of any word in any code $\mathcal{C}_{\sigma,n}$, or else s^2 is not a head of any word in any code $\mathcal{C}_{\sigma,n}$. Therefore, by induction, the theorem is true for all σ_0 and all odd n_0.

Q.E.D.

Corollary: The codes $\mathcal{C}_{\sigma,n}$ are comma-free for all σ and all odd n.

IV. Maximality

In this section, it will be shown that the codes $\mathcal{C}_{\sigma,n}$ of Definition 2 are maximal (in number of words), and that the conjecture of Golomb, Gordon, and Welch in [1] is true.

Let G_n be the cyclic group of transformations on the words generated by α, where

$$w_1w_2 \cdots w_n \xrightarrow{\alpha} w_2w_3 \cdots w_nw_1$$

Thus, G_n is the group of transformations which rotate the word. We say that two words w_1 and w_2 are equivalent if, and only if, there exists a transformation in G_n which maps w_1 into w_2. An equivalence class can contain at most n members. If a class contains exactly n (distinct) words, then the class will be called nondegenerate; otherwise it it will be called degenerate. A comma-free code can never contain a word from a degenerate equivalence class, nor more than one word from any nondegenerate equivalence class. A sufficient condition for maximality of a code $\mathcal{C}_{\sigma,n}$, then, is that it contain some word from every nondegenerate equivalence class. If this can be shown for all codes $\mathcal{C}_{\sigma,n}$ of Definition 2, then the conjecture of Golomb, et al., is true, for $f(\sigma, n)$ is the number of nondegenerate equivalence classes, as shown in [1].

Theorem 2

For any σ and any odd n, the code $\mathcal{C}_{\sigma,n}$ contains some word from every nondegenerate equivalence class.

Proof: If $n = 1$, the theorem is true; for, each letter of the alphabet $\{0, 1, \cdots, \sigma - 1\}$ constitutes a nondegenerate equivalence class, and there are no other classes. Suppose, now, the theorem is true for all (odd) $n < N$. Let w be any word in any nondegenerate equivalence class of words of length N constructed from an alphabet of σ letters. Construct a pattern p which generates some cyclic permutation of w. Now, $p = p_1p_2 \cdots p_i$ for some odd i in the range $1 \leq i \leq N/3$, and p clearly satisfies conditions 1(a) and 1(b) of Definition 2. Divide w into the i sections generated by the i components of p, and calculate the numerical values $v_j (j = 1, \cdots, i)$. Since w belongs to a nondegenerate class, $v = v_1v_2 \cdots v_i$ belongs to a nondegenerate class of words of length $i < N$. For some ρ, the code $\mathcal{C}_{\rho,i}$ contains a cyclic permutation of v, by the inductive hypothesis. Therefore, some cyclic permutation of w is included in the code $\mathcal{C}_{\sigma,N}$. Therefore, by induction on n, the theorem is true for all σ and all odd n.

Q.E.D.

Corollary: The codes $\mathcal{C}_{\sigma,n}$ are maximal (in number of words) for all σ and all odd n, and contain $f(\sigma, n)$ words, proving the conjecture of Golomb, Gordon, and Welch.

Acknowledgment

The author wishes to thank Dr. Shimon Even for his helpful criticism of the manuscript.

References

[1] Golomb, S. W., B. Gordon, and L. R. Welch, Comma-free codes, *Can. J. Math.*, vol. 10, 1958, pp 202–209.
[2] Eastman, W. L., and S. Even, On synchronizable and *PSK*-synchronizable block codes, *IEEE Trans. on Information Theory*, vol IT-10, Oct 1964, pp 351–356.
[3] Golomb, S. W., L. R. Welch, and M. Delbrück, Construction and properties of comma-free codes, *Biol. Medd. Dan. Vid. Selsk.*, vol 23, 1958, pp 3–34.
[4] Golomb, S. W., Efficient coding for the desoxyribonucleic channel, *Mathematical Problems in the Biological Sciences*, Am. Math. Soc., Providence, Rhode Island, 1962, pp 87–100.
[5] Jiggs, B. H., Recent results in comma-free codes, *Can. J. Math.*, vol 15, 1963, pp 178–187.
[6] Eastman, W. L., Defining patterns for comma-free codes for small prime word lengths, Sperry Rand Research Center Research Memo. 64–4, Apr 1964.

