XIII. SIGNAL PROCESSING*

Academic and Research Staff

Prof. A. G. Bose
Prof. J. D. Bruce
Prof. A. V. Oppenheim
Prof. C. L. Searle
Prof. H. J. Zimmermann
Dr. M. V. Cerillo

Graduate Students

J. B. Bourne
M. F. Davis
D. A. Feldman
J. M. Kates
R. M. Mersereau
J. R. Samson, Jr.
R. M. Stern, Jr.

A. RECENT ADVANCES IN THE THEORY OF RECONSTRUCTING MULTIDIMENSIONAL SIGNALS FROM PROJECTIONS

1. Introduction

The problem of reconstructing multidimensional signals from their projections is of interest because x-ray photographs and electron micrographs can be considered to be projections of three-dimensional objects. Thus mathematical techniques for performing such reconstructions will permit us to reconstruct visually opaque objects from their x-rays at different orientations and to determine the structure of macromolecules from electron micrographs. In a previous report[1] some techniques were discussed whereby we could perform such a reconstruction; in the present report, some other more powerful algorithms will be developed. One of these algorithms, in fact, permits the reconstruction of a broad class of multidimensional signals of any dimensionality from a single one-dimensional projection.

The idea of reconstructing functions from their projections can be applied to functions of any dimensionality; however, the most interesting problems, since they have useful applications, are the two-dimensional and the three-dimensional problems. By extension, there is a one-dimensional problem, but it is a trivial case because the projection of a one-dimensional function is the function itself. Most of the derivations in this report will be given in terms of the two-dimensional problem because it is notationally and conceptually simpler than the three-dimensional problem, but we shall also explore some of the issues that are unique to the three-dimensional case.

In both of the algorithms that are developed here it is assumed that the function which is being reconstructed is bandlimited, and if a further assumption is made they will yield exact reconstructions. If these assumptions are not appropriate for the problem at hand, there are other techniques that will yield approximate reconstructions.[1-6]

---

*This work was supported by the Joint Services Electronics Programs (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAB07-71-C-0300, by the U.S. Coast Guard (Contract DOT-CG-13446-A), and by M.I.T. Lincoln Laboratory Purchase Order CC-570.

Inasmuch as we shall deal with bandlimited functions exclusively, it is appropriate to begin with a discussion of the properties of the projections of multidimensional bandlimited functions.

2. Projections of Bandlimited Functions

The assumption of bandlimitedness is not especially harsh, for although most functions that we shall reconstruct are spacelimited, and hence strictly speaking not bandlimited, they are nearly so. Furthermore, if any algorithm is to be implemented on a computer, it is necessary to reconstruct a sampled multidimensional function from sampled projections. Thus bandlimitedness is implicitly assumed to a greater or lesser degree by all digital reconstruction algorithms. In these algorithms we shall explicitly assume bandlimitedness and then utilize this assumption in the design of our algorithms, with the hope that they will yield high-quality reconstructions for nearly bandlimited functions. The last premise must be verified experimentally.

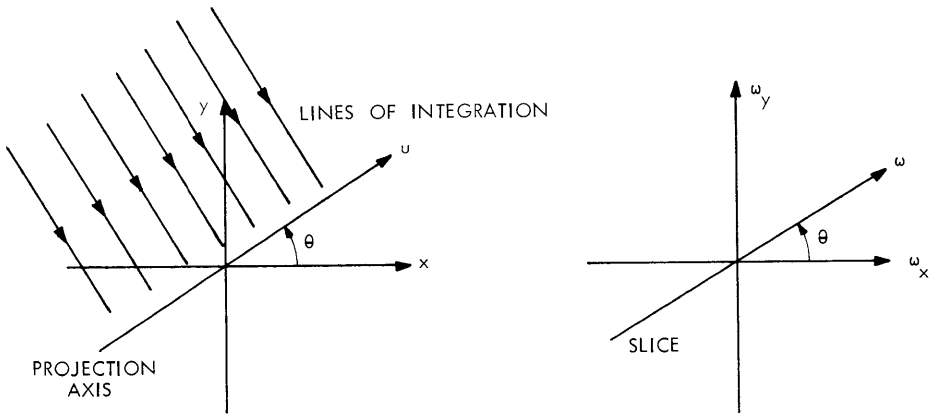

Fig. XIII-1. Relationship between a projection and a slice.

The projections of a two-dimensional function (picture) can be considered as a collection of line integrals taken perpendicular to an axis, which we call the projection axis. Thus the projection perpendicular to the x axis, $p_0(x)$, can be defined as

$$p_0(x) = \int_{-\infty}^{\infty} f(x, y)\, dy.$$

At a general angle θ, a projection can be similarly defined by

$$p_\theta(u) = \int_{-\infty}^{\infty} f(u \cdot \cos\theta + v \cdot \sin\theta, -u \cdot \sin\theta + v \cdot \cos\theta)\, dv, \tag{1}$$

and it satisfies the Fourier transform relation

$$p_\theta(u) \longleftrightarrow F(\omega\cos\theta, \omega\sin\theta), \tag{2}$$

where $F(\omega_x, \omega_y)$ represents the two-dimensional Fourier transform of $f(x, y)$. The right-hand side of Eq. 2 will be referred to as the slice of the two-dimensional Fourier transform at an angle θ. Thus the one-dimensional Fourier transform of the projection of a picture at an angle θ to the x axis is a slice of the two-dimensional Fourier transform of that picture at an angle θ with the ω_x axis. This relationship is illustrated in Fig. XIII-1.

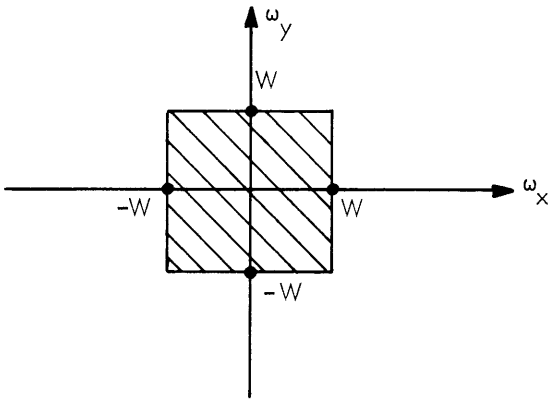

Fig. XIII-2. Region of Fourier plane over which a bandlimited picture is nonzero.

If we now assume that the picture is bandlimited, that is, that its frequency response is nonzero only in that region of the Fourier plane illustrated in Fig. XIII-2, then we can use the sampling theorem to express the picture in terms of its samples on a regular Cartesian raster as in

$$f(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right) \frac{\sin W\left(x - \frac{m\pi}{W}\right) \sin W\left(y - \frac{n\pi}{W}\right)}{W^2\left(x - \frac{m\pi}{W}\right)\left(y - \frac{n\pi}{W}\right)}. \tag{3}$$

Since all of the projections transform to slices, they too must be bandlimited (in one dimension) and each projection can thus be expanded in terms of its samples as in

$$p_\theta(u) = \sum_{n=-\infty}^{\infty} p_\theta\left(\frac{n\pi}{W_\theta}\right) \frac{\sin W_\theta\left(u - \frac{n\pi}{W_\theta}\right)}{\left(u - \frac{n\pi}{W_\theta}\right)}. \tag{4}$$

The bandwidth of each projection W_θ can be expressed as

$$W_\theta = \frac{W}{\max\{|\cos\theta|, |\sin\theta|\}}. \tag{5}$$

From Eqs. 4 and 5 we can ascertain the Nyquist sampling rate for each projection, which is observed to be a function of θ, the projection angle. Since we must work with sampled projections, this will prove to be an important quantity.

