PROCEEDINGS OF THE IEEE, VOL. 64, NO. 3, MARCH 1976

On Bandwidth

DAVID SLEPIAN, FELLOW, IEEE

Abstract—It is easy to argue that real signals must be bandlimited. It is also easy to argue that they cannot be so. This paper presents one possible resolution of this seeming paradox. A philosophical discussion of the role of mathematical models in the exact sciences is given and a new formulation of the 2 *WT* theorem is presented.

The paper is a written version of the second Shannon Lecture given at the 1974 International Symposium on Information Theory. An appendix giving proof of the 2 *WT* theorem has been added.

The Dilemma

Are signals really bandlimited? They seem to be, and yet they seem not to be.

On the one hand, a pair of solid copper wires will not propagate electromagnetic waves at optical frequencies, and so the signals I receive over such a pair must be bandlimited. In fact, it makes little physical sense to talk of energy received over wires at frequencies higher than some finite cutoff W, say 10^{20} Hz. It would seem, then, that signals must be bandlimited.

On the other hand, however, signals of limited bandwidth W are finite Fourier transforms,

$$s(t) = \int_{-W}^{W} e^{2\pi i f t} S(f)\, df$$

and irrefutable mathematical arguments show them to be extremely smooth. They possess derivatives of all orders. Indeed, such integrals are entire functions of t, completely predictable from any little piece, and they cannot vanish on any t interval unless they vanish everywhere. Such signals cannot start or stop, but must go on forever. Surely *real signals* start and stop, and so they cannot be bandlimited!

Thus we have a dilemma: to assume that real signals must go on forever in time (a consequence of bandlimitedness) seems just as unreasonable as to assume that real signals have energy at arbitrarily high frequencies (no bandlimitation). Yet one of these alternatives must hold if we are to avoid mathematical contradiction, for either signals are bandlimited or they are not: there is no other choice. Which do you think they are?

I have my own pet resolution of this seeming paradox, and that is what I plan to talk about this morning. The preliminary discussion is long and will take us rather far afield from Information Theory, but I will come back to touch on it later in a fundamental way if you will but bear with me. My solution to this dilemma will certainly not please all of you: it rests on matters I do not fully understand myself. But, then, perhaps this is the best function these Shannon Lectures can serve—to shake us all up a bit, to stir the waters with controversy. From such a jostling new ideas are often born.

Manuscript received August 15, 1975; revised October 4, 1975. The body of this paper was presented as the second Shannon Lecture at the International Symposium on Information Theory held at Notre Dame University, Notre Dame, IN, October 31, 1974.

The author is with Bell Laboratories, Murray Hill, NJ 07974, and the University of Hawaii, Honolulu, HI 96822.

On Models and Reality

My starting point is to recall to you that each of the quantitative physical sciences—such as physics, chemistry, and most branches of engineering—is comprised of an amalgam of two *distinctly different* components. That these two facets of each science are indeed distinct from one another, that they are made of totally different stuff, is rarely mentioned and certainly not emphasized in the traditional college training of the engineer or scientist. Separate concepts from the two components are continually confused. In fact, we even lack a convenient language for keeping them straight in our thinking. I shall call the two parts Facet A and Facet B.

Facet A consists of observations on, and manipulations of, the "real world." Do not ask me what this real world is: my thoughts become hopelessly muddled here. Let us assume that we all understand the term and agree on what it means. For the electrical engineer, this real world contains oscilloscopes and wires and voltmeters and coils and transistors and thousands of other tangible devices. These are fabricated, interconnected, energized, and studied with other real instruments. Numbers describing the state of this real world are derived from reading meters, thermometers, counters, and dial settings. They are recorded in notebooks as *rational real numbers*. (No other kind of number seems to be *directly* obtained in this real world.)

Facet B is something else again. It is a mathematical model and the means for operating with the model. It consists of papers and pencils and symbols and rules for manipulating the symbols. It also consists of the minds of the men and women who invent and interpret the rules and manipulate the symbols, for without the seeming consistency of their thinking processes there would be no single model to consider. When numerical values are given to some of the symbols, the rules prescribe numerical values for other symbols of the model.

Now, as you all know, we like to think that there is an intimate relationship between Facet A and Facet B of a given science. The numerical value associated with the symbol V_3 in the model should, in the right circumstances, agree with the reading of the voltmeter we have labeled #3 on the workbench over there, the meter we touch in Facet A. Indeed, so confident are we of this agreement that we use the very same name "the voltage across R_3" for these two very different quantities, thus confounding hopelessly the distinction between these constructs. I have carefully said that we "like to think" that there is an intimate relationship between the facets because, in fact, under closer scrutiny one sees the correspondence as tenuous, most incomplete, and imprecise. There is a myriad of detail in the laboratory ignored in the model. Worse yet, many key parts of the model—many of its concepts and operations—have no counterpart in Facet A. To the extent that there is some correspondence between Facets A and B, we have the miracle of modern science—the deepening understanding of our universe, and the bounty and ease of the technological society in which we live. A second-order miracle, little recognized or appreciated, is that this first miracle could arise from such a really ragged fit between the facets.

This gross mismatch goes in two directions. Details from Facet A do not appear in Facet B; details of Facet B may have no counterpart or meaning in Facet A. The first type of mismatch usually causes little trouble. The angle of the caster wheels on the little portable table supporting the oscilloscope with which we observe the waveform of a signal does not enter our circuit equations. We say that this angle has no effect on the circuit and we like to think that, if we wanted to, we could describe a completely comprehensive model that would include the angle of the caster wheels as a parameter and that this model would indeed show the voltage here or there to depend hardly at all on the inclination of the caster wheels.

Mismatches the other way are much more troublesome. Our mathematical models are full of concepts, operations, and symbols that have no counterpart in Facet A. Take the very fundamental notion of real number, for instance. In Facet B, rtain symbols take numerical values that are supposed to correspond to the readings of instruments in Facet A. Almost always in Facet B these numerical values are elements of the real-number continuum, the rationals *and irrationals*. This latter sort of number seems to have no counterpart in Facet A. In Facet B, irrational numbers are defined by limiting ,erations or Dedekind cuts—mental exercises that with some effort and practice we can be trained to "understand" and agree upon. After years of experience with them, we theoreticians find them very "real," but they do not seem to belong to the real world of Facet A. *The direct result of every instrument reading in the laboratory is a finite string of decimal digits—usually fewer than 6—and a small integer indicating the exponent of some power of 10 to be used as a factor.* Irrationals just cannot result directly from real measurements, as I understand them.[1]

Now there are several ways in which we can handle this fundamental lack of correspondence between symbol values in Facet B and measurements in Facet A. We could build a mathematical model in which only a finite number of numbers can occur, say those with 10 significant digits and one of a few hundred exponents. Differential equations would be replaced by difference equations, and complicated boundary conditions and rules would have to be added to treat the ındoff problem at every stage. The model would be exceedingly complex. Much simpler is the scheme usually adopted and known to you all. We admit the real-line continuum into Facet B and we impose yet another abstraction—continuity. In the end, if the model says the voltage is π, we pleased if the meter in Facet A reads 3.1417. We work ..ıth the abstract continuum in Facet B, and we round off to make the correspondence with Facet A.

