left| J_1 - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} I_{ij}(\underline{x}_0)\mathrm{cov}(X_i,X_j)\right| < \delta \sum_{i=1}^{N}\sum_{j=1}^{N} I_{ij}(\underline{x}_0)\mathrm{cov}(X_i,X_j). \qquad \text{(A.47)}$$

Next, since $\mid \ln x \mid < x + \frac{1}{x}$ for all $x > 0$, we have

$$\begin{aligned} \mid J_2 \mid &< \int_{B_{K\epsilon,\delta}(\underline{x}_0)} \frac{E(f^2(\underline{X},y))}{Ef(\underline{X},y)}\nu(dy) + \int_{B_{K\epsilon,\delta}(\underline{x}_0)} Ef(\underline{X},y)\nu(dy) \\ &\leq 2\int_{B_{K\epsilon,\delta}(\underline{x}_0)} \sup_{\|\underline{x}-\underline{x}_0\|<K\epsilon} f(\underline{x},y)\nu(dy) \\ &= o(D(\underline{X})) \end{aligned} \qquad \text{(A.48)}$$

by (A.23), since $\epsilon^2 = D(\underline{X})$. Combining (A.42), (A.46), and (A.48), we obtain, for $D(\underline{X})$ sufficiently small,

$$\left| I(\underline{X};Y) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} I_{ij}(\underline{x}_0)\mathrm{cov}(X_i,X_j)\right| < \delta O(D(\underline{X})). \qquad \text{(A.49)}$$

Now letting $\delta = \delta(D(\underline{X}))$ tend to zero sufficiently slow as $D(\underline{X}) \to 0$ (cf. Remark 2.3) we conclude that

$$\left| I(\underline{X};Y) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} I_{ij}(\underline{x}_0)\mathrm{cov}(X_i,X_j)\right| = o(D(\underline{X})), \qquad \text{(A.50)}$$

which completes the proof of Theorem 1.

REFERENCES

[1] R.G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[2] I.A. Ibragimov and R.Z. Khas'minskii, "Asymptotic behavior of certain statistical estimates in the smooth case (I)," *Theory Probab. Appl.*, vol. 17, pp. 445–462, 1972.

[3] ——, "Weak signal transmission in a memoryless channel," *Probl. Peredach. Inform.*, vol. 8, pp. 28–39, Oct.-Dec. 1972. English translation in *Probl. Inform. Transmission*, vol. 8, pp. 290–299, 1972.

[4] S. Kullback, *Information Theory and Statistics.* New York: Wiley, 1959, and New York: Dover, 1968.

[5] V.V. Prelov, "Asymptotic behavior of the capacity of a continuous channel with a large amount of noise," *Probl. Peredach. Inform.*, vol. 6, pp. 40–57, Apr.-June 1970. English translation in *Probl. Inform. Transmission*, vol. 6, pp. 122–135, 1970.

[6] C.R. Rao, *Linear Statistical Inference and its Applications.* New York: Wiley, 1973, 2nd ed.

[7] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. II, pp. 623–656, 1948.

[8] V.V. Prelov, "On the asymptotic behavior of the capacity of a continuous channel with large nonadditive noise," *Probl. Peredach. Inform.*, vol. 8, pp. 22–27, Oct.-Dec. 1972. English translation in *Probl. Inform. Transmission*, vol. 8, pp. 285–289, 1972.

[9] S. Verdu, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, 1990.

[10] A.J. Viterbi and J.K. Omura, *Principles of Digital Communications and Coding.* New York: McGraw-Hill, 1979.

[11] A.J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels", *IEEE J. Select. Areas Commun.*, vol. 8, pp. 641–649, May 1990.