We can get an alternative expression for $p_\theta(u)$, not in terms of the samples of the projections, but in terms of the samples of the picture itself. If we take the Fourier transform of Eq. 3, we get

$$F(\omega_x, \omega_y) = \frac{\pi^2}{W^2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right) \exp\left\{-j \frac{\pi}{W}(m\omega_x + n\omega_y)\right\} b_{WW}(\omega_x, \omega_y), \tag{6}$$

where

$$b_{WW}(\omega_x, \omega_y) = \begin{cases} 1, & \text{if } |\omega_x| \leq W \text{ and } |\omega_y| \leq W \\ 0, & \text{otherwise} \end{cases}$$

From Eq. 6 we can evaluate $F(\omega\cos\theta, \omega\sin\theta)$ which is the expression for a slice (from Eq. 2).

$$F(\omega\cos\theta, \omega\sin\theta) = \frac{\pi^2}{W^2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right) \exp\left\{-j \frac{\pi\omega}{W}(m\cos\theta + n\sin\theta)\right\}$$

$$\times b_{WW}(\omega\cos\theta, \omega\sin\theta). \tag{7}$$

Performing an inverse Fourier transform on Eq. 7 gives an expression for the projection at angle θ.

$$p_\theta(u) = \frac{\pi}{W^2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right) \frac{\sin W_\theta\left(u - \frac{m\pi}{W}\cos\theta - \frac{n\pi}{W}\sin\theta\right)}{u - \frac{m\pi}{W}\cos\theta - \frac{n\pi}{W}\sin\theta}. \tag{8}$$

In the two reconstruction techniques that follow, we must impose one further restriction on the picture in addition to bandlimitedness. We must assume that the digitized picture $f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right)$ be nonzero for integral values of m and n only when m and n are in the range $0 \leq m, n \leq N-1$, for some finite integer N. We call this assumption quasi-spacelimitedness, although note that we do not assume that f(x, y) is spacelimited (which would contradict the assumption that it is bandlimited), but only

that its samples are spacelimited. This assumption has the effect of making the double summations of Eqs. 3, 6, 7, and 8 finite. This, like the assumption of bandlimitedness, is implicit in most reconstruction techniques, since only a finite number of samples of the Fourier transform of the picture are generally computed, and only a finite number of picture samples are reconstructed.

3. An Algorithm for Reconstructing a Function from N+1 Projections

Equation 4 gives us the smallest sampling rate that can be employed for sampling a projection in order that information not be lost by sampling. Each projection, of course, can be sampled with a higher rate. The traditional approach for getting samples of the slices of a picture is to find a sampling rate that is large enough so that all of the projections can be sampled at the same rate. The resulting sequences can then be aliased to give M point sequences, and these M point sequences can then be Fourier-transformed by using a discrete Fourier transform (DFT) algorithm to yield M sample values along each slice. The M-point aliased sequence x(n) corresponding to the infinitely long sequence x(n) is defined by

$$x(n) = \sum_{m=-\infty}^{\infty} x(Mm+n).$$

If this procedure is followed, the Fourier transform of the picture will be known at points lying on a polar lattice. The points of such a lattice can be thought of as the intersections of the set of slices with a family of evenly spaced concentric circles, including one of zero radius at the origin. Once the transform of the picture is known at these points, the next step is to approximate the transform of the picture over the whole plane and then perform an inverse Fourier transform. There are no nice polar "sampling theorems" that will allow us to obtain directly the set $f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right)$.

As a different approach, let us therefore sample each projection at its own Nyquist rate, or at a rate proportional to its Nyquist rate, then alias the resulting sequences to N points (N is the width of the digitized picture) and use a DFT algorithm to get samples of the Fourier transform of the picture. If this procedure is followed, the Fourier samples which result lie at the intersection of the slices with a family of concentric squares, as illustrated in Fig. XIII-3.

In the special case of a bandlimited quasi-spacelimited (BLQSL) function, a concentric squares lattice has definite advantages over a polar one. Along any horizontal or vertical lines in the Fourier plane, the Fourier transform of a BLQSL function is a one-dimensional complex polynomial of degree N-1, and as a result any line in the Fourier plane is completely specified by N-1 samples that lie along that line. Furthermore, a BLQSL function is completely specified by its DFT, that is, by the

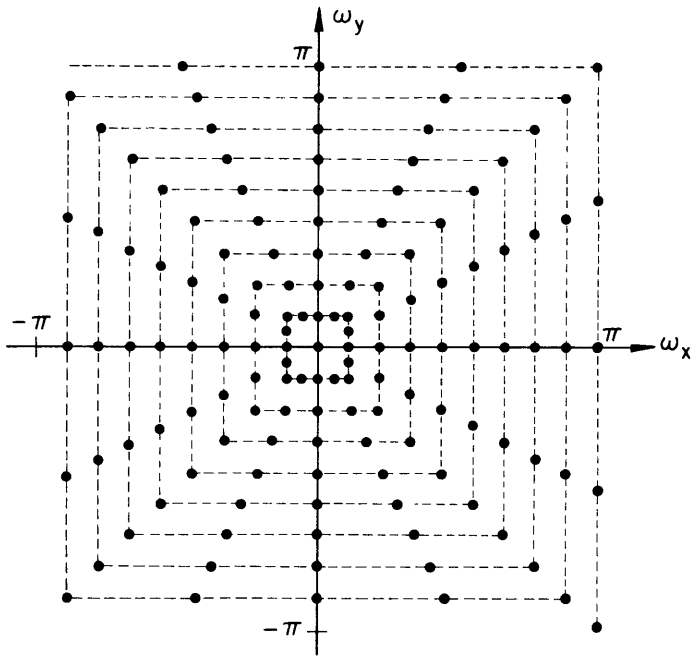

Fig. XIII-3. A set of samples of the Fourier transform of a bandlimited function obtained by sampling each projection at a rate proportional to its own Nyquist rate.

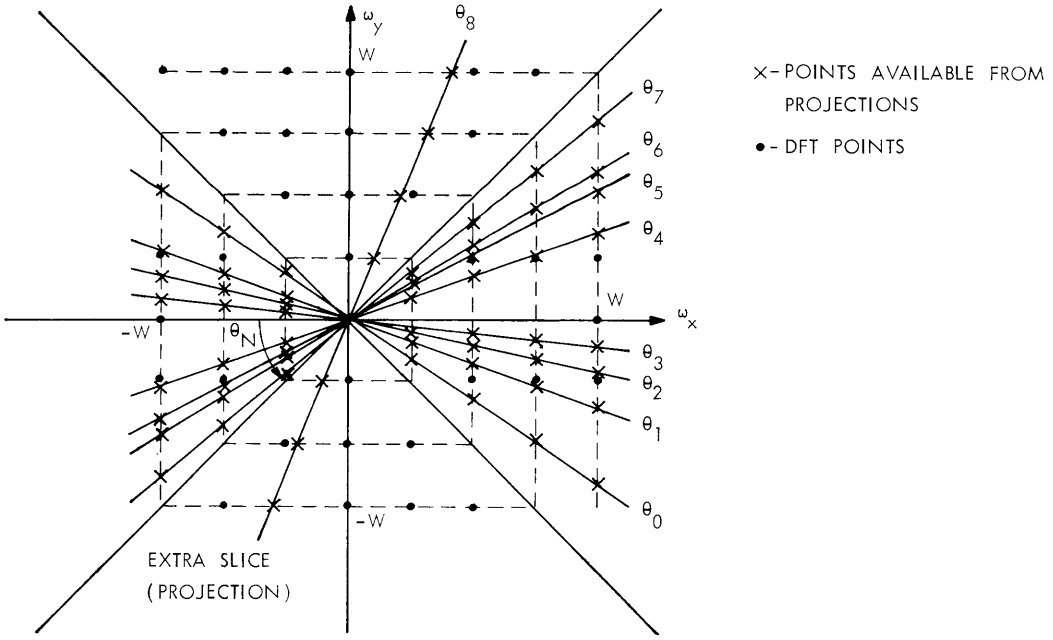

Fig. XIII-4. Set of Fourier plane samples by which an 8 × 8 picture can be reconstructed exactly, under the assumption that the picture is bandlimited and quasi-spacelimited.

samples of its Fourier transform at $F\left(\frac{W}{N} i, \frac{W}{N} j\right)$, $-\frac{N}{2}+1 \leq i, j \leq \frac{N}{2}$. These points all lie on the sides of the concentric squares or their extensions. These properties enable us to reconstruct a BLQSL function exactly from a set of N concentric-squares projections.