Mathematical continuity deserves a few words. It is another concept with no counterpart in the real world. It makes no sense at all to ask whether *in Facet A* the position of the voltmeter needle is a continuous function of time. Observing the position of the needle at millisecond or microsecond or even picosecond intervals comes no closer to answering the question than does measurement daily or annually. Yet continuity is a vital concept for Facet B. By invoking it, by demanding continuous solutions of the equations of our models, we make the parts of the model that correspond to measurements in Facet A insensitive to small changes in the parts of the model that do not correspond to anything in Facet A. Specifically, continuity means that the first five significant digits of our computed answers, those to which we do ultimately attribute real significance, will be dependent only weakly on the sixth to tenth significant digits of the numbers we assign to the parameters of the model. They will be essentially independent of the 100th or 1000th significant digit—constructs of importance to the working of Facet B but with no meaningful counterpart in Facet A.

The situation just exemplified by this discussion of numbers and continuity occurs in many different guises in the sciences. There are certain constructs in our models (such as the first few significant digits of some numerical variable) to which we attach physical significance. That is to say, we wish them to agree quantitatively with certain measurable quantities in a real-world experiment. Let us call these the *principal quantities* of Facet B. Other parts of our models have no direct meaningful counterparts in Facet A but are mathematical abstractions introduced into Facet B to make a tractable model. We call these *secondary constructs* or *secondary quantities*. One can, of course, consider and study any model that one chooses to. It is my contention, however, that a necessary and important condition for a model to be *useful* in science is that the *principal quantities of the model be insensitive to small changes in the secondary quantities*. Most of us would treat with great suspicion a model that predicts stable flight for an airplane if some parameter is irrational but predicts disaster if that parameter is a nearby rational number. Few of us would board a plane designed from such a model.

The Dilemma Resolved

What has this long digression to do with bandwidth? (You have been most patient.) I assert that, as *usually* used by members of this sophisticated audience, the words "bandlimited," "start," "stop," and even "frequency" describe secondary constructs from Facet B of our field. They are abstractions we have introduced into our paper and pencil game for our convenience in working with the model. They require precise specification of the signals in the model at times in the infinitely remote past and in the infinitely distant future. These notions have no meaningful counterpart in Facet A. We are no more able to determine by measurements whether a "real signal" was *always* "zero" before noon today than we are able to determine its continuity with time.

I shall soon discuss Facet B of communication theory in more detail and tell you how we can proceed to keep the principal quantities insensitive to small changes in the secondary ones. But first, we have come far enough to lay to rest the question with which I opened this discussion. Are signals really bandlimited? If you mean real signals, those of Facet A, and if by *bandlimited* you mean the usual definition of our trade in terms of the Fourier integral—a notion from Facet B—then I assert this is a nonsense question, one completely without meaning. If you are asking a question about the signals of Facet B, why, use whatever kind that suits your purposes in the model—bandlimited or not as you choose. In *useful* models, the principal quantities will be insensitive to this choice.

[1] An idiosyncratic scientist can, of course, use the symbols π or e or $\sqrt{2}$ in place of digits, but this does not alter the situation. The point is that what is recorded in the notebook is a "word" drawn from a finite (but possibly very large) list of words.

And so, you see, by purposely mixing concepts from Facets A and B, I set up a strawman to begin my lecture: after much talk, I have succeeded in tearing him down.

Models in Communication Theory

Let us turn now to look in more detail at the models of communication theory. Signals frequently are represented as functions of time that are defined on the whole real line. Many real-world devices are then represented in Facet B by linear time-invariant operators that transform one signal of the model into another. The notion of time invariance entails shifts of arbitrary duration. This introduction into Facet B of the infinitely remote past and the infinitely distant future is certainly worrisome from a philosophical point of view. Indeed, as was the case with the irrationals, it could be avoided. But, by introducing these abstractions, we obtain an enormous simplification, one so great as to override objections on mere philosophical grounds. On the infinite domain, *all* time-invariant linear operators have the same simple eigenfunctions—the complex exponentials. A sinusoid in gives a sinusoid of like frequency out. This simplification makes possible the elementary description of devices by transfer functions: it is the true genesis of the widespread applicability of Fourier analysis to electrical engineering.

But what about these infinities affecting the principal quantities of the model? For the common garden-variety principal quantities—the power dissipated in resistor R, the measured voltage at time t at the output terminals, etc.—it is clear that small changes made in the model in the signal behavior at very large times or, dually, at very high frequencies, cause correspondingly small changes in these principal quantities. If ϵ is small enough, the *numerical values* of

$$s_1(t) \equiv \int_{-W}^{W} e^{2\pi i f t} S(f)\, df$$

a bandlimited signal, and those of

$$s_2(t) \equiv s_1(t) + \epsilon \left[\int_{-\infty}^{-W} \frac{e^{2\pi i f t}}{1+f^2}\, df + \int_{W}^{\infty} \frac{e^{2\pi i f t}}{1+f^2}\, df \right]$$

which is not bandlimited, are very nearly the same for all values of t. Fortunately, most principal quantities depend just on such numerical values, for there *are* other ways in which s_1 and s_2 differ drastically. For example, s_1 is infinitely differentiable at $t = 0$, while s_2 is not differentiable there at all. Yet s_2 differs from s_1 only by arbitrarily small changes in its high-frequency tail. We should thus be wary of making the order of differentiability a principal quantity in any physical theory.

Now while most classical principal quantities we deal with do not seem to be sensitive to small changes in signal behavior at infinity, Information Theory has come along with some questions whose answers at first blush do seem to depend on these secondary quantities. Anyone who thinks deeply about data transmission asks sooner or later for the number of numbers we can transmit per second with "real" signals. The question is fuzzy,[2] and a fully satisfying answer to even a well-posed form of the question is elusive.

[2] If we can transmit one real number per second, say $s = .a_1a_2a_3 \cdots$, where the a's are decimal digits, then we can also transmit two real numbers $.b_1b_2 \cdots$ and $.c_1c_2$ per second by the trick of transmitting $s = .b_1c_1b_2c_2 \cdots$, etc.

