Case 1: If this number is 1, then

$$w(a,\underline{v}) = 2^{k-1} + nw[\,\mu_{\underline{u}}(x)] - 2\left[2^{t-1}\left(\frac{2^t+\eta(s-1)}{s}\right) + 2^{t-1}\left(\frac{2^t-\eta}{s}\right)(|S|-1)\right].$$

In this case, the number $M_1(\underline{v})$ of nonzero elements a in $\mathbf{F}$ is the cardinality of $E(v) = \bigcup_{j\in S}(\alpha^j G_n)$. That is, $M_l(\underline{v})) = nw[\,\mu_{\underline{u}}(x)]$.

Case 2: Otherwise,

$$w(a,\underline{v}) = 2^{k-1} + nw[\,\mu_{\underline{u}}(x)] - 2\left[2^{t-1}\left(\frac{2^t-\eta}{s}\right)|S|\right].$$

The number $M_2(\underline{v})$ of nonzero elements a in $\mathbf{F}$ is $M_2(\underline{v}) = 2^{2t} - 1 - M_1(\underline{v})$. Now using $|S| = w[\,\mu_{\underline{u}}(x)] = w$, $ns = 2^{2t} - 1$, and Proposition 4.6, we obtain the expected result. □

Corollary 4.7: If a) $\mathbf{F}_{2^k}$ is the splitting field of $x^n - 1$ over $\mathbf{F}_2$, b) If $k = 2t$, $s = 2^d + 1$, where d is a divisor of t, and if $g(x)$ is a *primitive* divisor of $x^s - 1$ over $\mathbf{F}_2$, then the weight distribution of C is given by Theorem 4.5 for all integer w such that w is even and

$$\left|w - \frac{(2^d+1)}{2}\right| \le 2^{d/2}$$

with $A_w = (2^d+1)A(2w-2^d)$ where $A(x) = \#\{x \in \mathbf{F}_{2^d} \setminus \{0\} : Kl_d(a) = 2^d - 1 - 4x\}$.

Proof: This is a direct consequence of Theorem 4.5 and Proposition 4.2. □

V. Examples

From [7], we can find the weight distributions for irreducible cyclic codes. These results can be used to apply Theorem 4.1 and Theorem 4.4. A numerical table obtained in this way is given in [14].

References

[1] L. D. Baumert and R. J. McEliece, "Weights of irreducible cyclic codes," *Inform. Control*, vol. 20, pp. 158–175, 1972.

[2] P. Delsarte and J. M. Goethals, "Irreducible binary cyclic codes of even dimension," U. North Carolina, Dept. Stat. Mineo, Ser. 600, p. 27, 1970.

[3] P. Delsarte, "On subfield subcodes of Reed-Solomon codes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 575–576, 1975.

[4] G. Lachaud and J. Wolfmann, "The weights of the orthogonals of the extended quadratic binary Goppa codes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 686–692, 1990.

[5] R. Lidl and H. Niederreiter, *Finite Fields, Encyclopedia of Mathematics and its Applications*, vol. 20, Reading, MA: Addison-Wesley, 1983.

[6] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-correcting Codes*. Amsterdam: North-Holland, 1977.

[7] F. J. MacWilliams and J. Seery, "The weight distribution of some minimal cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 796–806, 1981.

[8] H. F. Mattson, Jr. and G. Solomon, "A new treatment of Bose–Chaudhuri codes," *J. Soc. Indust. Appl. Math.*, vol. 9, pp. 654–669, 1961.

[9] S. Sh. Oganesyan and V. G. Yagdzhyan, "Class of optimum cyclic codes with base p," *Problems Inform. Tran.*, vol. 8, no. 2, pp. 167–169, 1972.

[10] A. Tietäväinen, personal communication (lecture notes on coding theory, in Finnish).

[11] J. Wolfmann, "Formes quadratiques et codes à deux poids," *C.R. Acad. Sc. Paris*, t. 281, pp. 533–535, 1975.

[12] ——, "New bounds on cyclic codes from algebraic curves," *Lecture Notes in Computer Science*. New York: Springer-Verlag, vol. 388, pp. 2055–2060, 1989.

[13] ——, "The number of points on certain algebraic curves over finite fields," *Commun. Algebra*, vol. 17, no. 8, pp. 2055–2060, 1989.

[14] ——, "Weights of primitive binary cyclic codes from non-primitive codes," *CISM Lecture Notes 339*. New York: Springer-Verlag, pp. 107–117, 1993.