The Convolution Inequality for Entropy Powers

NELSON M. BLACHMAN, SENIOR MEMBER, IEEE

***Abstract*—The entropy power of a band-limited random process is the power of white Gaussian noise having the same entropy rate. Shannon's convolution inequality for entropy power states that the entropy power of the sum of two independent random processes is at least the sum of their entropy powers. This paper presents an improved version of Stam's proof of this inequality, which is obtained by mathematical induction from the one-dimensional case.**

Manuscript received April 1, 1964; revised December 16, 1964.
The author is with Sylvania Electronic Defense Labs., Mountain View, Calif.

Introduction

ALTHOUGH it is generally difficult to determine the capacity C of a channel of bandwidth W which accepts signals of power S and adds to them statistically independent noise of power N, Shannon[1] has shown that

[1] See Shannon [1], Theorem 18, p 641.

$$W \log \frac{S + N^\dagger}{N^\dagger} \leq C \leq W \log \frac{S + N}{N^\dagger},$$

where $N^\dagger$ is the entropy power of the noise, i.e., the power of white Gaussian noise having the same entropy rate. If this rate is $\dot{H}$, then $N^\dagger = (\exp \dot{H}/W)/(2\pi e)$. If H is the entropy of a portion of duration T which is described by $n = 2WT$ values (or the n-dimensional vector which they comprise), $\dot{H} = H/T$, and the entropy power is

$$N^\dagger = \frac{\exp 2H/n}{2\pi e}.$$

We shall use only natural logarithms and exponentials.

The foregoing bounds on C are based on the convolution inequality for entropy powers[2], which asserts that the entropy power of the sum of two statistically independent band-limited random processes is not more than the sum of the powers of the processes nor less than the sum of their entropy powers. It attains the upper bound only when all are white Gaussian processes, and it attains the lower bound only when they are Gaussian processes with proportional covariance matrices or, equivalently, proportional power spectral densities.

The upper bound follows from the entropy-maximizing property of white Gaussian statistics for a given mean-squared value. In deriving the lower bound, Shannon[3] showed that the entropy power of the sum of two independent random processes of given entropy powers has a stationary point where the two processes are Gaussian with proportional covariance matrices, but his variational approach was not able to exclude the possibility that other statistics might yield an equal or lower entropy power for the sum.

More recently, Stam [2] put this theorem[4] on a rigorous footing through a somewhat involved chain of inequalities[5] The purpose of this note is to present his proof as simply and clearly as possible, eliminating the use of measure theory (which Stam [3] likewise eliminated in an outline of his proof), avoiding any reference to theorems concerning Fisher's quantity of information (which are likely to be unfamiliar to readers of this journal), and supplying some details omitted from Stam's argument as well as removing some unnecessary restrictions. For completeness, some parts of Stam's proof are reproduced here without any change.

This proof may be divided broadly into five parts.

1) We first define x_f to be the sum of a random variable x with probability density $p(x)$ plus an independent zero-mean normal random variable with variance f, and we show that the derivative of its entropy $H(X_f)$ with respect to f is

$$\tfrac{1}{2}J(X_f) = \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\partial p_f}{\partial x_f}\right)^2 \frac{dx_f}{p_f}, \tag{1}$$

where

$$p_f(x_f) = \frac{1}{\sqrt{2\pi f}} \int_{-\infty}^{\infty} p(x) \exp -\frac{(x_f - x)^2}{2f}\, dx \tag{2}$$

is the probability density of x_f.

2) We establish that, if x and y are statistically independent random variables with sum $z = x + y$, then

$$\frac{1}{J(Z)} \geq \frac{1}{J(X)} + \frac{1}{J(Y)}, \tag{3}$$

with equality only when x and y are normally distributed.