Suppose, for example, we have a two-dimensional BLQSL function of dimension N. Let us assume also that we have the capability of obtaining the projections of the picture at any angle we desire. We can thus take N projections at N distinct angles in the range $-45° \leq \theta \leq 45°$, and we can also take one projection at an angle outside this range. The known points in Fourier space will then correspond to those illustrated in Fig. XIII-4 for the special case N = 8. Along each vertical square side we thus have 8 samples and along these sides the Fourier transform is a 7^{th}-order polynomial in the variable $e^{-j\omega_y}$. Then, using Lagrange polynomials (or some other technique), we can evaluate the Fourier transform at all of the DFT points on each of the vertical lines, except for the one at $\omega_x = 0$. Now consider the horizontal sides. Along each of these lines we also have a polynomial of degree 7 and we also have 8 samples, seven computed from the column calculations, and the eighth provided by the remaining projection. Since this projection was taken outside the range $-45° \leq \theta \leq 45°$, it must intersect all of the horizontal square sides (and must also not pass through any of the DFT points whose value is already known). Thus we can apply Lagrange polynomials to the horizontal lines to fill in the remaining DFT values. Consequently, we know all of the DFT values exactly, and a BLQSL picture can be

(a)

(b)

(c)

Fig. XIII-5.

Comparison of reconstructions from a concentric squares grid and from a concentric circles grid. (a) Original picture. (b) Concentric squares reconstruction. (c) Polar (concentric circles) reconstruction.

reconstructed from its DFT so that we know the set of picture samples exactly. If the last projection had been taken perpendicular to the y axis, then the second round of interpolation would not have been necessary, for the remaining DFT samples would be available directly from the DFT of the last sampled projection.

In Fig. XIII-5 we show a concentric-squares reconstruction and compare it with the corresponding concentric-circles (polar) reconstruction. Instead of using Lagrange interpolation to exactly perform the reconstruction, a simpler approximate strategy was employed. Linear interpolation was used to approximate the DFT samples from the samples obtained from the projections in both reconstructions. Each reconstruction is a 64 × 64 picture which was obtained from 64 projections. The projections were computed from the original picture which is included for comparison. Note that the concentric-squares reconstruction is truer to the original.

4. Reconstructing a BLQSL Picture from a Single Projection

Let us now restrict ourselves to the slice at an angle $\theta = \tan^{-1} 1/N$. From Eq. 7 this slice can be written as

$$F\left(\frac{N\omega}{\sqrt{N^2+1}}, \frac{\omega}{\sqrt{N^2+1}}\right) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right) \exp\left\{-j \frac{\pi\omega}{\sqrt{N^2+1}} W(Nm+n)\right\},$$

$$\text{if} \quad |\omega| \leq \frac{W}{N}\sqrt{N^2+1}$$

$$= 0 \qquad \text{otherwise} \tag{9}$$

If we define $g(Nm+n) = f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right)$, then Eq. 9 becomes

$$F\left(\frac{N\omega}{\sqrt{N^2+1}}, \frac{\omega}{\sqrt{N^2+1}}\right) = \sum_{l=0}^{N^2-1} g(l) \exp\left(-j \frac{\pi\omega l}{W\sqrt{N^2+1}}\right), \quad \text{if } |\omega| \leq \frac{W}{N}\sqrt{N^2+1}$$

$$= 0, \qquad \text{otherwise} \tag{10}$$

Thus, over the region of interest, the slice at $\theta = \tan^{-1} 1/N$ is a one-dimensional polynomial of degree N^2-1 in the variable $\exp\left(-j \frac{\pi\omega}{W\sqrt{N^2+1}}\right)$, and the coefficients of that polynomial are simply the picture samples arranged as they would be if the picture were scanned column by column. Since $F\left(\frac{N\omega}{\sqrt{N^2+1}}, \frac{\omega}{\sqrt{N^2+1}}\right)$ can be computed from

N^2 samples taken along the slice, and knowledge of a polynomial implied knowledge of its coefficients, specification of N^2 samples along the slice at $\theta = \tan^{-1} 1/N$ implies knowledge of the whole set of picture samples. (Similar statements can be made about other slices in the Fourier plane.)

Let us now, for convenience, define $G(\omega) = F\left(\frac{N\omega}{\sqrt{N^2+1}}, \frac{\omega}{\sqrt{N^2+1}}\right)$, and let us set $\Delta\omega = \frac{2W\sqrt{N^2+1}}{N^3}$. Then if we compute $G(k\Delta\omega)$ for $k = -\frac{N^2}{2} + 1, \ldots, 0, 1, \ldots, \frac{N^2}{2}$, we shall have N^2 equally spaced samples of $G(\omega)$ which extend over the entire band. There is a strong reason for choosing this particular set of frequency samples on this slice. If the projection $p_{\tan^{-1}\frac{1}{N}}(u)$ is sampled at its Nyquist rate, if the infinite sequence that results is then aliased to give a sequence of length N^2, and if this sequence is then Fourier-transformed by means of the DFT, the resulting N^2 point sequence is $G(k\Delta\omega)$. Substituting in Eq. 10, we have

$$G(k\Delta\omega) = \sum_{l=0}^{N^2-1} g(l) \exp\left(-j \frac{2\pi kl}{N^3}\right) \qquad k = -\frac{N^2}{2} + 1, \ldots, \frac{N^2}{2}. \tag{11}$$

Examining Eq. 11, we see that $G(k\Delta\omega)$ corresponds to the first N^2 points of the N^3 point DFT of the sequence formed by taking the N^2 picture samples column by column and appending $N^3 - N^2$ zeros.

The sequence $G(k\Delta\omega)$ could be obtained from the sequence $g(l)$ by means of a chirp z-transform algorithm CZT.[7] To obtain $g(l)$ (the picture samples) from $G(k\Delta\omega)$, we thus need an inverse CZT, which will be developed.

These results have an interesting interpretation in terms of another problem. The impulse response of a two-dimensional nonrecursive digital filter behaves exactly like the set of rectangular samples of a BLQSL picture and thus the impulse response, or the two-dimensional frequency response, of such a filter is completely specified by its frequency response along the line $\theta = \tan^{-1} 1/N$. As well as providing an interesting property for such filters, this result suggests a mapping between one-dimensional nonrecursive and two-dimensional nonrecursive filter designs that may be useful in filter design. These implications are worthy of further study.

5. Reconstructing a Three-Dimensional BLQSL Function from a Single Projection

Probably the simplest way to reconstruct a three-dimensional sequence from its projections is to consider that three-dimensional sequence as a stack of two-dimensional

sequences. If we think of these two-dimensional sequences as lying parallel to the x-y plane, and then take a projection parallel to the x-y plane at an angle $\theta = \tan^{-1} 1/N$ with the x axis, then the resulting two-dimensional projection of the three-dimensional object will be a stack of one-dimensional projection functions, each of which is the projection of one member of the original stack of two-dimensional functions and each of which is taken at its critical angle. This is a straightforward extension of the two-dimensional problem and hardly requires elaboration. It would be a computationally efficient scheme, however, if a complete reconstruction were not desired, but only a limited number of cross sections.

From a theoretical point of view, a more interesting approach to the three-dimensional problem is to parallel the reasoning of the two-dimensional analysis. In that case we found a line in the Fourier plane; if we knew the Fourier transform of the picture along this line, then we knew the whole set of picture samples. Such a line also exists in the three-dimensional case. This is that line which is traced out by the vector $\vec{\omega}_c$, where

$$\vec{\omega}_c = \left(\frac{N^2}{\sqrt{N^4 + N^2 + 1}}, \frac{N}{\sqrt{N^4 + N^2 + 1}}, \frac{1}{\sqrt{N^4 + N^2 + 1}} \right) \omega.$$

Along this line the frequency response is a polynomial of degree N^3-1, and the coefficients of this polynomial are the function samples $f\left(\frac{m\pi}{W}, \frac{n\pi}{W}, \frac{p\pi}{W}\right)$, $0 \leq m, n, p \leq N-1$, where W is the bandwidth, defined as in the two-dimensional case. If we sample this line at N^3 evenly spaced points over the band, then

$$G(k\Delta\omega) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \sum_{p=0}^{N-1} f\left(\frac{m\pi}{W}, \frac{n\pi}{W}, \frac{p\pi}{W}\right) \exp\left\{-j \frac{2\pi k}{N^5} (N^2 m + Nn + p)\right\},$$

$$k = -\frac{N^3}{2} + 1, \ldots, 0, 1, \ldots, \frac{N^3}{2}. \qquad (12)$$

Thus we have the first N^3 points of an N^5 point sequence. Equation 12 can be solved by using the inverse chirp z-transform.