An Upper Bound on the Volume of Discrete Spheres

Hans-Andrea Loeliger, *Member, IEEE*

***Abstract*—Finite-length sequences over a finite alphabet with weights are considered. An information-theoretic upper bound on the number of such sequences whose weight does not exceed some given threshold is given.**

***Index Terms*—Volume bound, entropy.**

Let $A = \{a_1, a_2, \cdots, a_{|A|}\}$ be a finite alphabet, each element of which has a nonnegative real weight $w(a_i)$. We define the weight of an n-tuple over A (i.e., an element of A^n) as the sum of the weights of the components. Let

$$B_n(\rho) \triangleq \{\boldsymbol{v} \in A^n : w(\boldsymbol{v}) \le n\rho\}$$

be the "discrete ball" of normalized radius[1] ρ. The problem of upper bounding the volume $|B_n(\rho)|$ of such discrete spheres and of computing its asymptotic growth rate (or "entropy") $\lim_{n\to\infty}(1/n)\log|B_n(\rho)|$ is ubiquitous in coding theory. In particular, it arises in sphere packing and Gilbert–Varshamov type arguments in both Hamming space [1] and Euclidean space (e.g., [2]), and in spherical "shaping" of high-dimensional signal constellations (e.g., [3]).

The bound of this note is well known for the special case where A is the binary alphabet $\{0,1\}$ and $w(\cdot)$ is Hamming weight: it is stated in almost every textbook on coding that, for $0 < \rho \le 1/2$,

$$|B_n(\rho)| = \sum_{m=0}^{\lfloor n\rho\rfloor}\binom{n}{m} \le 2^{nh(\rho)} \tag{1}$$

where $h(\rho) \triangleq -\rho\log_2\rho - (1-\rho)\log_2(1-\rho)$ is the binary entropy function and $\lfloor\cdot\rfloor$ denotes rounding down to the nearest integer. Somewhat surprisingly, however, the corresponding bound for the general case seems not to have appeared in the literature other than as an asymptotic result (e.g., [2]); a suitable reference with a clear-cut, explicit formulation of the general version of (1) (with coefficient 1 and no "epsilon" term in the exponent) seems not to exist. The purpose of this note is to fill this gap.

Theorem: Let ρ be a real number such that $w_{\min} < \rho \le \overline{w}$, where $w_{\min} \triangleq \min_{a\in A} w(a)$ and $\overline{w} \triangleq |A|^{-1}\sum_{a\in A} w(a)$ are the minimum and the average weight, respectively, of the elements of A. Then

$$|B_n(\rho)| \le 2^{nH(p)} \tag{2}$$

where $p(\cdot)$ is the probability distribution on A defined by

$$p(a) = e^{-\lambda w(a)}/\mathcal{Z}(\lambda) \tag{3}$$

Manuscript received March 23, 1993; revised April 18, 1994.

The author is with the ISY, Linköping University, S-58183 Linköping, Sweden.

IEEE Log Number 9406003.

[1] In Euclidean-space applications, "weight" is usually taken to be squared norm, in which case ρ is actually the normalized *squared* "radius."

0018-9448/94$04.00 © 1994 IEEE

with $\mathcal{Z}(\lambda) \triangleq \sum_{a \in A} e^{-\lambda w(a)}$, where the nonnegative real number λ is uniquely determined by the condition

$$\sum_{a \in A} p(a)w(a) = \rho \tag{4}$$

and where $H(p)$ is the entropy $-\sum_{a \in A} p(a) \log_2 p(a)$. Moreover, the bound is asymptotically tight, i.e., $\lim_{n \to \infty} (1/n) \log_2 |B_n(\rho)| = H(p)$.

Example: Let A be a finite subset of N-dimensional Euclidean space, and let $w(a) = \|a\|^2$. Then $p(\cdot)$ is "discrete N-dimensional Gaussian" with expected energy ρ.

The reader will have recognized that $p(\cdot)$ is the maximum-entropy distribution on A subject to the constraint (4) [4, p. 266 ff.]. (Such distributions are abundant in statistical physics [5], where λ is the "inverse temperature" and $\mathcal{Z}(\lambda)$ is the partition function, or "Zustandssume.") The existence and uniqueness of $p(\cdot)$ follow from the following proposition which, moreover, shows that λ is easily computed numerically from either ρ or $H(p)$.

Proposition: Let m be the number of elements of $|A|$ of minimum weight, and assume that $m < |A|$. For any fixed λ, let $p(\cdot)$ be the probability distribution of (3). Then both the expected weight $E[w] \triangleq \sum_{a \in A} p(a)w(a)$ and the entropy $H(p)$ decrease monotonically (from $\overline{w}$ to $w_{\min}$ and from $\log_2 |A|$ to $\log_2 m$, respectively) as λ increases from 0 to ∞.