3) We make use of (3) to show that

$$e^{2H(Z)} \geq e^{2H(X)} + e^{2H(Y)}, \tag{4}$$

with equality only under the same condition. If both sides of (4) are divided by $2\pi e$, it becomes the convolution inequality for entropy power in one dimension.

4) Before proceeding to generalize (4) to the vector case, we prove an inequality based on the convexity of the function $\log \cosh u$.

5) Finally we use mathematical induction to show that, for statistically independent n-dimensional vectors $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$,

$$e^{2H(\mathbf{z})/n} \geq e^{2H(\mathbf{x})/n} + e^{2H(\mathbf{y})/n}, \tag{5}$$

with equality if, and only if, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z} = \mathbf{x} + \mathbf{y}$ are normally distributed with proportional covariance matrices. Dividing both sides of (5) by $2\pi e$, we see that the entropy power of $\mathbf{z}$ is at least the sum of the entropy powers of $\mathbf{x}$ and $\mathbf{y}$.

The Derivative of $H(X_f)$

Differentiating (2) inside the integral, we get the diffusion equation

$$\frac{\partial p_f(x_f)}{\partial f} = \frac{1}{2}\frac{\partial^2 p_f(x_f)}{\partial x_f^2}. \tag{6}$$

From (2) we see that $p_f \leq 1/\sqrt{2\pi f}$; hence,

$$H(X_f) = -\int_{-\infty}^{\infty} p_f(x_f) \log p_f(x_f)\, dx_f \tag{7}$$

either converges absolutely or is $+\infty$. Because the mutual information of x and x_f cannot increase with increasing f and because the mutual information of $x_{f_2} - x_{f_1}$ and x_{f_2} with $f_2 > f_1$ cannot be negative, we have $H(X_{f_1}) \leq H(X_{f_2}) \leq H(X_{f_1}) + \frac{1}{2}\log (f_2/f_1)$. Hence, $H(X_f) < \infty$ for all $f > 0$ or $H(X_f) = +\infty$ for all $f > 0$. Supposing it to be finite and differentiating inside the integral, we get

$$\begin{aligned}
\frac{dH(X_f)}{df} &= -\int_{-\infty}^{\infty} \frac{\partial p_f}{\partial f}\, dx_f - \int_{-\infty}^{\infty} \frac{\partial p_f}{\partial f} \log p_f\, dx_f \\
&= 0 - \frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial^2 p_f}{\partial x_f^2} \log p_f\, dx_f \\
&= -\frac{1}{2}\frac{\partial p_f}{\partial x_f} \log p_f \Big|_{-\infty}^{\infty} + \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\partial p_f}{\partial x_f}\right)^2 \frac{dx_f}{p_f} \\
&= \tfrac{1}{2}J(X_f).
\end{aligned} \tag{8}$$

[2] *Ibid.*, Theorem 15, p 636.
[3] *Ibid.*, Appendix 6, p 653.
[4] See Stam [2], sec. 4.4, Theorem 4, p 108.
[5] *Ibid.*, sec. 4.1, Theorem 1, p 91; sec. 4.3, Theorem 1, p 98; sec. 4.4, Theorems 1 and 3, p 102 ff; sec. III.I, Theorem 5, p 147.

Stam credits N. G. de Bruijn with the discovery of this relationship and its relevance to the convolution inequality for entropy power.

The term $-\frac{1}{2}(\partial p_f/\partial x_f) \log p_f \,|_{-\infty}^{\infty}$ can be shown to vanish for $f > 0$ by applying Schwarz's inequality to

$$\frac{\partial p_f}{\partial x_f} = -\frac{1}{\sqrt{2\pi f^3}} \int_{-\infty}^{\infty} p(x)(x_f - x) \exp -\frac{(x_f - x)^2}{2f}\, dx.$$

