STATE UNIVERSITY OF NEW YORK AT STONY BROOK

COLLEGE OF ENGINEERING

Report No. 38

ON THE TIME-BANDWIDTH CONCENTRATION OF SIGNAL FUNCTIONS FORMING GIVEN GEOMETRIC VECTOR CONFIGURATIONS

P. M. Dollard

March, 1965

On The Time-Bandwidth Concentration of Signal Functions Forming Given Geometric Vector Configurations

P. M. DOLLARD, MEMBER, IEEE

Summary—Landau, Pollak, and Slepian, [4]-[6] have shown that the prolate spheroidal wave functions play an important role in determining the approximate dimensionality of a space of functions whose energies are concentrated in a given time bandwith WT. They have also shown the extent to which this space may be assumed $2\,WT$ dimensional. The function space which they consider is actually infinite dimensional and a subset of $\mathcal{L}_\infty^2$, but it is not a *linear subspace* of $\mathcal{L}_\infty^2$, nor in general does it necessarily contain any linear subspace of dimensionality $2\,WT$.

However, in the problem of the discrete M-nary channel with additive Gaussian noise and perhaps other types of noise, one is mainly concerned with given n-dimensional linear subspaces of $\mathcal{L}_\infty^2$ and given geometric configurations of vectors in those subspaces. Thus to be conveniently applied to this problem, the results of Landau, Pollak and Slepian should be reformulated in terms of arbitrary given finite dimensional linear subspaces of $\mathcal{L}_\infty^2$, with given geometric configurations therein. This paper undertakes such a reformulation for some important special cases.

In particular, for the cases of orthogonal, biorthogonal and simplex configurations, it is shown that one can orient the configuration such that the time-bandwidth concentration of the least concentrated vector in the configuration is maximized. The maxi-min criterion is chosen because, as is also shown, the average concentration for these three configurations is always independent of orientation.

I. Introductory Remarks

MUCH OF MODERN detection theory was formulated under the impetus of wartime and early post-war researches into radar. The problems involved echoes from large, relatively slow, and widely separated targets. In the context of that problem, everyone was well aware that when the theory required the limits of a convolution integral over the time variable to be extended to plus and minus infinity, this really meant: "far enough to include virtually all the return associated with a single echo—but not so far as to include substantial parts of the returns from other echoes". Usually, this statement was practically meaningful. The separation between echos was normally many times the nominal duration of a single echo. Similar integrals in the frequency domain could have their limits extended over an "effectively infinite" bandwidth without picking up interference from other radars or man-made sources of electromagnetic radiation because the narrow antenna beamwidth, generally high pointing angles and often the geographical remoteness of the actual systems considered gave a high degree of spatial isolation that left the usable spectrum relatively uncrowded. Unfortunately, these convenient circumstances generally have never existed in the field of communications problems. With today's dense, high velocity "threat clouds" and the associated ambiguity and discrimination problems, spatial isolation can no longer so blithely be presumed in the context of many radar problems

In theory, the signal functions of the communications problem and the echo functions of the multitarget radar problem can be made identically zero outside some arbitrarily specified time interval. An integration of a single-signal function over limits just encompassing this interval then gives the desired result without encroaching on time slots otherwise occupied. However, in attempting to synthesize signal functions which are absolutely time limited, one necessarily causes the corresponding spectrum functions to spread beyond any given finite limits, by virtue of the uncertainty principle of Fourier transform theory.

Certainly there are interesting communications problems where by virtue of spatial isolation the entire usable spectrum is, in effect, available at all times to each channel (*e.g.*, the local telephone system). Effective bandwidth is then limited solely by transmission and noise properties of the medium or by practical problems in the design of arbitrarily wideband terminal equipment. However, this limit is not sharp, and leaves the bandwidth difficult to define in terms of a single numeric measure W. Then basic results of communication theory, including the very cornerstone relation

$$C = W \log\left(1 + \frac{S}{N}\right),$$

become correspondingly difficult to interpret.

The type of problem for which the sophisticated theory is at once more clear in its interpretation and more useful in its application is one in which the usable spectrum is shared by frequency multiplexing among a large number of channels with each assigned to a different one of a set of contiguous sharply defined frequency bands. Any spillover in frequency then causes the channels to interfere with one another, just as spill over in time causes the individual time-contiguous signals on one channel to interfere with one another. One must seek signal functions or sets of signal functions which are simultaneously concentrated both in time and bandwidth. This paper is concerned with sets of

Manuscript received August 1, 1963; revised June 26, 1964. The research reported in this paper was sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, under contract AF 19(628)-3865.

The author is with State University of New York, Stony Brook, N. Y. He was formerly with Bell Telephone Laboratories, Whippany, N. J.

functions which are time-bandwidth concentrationed and have certain desirable geometric properties. It is directed mainly at the communications problem just mentioned but may have some interesting implications for the many-target radar problem.

We present the results simply as a collection of theorems and proofs. Interpretations thereof in the way of practical applications would vary widely according to individual circumstances and are therefore left largely to the reader. The results, however, have been formulated with a particular model in mind and with the general intention of relating this model more closely to the actual systems one might like to represent with it.

The model has been discussed quite extensively in the literature (*e.g.*, [1], [2], [3]). It constitutes a discrete M-nary channel on which the M independent equiprobable channel symbols are each represented by one of a configuration of M vectors (usually of equal length) in Euclidean n space. Once each T seconds, a vector is received which consists of the sum of one of these symbol vectors and a spherical Gaussian-distributed random noise vector. The detection is by minimum distance (maximum likelihood).

In this model, for every doublet (M, n) there is an "optimum" configuration (see Shannon, [1]). In particular, for any n and $M = n + 1$ there is the "simplex," while for any n and $M = 2n$ there is the "biorthogonal" configuration.[1] Our results are specifically formulated in terms of these configurations (and of the plain orthogonal configurations, which are not optimum but which lead trivially to the biorthogonal).

The use of such configurations is meaningful only if the vector model is valid. Shannon [1] points out the obvious correspondence between the model and an actual system wherein signaling is accomplished by using waveforms of electromagnetic energy which have nominal duration T and nominal bandwidth W with $n = 2WT$. He also warns his readers to invoke this correspondence at their own peril. We eliminate one possible source of that peril simply by assuming that n, W, and T are given numbers, not necessarily (although perhaps desirably) obeying the $n = 2WT$ relation even approximately. Reduction of the remaining peril by reducing the spurious noiselike effects of intersymbol interference and interchannel interference is one end product of the results to follow. The model requires that the noise be "white Gaussian[2]," whereas the various spurious effects are unlikely to be so.

[1] A "simplex" configuration consists of the position vectors of the vertices of a regular simplex centered at the origin in n space. The $n + 1$ vectors are equi cross correlated with cross correlation of any pair equal to $-1/n$. A "biorthogonal" configuration in n space consists of n equal length mutually orthogonal vectors with the negative of each. It could be termed "cross polytope" by analogy to the term "simplex," but this is obviously too lacking in euphony.

[2] "White Gaussian" noise is defined as noise whose vector representation in the model has all its n orthogonal components given by independent zero mean Gaussian variates with equal variances. (The more usual notion of white Gaussian noise is sufficient but not necessary for this more restricted notion.) A process which is white Gaussian in this sense makes each event a spherical Gaussian n variate.

To the extent that they are not white Gaussian, effects must be kept negligibly small compared to wh ever true noise or noise-like effects are white Gauss and should, of course, be kept small in order to red the total noise.

Functions will be related to vectors in the follow manner. A set of n linearly independent functions $\{x_i($ will be given. We may assume without loss of genera that they are orthonormal such that

$$\int_{-\infty}^{\infty} x_i(t)x_j(t)\,dt = \delta_{ij} \quad \text{(Kronecker delta).}$$

These will be the basis functions which define the n sp Any square integrable function $\hat{s}(t)$ may be represen as a vector in the space by the ordered n tuple of numb

$$\sigma_j = \int_{-\infty}^{\infty} \hat{s}(t)x_j(t)\,dt.$$

Then the function

$$s(t) = \sum_{j=1}^{n} \sigma_j x_j(t)$$

is the orthogonal projection of $\hat{s}(t)$ onto the n space; remaining component $\hat{s}(t) - s(t)$ is orthogonal to space and is "lost" in the representation of $\hat{s}(t)$.