This is intuitively rather obvious and undoubtedly well known. For completeness, however, the proposition is proved in the Appendix.

As a last remark before proving the theorem, we note that $H(p)$ can be written as

$$\begin{aligned} H(p) &= -\sum_{a \in A} p(a)(-\lambda w(a) \log_2 e - \log_2 \mathcal{Z}(\lambda)) \\ &= \lambda \rho \log_2 e + \log_2 \mathcal{Z}(\lambda). \end{aligned}$$

The bound (2) then becomes

$$|B_n(\rho)| \le (e^{\lambda \rho} \cdot \mathcal{Z}(\lambda))^n, \tag{5}$$

which may be advantageous for numerical evaluation. We will see later (cf. proof via Chernoff bound) that (5) actually holds for all nonnegative λ, but the right side is minimized by choosing λ as in the theorem.

We now prove the theorem. In addition to an elementary first proof, we also sketch two interesting alternative proofs that were communicated to this author in response to an earlier version of this note.

First Proof of the Theorem: We extend $p(\cdot)$ to A^n by defining, for every $\boldsymbol{v} = (v_1, \cdots, v_n) \in A^n$, the probability $p(\boldsymbol{v})$ as $p(\boldsymbol{v}) = p(v_1)p(v_2) \cdots p(v_n)$. Consider the "high probability set"

$$B_n'(\rho) \triangleq \{\boldsymbol{v} \in A^n : p(\boldsymbol{v}) \ge 2^{-nH(p)}\}.$$

(Remember that $p(\cdot)$ depends on ρ.) Since $1 \ge \sum_{\boldsymbol{v} \in B_n'(\rho)} p(\boldsymbol{v}) \ge |B_n'(\rho)| \cdot 2^{-nH(p)}$, we have $|B_n'(\rho)| \le 2^{nH(p)}$.

We next show that $B_n'(\rho) = B_n(\rho)$. The following inequalities are equivalent for $\lambda \ge 0$:

$$\begin{aligned} p(\boldsymbol{v}) &\ge 2^{-nH(p)}; \\ \log_2 p(\boldsymbol{v}) &\ge -nH(p); \\ -\sum_{j=1}^{n} (\lambda w(v_j) \log_2 e + \log_2 \mathcal{Z}(\lambda)) &\\ \ge -n \sum_{a \in A} p(a)(\lambda w(a) \log_2 e + \log_2 \mathcal{Z}(\lambda)); &\\ w(\boldsymbol{v})\lambda \log_2 e + n \log_2 \mathcal{Z}(\lambda) \le n\rho\lambda \log_2 e + n \log_2 \mathcal{Z}(\lambda); &\\ w(\boldsymbol{v}) \le n\rho. & \end{aligned}$$

Thus, $B_n(\rho) = B_n'(\rho)$, which proves (2).

The asymptotic tightness follows from a standard argument. Without loss of generality, let a_1 be an element of A of minimum weight. Let $T_n(p)$ be the set of n-tuples $\boldsymbol{v}$ in A^n such that, for $i > 1$, the letter a_i occurs precisely $\lfloor p(a_i)n \rfloor$ times in $\boldsymbol{v}$. Since $w(\boldsymbol{v}) \le \sum_{a \in A} p(a)nw(a) = n\rho$ for any such $\boldsymbol{v}$, we have $T_n(p) \subseteq B_n(\rho)$. But the asymptotic growth rate $\lim_{n \to \infty} 1/n \log_2 |T_n(p)|$ of $T_n(p)$ is well known [4, p. 282] to be $H(p)$; together with (2), this establishes $\lim_{n \to \infty} (1/n) \log_2 |B_n(\rho)| = H(p)$. □

An alternative method to prove the theorem is to show first that

$$1/n \log_2 |B_n(\rho)| \le \lim_{n \to \infty} 1/n \log_2 |B_n(\rho)| \tag{6}$$

and then to calculate the right-hand limit with a standard variational technique. A particularly elegant way to carry out both of these steps is the following proof due to F. R. Kschischang (private communication).

Proof via Capacity of Noiseless Channel: Consider a noiseless channel with input/output alphabet A, per-symbol costs $w(\cdot)$, and a per-symbol cost constraint ρ. For any finite blocklength n, the set $B_n(\rho)$ is a code for this channel that satisfies the cost constraint, and the capacity $C(\rho)$ of this channel is $\lim_{n \to \infty} 1/n \log_2 |B_n(\rho)|$. Since the channel is noiseless, the rate of any finite-length code $B_n(\rho)$ cannot exceed $C(\rho)$, which proves (6).