The projection of a three-dimensional function is two-dimensional, whereas the critical line along which we desire the frequency response is one-dimensional. This frequency response can be evaluated directly from the two-dimensional projection samples (the projection is a bandlimited function) or equivalently a one-dimensional projection of the two-dimensional projection can be computed digitally and then transformed. If the angle of this projection is chosen properly, this slice of a slice will correspond to the desired line. It must be remembered however, when working with the

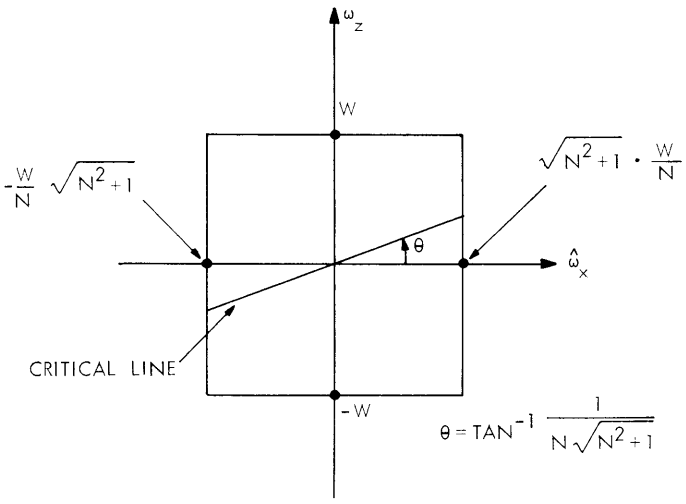

Fig. XIII-6. Two-dimensional slice of the Fourier transform of a three-dimensional function taken perpendicular to the plane $\omega_x = N\omega_y$. The bandwidths of the slice are shown, as well as the location of the critical line, whose critical frequency response determines the whole three-dimensional frequency response.

two-dimensional projection that although this is a bandlimited function, the bandwidth in the two orthogonal frequency variables is a function of the direction of that projection. In Fig. XIII-6 we show the relevant parameters for computing the frequency response along the critical line when the original projection was projected onto the plane $\omega_x = N\omega_y$.

6. Inverse Chirp z-Transform

The chirp z-transform (CZT) algorithm[7] is an efficient algorithm for evaluating the sum

$$X_k = \sum_{n=0}^{L-1} x(n)(AW^k)^n \qquad k = 0, 1, \ldots, K-1, \tag{13}$$

where

$$A = A_o \exp(j2\pi\theta_o)$$

$$W = W_o \exp(j2\pi\phi_o).$$

The CZT calculates the z-transform of the finite duration sequence x(n) at a set of points that are regularly spaced on a spiral in the z plane as illustrated in Fig. XIII-7.[7] Equation 11 can be seen to be of the same form as Eq. 13 if in place of the sequence x(n)

we substitute the sequence $g(l) = g(Nm+n) = f\left(\frac{m\pi}{W}, \frac{n\pi}{W}\right)$ and if we set $A = \exp\left(-j\frac{N^2}{2} + l\right)$ and if $W = \exp(j2\pi/N^3)$. The sequence X_k and A and W are known in this particular case, and we desire a means of calculating g(l). What we need, therefore, is a means of inverting Eq. 13 – an inverse CZT.

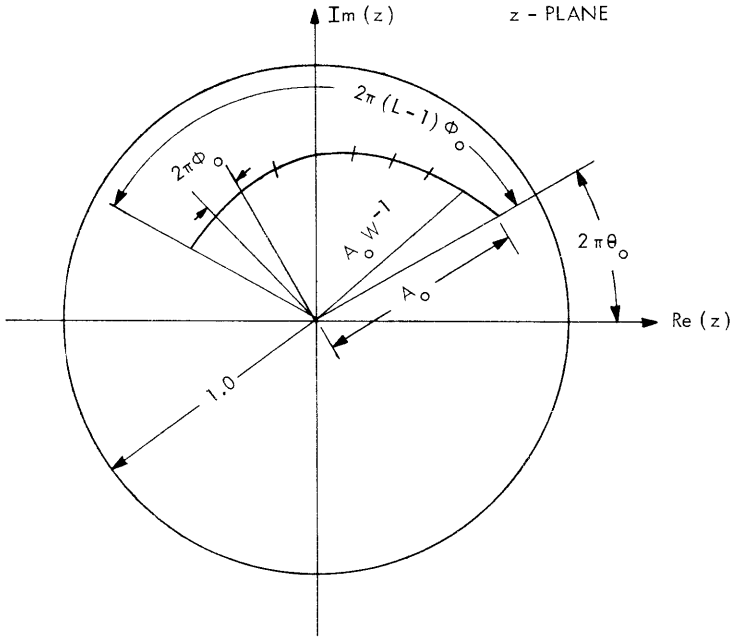

Fig. XIII-7. Illustration of the independent parameters of the CZT algorithm and the inverse CZT algorithm. (Modified from Rabiner et al.[8])

Since the sequence X_k corresponds to samples of a polynomial of degree L-1, we know that Eq. 13 can be inverted if there are more than L independent values of X_k, or if $K \geq L$. This follows from the fact that the matrix of coefficients $[(AW^k)^n]$ is a Vandermonde matrix. One possible technique to use is to invert (13) directly. For values of K of the order of several thousand, however, this is computationally not feasible.

Another approach which proves to be far more attractive computationally, although at first appearance it would not be so, is to use the Lagrange polynomial interpolation formula to reconstruct the complete polynomial over the whole z plane from the set of K samples and then perform an inverse z-transform integral of this polynomial to get the sequence x(n).

If X(z) is a polynomial of degree L-1 in z^{-1} and if X(z) is specified at the points $z_o^{-1}, z_1^{-1}, \ldots, z_{L-1}^{-1}$, then

$$X(z) = \sum_{m=0}^{L-1} X(z_m)\, \ell_m(z^{-1}), \tag{14}$$

where

$$z_m^{-1} = [AW^m]^{-1} \qquad m = 0, 1, \ldots, L-1$$

and $\ell_m(z^{-1})$ is a Lagrange interpolating polynomial

$$\ell_m(z^{-1}) = \frac{\left(z^{-1}-z_o^{-1}\right)\left(z^{-1}-z_1^{-1}\right)\cdots\left(z^{-1}-z_{m-1}^{-1}\right)\left(z^{-1}-z_{m+1}^{-1}\right)\cdots\left(z^{-1}-z_{L-1}^{-1}\right)}{\left(z_m^{-1}-z_o^{-1}\right)\left(z_m^{-1}-z_1^{-1}\right)\cdots\left(z_m^{-1}-z_{m-1}^{-1}\right)\left(z_m^{-1}-z_{m+1}^{-1}\right)\cdots\left(z_m^{-1}-z_{m+1}^{-1}\right)}.$$

Since the denominator of $\ell_m(z^{-1})$ is a constant, let us write it as $1/C_m$. Thus

$$X(z) = \left[\sum_{m=0}^{L-1} \frac{X(z_m)\, C_m}{\left(z^{-1}-z_m^{-1}\right)}\right]\left[\prod_{\ell=0}^{L-1}\left(z^{-1}-z_\ell^{-1}\right)\right]. \tag{15}$$

Equation 15 represents the z-transform of the sequence x(n) which we desire. Thus we see that the sequence x(n) can be regarded as the impulse response of a bank of resonators and a comb filter in cascade, as in Fig. XIII-8.