For simplicity of presentation, we will assume for remainder of the paper that the functions $x_j(t)$ identically zero outside the interval $(0, T)$. This illustrate the salient features of the more general and will require a somewhat briefer development. Am other simplifications, this permits a direct connec between the functions $x_j(t)$ and the impulse respo of n "sampling networks" in the receiver of an ac system. Thus by letting $x_j(t) = h_j(T - t)$ where $h_j(t)$ is impulse response of the jth sampling network, σ_j comes the sampled value at time T of the output of jth network when the input is $\hat{s}(t)$.[3]

We are interested primarily in the case where $\hat{s}($ a sequence of signal waveforms,

$$\hat{s}(t) = \sum_k \hat{s}_{i_k}(t - kT),$$

and where $\{\hat{s}_i(t)\}_1^M$ is the set of M channel symbol w forms as they appear at the inputs to the bank of rec sampling networks. Thus the ordered n tuple of ou sample values at $t = T$ represents $\hat{s}_{i_0}$, while the n of sample values at $t = (k + 1)\ T$ represents $\hat{s}_{i_k}$ course, if each of the $\hat{s}_i(t)$ are not absolutely time-lin to $(0, T)$, then the sample values at $t = (k + 1)T$ not be a "pure" representation of $\hat{s}_{i_k}(t - kT)$ but

[3] For further simplification, we will assume either 1) th signalling is at baseband for the channeli n question, with channels occupying modulation bands which start just abov baseband, or 2) that the signals are generated at baseband, r lated up to some assigned RF band, then coherently demodu back to baseband after reception. The modulation-demodu process is conceptually included in the channel, but *all* fil which follows demodulation is assumed to be included as p the sampling networks. These assumptions are necessary in that we may invoke certain prior results which apply or baseband signals.

include contributions from the transient "tails" of earlier arrivals and the anticipatory buildups to later arrivals. This is one form of intersymbol interference.[4] As mentioned above, however, we combine this with the noise and define the "pure" representation of $\hat{s}_i(t)$ by

$$\sigma_{ij} = \int_0^T \hat{s}_i(t) x_j(t)\, dt$$

$$s_i(t) = \sum_{j=1}^{n} \sigma_{ij} x_j(t).$$

For lack of any better or more reasonably simple measure of the intersymbol interference, we will use a measure of relative energy concentration within $(0, T)$ of each of the $\hat{s}_i(t)$. This will be most meaningful when the restriction of $\hat{s}_i(t)$ to $(0, T)$ is in fact equal to $s_i(t)$ so that none of the energy of $\hat{s}_i(t)$ within $(0, T)$ is lost in the projection. A similar measure of spectral energy concentration within $(-W, W)$ in the frequency domain will be used as a measure of interchannel interference.

If the signal that one must transmit in order to have $\hat{s}_i(t)$ appear at the receiver has spectral energy components which "spill over" beyond the assigned baseband $(-W, W)$, then this will cause a noise-like interference to appear within the modulation bands assigned to other channels. Suppose we were to confine our own transmission entirely to the assigned band. Then, if our solicitude for the other channel users was reciprocated in kind, we would have eliminated one form of interchannel interference within our own channel, *viz.*, that which arises from other channels' transmission appearing in our demodulated wave *within* our *assigned baseband* $(-W, W)$. However, we have already assumed that the impulse responses of our sampling networks are absolutely time limited and therefore cannot be absolutely band limited. They *must* admit some energy from outside the $(-W, W)$ band, and our demodulation of the other channels' transmitted energy will place it in bands starting just outside $(-W, W)$. This latter form could be measured by the relative spectral energy concentration of the $x_i(t)$, but it is measured equally well and more conveniently by the concentration of the $\hat{s}_i(t)$, since this more nearly represents the comparative disturbing effect as and when each different signal becomes the one whose presence one is attempting to detect. We will also assume by the above philosophy of reciprocation that the first kind of interchannel interference is estimated by the spectral concentration of our own transmitted signals, and that this in turn must be measured by the spectral concentration of the signals $\hat{s}_i(t)$ which appear in the receiver (since the transmission characteristics of the medium are not specified). That is, we assume the other users will avoid our assigned spectral band approximately to the same degree that we avoid theirs.

[4] Note that if the $x_i(t)$ were not identically zero outside of $(0, T)$, then a similar interference effect would appear even if the $\hat{s}_i$ were absolutely time limited.

With this background we may now proceed to the business at hand. In Section II, we begin with a brief review of some important recent results in the theory of time-bandwidth concentrated functions which will be the basis for our own results. Since the final results are simply set down in Section IV in the form of Theorem: Proof, with little intervening discussion, the results will be summarized and discussed beforehand in Section III.

II. Review of Background Theory

This paper is concerned with the extent to which sets of functions forming simplex and biorthogonal vector configurations in n space can be simultaneously concentrated in a time interval of width T and in a bandwidth interval $(-W, W)$. In particular, it seeks to explore the dependence of this concentration on n, W, T, and the particular n-dimensional function space in which the configuration is imbedded. Clearly this must be related in some way to the uncertainty principle of Fourier transform theory.

In a series of recent papers [4], [5], [6], Landau Pollak, and Slepian have shown that the prolate spheroidal wave functions (PSWF's) play a fundamental role in this uncertainty principle. Among the many significant results in this monumental work, they have at last provided a rigorous mathematical statement of the old engineering addage, that the space of functions time bandwidth concentrated in WT is $2WT$ dimensional. In view of this, it would seem that the question of dimensionality in relating the vector model to actual channels, as discussed in Section I, is now properly answered. Further, the manner in which they provide this answer in terms of time-bandwidth concentration would seem to provide answers simultaneously to the interference problems.

Unfortunately, their results are formulated from a different point of view and in terms of *subsets* of the set of functions in $\mathfrak{L}_\infty^2$ rather than in terms of finite dimensional *subspaces* of $\mathfrak{L}_\infty^2$ itself. These subsets 1) in general may not contain any simplex set of $n + 1$ functions or orthogonal set of n functions if $n = [2WT] + 1$ (one plus the largest integer in $2WT$), 2) almost surely will not contain any entire n-dimensional subspace if $n = [2WT] + 1$, and 3) in general may not even contain any subspace of dimensionality greater than 1. (Any which contain no one-dimensional subspace must be empty.) Thus we must reformulate their results if we are to apply them to finite subspaces. The following brief review of their more pertinent results is provided for the reader's convenience.

We will be interested mainly in two infinite dimensional linear subspaces of $\mathfrak{L}_\infty^2$, which are the ranges of two operators whose domain is all of $\mathfrak{L}_\infty^2$. Norms are defined throughout to be the ordinary Hilbert norms in $\mathfrak{L}_\infty^2$. Following Landau, Pollak, and Slepian, we define the operator D to be that of absolute time truncation to the interval $(-T/2, T/2)$, and operator B to be that of absolute bandwidth limiting to the interval $(-W, W)$.[5] Their ranges are denoted $\mathfrak{D}$ and $\mathfrak{B}$, respectively. Landau, Pollak, and Slepian focus their attention mainly on functions in $\mathfrak{B}$, but for reasons previously explained we choose rather to concentrate on $\mathfrak{D}$. Thus we seek the eigenvectors and eigenvalues of the combined operator DB whose eigenvectors lie in $\mathfrak{D}$, rather than that of BD whose eigenvectors lie in $\mathfrak{B}$. As they point out, however, our form follows directly from theirs by time-frequency duality.