Thus,

$$\left(\frac{\partial p_f}{\partial x_f}\right)^2 = \left[\int_{-\infty}^{\infty} \frac{\sqrt{p(x)} \exp -\frac{(x_f - x)^2}{4f}}{\sqrt[4]{2\pi f}} \cdot \frac{\sqrt{p(x)}\,(x_f - x) \exp -\frac{(x_f - x)^2}{4f}}{\sqrt[4]{2\pi f^5}}\, dx\right]^2$$

$$\leq \frac{p_f}{\sqrt{2\pi f^5}} \int_{-\infty}^{\infty} p(x)(x_f - x)^2 \exp -\frac{(x_f - x)^2}{2f}\, dx$$

$$\leq \sqrt{\frac{2}{\pi f^3}}\, \frac{p_f}{e}, \tag{9}$$

since $(x_f - x)^2 \exp -(x_f - x)^2/2f \leq 2f/e$. From (9) we see that $\partial p_f/\partial x_f$ is, at most, of the order of $\sqrt{p_f}$ as $p_f \to 0$, and so $(\partial p_f/\partial x_f) \log p_f \to 0$ as $x_f \to \pm\infty$.

Since the integrand of (1) is never negative, $J(X_f)$ either converges or is $+\infty$. Dividing the second line of (9) by p_f and integrating over all x_f, we see that $J(X_f) \leq 1/f$. Hence, $J(X_f)$ is finite for all $f > 0$, converging uniformly for $0 < \epsilon \leq f$ and thus justifying (8).

The Convolution Inequality for J

If the probability densities of x, y, and $z = x + y$ are $p(x)$, $q(y)$, and

$$r(z) = \int_{-\infty}^{\infty} p(x)q(z - x)\, dx, \tag{10}$$

respectively, and these density functions are differentiable, then dr/dz is

$$r'(z) = \int_{-\infty}^{\infty} p'(x)q(z - x)\, dx.$$

Thus, if p, q, and r never vanish,

$$\frac{r'(z)}{r(z)} = \int_{-\infty}^{\infty} \frac{p(x)q(z - x)}{r(z)} \frac{p'(x)}{p(x)}\, dx = E\left\{\frac{p'(x)}{p(x)} \,\middle|\, z\right\},$$

the conditional expectation of $p'(x)/p(x)$ for a given value of z. Likewise,

$$\frac{r'(z)}{r(z)} = E\left\{\frac{q'(y)}{q(y)} \,\middle|\, z\right\}.$$

Therefore, for any constants a and b,

$$E\left\{a\,\frac{p'(x)}{p(x)} + b\,\frac{q'(y)}{q(y)} \,\middle|\, z\right\} = (a + b)\,\frac{r'(z)}{r(z)}.$$

Hence,

$$(a + b)^2\left[\frac{r'(z)}{r(z)}\right]^2 = \left[E\left\{a\,\frac{p'(x)}{p(x)} + b\,\frac{q'(y)}{q(y)} \,\middle|\, z\right\}\right]^2 \leq E\left\{\left[a\,\frac{p'(x)}{p(x)} + b\,\frac{q'(y)}{q(y)}\right]^2 \,\middle|\, z\right\}, \tag{11}$$

since the second moment of any random variable is never less than the square of its mean, with equality only if

$$a\,\frac{p'(x)}{p(x)} + b\,\frac{q'(y)}{q(y)} = (a + b)\,\frac{r'(z)}{r(z)} \tag{12}$$

with probability one whenever $z = x + y$.

Averaging both sides of (11) over the distribution of z gives us

$$(a + b)^2 E\left\{\left[\frac{r'(z)}{r(z)}\right]^2\right\} \leq E\left\{\left[a\,\frac{p'(x)}{p(x)} + b\,\frac{q'(y)}{q(y)}\right]^2\right\} = a^2 E\left\{\left[\frac{p'(x)}{p(x)}\right]^2\right\} + b^2 E\left\{\left[\frac{q'(y)}{q(y)}\right]^2\right\}$$

or

$$(a + b)^2 J(Z) \leq a^2 J(X) + b^2 J(Y).$$

Setting $a = 1/J(X)$ and $b = 1/J(Y)$, we obtain (3), with equality only when (12) holds. Substituting $x = z - y$ into (12) and integrating with respect to y, we get