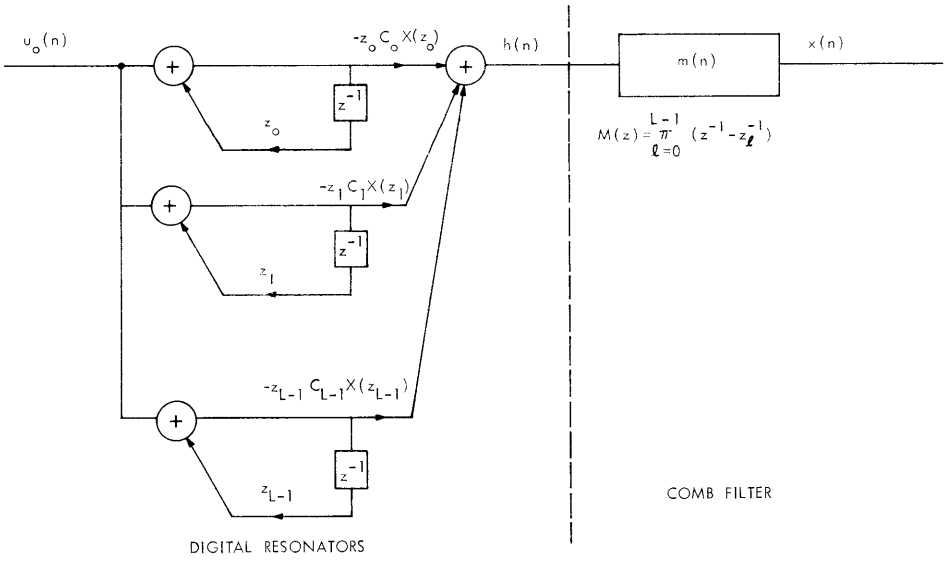

Fig. XIII-8. Digital network implementation of the inverse CZT.

Let us define h(n) to be the impulse response of the bank of resonators. From Eq. 15 we can write

$$h(n) = \sum_{i=0}^{L-1} -C_i X(z_i)\, z_i^{n+1} \qquad n = 0, 1, \ldots, L-1.$$

If we recognize that $z_i = AW^{-i}$, then

$$h(n) = \sum_{i=0}^{L-1} -C_i X(z_i)\, A^{n+1} W^{-i(n+1)} = A^{n+1} \sum_{i=0}^{L-1} -C_i X(z_i)\, W^{-i}\, W^{-in}$$

$$n = 0, 1, \ldots, L-1. \tag{16}$$

If we write

$$CZT(x(n), A, W, N) = \sum_{n=0}^{N-1} x(n)\, A^{-n} W^{nk}, \tag{17}$$

then we can write (16) in the form

$$h(n) = -A^{n+1} CZT(C_n X(z_n), W, W^{-1}, L) \tag{18}$$

and thus h(n) can be evaluated efficiently by using the CZT algorithm itself.

The output sequence x(n) is then

$$x(n) = h(n) \circledast m(n),$$

where $\circledast$ denotes convolution. Inasmuch as we only care about the first L values of the sequence x(n) and m(n) is a causal sequence of length L+1, only the first L values of the sequence h(n) are necessary. This fact allows us to evaluate h(n) using a CZT, and will further allow us to perform the convolution of (18) using high-speed convolution techniques.

Except for calculating the arrays C_k and m(n), the computation of the inverse CZT can all be done efficiently. In fact, the time required to calculate an inverse CZT is approximately twice that required to calculate a CZT, and thus is roughly proportional to $2L \log_2 2L$ if L is a power of two. To the best of my knowledge, there are no particularly convenient methods for calculating C_k and m(n). These quantities do not depend upon the sequence X_k, but only on the location of the samples of the z-transform in the z plane, and therefore they will be the same for all reconstructions

of a given size, which will allow these arrays to be precomputed and stored. In this sense, the calculation of these quantities can be overlooked when talking about computation times. To reconstruct a 32×32 array from a single projection requires approximately 10^5 operations (complex multiplies and adds) if the calculation of these initial arrays is overlooked, and it requires approximately 5000 complex storage locations. To solve Eq. 13 by direct inversion would require approximately 10^9 operations and roughly 10^6 complex storage locations.

A one-projection reconstruction algorithm is now being implemented.

R. M. Mersereau

References

1. R. M. Mersereau, "Reconstruction of Two-Dimensional Signals from Projections," Quarterly Progress Report No. 102, Research Laboratory of Electronics, M.I.T., July 15, 1971, pp. 190-196.
2. R. A. Crowther, D. J. DeRosier, and A. Klug, "The Reconstruction of a Three-Dimensional Structure and Its Applications to Electron Microscopy," Proc. Roy. Soc. (London) A 317, 319-340 (1970).
3. D. J. DeRosier and A. Klug, "Reconstruction of Three-Dimensional Structures from Electron Micrographs," Nature 217, 130-134 (1968).
4. R. Gordon, R. Bender, and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for Three-Dimensional Electron Microscopy and X-ray Photography," J. Theoret. Biol. 29, 471-481 (1970).
5. R. N. Bracewell and A. C. Riddle, "Inversion of Fan Beam Scans in Radio Astronomy," Astrophys. J., Vol. 150, No. 2, Part 1, pp. 427-434, November 1967.
6. O. Tretiak, D. Ozonoff, J. Klopping, and M. Eden, "Calculation of Internal Structure from Multiple Radiographs," Proc. Two-Dimensional Digital Signal Processing Conference, University of Missouri, Columbia, Missouri, October 7, 1971, pp. 6.2.1-6.2.3.
7. L. R. Rabiner, R. W. Schafer, and C. M. Rader, "The Chirp Z-Transform Algorithm and Its Application," Bell System Tech. J. 48, 1249-1292 (1969).
8. Ibid., Fig. 2, p. 1253. Original figure, copyright 1969, American Telephone and Telegraph Company. Printed by permission.

B. TRANSIENT RESPONSE OF A VARACTOR-CONTROLLED OSCILLATOR

A study has been made of the Q-related effects on the transient response of a voltage-controlled negative-resistance oscillator.[1] The equation governing the nonlinear oscillations of a second-order time-invariant circuit has the form

$$\frac{d^2x}{dt^2} + \omega_o^2 x = \delta f\left(x, \frac{dx}{dt}\right)$$

in which $f\left(x, \frac{dx}{dt}\right)$ is a general nonlinear function of the variable x and its derivative,

and δ is proportional to the reciprocal of the effective Q of the circuit. A perturbational analysis of this equation yields a solution in which the instantaneous frequency of oscillation ω_i is given by an expression of the form $\omega_i = \omega_o + F(a)$, where a corresponds to the magnitude of the amplitude envelope of the oscillations. This expression indicates a possible variation in frequency because of a variation in the amplitude envelope during a transient period of the oscillations. It can be argued that the frequency does not reach a steady-state value until the amplitude reaches a steady-state value.

1. Cause and Mode of Transient Operation

An analysis of an idealized step-change in one of the frequency-determining elements indicates that such a change could cause a disturbance from the equilibrium steady-state oscillation. Such a disturbance implies that the state of the oscillation, specified in the phase plane by x and dx/dt immediately after the change occurs, does not correspond in general to a state that is located on the steady-state limit cycle. Standard phase-plane analysis shows that if the state of the oscillator is described by a set of coordinates (the operating point) not located on the limit cycle, the oscillation will spiral to the stable limit cycle. This spiraling to the limit cycle corresponds to a variation in the amplitude envelope. The nonlinear mechanism that determines the steady-state limit cycle operation also controls this transient response back to the steady state.

2. Specific Case – Van der Pol Negative-Resistance Oscillator

A specific case of an oscillator with a Van der Pol type of nonlinearity was studied,

$$f\left(x, \frac{dx}{dt}\right) = (1-x^2)\frac{dx}{dt}.$$

The analytical solution to a second-order approximation indicated that the parameter $\delta\left(\propto \frac{1}{Q}\right)$ has a strong influence on the transient response, a first-order effect on the rate of amplitude variation, and a second-order effect on the frequency. These relations take the form $da/dt = \delta B(a)$, and $\omega_i = \omega_o + \delta^2 K(a)$, where the B(a) and K(a) are specific functions of a, corresponding to the solution of the Van der Pol equation.