The eigenvector solutions to

$$\lambda f = DBf$$

are the time truncated PSWF's $\{D\psi_i(t; c)\}_0^\infty$ in which c is a parameter equal to πWT, with eigenvalues $\{\lambda_i(c)\}_0^\infty$; $1 > \lambda_0 > \lambda_1 > \cdots > 0$. The $\lambda_i(c)$ are monotone increasing functions of c for all i. We normalize $||D\psi_i||^2 = 1$, and hereafter drop the prefixed operator D such that the notation ψ_i implies time truncated PSWF's. Then

$$(\psi_i, \psi_j) = \delta_{ij}$$

$$(B\psi_i, B\psi_j) = \lambda_i \delta_{ij}.$$

Relative frequency concentration $C_B(f)$ of any $f \in \mathfrak{D}$ is defined by

$$C_B(f) = ||f||^{-2}\, ||Bf||^2.$$

The function ψ_0 is the maximally concentrated function in $\mathfrak{D}$; ψ_1 is the maximally concentrated of all functions orthogonal to ψ_0 in $\mathfrak{D}$, etc.

In terms of the $\{\lambda_i\}_0^\infty$, $C_B(f)$ for arbitrary $f \in \mathfrak{D}$ is given by

$$C_B(f) = \left[\sum_0^\infty a_i^2\right]^{-1}\left[\sum_0^\infty a_i^2 \lambda_i\right]$$

in which the $\{a_i\}_0^\infty$ are the coefficients of the Fourier expansion of f in the PSWF's.

Landau, Pollak, and Slepian [4]–[6] prove the following two theorems ([6], Theorems 1 and 3):

Theorem: Let $\mathfrak{B}(\epsilon)$ be the subset of all functions $f \in \mathfrak{D}$ for which $C_B(f) \geq 1 - \epsilon^2$. Then for any N the first $N + 1$ PSWF's achieve the final minimum in

$$\min_{\{\varphi_i\}_0^N} \max_{f \in \mathfrak{B}(\epsilon)} \min_{\{a_i\}_0^N} ||f||^{-2} \left|\left| f - \sum_{i=0}^N a_i \varphi_i \right|\right|^2.$$

Theorem: For all $f \in \mathfrak{B}(\epsilon)$ and $N = [2WT]$,

$$\min_{\{a_i\}_0^N} ||f||^{-2} \left|\left| f - \sum_{i=0}^N a_i \psi_i \right|\right|^2 \leq \frac{\epsilon^2}{1 - \lambda_{N+1}} \leq 12\epsilon^2.$$

The latter states that for all $f \in \mathfrak{B}(\epsilon)$, a Fourier expansion in PSWF's using $[2WT] + 1$ degrees of freedom is sufficient to represent f to within a relative integrated square error of $12\epsilon^2$, and in this sense, the space of absolutely time-limited, nominally band-limited f tions is $2WT$ dimensional. (Similar results are sh for the more general case of nominally time-bandw limited functions.)

However, in the context of the problem as outl previously, n is a number already chosen, with the al result as one criterion but subject to other considerat as well. In fact the n-dimensional subspace itself w be chosen, and this would not likely be one spanned the first n or any n PSWF's. The PSWF's are not tabulated nor are they the impulse responses of known networks. Moreover, we would not be intere in any finite representation of signals or noise in te of the PSWF's, since the n space we have chosen fact the very space of representations to which we l decided to limit ourselves in the actual detection.

The pertinent question for the problem as state not, "What is the effective dimensionality of a subs all functions time bandwidth concentrated to a g degree?", but rather, "What is the effective concen tion of a given configuration in a given n-dimensi linear subspace of $\mathfrak{L}_\infty^2$?" We will define this effec concentration to be the concentration of the least centrated vector in the configuration when it is orie to maximize this least concentration. This maxi- criterion is used because, as shown below, the ave concentration for each of the particular configurat considered herein is independent of orientation.

III. Summary and Discussion

We are now prepared to summarize the result follow. Let X_n be an n-dimensional linear subspac $\mathfrak{D}$. Then it is shown (*Theorem III*) that the averag the concentrations of the vectors in any orthono set of n vectors in X_n is the same as for any other thonormal set of n vectors in X_n. From this it is sh (*Corollary III-a*) that the average of the concentrat of any simplex set of $n + 1$ vectors in X_n has the s value as for an orthonormal set.

Next it is shown (*Theorem IV*) that for every n $X_n \subset \mathfrak{D}$, there exists at least one orthonormal b each of whose vectors has concentration exactly e to the average concentration. From this it follows triv (*Corollary* (*IV-a*) that there exists a biorthogonal se $2n$ vectors with the same property. This is clearly maxi-minimally concentrated biorthogonal configura in X_n. Next it is shown (*Corollary IV-b*) that sim configurations, each of whose vectors has concentra equal to the average, do not exist in general. Howe special cases for which they do exist are noted, an method valid for "most" if not all other cases is outl for finding simplex configurations whose least con trated vectors very nearly achieve the average. configurations obtained by this method are believe be maxi-minimally concentrated, but at present th pure conjecture.

Vectors in these maxi-minimally concentrated figurations are expressed as n tuples in what is de

[5] At this point, in order to follow the notation of [4]–[6], we make a shift of $T/2$ in the zero reference time to center the T interval. This merely means that the impulses which yield the impulse responses discussed in the introduction occur at $-T/2$ in this new time scale.

as a "proper" basis of X_n. This basis $\{x_i\}_1^n$ has the double orthogonality property

$$(x_i, x_j) = \delta_{ij}$$

$$(Bx_i, Bx_j) = \mu_i \delta_{ij}.$$

Such a basis is shown (*Theorem I*) to exist for all $X_n \subset \mathfrak{D}$. In fact, $\{x_i\}_1^n$ and $\{\mu_i\}_1^n$, respectively, are the normalized eigenvectors and eigenvalues of a certain operator on X_n analogous to the operation DB on $\mathfrak{D}$ itself. The set function Tr (X_n) is defined as the trace of this operator on X_n, *i.e.*, Tr $(X_n) = \sum_1^n \mu_i$. Therefore, the average concentration for any orthonormal basis of X_n is $\bar{\mu} = 1/n$ Tr (X_n). For each given X_n, the μ_i, and therefore $\bar{\mu}$, are clearly monotone increasing functions of $c = \pi WT$ (the x_i themselves depend parametrically on the parameter c) and are semi-ordered. $1 > \mu_0(c) \geq \cdots \geq \mu_{n-1}(c) > 0$ for all $c > 0$.

Thus at least for biorthogonal configurations, the quantity of interest for determining ϵ from given WT and X_n is $\bar{\mu}(c) = 1 - \epsilon^2$. In the special cases noted, this same quantity determines ϵ from given WT and X_n for simplex configurations. That is, in line with the notation of [4]–[6], we measure concentration by the parameter ϵ using the relation maxi-min $C_B = 1 - \epsilon^2$. For these configurations, maxi-min $C_B = \bar{\mu}$. For simplex configurations in general, maxi-min C_B will equal a weighted average μ_i with weights fairly close to unity, and therefore $\bar{\mu}$ will be a good upper bound approximation to maxi-min C_B.

The following is suggested as a formal scheme for choosing a X_n, and finding $\{x_i\}_1^n$ and $\{\mu_i\}_1^n$ for it when WT is given. First find $x_1(t)$ as the most B-concentrated unit energetic function in $\mathfrak{D}$ which one is willing and able to implement.[6] Next find $x_2(t)$ as the most B-concentrated unit energetic function orthogonal to x_1 in $\mathfrak{D}$ which one is willing and able to implement. Continue in this manner to obtain $\{x_i\}_1^n$. Clearly, this spans a X_n and by construction, it is in fact a proper basis for X_n.[7] Except for sign changes, it is the unique proper basis if the concentrations of the x_i are strictly ordered. Presumably, the numerical values of the $\{\mu_i\}_1^n$ will be obtained as a by-product of the calculations leading to the choice of the $\{x_i\}_1^n$. As a guide in determining how much effort it is profitable to expend on the implementation of the $\{x_i\}_1^n$, it is shown (*Theorem II*) that for any $X_n \subset \mathfrak{D}$, $\mu_1(c) \leq \lambda_0(c)$, $\mu_2(c) \leq \lambda_1(c)$, $\cdots$, $\mu_n(c) \leq \lambda_{n-1}(c)$ for all c, where the λ_i are the eigenvalues of DB as noted previously. We might note, incidently, and without proof herein that equality can hold for some value of c and for some $i = k$ only if it holds for all $i < k$, in which case $x_i(t; c) \equiv \psi_{i-1}(t; c)$ for all $i \leq k$ at that value of c.