$$-a \log p(z - y) + b \log q(y) = (a + b)\,\frac{r'(z)}{r(z)}\, y + c(z).$$

Setting $y = 0$ shows that the "constant" of integration $c(z)$ is differentiable, and it follows that $r'(z)/r(z)$, too, is differentiable. Thus, differentiating with respect to z and setting $z = 0$, we have

$$-a\,\frac{p'(-y)}{p(-y)} = (a + b)\,\frac{r(0)r''(0) - r'^2(0)}{r^2(0)}\, y + c(0),$$

from which we see that $p(x)$ must be a normal density function. Similarly, the condition (12) for equality in (3) implies that $q(y)$ is normal.

The One-Dimensional Entropy-Power Inequality

We now define x_f, x_g, and z_h to be x, y, and z plus independent zero-mean normal random variables with variance $f(t)$, $g(t)$, and $h(t) = f(t) + g(t)$, respectively. We suppose that $f(0) = g(0) = h(0) = 0$, and we restrict t to nonnegative values. When f and g are positive, the probability density functions of x_f, y_g, and z_h are everywhere differentiable and positive. Thus, differentiating

$$s(t) = \frac{\exp 2H(X_f) + \exp 2H(Y_g)}{\exp 2H(Z_h)} \tag{13}$$

with respect to t with the help of (8), we have

$$e^{2H(Z_h)}s'(t) = e^{2H(X_f)}f'(t)J(X_f) + e^{2H(Y_g)}g'(t)J(Y_g) - [e^{2H(X_f)} + e^{2H(Y_g)}][f'(t) - g'(t)]J(Z_h).$$

Substituting the upper bound on $J(Z_h)$ given by (3), we find that

$$e^{2H(Z_h)}s'(t) \geq \frac{[f'(t)J(X_f) - g'(t)J(Y_g)][e^{2H(X_f)}J(X_f) - e^{2H(Y_g)}J(Y_g)]}{J(X_f) + J(Y_g)}.$$

Hence, by putting

$$f'(t) = \exp 2H(X_f) \quad \text{and} \quad g'(t) = \exp 2H(Y_g), \tag{14}$$

we ensure that $s'(t) \geq 0$, with equality only when equality holds in (3), i.e., when x_f and y_g are normal. Thus, $s(t)$ is a constant if x and y are normally distributed; otherwise $s(+\infty) > s(0)$.

If the entropy integrals $H(X_f)$, $H(Y_g)$, and $H(Z_h)$ converge uniformly near $t = 0$, then $s(t)$ is continuous at that point and

$$s(0) = \frac{\exp 2H(X) + \exp 2H(Y)}{\exp 2H(Z)}.$$

To evaluate $s(+\infty)$ we note that (14) implies $f(+\infty) = g(+\infty) = h(+\infty) = +\infty$, and we define x_F to be $x_f/\sqrt{f}$. Then (2) gives us for its probability density

$$p_F(x_F) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sqrt{f}\, p(\sqrt{f}\, u) \exp -\tfrac{1}{2}(x_F - u)^2\, du.$$

Its entropy is $H(X_F) = H(X_f) - \frac{1}{2}\log f$. As f grows infinite, $\sqrt{f}\, p(\sqrt{f}\, u)$ becomes a delta function, and $p_F(x_F)$ approaches $(\exp -\frac{1}{2}x_F^2)/\sqrt{2\pi}$. Thus, if $H(X_F)$ converges uniformly as f grows infinite, it becomes $\frac{1}{2}\log 2\pi e$ in the limit, and $H(X_f) - \frac{1}{2}\log 2\pi e f \to 0$ as $t \to +\infty$. Similarly, $H(Y_g) - \frac{1}{2}\log 2\pi e g \to 0$ and $H(Z_h) - \frac{1}{2}\log 2\pi e(f+g) \to 0$. From (12), then, we have $s(+\infty) = 1$, and (4) is proved.