It was noted that the lower the Q of the circuit, the faster the response back to the steady state following some disturbance, but at the expense of frequency stability and noise reduction properties of the oscillator in the steady state.

3. Lower Limit of Transient Response Time

Specific consideration of a varactor-controlled oscillator established a lower limit for the transient response time. First, consider the case of a circuit with an infinite Q, the harmonic oscillator, governed by the second-order differential equation

$$\frac{d^2x}{dt^2} + \omega^2 x = \frac{d^2x}{dt^2} + \frac{1}{LC}\, x = 0.$$

In this case, linear differential equation theory indicates that there is no transient response associated with the basic oscillator equation. If we wish to consider $\omega = (1/LC)^{1/2}$ as the instantaneous frequency of the oscillation, the only transient response of the oscillator would be the frequency transient associated with the capacitance transient of the varactor. Any discontinuity in the boundary conditions would cause no transient response because the natural frequencies lie on the imaginary axis and thus the value of amplitude corresponding to the initial conditions on the harmonic oscillator becomes the amplitude of the pure sinusoidal oscillation. Therefore, in this case, the lower limit of the response time is that of the varactor transient.

We now consider a low-Q case, and make the assumption that the circuit responds very rapidly to any disturbance from equilibrium. Again, we take $\omega = (1/LC)^{1/2}$, in the second-order nonlinear differential equation

$$\frac{d^2x}{dt^2} - \delta f\left(x, \frac{dx}{dt}\right) + \frac{1}{LC}\, x = 0,$$

as an approximate value of the instantaneous frequency. We can now approximate the varactor capacitance transient by a series of small step-changes to which the oscillator immediately responds. In this case, the varactor transient is again the lower limit of the oscillator's response time. This technique of approximating the variation in one of the frequency-determining elements as a series of small discrete steps can be used to examine the transient corresponding to a slowly varying input.

For intermediate values of Q, particularly at the high end of these values, the speed and nature of the transient effects are dominated by the nonlinear mechanism of the oscillator. Therefore, the lowest possible limit of the response time is that of the varactor itself.

4. Improvement of VCO Transient Operation

The seriousness of this transient problem depends on the nature of the application of such a voltage-controlled oscillator. Two suggestions can be made to speed up the period of the transient response:

(a) Since the total time of the transient depends on the size of the disturbance from the new limit cycle operation, make an effort to minimize the magnitude of this disturbance. The analytical results indicate that this minimization can be accomplished in general, and that in some cases it is possible to reduce the disturbance to zero, by proper choice of the precise time within a cycle to effect a change in frequency.

(b) In order to benefit from both high- and low-Q operation, the effective Q of the

circuit could be lowered during the period of the transient and switched back to the higher Q for steady-state operation.

5. Past and Future Experimental Work

A few qualitative experiments presented definite evidence of the Q-related effect on the amplitude envelope of the transient response and the possibility of a discontinuity of operation caused by a step-change in the capacitance. Areas of future work include a precise determination of the associated frequency transient and the practical application of the techniques suggested by this initial work.

J. R. Samson, Jr.

References

1. J. R. Samson, Jr., S. M. Thesis, Department of Electrical Engineering, M.I.T., January 1972.

C. MODEL FOR LOW-FREQUENCY ATMOSPHERIC NOISE

In a previous report[1] we gave the background and principal experimental results of a research program to develop a model for low-frequency atmospheric noise. The main results in that report are summarized as follows.

The conceptual form of a model for atmospheric noise, observed at the output of a bandlimiting filter, is

$$y(t) = A(t)\, n(t),$$

where n(t) is a bandpass Gaussian process, and A(t) is a lowpass random process. The variations of A(t) model the fluctuating power level of the atmospheric noise, which is caused by electromagnetic radiation from lightning discharges. Since this radiation is broadband, A(t) is common to nearby, but disjoint, frequency channels.

The measured probability density of y(t) exhibits wide variations in form, caused principally by local weather patterns. We classified the noise characteristics by weather pattern as "quiet," "tropical," and "frontal."

The measured joint probability density of the noise, y(t), in one frequency channel at 65 kHz and the noise envelope in a channel at 83 kHz exhibits a strong correlation, as suggested by the model. In particular, the plot of the variance of the noise, y(t), conditioned on the value of the disjoint channel envelope, is a straight line, above some threshold value.[1, 2]

The envelope of the autocorrelation function of the noise, y(t), was found to be essentially the autocorrelation of the bandlimiting filter.

Estimates of the fluctuating power level, A(t), were measured by integrating the

magnitude of the noise, y(t), for periods of 1 ms and 10 ms. Autocorrelation of this sample record shows that A(t) is dominated by an uncorrelated term, but exhibits two second-order correlations of 3-4 ms and 300-400 ms. We qualitatively related these second-order effects to physical mechanisms in the lightning-discharge process.

In the present report we shall present a mathematical model for the observed noise process.

1. First-Order Noise Model

First-order statistics of the atmospheric noise process are given by statistically independent samples of

$$y(t) = n_1(t) + x(t)\, a(t)\, n_2(t). \tag{1}$$

The process $n_1(t)$ is a narrow-band Gaussian process with variance σ_G^2. The process $n_2(t)$ is also Gaussian with variance σ_H^2. The autocorrelation of n_1 and n_2 is proportional to the autocorrelation of the bandlimiting filter, and they are statistically independent. The process x(t) is a two-state process with value one or zero, with probability p_T and q_T, respectively. The lowpass process, a(t), is the inverse of a χ process with m degrees of freedom and unity variance parameter. Both a(t) and x(t) are common to noise waveforms observed in different frequency channels.

The random process generated by $a(t)\, n_2(t)$ is the process suggested by Hall[3] as a model for atmospheric noise waveforms observed at the output of a bandlimiting filter. We have added features as a possible way of providing the noise model with additional degrees of freedom to account for the variability of the first-order observations and the joint channel statistics. In some sense, we have moved closer to the physics of the actual noise process, which exhibits a constant low-level background noise with discrete time periods when it goes into a non-Gaussian state corresponding to the occurrence of a discharge. This time discreteness is modeled by the two-state x(t) process, while the noise burst itself is modeled by the "Hall component" of the noise.

It is not possible to express the envelope of the atmospheric noise model in a simple form by proceeding formally from (1) and using the Hilbert transform definition of an envelope. A suitable approximation, paralleling the previous definition, is given by

$$v(t) = v_1(t) + x(t)\, a(t)\, v_2(t). \tag{2}$$

The processes $v_1(t)$ and $v_2(t)$ are Rayleigh, paralleling the characteristics of n_1 and n_2 above. The processes x(t) and a(t) are those defined above, whether v(t) is observed in the same channel with y(t), or in a disjoint frequency channel.

2. Comparison of Noise Model with Measured First-Order Noise Characteristics

The probability density of the Hall noise component, both bandpass and envelope, can be expressed in closed form.[3] The composite probability density for random variables y and v is then given by a weighted sum of a Gaussian or Rayleigh density and the convolution of the Gaussian/Rayleigh with the Hall component. We have used numerical methods to evaluate this convolution. To compare the model with observed data, we have used the unconditional density of the bandpass noise and the envelope in disjoint channels, and the conditional variance of y on v as a "link" between these distributions. These three variables are shown in a composite plot in Fig. XIII-9, where we have used RF and E to indicate observed variables, corresponding to the model variables y and v, respectively.