It is clear that the capacity-achieving input distribution of this channel is the maximum-entropy distribution on A subject to the constraint $\sum_{a \in A} p(a)w(a) \le \rho$, which is the distribution $p(\cdot)$ of the theorem. Thus, $C(\rho) = H(p)$, which completes the proof. □

Another interesting proof was suggested by J. L. Massey (private communication).

Proof via Chernoff Bound: Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random variable that is uniform over A^n. Then

$$P(w(\boldsymbol{X}) \le n\rho) = |B_n(\rho)| \cdot |A|^{-n}. \tag{7}$$

Bounding the left side by the Chernoff bound [6, p. 97 ff.] gives

$$P(w(\boldsymbol{X}) \le n\rho) \le E[e^{-\lambda(w(\boldsymbol{X}) - n\rho)}], \tag{8}$$

which holds for all $\lambda \ge 0$. After some calculations, the right side of (8) reduces to $|A|^{-n}(e^{\lambda\rho} \cdot \mathcal{Z}(\lambda))^n$. Together with (7), this proves our earlier claim that (5) holds for all $\lambda \ge 0$. It remains to determine the optimal value of λ. To this end, consider

$$\begin{aligned} \frac{d}{d\lambda}[e^{\lambda\rho} \cdot \mathcal{Z}(\lambda)] &= \rho e^{\lambda\rho} \mathcal{Z}(\lambda) - e^{\lambda\rho} \sum_{a \in A} w(a) e^{-\lambda w(a)} \\ &= e^{\lambda\rho} \mathcal{Z}(\lambda) \left(\rho - \sum_{a \in A} w(a) e^{-\lambda w(a)} / \mathcal{Z}(\lambda) \right) \\ &= e^{\lambda\rho} \mathcal{Z}(\lambda)(\rho - E[w]) \end{aligned} \tag{9}$$

where $E[w]$ is defined as in the proposition. Since $e^{\lambda\rho}\mathcal{Z}(\lambda)$ is positive for $\lambda \ge 0$, it is clear from (9) and from the monotonic decrease of $E[w]$ (for increasing λ) that the unique minimizing λ is determined by condition (4) of the theorem. □

The bound of this note also holds for infinite discrete alphabets provided that the probability distribution $p(\cdot)$ of the theorem exists, as it does for most cases of practical interest. However, the determination of satisfactory conditions that are sufficient to guarantee the existence of $p(\cdot)$, as well as the study of asymptotic tightness, leads to questions outside the scope of this note.

The bound can also be adapted to continuous alphabets by replacing the probability distribution $p(\cdot)$ by a density, the cardinality $|B_n(\rho)|$ by a volume, and the entropy $H(p)$ by the corresponding differential entropy. With these substitutions—and provided that a density $p(\cdot)$ of the form (3) and satisfying (4) exists—the nonasymptotic part of the first proof, and thus the bound (2), is still valid. We conclude with the following example due to G. D. Forney, Jr., (private communication).

Example: Let A be the real line with weight $w(a) = a^2$; then $B_n(\rho)$ is the n-dimensional sphere (ball) of radius $\sqrt{n\rho}$ around the origin. The probability density $p(\cdot)$ is Gaussian with variance ρ, whose differential entropy is $\log_2 \sqrt{2\pi e\rho}$. According to (the continuous version of) (2), the volume of $B_n(\rho)$ is upper bounded by $(2\pi e\rho)^{n/2}$. The comparison of this bound, for $n = 2m$, with the exact formula $(2m\rho\pi)^m/m!$ for the volume yields the Stirling-type bound

$$m! \geq (m/e)^m,$$

derived purely from information theory and geometry. (The Stirling approximation is $m! \approx \sqrt{2\pi m}(m/e)^m$.)

Appendix
Proof of the Proposition

To simplify notation, we write w_i and p_i instead of $w(a_i)$ and $p(a_i)$, respectively. All logarithms are to the base 2.

We assume, without loss of essential generality, that $w_1, \cdots, w_m$ are the elements of A that have minimal weight. For $\lambda = 0$, $p(\cdot)$ is uniform over A, and thus $E[w] = \bar{w}$ and $H(p) = \log|A|$. The limits as $\lambda \to \infty$ of $p(\cdot)$ is the distribution $p_i = 1/m$ for $1 \leq i \leq m$ and $p_i = 0$ otherwise, which makes it clear that $\lim_{\lambda\to\infty} E[w] = w_{\min}$ and $\lim_{\lambda\to\infty} H(p) = \log m$.