As all of you in this audience know, most of Shannon's great 1948 paper founding Information Theory, and the bulk of his succeeding work, dealt with time-discrete communication systems. His channel models accepted inputs and delivered outputs *at discrete instants in time:* the inputs and outputs were either real numbers or symbols drawn from a countable list. By means of coding theorems, he rigorously established the limits of communication possible over such channels. The results can all be expressed in terms of bits transmitted *per channel use*. Time, the infinitely remote past and infinitely distant future, need never be considered. Later papers set tight bounds on how many bits could be transmitted and with what accuracy with N uses of the channel. These could equally well be N independent identical channels used at the same time, or they could refer to uses of a single channel at different times.

The extension of these ideas to create a model that describes accurately the limits of electrical communication systems in the real world is fraught with difficulties. I cannot possibly go into the details of these difficulties here. Two different approaches to overcoming them are to be found in Chapter 8 of Gallager's book on Information Theory [3] and in Wyner's paper entitled "Capacity of the band-limited Gaussian channel" [4]. Shannon, by an adroit hand-waving argument, extended his discrete channel work to the time-continuous case and, of course, came up with the correct answer. He invoked the sampling theorem and argued in an imprecise way that "using signals of bandwidth W one can transmit only $2WT$ independent numbers in time T." Shannon himself was unhappy with his method of bridging the gap from the time-discrete to the time-continuous case. Indeed, it was as a result of questions he raised in trying to make rigorous this notion of $2WT$ degrees of freedom for signals of duration T and bandwidth W that the research leading to the Landau–Pollak theorem got under way. I want to close my talk with a short discussion of this important result. I put a slightly different emphasis on the theorem which makes it fit particularly well into the framework I have constructed for models and their correspondence with the real world.

A New Version of the $2WT$ Theorem

As a first step, I shall give a special significance to energy. In Facet B, let us define the energy of a signal $s(t)$ by

$$E[s] = \int_{-\infty}^{\infty} s^2(t)\, dt$$

and let us restrict our attention to signals for which this integral exists, i.e., to signals of finite energy. We take this energy to be a principal quantity of the model and so assume that a direct reading of a laboratory instrument will provide us with a "corresponding energy of the corresponding real signal." We have the usual fuzziness in this correspondence: Facet B in general gives us an irrational for E; in the laboratory our instrument measures energy only to a few decimal places. We suppose that the units are fixed in which we measure various quantities and that in these units there is a minimum energy $\epsilon > 0$ that we can just detect with our meters in Facet A.

I have already commented on the lack of precise correspondence between signals in Facet B and Facet A. Since small enough changes in the signals of the model are not to affect quantities with meaning in Facet A, it seems natural to attempt to make the correspondence many-to-one. We wish to say that two Facet B signals correspond to the same Facet A signal if they are enough alike in form. If they do correspond to the same Facet A signal, we shall also say they are "really indistinguishable." But what should we take for this criterion of distinguishability? The energy of the difference, $E[s_1 - s_2]$, of course. Thus we adopt the Facet B definition:

Two signals, $s_1(t)$ and $s_2(t)$, are *really indistinguishable at level* ϵ if

$$E[s_1(t) - s_2(t)] \equiv \int_{-\infty}^{\infty} [s_1(t) - s_2(t)]^2 \, dt \leqslant \epsilon.$$

Thus if, in the real world, we cannot measure the energy of the difference of the corresponding signals, the signals must be considered "the same." Notice that, at level ϵ, $s_1(t)$ may be really indistinguishable from $s_2(t)$, and $s_2(t)$ may be really indistinguishable from $s_3(t)$, while $s_1(t)$ and $s_3(t)$ are not really indistinguishable from one another.

Having adopted this definition, we now say that a signal $g(t)$ ... Facet B is timelimited to the interval $(-T/2, T/2)$ at level ϵ if $g(t)$ is really indistinguishable from its time truncation to this interval. That is, we make the definition

$g(t)$ is *timelimited to* $(-T/2, T/2)$ *at level* ϵ if

$$s_1(t) \equiv g(t), \qquad -\infty \leqslant t \leqslant \infty$$

and

$$s_2(t) \equiv \begin{cases} g(t), & |t| \leqslant T/2 \\ 0, & |t| > T/2 \end{cases}$$

are really indistinguishable at level ϵ.

If T_0 is the smallest value of T for which s_1 and s_2 are really indistinguishable at level ϵ, we say that $g(t)$ is of *duration* T_0 *at level* ϵ. Similarly, we make the definition

$g(t)$ *is bandlimited to* $(-W, W)$ *at level* ϵ if $u_1(t)$ and $u_2(t)$ are really indistinguishable at level ϵ, where

$$U_1(f) = G(f), \qquad -\infty \leqslant f \leqslant \infty$$

and

$$U_2(f) = \begin{cases} G(f), & |f| \leqslant W \\ 0, & |f| > W. \end{cases}$$

Here, of course, U_1, U_2, and G are the Fourier transforms of $u_1(t)$, $u_2(t)$, and $g(t)$, respectively. If W_0 is the smallest value of W for which u_1 and u_2 are really indistinguishable at level ϵ, we say that $g(t)$ is a signal of *bandwidth* W_0 *at level* ϵ.

Thus a Facet B signal is bandlimited to $(-W, W)$ at level ϵ .. it is really indistinguishable at level ϵ from the signal obtained by cutting off the high-frequency tails beyond W. We do not require there to be no energy at frequencies higher than W; we require only that the energy there be smaller than the quantity we can just measure in Facet A. Note that with these definitions, doubling the strength of a signal may well increase its bandwidth. Similar remarks hold for the time duration of a signal. A consequence of these definitions is that *all* signals of finite energy are both bandlimited to some finite bandwidth W and timelimited to some finite duration T. This is in distinct contrast to the situation that obtains with the usual definitions, where only the always-zero signal can be both bandlimited and timelimited.

One more definition is now needed to complete the picture. We shall say that a set $\mathcal{F}$ of signals has *approximate dimension N at level ϵ during the interval* $(-T/2, T/2)$ if there is a fixed collection of $N = N(T, \epsilon)$ signals, say $\Psi_1, \Psi_2, \cdots, \Psi_N$, such that every signal in $\mathcal{F}$ is really indistinguishable at level ϵ during the interval from some signal of form $\Sigma_1^N a_i \Psi_i(t)$. That is, we require for each $f \in \mathcal{F}$ that there exist a's such that

$$\int_{-T/2}^{T/2} \left[f(t) - \sum_1^N a_i \Psi_i(t) \right]^2 dt \leqslant \epsilon.$$

We further require that there be no set of $N - 1$ functions whose linear combinations can furnish signals really indistinguishable in this sense from every member of $\mathcal{F}$.