Four parameters characterize either the radio-frequency or envelope density: σ_G, σ_H, p_T and m. The variance of the background noise component, σ_G^2, can be estimated from the conditional-variance plot as the intersection of the horizontal portion with the radio-frequency axis. This horizontal portion is caused by the uncorrelated background noise component, $n_1(t)$. The value of p_T can be estimated from the conditional plot as the intersection of the value of the envelope, at which the conditional variance begins to depend linearly on the envelope, with the cumulative envelope probability, $P(v \leq v_o)$. This value of p_T is common to both channels. For large values of the argument in either the y or v density, the Hall noise component dominates, and the log of this density is given by

$$\log [p(y)] \propto -(m+1) \log (y) \quad \text{and} \quad \log [p(v)] \propto -(m+2) \log (v).$$

Thus m, common to both channels, can be estimated from the slope of the tails of the observed unconditional densities. Although the probability density of the Hall noise component is valid for any $m > 0$, we have restricted m to be an integer so that the χ-process generator can be used in simulation applications.

With these parameter values given approximately, the remaining parameters were chosen, with the aid of a computer, to achieve the match of model and measured characteristics shown in Fig. XIII-9. We found that the resulting match was quite sensitive to the entire set of parameters, and that a 10-20% change in any single parameter was reflected in the quality of the three-curve match. Similar results were achieved for "quiet" noise conditions and "quiet-tropical" transition conditions, characterized by a value for p_T of 0.11 and 0.5, respectively. The match to "frontal" observations, characterized by a p_T of 0.9, was not as good, although apparently adequate for design use.

Several remarks should be made concerning the model and its application to

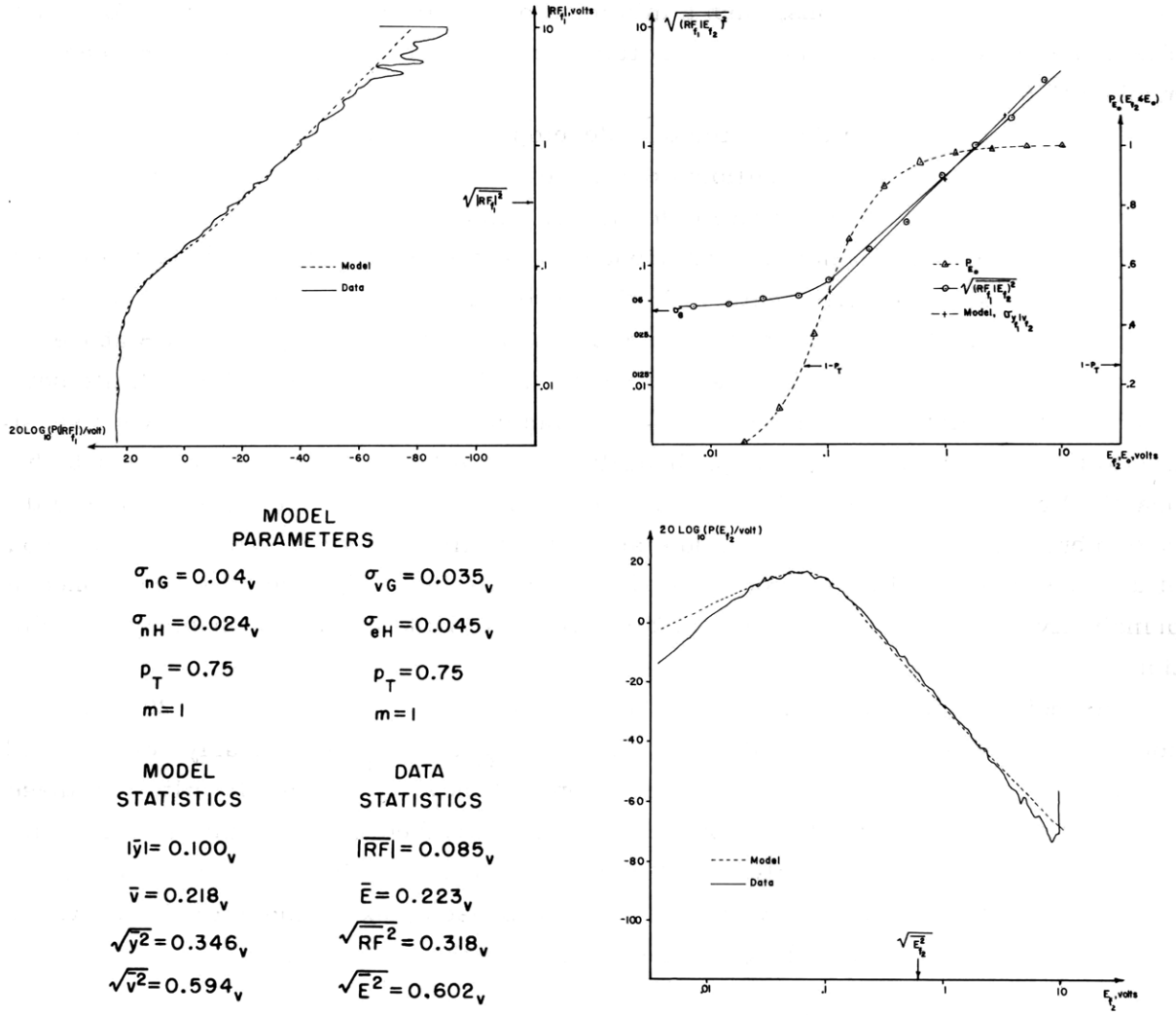

Fig. XIII-9. Comparison of measured atmospheric noise characteristics with noise-model characteristics. Measurements were made at Wildwood, New Jersey, August 20, 1971, 1500 GMT, under tropical conditions.
E_{f2} = 83 kHz, 1 kHz BW.
RF_{f1} = 65 kHz, 1 kHz BW.

the observed data.

If we consider the three data plots in Fig. XIII-9 as three curves to be represented by mathematical expressions, then a minimum of 10 parameters is required for a set. The noise model and associated assumptions on cross-channel dependence reduce this number to 6.

The noise model with the parameters developed here also predicts other joint channel statistics, such as the conditional density of the bandpass noise, given an envelope value, or the ratio of mean deviation to rms of this density.

The Hall noise component has an unbounded variance for $m \leq 2$, and hence the density of y or v does. We have dealt with this problem in a manner similar to Hall's by truncating the density at a point of maximum validity, generally 50-60 dB above σ_G. The justification for this is its practicalness. Variance is only one of an infinite set of possible parameters to describe a non-Gaussian distribution. Variance is very sensitive to the large-amplitude, low-probability region of the distribution, whereas both practical and desirable signal-processing realizations are relatively insensitive to this region but very sensitive to the middle and low-amplitude region. To achieve our goal of developing a useful noise model for signal-processing design, we have been concerned principally with achieving a good match over the bulk of the probability distribution.

Our noise observations have been in RF bandwidths of 1, 10, and 20 kHz. Our noise-model parameters show that the values of σ_G and σ_H are linearly related, with constant p_T and m values, by the noise power bandwidth of the RF filter at these observed bandwidths. We know, by the Central Limit theorem, that this relationship cannot continue for arbitrarily small bandwidths because the noise waveform must become Gaussian in nature. We do not have, at present, experimental data on which to base changes in our model parameters to reflect this behavior.