This entropy-power inequality is automatically satisfied if any of the entropies $H(X)$, $H(Y)$, $H(Z)$ are $\pm\infty$, for $I(X;Z) = H(Z) - H(Y)$ and $I(Y;Z) = H(Z) - H(X)$ cannot be negative. If any of the other entropies are infinite or do not converge uniformly or if $p(x)$ and $q(y)$ are not always finite, we can establish (4) in the following manner provided only that the integrals $H(X)$ and $H(Y)$ converge absolutely. We begin with a sequence of independent random channel inputs $\{x_i\}$, all having probability density function $p(x)$, and a sequence of independent additive random disturbances $\{y_i\}$, all having probability density function $q(y)$. If $z_i = x_i + y_i$ is the channel output at time i, the mutual information rate for this channel is $I(X;Z) = H(Z) - H(Y)$.

Next, we consider a channel with the same input and disturbance sequences which, however, skips over x_i and y_i without delay and goes on to x_{i+1} and y_{i+1} whenever $|x_i|$, $|y_i|$, $p(x_i)$, or $q(y_i)$ exceeds some large value L. If $1 - P(L)$ is the probability that any x_i and y_i are skipped, the mutual information rate of the second channel cannot exceed $I(X;Z)/P(L)$, for it conveys no more information than the first channel, and it takes $P(L)$ times as long to do so.

In effect, the second channel has a sequence of independent inputs with the probability density function $P_L(x) = a_L p(x)$ for $|x| \leq L \geq p(x)$ and $p_L(x) = 0$ otherwise, to which it adds statistically independent disturbances having the probability density function $q_L(y) = b_L q(y)$ for $|y| \leq L \geq q(y)$ and $q_L(y) = 0$ otherwise, a_L and b_L being normalizing constants lying between 1 and $1/P(L)$. If we call the input, disturbance, and output of this channel x_L, y_L, and z_L, respectively, its mutual information rate is $I(X_L; Z_L) = H(Z_L) - H(Y_L)$. Hence,

$$H(Z_L) \leq \frac{H(Z) - H(Y)}{P(L)} + H(Y_L).$$

The foregoing proof of (4) is applicable without difficulty to the truncated distributions $p_L(x_L)$, $q_L(y_L)$, and their convolution $r_L(z_L)$, and we have

$$\exp [2H(X_L)] + \exp [2H(Y_L)] \leq \exp\left[\frac{2H(Z) - 2H(Y)}{P(L)} + 2H(Y_L)\right].$$

As $L \to \infty$, $P(L) \to 1$ and, if $H(X)$ and $H(Y)$ converge absolutely, $H(X_L) \to H(X)$ and $H(Y_L) \to H(Y)$. This thus becomes (4) in the limit, completing the proof of the one-dimensional entropy-power inequality.

The Log Cosh Inequality

Before extending this result to the case of n-dimensional vectors, we must obtain an inequality that we shall use twice in deriving that generalization. Since $(d^2/du^2)\log(2\cosh u) = \operatorname{sech}^2 u > 0$, the function $\log(2\cosh u)$ is strictly convex and thus lies entirely above any tangent except at the point of tangency u_0:

$$\log(2\cosh u) \geq \log(2\cosh u_0) + (u - u_0)\tanh u_0.$$

Hence, if E is an expectation operator and we take $u_0 = Eu$, we have

$$E\{\log(2\cosh u)\} \geq \log(2\cosh Eu),$$

with equality only when u always (with probability 1) takes the same value. Since

$$E\{\log(\exp v)\} = \log(\exp Ev),$$

we have, on adding,

$$E\{\log(2\cosh u \exp v)\} \geq \log(2\cosh E\{u\} \exp E\{v\})$$

or

$$E\{\log(e^{v+u} + e^{v-u})\} \geq \log(e^{E\{v+u\}} + e^{E\{v-u\}}).$$

Putting $v = H_1 + H_2$ and $u = H_1 - H_2$, we finally get

$$E\{\log(e^{2H_1} + e^{2H_2})\} \geq \log(e^{2EH_1} + e^{2EH_2}), \tag{15}$$

with equality if, and only if, $H_1 - H_2$ always takes the same value.