We can now state a version of the Landau-Pollak theorem [5, theorem 12] suited to our point of view. Let $\mathcal{G}_\epsilon$ be the set of all signals bandlimited to $(-W, W)$ and timelimited to $(-T/2, T/2)$ at level ϵ. Let $N(W, T, \epsilon, \epsilon')$ be the approximate dimension of $\mathcal{G}_\epsilon$ at level ϵ' during the interval $(-T/2, T/2)$. Then, for every $\epsilon' > \epsilon$,

$$\lim_{T \to \infty} \frac{N(W, T, \epsilon, \epsilon')}{T} = 2W$$

$$\lim_{W \to \infty} \frac{N(W, T, \epsilon, \epsilon')}{W} = 2T.$$

A proof is given in the Appendix.

It would be satisfying, of course, if in the theorem we could take $\epsilon' = \epsilon$ and still draw the same conclusions, but this is not the case. The situation is delicate. We must be a little more stringent in bandlimiting and timelimiting than in fitting the signals of $\mathcal{G}_\epsilon$ with those of a finite-dimensional function space, but only infinitesimally more stringent. The nuance is a mathematical one, of no significance in the Facet A interpretation of results.

Note that the theorem holds for every ϵ and ϵ' provided only that $\epsilon' > \epsilon$. Thus the result is not really dependent on the precision with which we can measure energy. If in future years we refine our instruments and so decrease ϵ, it will still be true that the approximate dimension of the set of bandlimited and timelimited functions is asymptotically $2WT$ as W or T becomes large. A completely dual form of the theorem exists, of course, in which the signals of $\mathcal{G}_\epsilon$ are approximated in the frequency domain throughout the interval $(-W, W)$.

The foregoing theorem differs in many ways from the Landau-Pollak theorem as stated by those authors, but it is readily established by minor modifications of their techniques and results. The version presented here is particularly well suited to the point of view I have adopted regarding models and their correspondence to the real world. I have made bandwidth and duration of signals principal quantities of the model and independent of such secondary quantities as signal behavior at infinite times or infinite frequencies. The definitions correspond to measurements that can be made in Facet A. The approximate dimension N of the set of bandlimited and timelimited signals is also a principal quantity. We find the robust result that $N = 2WT + o(WT)$, independently of how well we can measure energy.

There are many other areas where these ideas can be used to clarify apparent paradoxes. Singular detection of signals and hyperresolution of optical images are two of the most important of these. Time limitation (in a different sense) keeps me from discussing these subjects now. Let me just say that proper application of the principle of making principal quantities insensitive to secondary ones precludes perfect detection and prevents resolution far beyond the Rayleigh limit.

Appendix
Proof of Dimension Theorem

A. Review

We first recall some properties of a set of special functions discussed in detail in [5] and [6].

Let $W > 0$ and $T > 0$ be given. Let $\Psi_0(t), \Psi_1(t), \cdots$ be a complete set of solution of the integral equation

$$\int_{-T/2}^{T/2} \frac{\sin 2\pi W(t - t')}{\pi(t - t')} \Psi(t')\, dt' = \lambda \Psi(t), \qquad -\infty < t < \infty \tag{1}$$

and let λ_j be the eigenvalue corresponding to $\Psi_j(t)$, $j = 0, 1, \cdots$. We suppose the solutions ordered so that $\lambda_0 \geqslant \lambda_1 \geqslant \lambda_2 \geqslant \cdots$. The $\Psi_j(t)$ can be chosen real and can be normalized so that the following statements hold:

$$\int_{-\infty}^{\infty} \Psi_j(t)\Psi_k(t)\, dt = \delta_{jk} \tag{2}$$

$$\int_{-T/2}^{T/2} \Psi_j(t)\Psi_k(t)\, dt = \delta_{jk}\lambda_j \tag{3}$$

$$\int_{-\infty}^{\infty} e^{2\pi i f t}\, \Psi_j(t)\, dt = \frac{1}{\gamma_j} \Psi_j\left(\frac{T}{2}\frac{f}{W}\right) \chi\left(\frac{f}{W}\right) \tag{4}$$

$$\int_{-T/2}^{T/2} e^{2\pi i f t}\, \Psi_j(t)\, dt = \frac{T}{2W} \gamma_j^* \Psi_j\left(\frac{T}{2}\frac{f}{W}\right) \tag{5}$$

$$\int_{-T/2}^{T/2} \frac{\sin 2\pi W(t' - t)}{\pi(t' - t)} \Psi_j(t)\, dt = \lambda_j \Psi_j(t'), \qquad j, k = 0, 1, 2, \cdots; \quad -\infty < t', f < \infty. \tag{6}$$

Here

$$\chi(t) \equiv \begin{cases} 1, & |t| \leqslant 1 \\ 0, & |t| > 1 \end{cases} \tag{7}$$

and

$$\gamma_j = i^j \sqrt{\frac{2W}{T}\lambda_j}, \qquad j = 0, 1, 2, \cdots$$

where $i = \sqrt{-1}$ and the $*$ denotes complex conjugate.

The quantities λ_j satisfy the inequalities

$$1 > \lambda_j > \lambda_{j+1} > 0, \qquad j = 0, 1, 2, \cdots, \ \lim_{j \to \infty} \lambda_j = 0. \tag{8}$$

Furthermore, for every fixed $\eta > 0$, we have the limits

$$\lim_{WT \to \infty} \lambda_n = \begin{cases} 0, & n = [(1 + \eta)2WT] \\ [1 + e^{\pi b}]^{-1}, & n = [2WT + \frac{b}{\pi}\log WT] \\ 1, & n = [(1 - \eta)2WT] \end{cases} \qquad \eta > 0. \tag{9}$$

Here b and η are numbers independent of W and T and the square brackets denote "largest integer not exceeding."