In this work, we have made no attempt at a direct clarification of the significance of Landau and Pollak's Theorem 3 [6] for the class of problems considered. We have, in fact, taken a directly opposite viewpoint in seeking to determine the extent to which a particular n-dimensional signal space (a linear subspace of $\mathfrak{D}$) can be considered WT-limited in a particular problem, rather than seeking to determine the extent to which the entire subset, of functions in $\mathfrak{D}$ which are B concentrated by a given amount (*not* a *linear subspace* of $\mathfrak{D}$) can be considered to be $2WT$ dimensional. Thus we virtually abandon the *a priori* assumption that $n = 2WT$. Nevertheless, our results are for the most part a straightforward reformulation of Landau and Pollak's from the alternate viewpoint and therefore retain a close relation to theirs. When $n = 2WT$, one should expect that the task of finding and implementing a X_n for which $1/n$ Tr (X_n) represents a high degree of concentration should prove relatively easy. As n is decreased below $2WT$, it should become progressively easier. Conversely, if $n = 2WT/(1-\epsilon^2)$, then $1/n$ Tr $(X_n) < 1 - \epsilon^2$ for all $X_n \subset \mathfrak{D}$. This follows as a direct consequence of *Corollary II-a* which states Tr $(X_n) \leq$ Tr $(\Psi_n) <$ Tr $(\mathfrak{D}) = 2WT$.

The numerical significance of the differences between the two approaches may be illustrated by some examples taken from Table I.[8] One might, for example, require that all signal functions be at least 90 per cent concentrated. But clearly from the table, $\mathcal{B}(\sqrt{0.1})$ for $2WT = 2.55$ contains only two mutually orthogonal vectors and for $2WT = 5.10$ contains only five mutually orthogonal vectors, whereas, $[2WT] + 1$ equals 3 and 6 respectively, for the two cases. Unfortunately, data are not available on the behavior of $\lambda_n(c)$ for larger n and larger $c = \pi WT$, but it appears quite possible that $\mathcal{B}(\sqrt{0.1})$ for large $2WT$ contains more than $[2WT] + 1$ mutually orthogonal vectors. An upper bound is $[2WT/0.9]$, and it is believed that this upper bound is approached asymptotically with increasing $2WT$. Thus if one is interested in orthogonal sets of signal functions WT-limited to a prescribed degree, $[2WT] + 1$ is no more than an estimate of the total number of such functions theoretically available. The approach adopted herein can be used to determine the actual number available in practice. It is interesting also to note that the space spanned by the orthogonal set can contain a vector considerably less concentrated than the vectors in the set. For example, $\mathcal{B}(\sqrt{0.1})$ for $2WT = 5.10$ contains five mutually orthogonal vectors whose actual concentrations are all equal to 0.94^+, yet the 5 space they span contains a vector (not in $\mathcal{B}(\sqrt{0.1})$) whose concentration is less than 0.75.

[6] That is, implement with a network whose impulse response is $h(t) = x_1(T - t)$.

[7] Implied in this is the premise that if one is willing and able to implement the $\{x_i\}_1^n$, one is willing and able to implement any linear combination of them. The construction therefore imparts to the chosen $\{x_i\}_1^n$ a property unique to the proper basis, *viz.*, that x_j is the most concentrated vector in X_n orthogonal to x_i for all $i < j$ and all $j = 1, 2, \cdots, n$.

[8] Data for this table are taken from Slepian and Pollak [4], Table 1.

TABLE I

Tabulations of λ_{n-1} and $\bar{\lambda}$ vs n for Two Values of $2WT = 2c/\pi$.

n	$2WT = 2.55$		$2WT = 5.10$	
	λ_{n-1}	$\bar{\lambda}$	λ_{n-1}	$\bar{\lambda}$
1	0.99589	0.99589	1.00000^-	1.00000^-
2	0.91211	0.95400	0.99988	0.99994
3	0.51905	0.80901	0.99700	0.99896
4	0.11021	0.63431	0.96055	0.98936
5	0.00883	0.50922	0.74790	0.94107
6	0.00038	0.42441	0.32028	0.83760
7	0.00001	0.36378	0.06078	0.72663
8	0.0^62	0.31831	0.00613	0.63656
9	0.0^83	0.28294	0.00042	0.56588

Thus far we have considered only an optimization with respect to the receiver *representations* $\{s_i\}$ of the received signals. Clearly, if the actual received signals are themselves virtually time-limited (no intersymbol interference), then they are identical to the $\{s_i\}$. In this case, both types of interchannel interference described in the introduction are mini-maxed. If, alternatively, the actual received signals are virtually band-limited (no interchannel interference of the first type) then it is shown (*Corollary I-a*) that for every element $x \in X_n$ there exists a unique element $\tilde{x} \in \mathcal{B}$ such that $D\tilde{x} = x$. From this and *Theorem III* it follows (*Corollary III-b*) that for any orthonormal set $\{s_i(t)\}_1^n$ of X_n the average reciprocal *time* concentration $\overline{C_D^{-1}(\tilde{s}_i)}$ of the corresponding set $\{\tilde{s}_i(t)\}_1^n$ is the same as for any other orthonormal basis of X_n where $C_D(f) \equiv \|f\|^{-2}\,\|Df\|^2$. The same is shown (*Corollary III-c*) to be true of any simplex set of $n + 1$ vectors in X_n. Finally it is shown (*Corollary IV-c*) that for any n and any $X_n \subset \mathfrak{D}$ there exists an orthonormal basis $\{s_i(t)\}_1^n$ for which

$$\min_i C_D(\tilde{s}_i) = \overline{C_D^{-1}(\tilde{s}_i)}^{-1},$$

and that this achieves the maxi-min $C_D(\tilde{s}_i)$ over all orthonormal bases of X_n. In this paper, we do not attempt either to prove or disprove the existence in general of an orthonormal basis which simultaneously achieves maxi-min $C_B(s_i)$ and maxi-min $C_D(\tilde{s}_i)$. However, interesting special cases in which they are achieved simultaneously are noted. Similar special cases lead to simultaneous achievement of both maxi-mins for simplex configurations.

With regard to the more general case of transmitted signals which are neither virtually time-limited nor virtually band-limited, little can be said in line with the above. The signals $\tilde{s}_i$ for which $D\tilde{s}_i = s_i$ are no longer unique. One might investigate transmitted signals from the space $\mathfrak{D} + \mathcal{B}$, such as $as_i + (1 - a)\tilde{s}_i$ where $\tilde{s}_i \in \mathcal{B}$ as in the preceding paragraph and $0 \le a \le 1$. For specific problems this may indicate a desirable compromise between no intersymbol interference and no interchannel interference of the first type. However, a suffici accurate model of a given physical situation may clude a choice of transmitter functions from $\mathcal{B}$ - A theory sufficiently general to cover this case is be the scope of the present effort.

IV. Detailed Results

Theorem I: *For each $X_n \subset \mathfrak{D}$, there exists a basis $\{x_i(t$ with the double orthogonality property $(x_i, x_j) =$ $(Bx_i, Bx_j) = \mu_i\,\delta_{ij}$. (It is termed the "proper ba*

Proof: The proof follows trivially from the obs tion[9] that any projection operator P is characterize

P is self adjoint
$P^2 = P$ (P is idempotent).

B is such a projection (of $\mathfrak{L}_\infty^2$ onto $\mathcal{B}$). Define P_X a projection of $\mathfrak{L}_\infty^2$ onto X_n and note that $P_X x = x$ fo $x \in X_n$. Then the combined operator $P_X B$ is compl continuous with n independent solutions to

$$\mu x = P_X B x.$$

The eigenvectors associated with unequal eigenv are orthogonal, and those associated with equal e values can be chosen as orthogonal. All can of cours normalized in $\mathfrak{L}_\infty^2$. Denote these eigenvectors and e values by $\{x_i\}_1^n$ and $\{\mu_i\}_1^n$, respectively. Then

$$\begin{aligned}(Bx_i, Bx_j) &= (B^2x_i, x_j) = (Bx_i, x_j)\\ &= (Bx_i, P_X x_j) = (P_X B x_i, x_j)\\ &= \mu_i(x_i, x_j) = \mu_i\delta_{ij}.\end{aligned}$$ Q.