The Entropy-Power Inequality for Vectors

To establish (5) now, we use mathematical induction. We have already proved it for $n = 1$. We suppose it proved, along with the condition for equality, for $n = m$, and we thence show it to hold for $n = m + 1$. Letting

$$\mathbf{u} = (x_1, \cdots, x_m), \quad x = x_{m+1},$$
$$\mathbf{v} = (y_1, \cdots, y_m), \quad y = y_{m+1},$$
$$\mathbf{w} = (z_1, \cdots, z_m), \quad z = z_{m+1},$$

we have, from (4), for any given $\mathbf{u}$ and $\mathbf{v}$,

$$e^{2H(Z|\mathbf{u},\mathbf{v})} \geq e^{2H(X|\mathbf{u})} + e^{2H(Y|\mathbf{v})}, \tag{16}$$

with equality only when the probability densities $p(x \mid \mathbf{u})$ and $q(y \mid \mathbf{v})$ are normal. The conditional entropies in (16) are for fixed $\mathbf{u}$ and $\mathbf{v}$ and are not averaged over $\mathbf{u}$ and $\mathbf{v}$; the random variables over which averages have been taken are indicated by capital letters.

Taking logarithms now and averaging over $\mathbf{u}$ and $\mathbf{v}$ by means of the operator E, we find

$$2H(Z \mid \mathbf{U}, \mathbf{V}) \geq E\{\log [e^{2H(X|\mathbf{u})} + e^{2H(Y|\mathbf{v})}]\}$$
$$\geq \log [e^{2H(X|\mathbf{U})} + e^{2H(Y|\mathbf{V})}] \tag{17}$$

by (15), with equality only if $H(X \mid \mathbf{u}) - H(Y \mid \mathbf{v})$ has the same value for all $\mathbf{u}$ and $\mathbf{v}$ (with probability 1), i.e., if $H(X \mid \mathbf{u})$ is independent of $\mathbf{u}$ and $H(Y \mid \mathbf{v})$ is independent of $\mathbf{v}$. Since the first inequality in (17) becomes an equality only when $p(x \mid \mathbf{u})$ and $q(y \mid \mathbf{v})$ are normal, equality holds throughout (17) only if, in addition, the variance of x for a given $\mathbf{u}$ does not depend on $\mathbf{u}$ and the variance of y for a given $\mathbf{v}$ does not depend on $\mathbf{v}$.

Because the mutual information of $\mathbf{u}$ and z for a given value of $\mathbf{w}$ is not negative, $H(Z \mid \mathbf{W}) - H(Z \mid \mathbf{U}, \mathbf{W}) \geq 0$ or, since $\mathbf{u}$ and $\mathbf{v}$ together determine $\mathbf{w} = \mathbf{u} + \mathbf{v}$,

$$H(Z \mid \mathbf{W}) \geq H(Z \mid \mathbf{U}, \mathbf{V}). \tag{18}$$

with equality only when the distribution of z for any given $\mathbf{u}$ and $\mathbf{v}$ depends only upon $\mathbf{u} + \mathbf{v}$. Since this is a normal distribution with fixed variance for equality in (17), we have equality in both (17) and (18) only when the conditional mean

$$E\{z \mid \mathbf{u}, \mathbf{v}\} = E\{x + y \mid \mathbf{u}, \mathbf{v}\} = E\{x \mid \mathbf{u}\} + E\{y \mid \mathbf{v}\}$$

is a function of $\mathbf{u} + \mathbf{v}$ alone. It follows directly that this function must be linear. Hence, $E\{x \mid \mathbf{u}\}$ is a linear function of $\mathbf{u}$, and $E\{y \mid \mathbf{v}\}$ is the same linear function of $\mathbf{v}$ to within an additive constant. From this, from the constancy of the conditional variances, and from the inductive hypothesis that $\mathbf{u}$ and $\mathbf{v}$ must be normally distributed for equality, it follows that $(\mathbf{u}, x) = (x_1, \cdots, x_{m+1})$ and $(\mathbf{v}, y) = (y_1, \cdots, y_{m+1})$ are multivariate-normal when equality obtains in (17) and (18) and in (5) for $n = m$.