The functions $\Psi_0(t), \Psi_1(t), \cdots$ are complete in $(-T/2, T/2)$ among all functions square-integrable on that interval. They are also complete in $(-\infty, \infty)$ among all functions in $\mathcal{B}$, defined as the set of all functions $b(t)$ whose Fourier transform

$$B(f) = \int_{-\infty}^{\infty} e^{2\pi i f t} b(t)\, dt$$

vanishes for $|f| > W$. We call members of $\mathcal{B}$ *strictly* bandlimited. From (4) and (7), it follows that the $\Psi_j(t)$ are strictly bandlimited and hence unaltered by low-pass filtering so that

$$\int_{-\infty}^{\infty} \frac{\sin 2\pi W(t - t')}{\pi(t - t')} \Psi_j(t')\, dt' = \Psi_j(t), \qquad j = 0, 1, 2, \cdots. \tag{10}$$

B. The Functions $g_j(t)$

We denote by $\mathcal{G}_\epsilon$ the set of all real functions that are of duration T and bandwidth W at level ϵ. That is, $\mathcal{G}_\epsilon$ is the set of all real functions $g(t)$ square-integrable in $(-\infty, \infty)$ for which

$$E_{\overline{T}}[g] \equiv \int_{|t| > T/2} g^2(t)\, dt \leqslant \epsilon \tag{11}$$

and

$$E_{\overline{W}}[g] \equiv \int_{|f| > W} |G(f)|^2\, df \leqslant \epsilon \tag{12}$$

where

$$G(f) = \int_{-\infty}^{\infty} e^{2\pi i f t} g(t)\, dt$$

is the Fourier transform of $g(t)$. We note that (12) can also be written as

$$\begin{aligned} E_{\overline{W}}[g] &= \int_{-\infty}^{\infty} |G(f)|^2\, df - \int_{-W}^{W} |G(f)|^2\, df \\ &= \int_{-\infty}^{\infty} g^2(t)\, dt - \int_{-W}^{W} df \int_{-\infty}^{\infty} dt\, e^{2\pi i f t} g(t) \int_{-\infty}^{\infty} dt' \cdot e^{-2\pi i f t'} g(t') \\ &= \int_{-\infty}^{\infty} g^2(t)\, dt - \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' \frac{\sin 2\pi W(t - t')}{\pi(t - t')} \cdot g(t) g(t'). \end{aligned} \tag{13}$$

Functions in $\mathcal{G}_\epsilon$ cannot be arbitrarily energetic. What member of $\mathcal{G}_\epsilon$ has the largest energy? We seek to maximize

$$E[g] \equiv \int_{-\infty}^{\infty} g(t)^2\, dt \tag{14}$$

subject to the inequality constraints (11) and (12).

To solve this problem, let us first replace (11) and (12) by the exact constraints

$$E_{\overline{T}}[g] = \alpha, \qquad E_{\overline{W}}[g] = \beta \tag{15}$$

where $\alpha > 0$ and $\beta > 0$ are given numbers. We now seek to maximize (14) subject to (15) by choice of a square-integrable $g(t)$. Introducing Lagrange multipliers μ and ν and using (13), we see that

$$I \equiv \int_{-\infty}^{\infty} g^2(t)\,dt + \mu \int_{-\infty}^{\infty} \left[1 - \chi\left(\frac{2t}{T}\right)\right] g^2(t)\,dt$$

$$+ \nu \left[\int_{-\infty}^{\infty} g^2(t)\,dt - \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' \,\frac{\sin 2\pi W(t-t')}{\pi(t-t')}\, g(t)g(t')\right]$$

must be stationary with respect to small changes in $g(t)$ for the most energetic function. Taking the first variation of I, we find that $g(t)$ must satisfy

$$Ag(t) + B\chi\left(\frac{2t}{T}\right) g(t) = \int_{-\infty}^{\infty} \frac{\sin 2\pi W(t-t')}{\pi(t-t')}\, g(t')\,dt' \quad (16)$$

ere A and B are independent of t.

Now the right side of (16) is a strictly bandlimited function of t since its Fourier transform is $\chi(f/W)G(f)$. The left side of (16) must also be a smooth function of this sort, so that we must have

$$g(t) = \begin{cases} b(t), & |t| \leqslant T/2 \\ \dfrac{A+B}{A}\, b(t), & |t| > T/2 \end{cases} \quad (17)$$

with $b(t) \in \mathcal{B}$. Substituting (17) for g in (16) yields

$$(A+B)b(t) = \int_{-T/2}^{T/2} \frac{\sin 2\pi W(t-t')}{\pi(t-t')}\, b(t')\,dt'$$

$$+ \frac{A+B}{A} \left[\int_{-\infty}^{\infty} dt' - \int_{-T/2}^{T/2} dt'\right] \frac{\sin 2\pi W(t-t')}{\pi(t-t')}\, b(t')$$

$$= \frac{A+B}{A}\, b(t) + \left[1 - \frac{A+B}{A}\right] \int_{-T/2}^{T/2} \frac{\sin \pi W(t-t')}{\pi(t-t')} \cdot b(t')\,dt'$$

$$\int_{-T/2}^{T/2} \frac{\sin \pi W(t-t')}{\pi(t-t')}\, b(t')\,dt' = \frac{(A+B)(1-A)}{B}\, b(t).$$

Comparison with (1) shows that we must have $b(t) = k\Psi_j(t)$ some j, so that, from (17), $g(t)$ is of the form $k_1\Psi_j(t) + k_2\chi(2t/T)\Psi_j(t)$.

By using the results of Subsection A, it is now a simple matter to determine k_1 and k_2 to meet the constraints (15). In this way, we find that the functions

$$\bar{g}_j(t) \equiv \sqrt{\frac{\alpha}{1-\lambda_j}}\,\Psi_j(t) + \sqrt{\frac{\beta}{\lambda_j(1-\lambda_j)}}\,\chi\, \frac{2t}{T}\, \Psi_j(t), \qquad j = 0, 1, 2, \cdots \quad (18)$$

are the only solutions of (15) and (16). A further calculation shows that

$$E[\bar{g}_j] = \frac{\alpha + 2\sqrt{\lambda_j \alpha\beta} + \beta}{1 - \lambda_j}. \quad (19)$$

Since the right member here is monotone increasing in λ_j for $0 \leqslant \lambda_j < 1$, we have now shown that $\bar{g}_0$ has the greatest energy among all functions satisfying (15). Note also that under the constraints $0 \leqslant \alpha < \epsilon$, $0 \leqslant \beta < \epsilon$, this greatest energy

$$E[\bar{g}_0] = \frac{\alpha + 2\sqrt{\lambda_0\alpha\beta} + \beta}{1 - \lambda_0}$$

is maximized when $\alpha = \beta = \epsilon$.