Corollary I-a: *For every element $x \in X_n$ there is a u element $\tilde{x} \in \mathcal{B}$ such that $D\tilde{x} = x$. Moreover, $\tilde{x} \in BX_n$.*

Proof: Uniqueness follows from the fact that element $f(t) \in \mathcal{B}$ is an entire function of time therefore completely determined by its behavio $(-T/2, T/2)$ or any other finite interval. Existen shown by construction. Note that the operator and D are naturally ordered[10] $P_X \le D$ which im $P_X D = DP_X = P_X$. Since $x \in X_n$, $P_X x = x$. There if $D\tilde{x} = x$, $P_X D\tilde{x} = P_X\tilde{x} = x$. Let the expansion of $\{x_i\}_1^n$ be given by $x = \sum_1^n a_i x_i$. Then $\tilde{x} = \sum_1^n (a_i/\mu_i$ clearly satisfies $P_X\tilde{x} = x$, and indeed $\tilde{x} \in BX_n$. Sin $\mu_i > 0$, $\tilde{x}$ always exists in the form given. Q.

Theorem II: *For any $X_n \subset \mathfrak{D}$, the eigenvalues $\{\mu_i(c$ $P_X B$ are majorized by the first n eigenvalues $\{\lambda_i($ of DB. That is,*

$$\mu_1(c) \le \lambda_0(c), \quad \mu_2(c) \le \lambda_1(c), \quad \cdots, \quad \mu_n(c) \le \lambda_{n-1}($$

[9] See Landau and Pollak [6], p. 1298. See also Dunfor Schwartz (7), p. 480. Dunford and Schwartz define a more g projection (nonorthogonal) which does not require self-adjoi However, the present work deals only with orthogonal proje

[10] *See* Dunford and Schwartz [7], p. 481.

Proof: This theorem in more general form was apparently first proved by Weyl [8]. The following simple proof of the above form is given for the reader's convenience. Let $\Psi_n \subset \mathfrak{D}$ be the space spanned by the first n PSWF's $\{\psi_i\}_0^{n-1}$, and let P_Ψ be the projection of $\mathfrak{L}_\infty^2$ onto Ψ_n. Let $\hat{x}_i$ denote $P_\Psi x_i$ where $x_i \in \{x_i\}_1^n$. Then either (*Case 1*) $\{\hat{x}_i\}_1^n$ is an independent set, or (*Case 2*) it is a dependent set, *i.e.*, $\{\hat{x}_i\}_1^n$ either spans or does not span Ψ_n.

Thus, *Case 1*: There exists a unique linear combination of the $\hat{x}_i$ satisfying

$$\sum_1^n a_i\hat{x}_i = \alpha\psi_{n-1}$$

$$\sum_1^n a_i^2 = 1; \qquad 0 < \alpha \leq 1.$$

The vector $x = \sum_1^n a_i x_i$ is a unit energetic vector in X_n. Since $\{\psi_i\}_0^\infty$ is complete in $\mathfrak{D}$ and $X_n \subset \mathfrak{D}$, x may be expanded in a Fourier series in the $\{\psi_i\}_0^\infty$. This yields

$$x = \alpha\psi_{n-1} + \sum_{i=n}^\infty \alpha_i\psi_i$$

$$\alpha^2 + \sum_{i=n}^\infty \alpha_i^2 = 1.$$

From this

$$C_B(x) = \alpha^2\lambda_{n-1} + \sum_{i=n}^\infty \alpha_i^2\lambda_i \leq \lambda_{n-1}.$$

However, by virtue of the double orthogonality of $\{x_i\}_1^n$,

$$C_B(x) = \sum_1^n a_i^2\mu_i \geq \mu_n$$

$$\therefore \mu_n \leq \lambda_{n-1}.$$

Case 2: There exists at least one linear combination of the $\hat{x}_i$ for which

$$\sum_1^n a_i\hat{x}_i(t) \equiv 0$$

$$\sum_1^n a_i^2 = 1.$$

The vector $x = \sum_1^n a_i x_i \in X_n \subset \mathfrak{D}$. Therefore

$$x = \sum_{i=n}^\infty \alpha_i\psi_i; \qquad \sum_{i=n}^\infty \alpha_i^2 = 1$$

$$C_B(x) = \sum_{i=n}^\infty \lambda_i\alpha_i^2 < \lambda_{n-1}$$

$$C_B(x) = \sum_1^n a_i^2\mu_i \geq \mu_n$$

$$\therefore \mu_n < \lambda_{n-1}.$$

In either case, $\mu_n \leq \lambda_{n-1}$. Now let X_{n-1} denote the space spanned by $\{x_i\}_1^{n-1}$, and Ψ_{n-1} denote the spaced spanned by $\{\psi_i\}_0^{n-2}$. Repeat the above argument to show $\mu_{n-1} \leq \lambda_{n-2}$. Repeat again for $\{\mu_i\}_2^{n-2}$. The final statement $\mu_1 \leq \lambda_0$ is obvious since ψ_0 is the most concentrated vector in $\mathfrak{D}$.

Q.E.D.

We now define the set function Tr (X_n) to be the trace of the operator $P_x B$ on X_n. That is, Tr $(X_n) \equiv \sum_1^n \mu_i$. It follows from *Theorem II* that Tr $(X_n) \leq$ Tr (Ψ_n). Clearly Tr $(\Psi_n) <$ TR $(\mathfrak{D}) = \sum_0^\infty \lambda_i$ since $\lambda_i > 0$ for all i. It is also clear from the results of [4], [5] and [6] that Tr $(\mathfrak{D})$ (which is the trace of DB on $\mathfrak{D}$) equals the trace of BD on $\mathfrak{B}$. The operation BDf is defined as

$$BDf(t) \equiv \int_{-T/2}^{T/2} \frac{\sin 2\pi W(t-t')}{\pi(t-t')} f(t')\, dt'$$

from which

$$\text{Trace } BD = \int_{-T/2}^{T/2} dt' \left[\lim_{t\to t'} \frac{\sin 2\pi W(t-t')}{\pi(t-t')}\right]$$

$$= 2WT.$$

We have thus proven:

Corollary II-a: TR $(X_n) \leq$ Tr$(\Psi_n) <$ TR $(\mathfrak{D}) = 2WT$ *for all* $X_n \subset \mathfrak{D}$.

We next state and prove:

Theorem III: *Given any orthonormal basis* $\{s_i(t)\}_1^n$ *for* $X_n \subset \mathfrak{D}$, *the average concentration* $\overline{C_B(s_i)} = 1/n \sum_1^n C_B(s_i)$ *exactly equals* $1/n$ Tr (X_n) *and is therefore the same for all orthonormal bases of given* X_n.

Proof: Each of the functions $s_i \in \{s_i\}_1^n$ may be expressed as

$$s_i(t) = \sum_{j=1}^n \sigma_{ij}x_j(t)$$

wherein

$$\sum_{k=1}^n \sigma_{ik}\sigma_{jk} = \delta_{ij}.$$

Let S denote the $n \times n$ matrix (σ_{ij}); $i, j = 1, 2, \cdots, n$. S is orthogonal by definition. Now let M denote the $n \times n$ matrix (μ_{ij}); $i, j = 1, 2, \cdots, n$ wherein

$$\mu_{ij} = \begin{Bmatrix} 0, & i \neq j \\ \mu_i, & i = j \end{Bmatrix}.$$

Clearly Trace M = Tr (X_n). The transformation

$$SMS^t = C.$$

is a similarity transform and therefore trace invariant. Note that the diagonal element c_{ii} $(i = 1, 2, \cdots, n)$

of C is given by

$$c_{ii} = \sum_{j=1}^{n} \sigma_{ij}^2 \mu_j = C_B(s_i).$$

Therefore

$$\text{Trace } C = \text{Tr }(\mathrm{X}_n) = \sum_{1}^{n} C_B(s_i)$$

or

$$\overline{C_B(s_i)} = \frac{1}{n} \text{Tr }(\mathrm{X}_n). \qquad \text{Q.E.D.}$$

Corollary III-a: Given any $\mathrm{X}_n \subset \mathfrak{D}$ *and any simplex set of* $n + 1$ *equi-energetic equicrosscorrelated vectors* $\{v_i(t)\}_1^{n+1}$, *the average concentration*

$$\overline{C_B(v_i)} = 1/(n+1) \sum_{1}^{n+1} C_B(v_i)$$

exactly equals $1/n$ Tr (X_n) *and is therefore the same for all simplex sets in the given* X_n.