Corrected Proof of de Buda's Theorem

Tamás Linder, Christian Schlegel, Member, IEEE, and Kenneth Zeger, Member, IEEE

***Abstract*—An error in de Buda's proof on the asymptotic optimality of lattice channel codes [1] is pointed out and corrected using a modification of de Buda's approach. Comments are given on the correct interpretation and the limitations of this result.**

***Index Terms*—Additive white Gaussian noise (AWGN) channel, lattice based channel codes.**

I. ERROR IN DE BUDA'S PROOF

The purpose of this correspondence is to correct, clarify, and interpret a recent paper [1] by R. de Buda, in which he states that there exist lattice based channel codes which meet Shannon's bound for optimal codes [2]. Unfortunately, there appears to be an error with the clever proof presented by de Buda. In this correspondence, we carefully examine de Buda's proof and discuss the problems. We show that de Buda's proof can be mended, but the resulting optimal lattice code is degenerate in the sense that its "structure" appears to be lost. More precisely, the result in [1] is valid only for lattice codes whose code points lie within a thin spherical shell. Such a code resembles more a random spherical code than a lattice code.

In order to proceed we need the careful definition of some concepts used in [1] and [2]. Consider the additive white Gaussian noise (AWGN) channel, with peak signal-power constraint S, i.e., each codeword x of an n-dimensional code for this channel must satisfy $\frac{1}{n}\|x\|^2 \leq S$, where $\|\cdot\|$ denotes the Euclidean norm. During transmission the i.i.d. zero-mean Gaussian random vector $Z = (Z_1,\ldots,Z_n)$ is added to the transmitted codeword. The common variance of the $Z_i's$ is denoted by N.

Manuscript received November 12, 1992. This work was supported in part by the National Science Foundation under Grant NCR-91-57770.

T. Linder is with the Faculty of Electrical Engineering, H-1521 Stoczek u. 2, Technical University of Budapest, Hungary.

C. Shlegel is with the University of South Australia, The Levels, South Australia, 5095, Australia.

K. Zeger is with the Coordniated Science Laboratory, Department of Electrical Engineering, University of Illinois, Urbana-Champaign, IL 61801.

IEEE Log Number 9210705.

Shannon in [2] developed tight upper and lower bounds on the error probability of optimal codes for the AWGN channel. His random coding argument used n-dimensional codes whose M_n codewords are drawn from a uniform distribution on the surface of a sphere of radius $\sqrt{nS}$ centered at the origin. Such codes have transmission rate $R = \frac{1}{n}\log M_n$. In order to quote Shannon's well known result, we need to introduce some notations. Let

$$Q_n(\theta) = Q_n(\theta, S, N) \tag{1}$$

be the probability of the event that for an n-dimensional vector x with $\|x\| = \sqrt{nS}$, the sum $x + Z$ lies outside the circular cone of half angle θ, with vertex at the origin, and x as its central axis. Define furthermore $\omega_n(\alpha)$ as the solid angle of the n-dimensional cone with half angle α, where $\omega_n(\alpha)$ is normalized so that $\omega_n(\pi) = 1$. That is, $\omega_n(\alpha)$ is the volume this cone cuts out from the sphere of unit volume. In [2] Shannon proved that there exists a code among these randomly chosen codes whose average error probability $P_e^*(n) = \frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(n)$, where $P_{e,i}$ is the probability of the incorrect decoding if the ith codeword is sent, is upper bounded by

$$P_e^*(n) \le Q_n(\theta_b) - M_n \int_0^{\theta_b} \omega_n(\theta)\, dQ_n(\theta), \tag{2}$$

where the angle θ_b is defined by the equation $\omega_n(\theta_b) = 1/M_n$. After an involved analysis of the asymptotic behavior of the sequence of functions

$$F_n(\theta_b, R, S/N) = Q_n(\theta_b) - M_n \int_0^{\theta_b} \omega_n(\theta)\, dQ_n(\theta), \tag{3}$$

Shannon found that the limit

$$E(R, S/N) = \lim_{n\to\infty}\left(-\frac{1}{n}\log F_n(\theta_b, R, S/N)\right) \tag{4}$$

existed and is positive for all $R < C = \frac{1}{2}\log(\frac{S}{N} + 1)$. Furthermore, in a certain rate region $R_c < R < C$ (the definition of the critical rate R_c is given in [2]) the exponent $E(R, S/N)$ coincides with the reliability exponent $E^*(R, S/N)$ defined by

$$E^*(R, S/N) = \liminf_{n\to\infty}\left(-\frac{1}{n}\log P_{e,\mathrm{opt}}(n)\right), \tag{5}$$

where $P_{e,opt}(n)$ is the error probability of the best n-dimensional code of rate R for the given channel.