It is now convenient to define the functions

$$g_j(t) = \sqrt{\frac{\epsilon}{1-\lambda_j}}\,\Psi_j(t) + \sqrt{\frac{\epsilon}{\lambda_j(1-\lambda_j)}}\,\chi\left(\frac{2t}{T}\right)\Psi_j(t), \qquad j = 0, 1, 2, \cdots \quad (20)$$

obtained from the $\bar{g}_j$ by setting $\alpha = \beta = \epsilon$. They have the following important properties:

$$E_{\overline{T}}[g_j] = E_{\overline{W}}[g_j] = \epsilon \quad (21)$$

$$E[g_j] \equiv \int_{-\infty}^{\infty} g_j^2(t)\,dt = \frac{2\epsilon}{1 - \sqrt{\lambda_j}} \quad (22)$$

$$\int_{-\infty}^{\infty} g_j(t)g_k(t)\,dt = \delta_{jk}\,\frac{2\epsilon}{1-\sqrt{\lambda_j}} \quad (23)$$

$$\int_{-T/2}^{T/2} g_j(t)g_k(t)\,dt = \delta_{jk}\,\frac{1+\sqrt{\lambda_j}}{1-\sqrt{\lambda_j}}\,\epsilon, \qquad j,k = 0, 1, 2, \cdots. \quad (24)$$

The $g_j(t)$ belong to $\mathcal{G}_\epsilon$ and we have shown that among all members of $\mathcal{G}_\epsilon$, $g_0(t)$ has the greatest energy.

We now seek that member of $\mathcal{G}_\epsilon$ orthogonal to $g_0(t)$ on $(-\infty, \infty)$ that has the greatest energy. To proceed, we again replace the constraints (11) and (12) by (15), i.e., we first find the function $\bar{g}$ of the largest energy orthogonal to $g_0(t)$ and satisfying (15). As before, we find the function must be a $\bar{g}_j$ given by (18). Equation (19) and the properties of the Ψ_j and λ_j show that $\bar{g}_1(t)$ is the desired $\bar{g}$. Maximizing on α and β subject to $0 \leqslant \alpha, \beta < \epsilon$ shows $g_1(t)$ to be the member of $\mathcal{G}_\epsilon$ of largest energy orthogonal to $g_0(t)$. In the same manner, we see that for $k = 1, 2, \cdots, g_k(t)$ is that member of $\mathcal{G}_\epsilon$ orthogonal to $g_0(t), g_1(t), \cdots, g_{k-1}(t)$ that has the greatest energy.

In exactly the same way, it can be shown that, among all functions of $\mathcal{G}_\epsilon$, $g_0(t)$ has the largest energy in $(-T/2, T/2)$. For $k = 1, 2, \cdots, g_k(t)$ is the member of $\mathcal{G}_\epsilon$ orthogonal on $(-T/2, T/2)$ to $g_0(t), g_1(t), \cdots, g_{k-1}(t)$ that has the greatest energy in $(-T/2, T/2)$.

The $g_i(t)$ are not complete in $\mathcal{G}_\epsilon$ but, as we shall see, they are the best functions to use in economically approximating functions in $\mathcal{G}_\epsilon$. We note from (21), (22), and (8) that, as $j \to \infty$, g_j tends to have as much energy in band as out of band and as much energy inside $(-T/2, T/2)$ as outside of this interval.

C. The Dimension Theorem

In this section we wish to show that, for every $\eta > 0$ and every $\epsilon' > \epsilon > 0$, there exists a number $c(\eta, \epsilon, \epsilon')$ such that

$$1 - \eta \leqslant \frac{N(W, T, \epsilon, \epsilon')}{2WT} \leqslant 1 + \eta \quad (25)$$

whenever

$$WT > c(\eta, \epsilon, \epsilon').$$

Here, as in the main text, $N(W, T, \epsilon, \epsilon')$ is the approximate dimension of $\mathcal{G}_\epsilon$ at level ϵ'.

1) The Right Inequality of (25): Consider approximating in the interval $(-T/2, T/2)$ any $g \in \mathcal{G}_\epsilon$ by its projection

$$\hat{g}(t) = \sum_0^m a_j \Psi_j(t) \tag{26}$$

$$a_j = \frac{1}{\lambda_j} \int_{-T/2}^{T/2} g(t)\Psi_j(t)\, dt, \qquad j = 0, 1, \cdots, m \tag{27}$$

on the space spanned by the first $(m + 1)$ Ψ's. The energy measure of the error of this approximation is

$$\sigma_m^2 = \int_{-T/2}^{T/2} [g - \hat{g}]^2\, dt = \int_{-T/2}^{T/2} g^2\, dt - \sum_0^m \frac{1}{\lambda_j} \left[\int_{-T/2}^{T/2} g(t)\Psi_j(t)\, dt \right]^2. \tag{28}$$

We shall show below that if $\epsilon' > \epsilon$, then for all $g \in \mathcal{G}_\epsilon$

$$\sigma_m^2 < \epsilon', \qquad \text{when } m = [2WT + k_1(\epsilon') \log WT] \tag{29}$$

provided $WT > k_2(\epsilon')$. Here k_1 and k_2 are independent of W and T. For $WT > k_2$, therefore, we have

$$N(W, T, \epsilon, \epsilon') \leqslant m + 1 \leqslant 2WT + 1 + k_1(\epsilon') \log WT$$

and so

$$N/2WT \leqslant 1 + \frac{1}{2}\left(\frac{1}{WT} + k_1 \frac{\log WT}{WT}\right).$$

Now let $x_0(\eta)$ be the smallest value of x for which $(1/x) + k_1 \cdot (\log x)/x = 2\eta$. Define $c = \max(k_2, x_0)$. Then, whenever $WT > c$, we have $N/2WT \leqslant 1 + \eta$ and the right side of (25) is established.

To show that $\sigma_m^2 < \epsilon'$ when (29) holds, we consider maximizing (28) over all $g \in \mathcal{G}_\epsilon$ that satisfy (15) where we assume $0 < \alpha, \beta \leqslant \epsilon$. Introducing Lagrange multipliers A and B, we see from (28) and (13) that

$$J \equiv \int_{-\infty}^{\infty} \chi\left(\frac{2T}{T}\right) g^2(t)\, dt - \sum_0^m \frac{1}{\lambda_j}\left[\int_{-\infty}^{\infty} \chi\left(\frac{2t}{T}\right) g(t)\Psi_j(t)\, dt\right]^2 + A \int_{-\infty}^{\infty}\left[1 - \chi\left(\frac{2t}{T}\right)\right] g^2(t)\, dt + B\left[\int_{-\infty}^{\infty} g^2(t)\, dt - \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' \frac{\sin 2\pi W(t - t')}{\pi(t - t')} g(t)g(t')\right] \tag{30}$$

must be stationary for the maximizing g. On setting the first variation of J equal to zero, we find that

$$(A + B)g(t) + (1 - A)\chi\left(\frac{2t}{T}\right) g(t) - \sum_0^m a_i \chi\left(\frac{2t}{T}\right)\Psi_i(t) = B \int_{-\infty}^{\infty} dt' \frac{\sin 2\pi W(t - t')}{\pi(t - t')} g(t'). \tag{31}$$