3. Time Structure of the Model

Inspection of the sample records of the integrated estimates of A(t), and consideration of the physical characteristics of the lightning discharges, show that the occurrence rate of discharges is nonstationary over periods of tenths of seconds, and that the amplitude of A(t) is correlated with this nonstationarity. We have approximated this complex behavior by modeling x(t) as a two-state Markov process with an occurrence rate which is also stochastic. The a(t) process is a nonlinear function of x(t), in that it assumes a fixed random value (described by the density of a) for each transition of x(t) from zero to one. The value of a(t) is independent of the occurrence rate of x(t), which preserves the first-order noise model previously described. We are, then, approximating each discharge event by a burst of Gaussian noise, added to the background level, where the rms value of each noise burst is described by the random

variable a. Thus the model ignores the actual correlation of noise-burst power level with event rate which is characteristic of multiple discharges.[1]

The duration of a given model discharge event is random, with the average value given by the transition rate of the two-state x(t) generator, λ_{10}. This rate can be chosen to approximate the behavior of the autocorrelation of the A(t) estimates. We have found that a rate of 850 Hz provides reasonable agreement with all of the

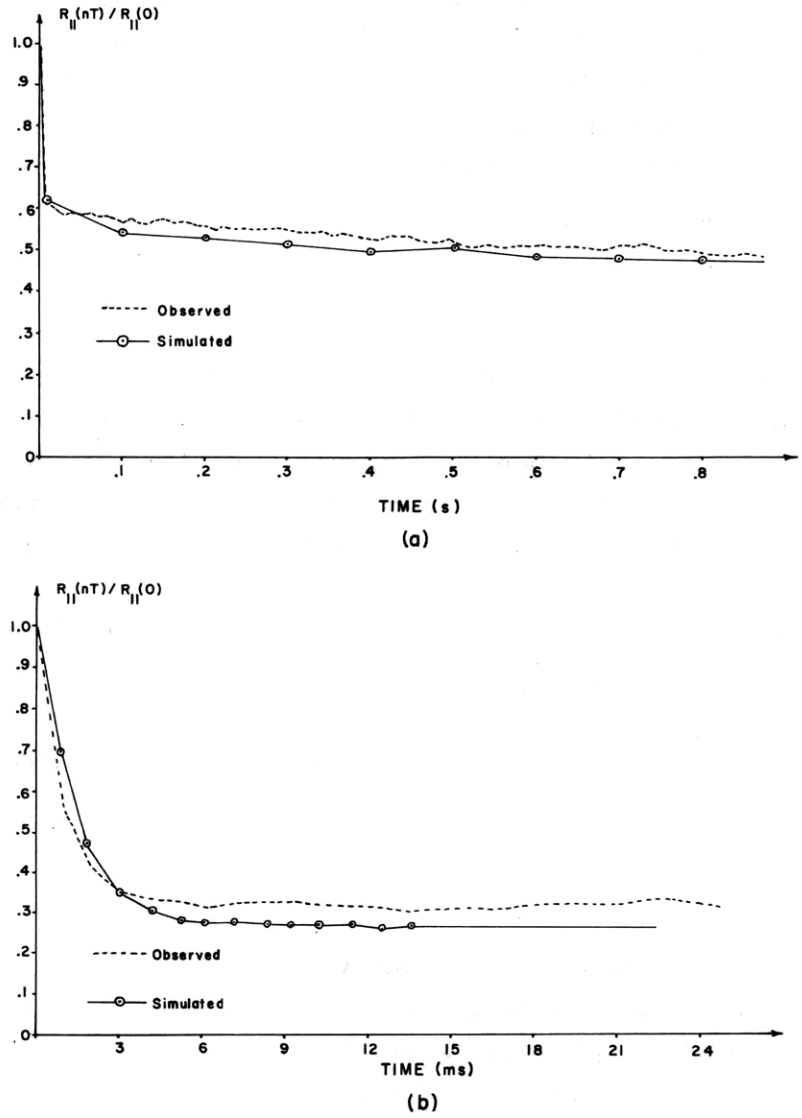

Fig. XIII-10. Comparison of the autocorrelation of observed and simulated 1-ms and 10-ms estimates ($A(t_i)$) of the atmospheric noise envelope. Tropical conditions and model parameters.
(a) 10-ms estimates. (b) 1-ms estimates.

observed data. This value agrees with the reported observation[1] that the duration of a typical single discharge event is approximately 1 ms. Figure XIII-10b shows the autocorrelation of simulated estimates of A(t) compared with observed data. The first-order model parameters of Fig. XIII-9 were used for the distributions required in this simulation.

Various methods might be considered for stochastically modulating the occurrence rate of x(t), which is the λ_{01} transition rate of the x(t) generator. For the noise conditions characterized by p_T of 0.75 and larger, any technique tends to "switch" between a maximum rate, thereby producing a near continuum of x(t) events, and a minimum rate. We have selected a second two-state Markov generator to modulate λ_{01} and to provide this behavior. The principal advantage of this approach is that simple analytic expressions can be formulated to guarantee that the average value of p_T, given by

$$\bar{p}_T = \left[\overline{\frac{\lambda_{01}(t)}{\lambda_{01}(t) + \lambda_{10}}}\right],$$

equals that required by the first-order model parameters. The average length of an intense burst of x(t) is controlled by the 10-transition rate of the second two-state generator, and this is chosen to imitate the long-term correlation of the A(t) estimates. The 01 transition rate is then fixed to provide the required $\bar{p}_T$. A comparison of simulated estimates of A(t) and observed data for the long-term correlation is shown in Fig. XIII-10a.

We incorporate our description of the x(t) a(t) time structure in (1), and see that the time variation of x(t) a(t) is significantly longer than the variations $n_1(t)$ and $n_2(t)$ which are given by the correlation time of the bandlimiting filter. Hence our model also provides the behavior required of the autocorrelation of y(t), that it be proportional to the bandlimiting filter autocorrelation. This does not hold for bandwidths significantly less than 1 kHz.

4. Conclusion

The complete atmospheric noise model that we have evolved can be synthesized with uniform and Gaussian random-number generators for Monte Carlo simulations. These, together with a table of model parameters describing the various noise conditions, allow us to achieve a wide degree of simulation complexity. This would include the range from simulation of independent samples of bandlimited noise, or an envelope, to joint channel simulation including time structure.

Our noise model has retained Hall's basic model as a description of the non-Gaussian noise component of atmospheric noise, while extending its usefulness to the modeling of the statistical dependence between disjoint frequency channels. Our final model

has also incorporated several features proposed by others. In particular, the idea of a switched process (the x(t) multiplier) was given by Kapp and Kurz,[4] while a time-variant Poisson process (for an impulse wave exciting a bandlimiting filter) was originally suggested by Furutsu and Ishida,[5] and recently incorporated in an analog atmospheric noise simulator by Coon et al.[6] We have demonstrated that the parameters of our model can be chosen to represent adequately some statistical characteristics of low-frequency atmospheric noise. We believe that the model can be extended to much higher frequencies, since the non-Gaussian noise component at these frequencies is dominated by the lightning-discharge leader structure which appears as a burst of noise.

We have noted at several points that our model becomes invalid for RF bandwidths significantly less than 1 kHz. This bandwidth plays a central role in atmospheric noise waveforms because it represents a lower limit at which the effect of individual discharge events can be delineated at the output of the bandlimiting filter. The practical significance of this fact is that if we use a bandwidth much less than 1 kHz, it becomes increasingly difficult for any nonlinear or time-variant signal-processing structure to mitigate the effect of individual noise bursts. In the limit of vanishingly small bandwidths, the output noise becomes Gaussian and the optimum signal processing structure becomes linear for linear signal modulation. In contrast to this, we expect to show in a subsequent report that a zero memory nonlinearity applied to atmospheric noise waveforms of 1-kHz bandwidth or larger results in improvements in signal-to-noise ratio of 5-17 dB compared with a linear-processing structure of equivalent post-detection bandwidth. This improvement is independent of the RF bandwidth, above 1 kHz. Thus our model describes atmospheric noise waveforms in a bandwidth region that is most important for signal-processing design.

D. A. Feldman

References

1. D. A. Feldman, "Model for Low-Frequency Atmospheric Noise," Quarterly Progress Report No. 104, Research Laboratory of Electronics, M. I. T., January 15, 1972, pp. 351-361.
2. Ibid., see Fig. XXIV-4, p. 358.
3. H. M. Hall, "A New Model for Impulsive Phenomena: Applications to Atmospheric-Noise Communication Channels," Technical Report 3412-8, Stanford Electronics Laboratory, Stanford, California, August 1966.
4. J. Kapp and L. Kurz, "Performance of Two Suboptimum Detectors and Signal Selection in Gaussian and Impulsive Noise," Proc. Seventh Allerton Conference on Circuits and System Theory (IEEE, New York, 1969), pp. 459-465.
5. K. Furutsu and T. Ishida, "On the Theory of Amplitude Distribution of Impulsive Random Noise," J. Appl. Phys. 32, 1206-1221 (1961).
6. R. M. Coon, E. Bolton, and W. Bensema, "A Simulator for H. F. Atmospheric Radio Noise," ESSA Technical Report ERL-128-ITS-90, July 1969.