In [1], de Buda aimed at showing that there exist structured (namely lattice based) codes for the AWGN channel that have the same near-optimal error probability properties as Shannon's "random" codes. To this end, de Buda considers an n-dimensional lattice Λ, which is translated by a vector $\hat{s}$. The bounding region of the code is a shell (or annulus), i.e., the region T between an outer sphere and an inner sphere both centered at the origin. Formally, the code $C(\Lambda, \hat{s})$ is defined by $C(\Lambda, \hat{s}) = (\Lambda + \hat{s}) \cap T$. Although it is not explicitly stated in [1], the reader can infer that the radius of the outer bounding sphere is $\sqrt{nS}$, this being the only choice to give the desired peak signal energy constraint. De Buda asserts that the only parameter of the "thick shell" T that appears in the development is its volume V_T. The main result in the paper claims that for each dimension n, there exists a lattice code of the above type with at least 2^{nR} codepoints such that its error probability $P_e(n)$ (assuming uniform prior distribution on the set of messages) satisfies

$$P_e(n) \le 4F_n(\theta_b, R, S/N), \tag{6}$$

where the right side is defined in (3). This implies that essentially the same upper bounds are valid on the decrease of the error probability for rates below the channel's capacity as the ones Shannon derived for random codes. The only requirement on the volume V_T of T needed was that it must satisfy the equation $\frac{V_T}{\det\Lambda} = 2^{nR+1}$, where $\det\Lambda$, the determinant of the lattice, can be chosen freely,

Unfortunately there seems to be a technical error in [1] in the proof of (6), which has important consequences and changes the scope of the result. As it turns out, to correct the error we have to use a bounding region T which is more appropriately described as a *thin shell.*

In the course of the proof de Buda introduces the function $f(x, s)$ as an upper bound on the probability that $x + s$ is decoded given that s was transmitted, and the noise x is "small" in the spherical sense. $f(x, s)$ is an upper bound that is valid for all lattice codes containing both of the points s and $s + x$. By small spherical noise the author means that $s + Z$ is inside the cone of half angle θ_{b_2} (an angle specified later) with vertex at the origin and with s as its axis. Denoting the angle between s and $s+Z$ by θ, this "small noise" event can be described as $\{\theta \le \theta_{b_2}\}$. The error in the development appears with the definition of "small noise." For a given lattice, the sum

$$f(\Lambda, s) = \sum_{\substack{x\in\Lambda \\ x\neq 0}} f(x, s) \tag{7}$$

is an upper bound on the probability of incorrect decoding when s is sent, given that $\{\theta \le \theta_{b_2}\}$. From this upper bound it follows, by a standard argument, that the probability $\mathrm{P}_e(s)$ of incorrectly decoding s is upper bounded by

$$P_e(s) \le \Pr\{\theta > \theta_{b_2}\} + f(\Lambda, s). \tag{8}$$

Denoting the number of codepoints in $C(\Lambda, \hat{s})$ by $M(\Lambda, \hat{s})$, de Buda calculates the average error probability ([1, p. 895, (9)]) and obtains

$$\begin{aligned}\mathrm{Avg}_{\mathrm{code}} P_e(s) &\stackrel{\mathrm{def}}{=} \frac{1}{M(\Lambda, \hat{s})}\sum_{s+x\in C(\Lambda,\hat{s})} P_e(s + x) \\ &\le \Pr\{\theta > \theta_{b_2}\} + \frac{1}{M(\Lambda, \hat{s})}\sum_{s+x\in C(\Lambda,\hat{s})} f(\Lambda, s + x) \\ &= \Pr\{\theta > \theta_{b_2}\} + \mathrm{Avg}_{\mathrm{code}} f(\Lambda, s).\end{aligned} \tag{9}$$