Since the entropy of $(\mathbf{w}, z) = (z_1, \cdots, z_{m+1})$ is

$$H(\mathbf{W}, Z) = H(\mathbf{W}) + H(Z \mid \mathbf{W}),$$

we have, from (17), (18), and the inductive hypothesis (5),

$$\frac{2}{m+1} H(\mathbf{W}, Z)$$
$$\geq \frac{m}{m+1} \frac{2}{m} H(\mathbf{W}) + \frac{1}{m+1} \cdot 2H(Z \mid \mathbf{U}, \mathbf{V})$$
$$\geq \frac{m}{m+1} \log (e^{2H(\mathbf{U})/m} + e^{2H(\mathbf{V})/m})$$
$$+ \frac{1}{m+1} \log (e^{2H(X|\mathbf{U})} + e^{2H(Y|\mathbf{V})}), \tag{19}$$

with equality only if, in addition to the foregoing conditions, $\mathbf{u}$ and $\mathbf{v}$ are normally distributed with proportional covariance matrices.

The right-hand side here is of the form of the left-hand side of (15) if, with probability $m/(m+1)$, $H_1 = H(\mathbf{U})/m$ and $H_2 = H(\mathbf{V})/m$ and, with probability $1/(m+1)$, $H_1 = H(X \mid \mathbf{U})$ and $H_2 = H(Y \mid \mathbf{V})$. Hence,

$$\frac{2}{m+1} H(\mathbf{W}, Z) \geq \log (e^{[2/(m+1)]H(\mathbf{U}) + [2/(m+1)]H(X|\mathbf{U})}$$
$$+ e^{[2/(m+1)]H(\mathbf{V}) + [2/(m+1)]H(Y|\mathbf{V})})$$
$$= \log (e^{[2/(m+1)]H(\mathbf{U},X)} + e^{[2/(m+1)]H(\mathbf{V},Y)})$$

which establishes (5) for $n = m + 1$. We have equality in (15) in this instance only if

$$\frac{H(\mathbf{U}) - H(\mathbf{V})}{m} = H(X \mid \mathbf{U}) - H(Y \mid \mathbf{V}).$$

This condition is fulfilled along with the conditions for equality in (17), (18), and (19) only if the variances of x and y for fixed $\mathbf{u}$ and $\mathbf{v}$ are in the same ratio as the mth roots of the determinants of the covariance matrices of $\mathbf{u}$ and $\mathbf{v}$. From the fact that equality requires $\mathbf{u}$ and $\mathbf{v}$ to have proportional covariance matrices and the fact that $E\{x \mid \mathbf{u}\}$ and $E\{y \mid \mathbf{v}\}$ are, to within an additive constant, both the same linear function it follows that the covariance matrices of $(x_1, \cdots, x_{m+1})$ and $(y_1, \cdots, y_{m+1})$ must be proportional.

Thus, provided the integrals for the conditional entropies $H(X \mid \mathbf{u})$ and $H(Y \mid \mathbf{v})$ converge absolutely, we conclude that the entropy power of the sum of two independent band-limited random processes is no less than the sum of the entropy powers of the two processes, with equality if, and only if, they are Gaussian and have proportional spectral densities.

References

[1] Shannon, C. E., A mathematical theory of communication, *Bell Sys. Tech. J.*, vol 27, Oct 1948, pp 623–656.

[2] Stam, A. J., *Some Mathematical Properties of Quantities of Information*, Ph.D. dissertation, Technische Hogeschool te Delft, Uitgeverij Excelsior, the Hague, The Netherlands, Apr 1959. I am indebted to Professor Shannon for referring me to this monograph.

[3] Stam, A. J., Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Information and Control*, vol 2, Jun 1959, pp 101–112.