We next show that $(d/d\lambda)E[w] < 0$ for all λ. Let $f(\lambda) \triangleq \sum_i w_i e^{-\lambda w_i}$.

$$\begin{aligned}
&[\mathcal{Z}(\lambda)]^2 \frac{d}{d\lambda} E[w] \\
&\quad = [\mathcal{Z}(\lambda)]^2 \frac{d}{d\lambda}[f(\lambda)/\mathcal{Z}(\lambda)] \\
&\quad = \mathcal{Z}(\lambda)\frac{d}{d\lambda} f(\lambda) - f(\lambda)\frac{d}{d\lambda}\mathcal{Z}(\lambda) \\
&\quad = -\sum_i e^{-\lambda w_i} \sum_j w_j^2 e^{-\lambda w_j} + \sum_i w_i e^{-\lambda w_i} \sum_j w_j e^{-\lambda w_j} \\
&\quad = -\sum_i \sum_j e^{-\lambda(w_i+w_j)} w_j(w_j - w_i) \\
&\quad = -\sum_i \sum_{j>i} e^{-\lambda(w_i+w_j)}[w_j(w_j - w_i) + w_i(w_i - w_j)] \\
&\quad = -\sum_i \sum_{j>i} e^{-\lambda(w_i+w_j)}(w_i - w_j)^2,
\end{aligned}$$

which is negative unless all weights are equal. Since $\mathcal{Z}(\lambda) > 0$, we have proved that $(d/d\lambda)E[w] < 0$ for all λ.

The monotonic decrease of $H(p)$ follows from the relation $(d/d\lambda)H(p) = \lambda \log e\,(d/d\lambda)E[w]$, which results from the following calculation:

$$\begin{aligned}
\frac{d}{d\lambda}H(p) &= \sum_i \frac{\partial}{\partial p_i} H(p) \frac{dp_i}{d\lambda} \\
&= -\sum_i \frac{\partial}{\partial p_i}(p_i \log p_i)\frac{dp_i}{d\lambda} \\
&= -\sum_i (\log p_i + \log e)\frac{dp_i}{d\lambda} \\
&= \sum_i (\lambda w_i \log e + \log \mathcal{Z}(\lambda) - \log e)\frac{dp_i}{d\lambda} \\
&= \lambda \log e \sum_i w_i \frac{dp_i}{d\lambda} \\
&= \lambda \log e \sum_i \frac{\partial}{\partial p_i}(p_i w_i)\frac{dp_i}{d\lambda} \\
&= \lambda \log e \frac{d}{d\lambda} E[w].
\end{aligned}$$

Acknowledgment

This note benefitted greatly from detailed comments by J. L. Massey, F. R. Kschischang, G. D. Forney, Jr., and T. Ericson.

References

[1] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam: Elsevier, 1988.
[2] Ph. Piret, "Bounds for codes over the unit circle," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 760–767, Nov. 1986.
[3] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877–892, Aug. 1989.
[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[5] E. Schrödinger, *Statistical Mechanics*. Cambridge: Cambridge Univ. Press, 1962.
[6] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

Asymptotic Results on Codes for Symmetric, Unidirectional, and Asymmetric Error Control

Jos H. Weber

***Abstract*—The asymptotic behavior of the rates of optimal codes correcting and/or detecting combinations of symmetric, unidirectional, and/or asymmetric errors is studied. These rates are expressed in terms of the rate of optimal codes with a certain Hamming distance. As a consequence, well-known bounds on the latter rate can also be applied to bound the former rates. Furthermore, it turns out that, without losing rate asymptotically, any error control combination can be upgraded to simultaneous symmetric error correction/detection and all unidirectional error detection.**

***Index Terms*—Asymmetric errors, code rate, error correction, error detection, symmetric errors, unidirectional errors.**

I. Introduction

We consider binary channels over which codewords from a block code $\mathcal{C}$ are sent. If a received word differs in e coordinates from the transmitted word, we say that e (symmetric) errors have occurred. If these transitions are all of the same type (either $1 \to 0$ or $0 \to 1$), the error pattern is said to be unidirectional, while if all transitions are of the $1 \to 0$ type, the error pattern is said to be asymmetric. So any asymmetric error pattern is also unidirectional, and any unidirectional error pattern is also symmetric. We call e the weight of the error pattern.

Manuscript received December 23, 1993; revised May 9, 1994. This paper was presented at the 15th Symposium on Information Theory in the Benelux, Louvain-la-Neuve, Belgium, May 1994.

The author is with the Department of Electrical Engineering, Delft University of Technology, 2600 GA Delft, The Netherlands.

IEEE Log Number 9406221.

0018-9448/94$04.00 © 1994 IEEE