However, with de Buda's definition of the event $\{\theta \le \theta_{b_2}\}$ (see e.g., Fig. 2 on p. 895), its probability varies with the magnitude of s, thus treating this probability as a constant independent of s results in an error in the proof. The same problem arises when the upper bound $f(x, s)$ is defined in terms of an angular decision scheme. De Buda derives an upper bound on the probability of decoding $x + s$ when s was sent given that the spherical noise is small, i.e., the angle θ between $s + Z$ and s is smaller than a fixed angle θ_{b_2}. This upper bound is established using a decoding scheme, where a received vector $\tilde{y}$ is decoded as codevector y if the angle between y and $\tilde{y}$ is smaller than the angle between $\tilde{y}$ and any other codevector. The essential feature of the bound $f(x, s) = f(\alpha)$ is that it depends only on the angle α between s and $x + s$. But the definition (21) on pg. 897 of $f(\alpha)$ (see also last paragraph on pg. 896 and Fig. 4.) clearly indicates that the author treats the distribution of the angle θ between s and $s + Z$ as the same for all codepoints s in T, which is true only if all codepoints have the same magnitude.

This condition, in general, does not hold for lattice codes. Therefore one is forced to consider the remark made on pg. 894: "In effect, all points of the code are radially projected on the outer hypersphere of the thick shell T", which is supported by the fact that the author uses (1) for the distribution of θ, indicating that all codewords have the same magnitude $\sqrt{nS}$, the radius of the outer sphere. However, Fig. 3 and the argument in the last paragraph on p. 896 seem to contradict this interpretation. But this projection will not solve the problem either, since the resulting code with codepoints on the outer sphere can clearly have smaller error probability than that of the lattice code.

Thus an upper bound on this error probability is not necessarily an upper bound on the error probability of the original code.

II. Correction to the Proof

Fortunately, there is a way to modify de Buda's proof so that essentially all his steps remain valid. The conclusion, however will be somewhat different. The idea is to consider the code that results from the radial projection of the lattice code onto the *inner* sphere. In this way we get a code whose error probability is *larger* then that of the lattice code. To see this consider the suboptimal angular decision scheme, where instead of nearest neighbor decoding (i.e. maximum likelihood decoding for AWGN channels) the above described "minimum angle" decision is used. It is not hard to verify that the nearest neighbor decoding region for the projected version s' of a lattice codeword s is the same as the decoding region for s in the angular decoding scheme, both being the same pyramidal sector of the space with vertex at the origin. Furthermore, if $s+Z$ is outside this sector so is $s'+Z$, but if $s'+Z$ is inside, then $s+Z$ is also inside. It follows that the projected code, whose codewords are all s', resulting from projecting s onto the inner sphere, has larger error probability than the suboptimal decoding scheme for the lattice code, and any upper bound on the error probability of the projected code can serve as an upper bound on the error probability of the lattice code.

Since the projected code has codepoints of the same magnitude, the definition of $f(\alpha)$ for this code becomes consistent. In fact, it can be checked step-by-step that de Buda's proof applies to this code with the only change needed that the distribution θ is defined with the radius of the inner sphere as signal power parameter. Formally, instead of (1), the distribution of θ used by de Buda, we are forced to redefine $Q_n(\theta)$ as

$$Q_n(\theta) = Q_n(\theta, S_n, N), \tag{10}$$

where $S_n = R_n^2/n < S$ is the signal power associated with the radius R_n of the inner sphere. This way de Buda's result (6) is modified to the statement that for each dimension n there exist a lattice Λ_n and a translating vector $\hat{s}_n$, such that the error probability of the lattice code $C(\Lambda_n, \hat{s}_n) = T_n \cap (\Lambda_n + \hat{s}_n)$, where T_n is the shell between the two spheres of radii $\sqrt{nS_n}$ and $\sqrt{nS}$, ($S_n < S$), satisfies

$$P_e(n) \leq 4F_n(\theta_b, R, S_n/N). \tag{11}$$

It remains to show that the R_n can be chosen such that de Buda's lattice code has the same exponential rate of decrease for rates below channel capacity and the same reliability exponent for rates satisfying $R_c < R < C$ as Shannon's random code. Fortunately de Buda's proof allows the choice of the inner radius arbitrarily (see eq. (22) on pg. 897). First, let us fix the signal power associated with the inner sphere : $S_n = S' < S$ for all n. Then by (3) we have

$$\lim_{n\to\infty}\left(-\frac{1}{n}\log 4F_n(\theta_b, R, S'/N)\right) = E(R, S'/N). \tag{12}$$