Proof: We assume without loss of generality that the v_i are unit energetic. Then each of the functions $v_i \in \{v_i\}_1^{n+1}$ may be expressed as

$$v_i(t) = \sum_{j=1}^{n} \xi_{ij} x_j(t)$$

wherein (from the definition of simplex set)[1]

$$\sum_{k=1}^{n} \xi_{ik}\xi_{jk} = \begin{cases} 1; & i = j \\ -\dfrac{1}{n}; & i \neq j \end{cases}.$$

Let V denote the $(n + 1) \times n$ matrix (ξ_{ij}). then form $\Gamma = VMV^t$ where M is the diagonal $n \times n$ matrix (μ_i) as above. Note that the diagonal element γ_{ii} $(i = 1, 2, \cdots, n + 1)$ of Γ is given by

$$\gamma_{ii} = \sum_{j=1}^{n} \xi_{ij}^2 \mu_j = C_B(v_i).$$

Now from V form the $(n + 1) \times (n + 1)$ orthogonal matrix $\hat{V}$ by adding an $(n + 1)$th column all of whose elements are $+1/\sqrt{n}$, (thus making the row vectors mutually orthogonal), then scaling by $\sqrt{n/(n+1)}$ (to renormalize). Also form the $(n + 1) \times (n + 1)$ diagonal matrix $\hat{\mathrm{M}}$ as the direct sum of $\hat{\mathrm{M}}$ and the 1×1 matrix $(\bar{\mu})$ where $\bar{\mu} = 1/n$ Tr (X_n). Then trace $\hat{\mathrm{M}} = (n + 1)/n$ Tr (X_n), and the similarity transform $\hat{V}\hat{\mathrm{M}}\hat{V}^t = \hat{\Gamma}$ is trace invariant. Note that the diagonal element $\hat{\gamma}_{ii}$ $(i = 1, 2, \cdots, n + 1)$ of $\hat{\Gamma}$ is now given by

$$\hat{\gamma}_{ii} = \frac{n}{n+1} \sum_{j=1}^{n} \xi_{ij}^2 \mu_j + \frac{1}{n+1} \bar{\mu}$$

$$= \frac{n}{n+1} C_B(v_i) + \frac{\bar{\mu}}{n+1}.$$

Then

$$\text{Trace } \hat{\Gamma} = \frac{n+1}{n} \text{Tr }(\mathrm{X}_n) = \sum_{i=1}^{n+1} \hat{\gamma}_{ii}$$

$$= \frac{n}{n+1} \sum_{i=1}^{n+1} C_B(v_i) + \frac{1}{n} \text{Tr }(\mathrm{X}_n).$$

Therefore

$$\frac{n}{n+1} \sum_{i=1}^{n+1} C_B(v_i) = \text{Tr }(\mathrm{X}_n)$$

or

$$\overline{C_B(v_i)} = \frac{1}{n} \text{Tr }(\mathrm{X}_n). \qquad \text{Q.}$$

Corollary III-b: Given any orthonormal basis $\{s_i(t)\}$ *[...]* X_n, *denote by* $\{\tilde{s}_i(t)\}_1^n$ *the unique set in* $B\mathrm{X}_n$ *for u[...]* $D\tilde{s}_i = s_i$. *Then the average reciprocal time concentr[...]* $\overline{C_D^{-1}(\tilde{s}_i)} = 1/n \sum_1^n C_D^{-1}(\tilde{s}_i)$ *exactly equals the averag[...] ciprocal eigenvalue* $\overline{\mu^{-1}}$, *and is therefore the same fo[...] orthonormal bases of* X_n.

Proof: By definition,

$$C_D^{-1}(\tilde{s}_i) = ||D\tilde{s}_i||^{-2}\, ||\tilde{s}_i||^2 = ||s_i||^{-2}\, ||\tilde{s}_i||^2 = ||\tilde{s}_i||^2$$

since $||s_i||^2 = 1$. But from *Theorem I*,

$$||\tilde{s}_i||^2 = \sum_{j=1}^{n} \frac{\sigma_{ij}^2}{\mu_j}.$$

Now denote by $\tilde{\mathrm{M}}$ the diagonal matrix $(1/\mu_j)$[11] and $\tilde{C}$ the similarity transform $S\tilde{\mathrm{M}}S^t = \tilde{C}$. Note that [...] diagonal element $\tilde{c}_{ii}$ of $\tilde{C}$ is given by

$$\tilde{c}_{ii} = \sum_{j=1}^{n} \frac{\sigma_{ij}^2}{\mu_j} = ||\tilde{s}_i||^2.$$

Therefore

$$\overline{||\tilde{s}_i||^2} = \frac{1}{n} \text{Trace } \tilde{C} = \frac{1}{n} \text{Trace } \tilde{\mathrm{M}} = \overline{\mu^{-1}}. \qquad \text{Q.E}$$

Corollary III-c: Given any simplex set $\{v_i(t)\}_1^{n+1}$ *[...]* X_n, *the average recirpocal time concentration* $\overline{C_D^{-1}(\tilde{v}_i)}$ *exc[...] equals the average reciprocal eigenvalue* μ^{-1} *and is there[...] the same for all simplex sets in* X_n.

Proof: Following the proofs of *III-a* and *III-b*, [...] that the diagonal element $\tilde{\gamma}_{ii}$ of $\tilde{\Gamma} = V\tilde{\mathrm{M}}V^t$ is given

$$\tilde{\gamma}_{ii} = \sum_{j=1}^{n} \frac{\xi_{ij}^2}{\mu_j} = C_D^{-1}(\tilde{v}_i).$$

[11] $\tilde{\mathrm{M}}$ is in fact M^{-1}.

Form $\hat{V}$ as above, and $\hat{\tilde{M}}$ as the direct sum of $\tilde{M}$ and the 1×1 matrix $(\overline{\mu^{-1}})$. Then as above,

$$\text{Trace } \hat{\Gamma} = \frac{n}{n+1} \sum_{i=1}^{n+1} C_D^{-1}(v_i) + \frac{1}{n} \text{Trace } \tilde{M}$$

$$= \text{Trace } \hat{\tilde{M}} = \frac{n+1}{n} \text{Trace } \tilde{M}.$$

Therefore,

$$\overline{C_D^{-1}(\hat{v}_i)} = \frac{1}{n} \text{Trace } \hat{\tilde{M}} = \overline{\mu^{-1}}. \qquad \text{Q.E.D.}$$

We have proven that simplex and orthogonal (and therefore, as we shall see, biorthogonal) configurations, of $n+1$ and n (and $2n$) vectors, respectively, in a given X_n, all have the same average concentration. One might advance heuristic geometric arguments to show that any configuration with sufficient symmetry would have this property; indeed that any optimum code configuration as defined in [1] or [3], even those with such misfit values of M and n that symmetry is totally lacking, would have "very nearly" this property. That is, the average concentration is at most weakly dependent on orientation of the configuration. We therefore adopt a maxi-minimal criterion, and proceed to show how one can find maxi-minimally concentrated simplex and biorthogonal configurations.

Theorem IV: Given any $X_n \subset \mathfrak{D}$, there exists at least one orthonormal basis $\{s_i(t)\}_1^n$ such that $C_B(s_i) = 1/n$ Tr $(X_n$ for all $i = 1, 2, \cdots, n$.