Now assume for the moment that

$$A \neq 0, \qquad B \neq 0, \qquad B \neq -1. \tag{32}$$

Since the right of (31) is a strictly bandlimited function of we must have

$$g(t) = \begin{cases} b(t), & |t| > T/2 \\ \dfrac{A + B}{B + 1} b(t) + \dfrac{1}{B + 1} \displaystyle\sum_0^m a_j \Psi_j(t), & |t| \leqslant T/2 \end{cases} \tag{33}$$

for some strictly bandlimited function $b(t)$. But now we ca write

$$b(t) = \sum_0^\infty b_i \Psi_i(t) \tag{34}$$

since the Ψ_i are complete in $\mathcal{B}$. Inserting (33) and (34) int (31) and using the independence of the Ψ's, we find that

$$[A(1 + B) - \lambda_j B(A - 1)] b_j = a_j \lambda_j B, \qquad j = 0, 1, \cdots, m - 1 \tag{35}$$

and that

$$[A(1 + B) - \lambda_j B(A - 1)] b_j = 0, \qquad j \geqslant m. \tag{36}$$

Now in (27), for $g(t)$, insert the value given by (33) and (34) There results

$$a_j \lambda_j B = (A + B)\lambda_j b_j, \qquad j = 0, 1, \cdots, m - 1. \tag{37}$$

We now use (37) to eliminate a_j from (35), and so find that

$$A(1 + B)(1 - \lambda_j) b_j = 0, \qquad j = 0, 1, \cdots, m - 1.$$

By virtue of (32) and (8), we conclude that $b_j = 0$, $j = 0$, 1 $\cdots, m - 1$, and hence from (37) that also $a_j = 0$, $j = 0, 1, \cdots$ $m - 1$. From (27), we now see that g is orthogonal to Ψ_0, Ψ_1 $\cdots, \Psi_{m-1}$ in $(-T/2, T/2)$ and hence from (20) is also orthog onal to $g_0, g_1, \cdots, g_{m-1}$ in $(-T/2, T/2)$. Note that when th a's are zero,

$$\sigma_m^2 = \int_{-T/2}^{T/2} g^2(t)\, dt.$$

We have now shown that when (32) holds, the maximum o σ_m^2 for all g satisfying (15) is the largest energy in $(-T/2, T/2$ of any g satisfying (15) that is also orthogonal on $(-T/2, T/2$ to $g_0, g_1, \cdots, g_{m-1}$. We saw in Subsection B that this larges energy is attained by $\bar{g}_m$ and has the value $(\alpha + 2\sqrt{\lambda_m \alpha\beta}$ - $\beta)/(1 - \lambda_m)$. Further maximization over α and β then gives

$$\sigma_m^2 \leqslant \frac{1 + \sqrt{\lambda_m}}{1 - \sqrt{\lambda_m}}\, \epsilon \tag{38}$$

for all $g \in \mathcal{G}_\epsilon$, with equality when $g = g_m$. It is not hard t show that, if any of the conditions in (32) is violated, ever solution of (31) contained in $\mathcal{G}_\epsilon$ also satisfies (38).

Equation (29) now follows easily from (38) and (9). Th limit in (9) implies the existence of a function $n_0(\delta)$ such that for $WT > n_0(\delta)$,

$$\lambda_m \leqslant \frac{1}{1+e^{\pi b}} + \delta, \qquad \text{when } m = \left[2WT + \frac{b}{\pi}\log WT\right].$$

Let $\delta = \frac{1}{2}[(\epsilon' - \epsilon)/(\epsilon' + \epsilon)]^2$ and set $b = (1/\pi)\log[(1-\delta)/\delta]$. We then have $\lambda_m \leqslant 2\delta = [(\epsilon'-\epsilon)/(\epsilon'+\epsilon)]^2$, which implies $[(1+\sqrt{\lambda_m})/(1-\sqrt{\lambda_m})]\ \epsilon \leqslant \epsilon'$. Thus, for the two constants k_1 and k_2 introduced at (29), we have

$$k_1 = \frac{1}{\pi^2}\log\left[2\left(\frac{\epsilon'+\epsilon}{\epsilon'-\epsilon}\right)^2 - 1\right]$$

and

$$k_2 = n_0\left[\frac{1}{2}\left(\frac{\epsilon'-\epsilon}{\epsilon'+\epsilon}\right)^2\right].$$

2) The Left Inequality of (25): To establish the left inequality of (25), we must show that for any $\varphi_0, \varphi_1, \cdots, \varphi_m$, nd every $\epsilon' > \epsilon > 0$, and for every $\eta > 0$, for sufficiently large WT there exists a $\hat{g} \in \mathcal{G}_\epsilon$ such that

$$\min_{a_i} \int_{-T/2}^{T/2} \left[\hat{g} - \sum_0^m a_i\varphi_i(t)\right]^2 dt > \epsilon' \tag{39}$$

where

$$m = [(1-\eta)2WT]. \tag{40}$$

For then, from the definition of $N(W, T, \epsilon, \epsilon')$, we have

$$N(W, T, \epsilon, \epsilon') > m + 1 \geqslant (1-\eta)2WT$$

which is the left inequality of (25).

Let $\hat{\mathcal{G}}_\epsilon$ be the set of strictly bandlimited functions $\hat{g}$ for which

$$E_{\overline{T}} \equiv \int_{|t|>T/2} \hat{g}^2(t)\,dt = \epsilon. \tag{41}$$

$\hat{\mathcal{G}}_\epsilon$ is a subset of $\mathcal{G}_\epsilon$. We shall find a $\hat{g}$ in $\hat{\mathcal{G}}_\epsilon$ for which (39) and (40) hold.

We shall show below that the quantity

$$K = \sup_{\hat{g}\in\hat{\mathcal{G}}_\epsilon} \min_{a_i} \int_{-T/2}^{T/2} \left[\hat{g} - \sum_0^m a_i\varphi_i(t)\right]^2 dt \tag{42}$$

is smallest when the φ_i are the functions $\Psi_0, \Psi_1, \cdots, \Psi_m$. Thus

$$K \geqslant \sup_{\hat{g}\in\hat{\mathcal{G}}_\epsilon} \min_{a_i} \int_{-T/2}^{T/2} \left[\hat{g} - \sum_0^m a_i\Psi_i(t)\right]^2 dt$$

$$\geqslant \min_{a_i} \int_{-T/2}^{T/2} \left[f - \sum_0^m a_i\Psi_i(t)\right]^2 dt \equiv L$$