From Shannon's work [2] it is apparent that the function $E(R,S,N)$ is continuous in S for $R < C = \frac{1}{2}\log(\frac{S}{N}+1)$. It follows from a standard continuity argument that there exists a sequence of signal powers S_n associated with the inner sphere such that $S_n \to S$ as $n \to \infty$, and

$$\lim_{n\to\infty}\left(-\frac{1}{n}\log 4F_n(\theta_b, R, S_n/N)\right) = E(R, S, N). \tag{13}$$

Thus choosing the inner radius for each dimension n as $R_n = \sqrt{nS_n}$, the above argument and de Buda's corrected result (11) show that there exists a sequence of n-dimensional lattice codes with error probability $P_e(n)$ for which

$$P_e(n) \leq e^{-n[E(R,S/N)-o(1)]}. \tag{14}$$

holds. This means that for rates satisfying $R_c < R < C$, de Buda's lattice codes have the same reliability exponent as that of optimal codes, and for rates below the critical rate R_c the error probability of these lattice codes has essentially the same exponential upper bond as Shannon's code. Note, however, that (14) does not imply that the ratio $P_e(n)/P_{e,\text{opt}}$ is bounded with increasing n.

III. Discussion

The condition $S_n \to S$ means that

$$\lim_{n\to\infty}\frac{R_n}{\sqrt{nS}} = \frac{S_n}{S} = 1, \tag{15}$$

thus the shell that contains the codepoints can no longer be be called a "thick shell." The more appropriate description is "thin shell." It is worth noting, that since the function $F_n(\theta_b, R, S'/N)$ is clearly continuous in S', by choosing $S' < S$ close enough to S, de Buda's result guarantees the existence of an n-dimensional lattice code whose error probability is upper bounded by a quantity arbitrarily close to the upper bound (2) for Shannon's code. However, the better this approximation is the less the thin shell bounded lattice code resembles a lattice code in the usual sense, and the more it looks like a "random" spherical code, for which Shannon originally proved the error bounds.

Were de Buda's original proof to be correct, one might argue that the class of sphere bounded lattice codes or even lattice bounded lattice codes are asymptotically optimal as the dimension of the signal constellation grows. However, this conclusion appears not to directly follow from our corrected version of the proof since the codepoints derived from the lattice are those which lie in a thin spherical shell, and specifically exclude the lattice points interior to the inner sphere. Adding these points to the code would invalidate our presented proof. In effect, the radius of the thin spherical shell is made to be large enough that enough lattice points fall within the sphere as needed.

Finally, we mention an interesting observation (due to A. Loeliger). Both de Buda's result and our correction deal with the average error probability $\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e,i}(n)$, where $P_{e,i}(n)$ is the conditional probability that a decoding error is made given that the ith codeword was sent. The usual method to obtain a code with a small maximal conditional error probability is the deletion of the worst half of the codewords from a code with M codewords and average error probability P_{av}. The resulting code has $M/2$ codewords and its maximal conditional error probability is at most $2P_{\text{av}}$. Now unlike for full lattice codes, the codewords of de Buda's code can have different individual error probabilities. Therefore, this technique must be used to obtain codes with conditional error probabilities individually upper bounded by the right-hand side of (14). But the throwing away of some codewords according to their conditional error probability can destroy the algebraic structure of the original code. It remains to be seen if the structure of de Buda's code assures that "bad" codewords can be removed without essentially effecting the lattice structure.

References

[1] R. de Buda, "Some optimal codes have structure," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 893–899, Aug. 1989.

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