Proof: We have only to prove that there exists an $n \times n$ orthogonal S matrix such that the similarity transform $SMS^t = C$ yields $c_{ii} = c_{jj}$ for all $i, j = 1, 2, \cdots, n$. The following constructive proof was suggested by Landau.[12] Consider the pair of row vectors (n tuples) in the proper basis $\{x_i\}_1^n$:

$$\left(+\sqrt{\frac{\bar{\mu} - \mu_n}{\mu_1 - \mu_n}}, 0, 0, \cdots, 0, 0, +\sqrt{\frac{\mu_1 - \bar{\mu}}{\mu_1 - \mu_n}}\right)$$

$$\left(-\sqrt{\frac{\mu_1 - \bar{\mu}}{\mu_1 - \mu_n}}, 0, 0, \cdots, 0, 0, +\sqrt{\frac{\bar{\mu} - \mu_n}{\mu_1 - \mu_n}}\right).$$

These are mutually orthogonal and unit energetic. The first has concentration

$$\left(\frac{\bar{\mu} - \mu_n}{\mu_1 - \mu_n}\right)\mu_1 + \left(\frac{\mu_1 - \bar{\mu}}{\mu_1 - \mu_n}\right)\mu_n = \bar{\mu} = \frac{1}{n} \text{Tr } (X_n).$$

The second has concentration

$$\left(\frac{\mu_1 - \bar{\mu}}{\mu_1 - \mu_n}\right)\mu_1 + \left(\frac{\bar{\mu} - \mu_n}{\mu_1 - \mu_n}\right)\mu_n = \mu_1 + \mu_n - \bar{\mu}.$$

[12] In a private conversation with the author.

The first will constitute the first row of the final S matrix. From the second form, the vector

$$\hat{x}_1 = -\sqrt{\frac{\mu_1 - \bar{\mu}}{\mu_1 - \mu_n}}\, x_1 + \sqrt{\frac{\bar{\mu} - \mu_n}{\mu_1 - \mu_n}}\, x_n.$$

Then $\hat{x}_1$ is unit energetic and orthogonal to all the $x_i \varepsilon \{x_i\}_2^{n-1}$. Moreover $B\hat{x}_1$ is orthogonal to all the $Bx_i \varepsilon \{Bx_i\}_2^{n-1}$. Let $X_{n-1} \subset \mathfrak{D}$ be the space spanned by $\hat{x}_1 \cup \{x_i\}_2^{n-1}$ and let $\{x_i'\}_1^{n-1}$ be the reordering of this orthonormal basis to place $\hat{\mu}_1 = \mu_1 + \mu_n - \bar{\mu}$ in its correct position in $\{\mu_i'\}_1^{n-1}$. Note that now

$$\text{Tr } (X_{n-1}) = \mu_1 + \mu_n - \bar{\mu} + \sum_2^{n-1} \mu_1$$

$$= \text{Tr } (X_n) - \bar{\mu}$$

$$= \frac{n-1}{n} \text{Tr } (X_n).$$

The entire process is then repeated to form

1) An $(n-1)$ tuple in the basis $\{x_i'\}_1^{n-1}$ which is unit energetic, has concentration $1/(n-1)$ Tr $(X_{n-1}1) = 1/n$ Tr (X_n), and becomes an n tuple in the original basis $\{x_i\}_1^n$ orthogonal to the one previously found. It constitutes the second-row of S
2) A new space X_{n-2} for which Tr $(X_{n-2}) = (n-2)/(n-1)$ Tr $(X_{n-1}) = (n-2)/n$ Tr (X_n).
3) A proper basis $\{x_i''\}_1^{n-2}$ for X_{n-2} with concentrations $\{\mu_i''\}_1^{n-2}$ in proper sequence.

The process is repeated $n-3$ more times, thereby generating a total of $n-1$ n tuples (rows of S) which are all mutually orthogonal, unit energetic, and have concentrations $1/n$ Tr (X_n). The "remainder space" X_1 has Tr $(X_1) = 1/n$ Tr (X_n). Its single unit energetic basis vector $x_1^{(n-1)}$ therefore has concentration $1/n$ Tr (X_n) also. When expressed as an n tuple in $\{x_i\}_1^n$ it is orthogonal to the first $n-1$ rows of S just found, and therefore can form the nth row to complete an orthogonal matrix $S = (\sigma_{ij})$, $i, j = 1, 2, \cdots, n$, with the desired property. The vectors $s_i \varepsilon \{s_i\}_1^n$ are of course formed by

$$s_i = \sum_{j=1}^n \sigma_{ij} x_j. \qquad \text{Q.E.D.}$$

Corollary IV-a: Given any $X_n \subset \mathfrak{D}$, there exists an equi-energetic biorthogonal code set of $2n$ vectors in X_n each of which has concentration $1/n$ Tr (X_n).

Proof: The proof is trivial. We need only form the set of $2n$ vectors $\{\pm\alpha s_i\}_1^n$ from $\{s_i\}_1^n$ just found and note that $C_B(\alpha s_i) = C_B(-\alpha s_i) = C_B(s_i)$.

Corollary IV-b: It is not true in general that given an X_n there exists a simplex set of $n+1$ code vectors in X_n, each of which has concentration $1/n$ Tr (X_n).

Proof: To prove the corollary as stated, we need only cite a single example to contradict the converse. We choose the trivial example of a "simplex" (triangle) in X_2. The code set consists of three vectors in the plane, separated by 120°. Given any μ_1, μ_2 such that $1 > \mu_1 > \mu_2 > 0$, the orientation which maximizes the least concentration is that which places one of the vectors along the x_1 axis. Then the three concentrations are μ_1, $\frac{1}{4}\mu_1 + \frac{3}{4}\mu_2$, and again $\frac{1}{4}\mu_1 + \frac{3}{4}\mu_2$. Thus the maxi-minimal concentration is $\frac{1}{4}\mu_1 + \frac{3}{4}\mu_2 < \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 = \frac{1}{2}$ Tr (X_2). (Of course if $\mu_1 = \mu_2$, then every vector in the space has concentration $\frac{1}{2}$ Tr (X_2), but this does not occur in general). Q.E.D.

The corollary is intended to imply the existence of a wide class of cases for which a maxi-min equal to the average can be achieved, and a correspondingly wide or wider class for which it cannot. Unfortunately, the proof given contributes little to the implication. Following the proof of *Corollary III-a,* one can easily show that a necessary and sufficient condition for the existence of an equiconcentrated simplex code in X_n is the existence of an orthogonal $(n + 1) \times (n + 1)$ matrix $\hat{V}$ having the same properties as those of the matrix S in *Theorem IV* (but not necessarily constructed by the method given in the proof of *Theorem IV*). That is, to be a $\hat{V}$ matrix it must contain a column all of whose elements are $+1/(\sqrt{n+1})$, while to be simultaneously an S matrix from *Theorem IV* for the ficticious space $\hat{X}_{n+1}$, for which the eigenvalues of $P_{\hat{X}}B$ are $\bar{\mu} \cup \{\mu_i\}_1^n$, it must equalize the diagonal elements of $\hat{V}\hat{M}\hat{V}^t = \hat{\Gamma}$; however, this merely translates the problem to that of finding necessary and sufficient conditions for the existence of such a $\hat{V}$, and these are not known. A sufficient condition, independent of the actual eigenvalues $\{\mu_i\}_1^n$ or of the actual space X_n itself, is the existence of a Hadamard matrix H of order $n + 1$. (See Paley [9] and Peterson [10].)

A Hadamard matrix is a square matrix all of whose elements are ± 1, and whose row vectors are mutually orthogonal. It remains Hadamard if the signs of all elements in any row or column are changed. Thus it may always be transformed to have all plus elements in the final column. If a Hadamard matrix H of order $n + 1$ exists, then $1/(\sqrt{n+1})$ times the H in this form will always yield the desired $\hat{V}$. (The proof is by inspection).