where f is a particular member of $\hat{\mathcal{G}}_\epsilon$. We now choose

$$f(t) = \sqrt{\frac{\epsilon}{1-\lambda_p}}\,\Psi_p(t)$$

$$p = [(1 - \tfrac{1}{2}\eta)2WT] > m.$$

This $f \in \hat{\mathcal{G}}_\epsilon$ since

$$\int_{|t|>T/2} f^2(t)\,dt = \frac{\epsilon}{1-\lambda_p}\left[\int_{-\infty}^{\infty} \Psi_p^2\,dt - \int_{-T/2}^{T/2} \Psi_p^2\,dt\right] = \epsilon$$

by (2) and (3). Furthermore, since $p > m$, then f is orthogonal to $\Psi_0, \Psi_1, \cdots, \Psi_m$, so that

$$L = \int_{-T/2}^{T/2} f^2\,dt = \epsilon\,\frac{\lambda_p}{1-\lambda_p}.$$

But by (9), for large enough WT, λ_p is arbitrarily close to 1 so that L becomes arbitrarily large. Thus we have shown that for every $\epsilon' > 0$ for large enough WT we have $\epsilon' < L \leqslant K$. There then exists a $\hat{g} \in \mathcal{G}_\epsilon$ for which (39) is true and the left of (25) is established.

There remains now only the task of showing that K attains its smallest value, say K_0, when the φ's in (42) are the m functions $\Psi_i/\sqrt{\lambda_i}$, $i = 0, 1, \cdots, m-1$. We note first that the functions $\Psi_j/\sqrt{\lambda_j}$ are orthonormal in $(-T/2, T/2)$ and complete in $\mathcal{B}$, the class of strictly bandlimited functions. Thus for any $\hat{g} \in \hat{\mathcal{G}}_\epsilon$, we write

$$\hat{g} = \sum_0^\infty \hat{g}_i \frac{\Psi_i(t)}{\sqrt{\lambda_i}}, \qquad \hat{g}_i = \int_{-T/2}^{T/2} \hat{g}(t)\frac{\Psi_i(t)}{\sqrt{\lambda_i}}\,dt \tag{43}$$

and the condition (41) is

$$\sum_0^\infty \hat{g}_i^2\gamma_i = \epsilon$$

$$\gamma_i \equiv \frac{1-\lambda_i}{\lambda_i}, \qquad i = 0, 1, \cdots. \tag{44}$$

From (8) we see that

$$0 < \gamma_0 < \gamma_1 < \cdots. \tag{45}$$

Suppose now that the linear span of $\varphi_0, \varphi_1, \cdots, \varphi_m$ in (42) does not coincide with the linear span $\mathcal{S}$ of $\Psi_0/\sqrt{\lambda_0}, \Psi_1/\sqrt{\lambda_1}, \cdots, \Psi_m/\sqrt{\lambda_m}$. We can then find a function in $\mathcal{S}$, say $h(t) = \Sigma_0^m h_i\Psi_i(t)/\sqrt{\lambda_i}$ that is orthogonal on $(-T/2, T/2)$ to each of $\varphi_0, \varphi_1, \cdots, \varphi_m$. Assume $h(t)$ scaled so that

$$\sum_0^m h_i^2\gamma_i = \epsilon \tag{46}$$

so that $h \in \hat{\mathcal{G}}_\epsilon$. For this function we have

$$M \equiv \min_{a_i} \int_{-T/2}^{T/2} \left[h - \sum_0^m a_i\varphi_i(t)\right]^2 dt = \int_{-T/2}^{T/2} h^2(t)\,dt = \sum_0^m h_i^2. \tag{47}$$

From (46) and (47), however,

$$\epsilon = \sum_0^m h_i^2\gamma_i \leqslant \gamma_{m+1}\sum_0^m h_i^2 = \gamma_{m+1}M. \tag{48}$$

Therefore,

$$K \geqslant M \geqslant \frac{\epsilon}{\gamma_{m+1}}$$

from (42) and (43) when the φ's do not span $\mathcal{S}$. When the φ's are the functions $\Psi_i/\sqrt{\lambda_i}$, $i = 0, 1, \cdots, m$, we compute that

$$\min_{a_i} \int_{-T/2}^{T/2} \left[\sqrt{\frac{\epsilon}{1-\lambda_{m+1}}}\, \Psi_{m+1}(t) - \sum_0^m a_i\varphi_i(t)\right]^2 dt$$

$$= \int_{-T/2}^{T/2} \frac{\epsilon}{1-\lambda_{m+1}}\, \Psi_{m+1}^2(t)\, dt = \frac{\epsilon}{\gamma_{m+1}}.$$

Since

$$\sqrt{\frac{\epsilon}{1-\lambda_{m+1}}}\, \Psi_{m+1}(t) \in \hat{\mathcal{G}}_\epsilon$$

we find $K \geqslant \epsilon/\gamma_{m+1}$ also in this case where the φ's are the Ψ's. We now have

$$K_0 \equiv \min_{\varphi\text{'s}} K \geqslant \frac{\epsilon}{\gamma_{m+1}}. \tag{49}$$

Now it is easy to see from (43) and (44) that

$$\sup_{\hat{g}\in\hat{\mathcal{G}}_\epsilon} \min_{a_i} \int_{-T/2}^{T/2} \left[\hat{g} - \sum_0^m a_i \frac{\Psi_i(t)}{\sqrt{\lambda_i}}\right]^2 dt = \frac{\epsilon}{\gamma_{m+1}}$$

and is attained when

$$\hat{g} = \sqrt{\frac{\epsilon}{1-\lambda_{m+1}}}\, \Psi_{m+1}.$$

Since K_0 is the minimum of (42) over all φ's, we now see that $K_0 \leqslant \epsilon/\gamma_{m+1}$. Combined with (49), this yields $K_0 = \epsilon/\gamma_{m+1}$, a value achieved when $\varphi_i = \Psi_i/\sqrt{\lambda_i}$, $i = 0, 1, \cdots, m$.

The proof of (25) is thus completed.

Acknowledgment

It is a pleasure to give thanks here to my friend and colleague H. Landau for many fruitful discussions on the ideas contained in this paper and for his contributions to them.

References

[1] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10-21, June 1949.

[2] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, pp. 1701-1706, Oct. 1967.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, ch. 8.

[4] A. D. Wyner, "Capacity of the band-limited Gaussian channel," *Bell Syst. Tech. J.*, vol. 45, pp. 359-395, Mar. 1965.

[5] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—I," *Bell Syst. Tech. J.*, vol. 40, pp. 43-64, 1961.
H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—II and III," *ibid.*, vol. 40, pp. 65-84, 1961; vol. 41, pp. 1295-1336, 1962.

[6] D. Slepian, "Some asymptotic expansions for prolate spheroidal wave functions," *J. Math. Phys.*, vol. 44, pp. 99-143, June 1965.