Another set of sufficient conditions is that the set $\{\mu_i\}_1^n$ of eigenvalues of $P_X B$ on X_n contain one value $\mu_j = \bar{\mu}$, and that there exist a Hadamard matrix of order n. Cast the Hadamard matrix in a form such that the jth column has all elements negative. Then scale by $\sqrt{n+1}/n$ and replace the jth column by a column all of whose elements are $-1/n$. Finally add an additional row vector containing all zeros except for a $+1$ in this jth column. This yields an $(n + 1) \times n$ V matrix directly, and the reader may easily verify that $VMV^t = \Gamma$ with $\gamma_{ii} = 1/n$ Tr (X_n) for all $i = 1, 2, \cdots, n + 1$.

If the second but not the first condition is satisfied, *i.e.*, if $\mu_j > \bar{\mu} > \mu_{j+1}$ when the set $\{\mu_i\}_1^n$ is properly ordered, then this same manipulation on the jth or $(j + 1)$th column of H yields a V matrix for X_n whose least centrated vector has concentration $\bar{\mu} - (\mu_j - \bar{\mu})$ μ_{j+1}, respectively. The former cannot be improve small perturbations in the orientation of the confi tion. The latter can, however, be improved by app to the V matrix the simple rotation operation

$$R = \begin{bmatrix} I_j & 0 & 0 \\ 0 & \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} & 0 \\ 0 & 0 & I_{n-j-2} \end{bmatrix}$$

in which I_k is the identity matrix of order k, and s is given by

$$\sin^2\theta = \rho - 2\,\frac{2\rho - 1 + \sqrt{1 + n^2(n+1)\rho - n^2(n+}}{n^2(n+1) + 4}$$

$$\rho = \frac{\bar{\mu} - \mu_{j+1}}{\mu_j - \mu_{j+1}}.$$

Then $V_{(\text{new})} = V_{(\text{old})}R$ contains $n/2 + 1$ row vect with concentrations all equal and slightly below $\bar{\mu}$, v the remaining $n/2$ row vectors have concentration equal and slightly above $\bar{\mu}$. The lower value is

$$C_B(v_{j+1}) = \bar{\mu} - 2(\mu_j - \mu_{j+1}) \cdot \frac{2\rho - 1 + \sqrt{1 + n^2(n+1)\rho - n^2(n+}}{n^2(n+1) + 4}$$

$$= \min_i C_B(v_i).$$

The greater of this number and $\bar{\mu} - (\mu_j - \bar{\mu})/n$ minimal concentration when the original manipula is applied to the index-j row of H) is believed to b maxi-minimal concentration for simplex codes in X_n. Similar but increasingly complex procedures ma used to derive simplex codes whose least concentr vector is "almost" $1/n$ Tr (X_n) in spaces X_n of di sionality one, two, or more above the order of a Hadar matrix.

It should also be noted that the use of Hadar matrices leads to a considerable simplification of procedure for constructing the S matrices of *The IV*. The matrix $1/\sqrt{n}$ times any Hadamard matr order n is clearly an S matrix satisfying the thec provided of course that a Hadamard matrix of orc exists.[14] If none exists of order n, determine the smal for which one of order $n - j$ exists. After the jth in the construction suggested in the proof, the rema

[13] n is even, since by hypothesis there exists an H mat order n.

[14] This fact was apparently noted concurrently by the and by Petrich [11] working independently. However Petrich to have overlooked the generalizations to other dimension and to the simplex sets.

task is to find an S matrix in the basis $\{x_1^{(j)}\}_1^{n-j}$ for the space X_{n-j}. Then $1/(\sqrt{n-j})$ times any Hadamard matrix of order $n - j$ is such an S matrix, and its row vectors expressed as n tuples in the original basis $\{x_i\}_1^n$, along with the j n tuples already found, comprise the rows of the desired S matrix for the original X_n. Rules for generating Hadamard matrices, of virtually all orders for which they are known to exist, are given by Paley [9].

The existence of Hadamard matrices of order n leads to interesting consequences of a further corollary to *Theorem IV*. We first state and prove the corollary.

Corollary IV-c: For every $X_n \subset \mathfrak{D}$ *there exists at least one orthonormal basis* $\{s_i(t)\}_1^n$ *(with the corresponding pre-image set* $\{\hat{s}_i(t)\}_1^n$ *in* BX_n*) for which*

$$\min_i C_D(\tilde{s}_i) = \overline{C_D^{-1}(\tilde{s}_i)}^{-1},$$

and this achieves the maxi-min $C_D(\tilde{s}_i)$ *over all possible orthonormal bases of* X_n.

Proof: The proof directly follows the construction in the proof of *Theorem IV*, with μ_j replaced by μ_j^{-1} for all j and $\bar{\mu}$ replaced by $\overline{\mu^{-1}}$. That construction determines for any given positive definite $n \times n$ diagonal matrix A, with diagonal elements ordered *either* nonincreasing *or* nondecreasing, an orthogonal matrix S such that the diagonal elements of $SAS^t = C$ are equalized. The S so constructed for $A = \tilde{M}$ has row vectors s_j for each of which $C_D^{-1}(\tilde{s}_j) = \overline{C_D^{-1}(\tilde{s}_i)}$. This proves the existence. We prove that this achieves the maxi-min by contradiction. Assume there exists an orthonormal basis $\{s_i'\}_1^n$ for X_n such that

$$\min_i C_D(\tilde{s}_i') > \overline{C_D^{-1}(\tilde{s}_i')}^{-1}.$$

Then

$$\max_i C_D^{-1}(\tilde{s}_i') < \overline{C_D^{-1}(\tilde{s}_i')}.$$

However it is clearly impossible for the largest of any set of real numbers to be less than the average. Q.E.D.

Now note that if there exists a Hadamard matrix of order n, the S matrix obtained from it simultaneously equalizes the diagonal elements of $SMS^t = C$ and of $SMS^t = \tilde{C}$, and therefore simultaneously achieves maxi-min $C_B(s_i)$ and maxi-min $C_D(\tilde{s}_i)$ for any given X_n in $\mathfrak{D}$. Similarly, if there exists a Hadamard matrix of order $n + 1$, then the $\tilde{V}$ matrix obtained from it specifies a simplex set $\{v_i\}_1^{n+1}$ in any given X_n such that the two concentrations are simultaneously maxi-minimized. It is significant to note that since $M = n + 1$ for simplex and $M = 2n$ for biorthogonal encodings, Hadamard matrices of the appropriate orders for both types of code will exist whenever M is a power of 2.

Acknowledgment

The author wishes to acknowledge the patience and kind indulgence of Drs. H. J. Landau and H. O. Pollak in explaining the fine points of their work, and also Dr. Landau's invaluable suggestions and discussions.

References

[1] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Sys. Tech. J.*, vol. 38, pp. 611–656; May, 1959.
[2] D. Slepian, "The threshold effect in modulation systems that expand bandwidth," IRE Trans. on Information Theory, vol. IT-8, pp. 122–127; September, 1962.
[3] G. Lachs, "Optimization of signal waveforms," IEEE Trans. on Information Theory, vol. IT-9, pp. 95–97; April, 1963.
[4] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—I," *Bell Sys. Tech. J.*, vol. 40, pp. 43–64; January, 1961.
[5] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—II," *Bell Sys. Tech. J.*, vol. 40, pp. 65–84; January, 1961.
[6] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—III: The dimension of the space of essentially time- and band-limited signals," *Bell Sys. Tech. J.*, vol. 41, pp. 1295–1336; July, 1962.
[7] N. Dunford and J. T. Schwartz, "Linear Operators—Part I," Interscience Publishers, Inc., New York, N. Y.; 1958.
[8] H. Weyl, *Math. Ann.*, vol. 71, pp. 441–479; 1912.
[9] R. E. A. Paley, "On orthogonal matrices," *J. Math. Phys.*, vol. 12, pp. 311–390; 1933.
[10] W. W. Peterson, "Error Correcting Codes," John Wiley and Sons, New York, N. Y.; 1961.
[11] M. Petrich, "On the number of orthogonal signals which can be placed in a *WT*-product," *J. Siam.*, vol. 11, pp. 936–940; December, 1963.

Reprinted from IEEE TRANSACTIONS ON *INFORMATION THEORY*
Vol. IT-10, Number 4, October 1964
Pp. 328-338