NASA CR 73,442
AVAILABLE TO THE PUBLIC

Final Report

CHARACTERIZATION OF CAUSES OF SIGNAL PHASE AND FREQUENCY INSTABILITY

By Elie J. Baghdady and Southard Lippincott

January 8, 1970

Prepared Under

Contract NAS2-5337

Design, Develop, and Fabricate an Oscillator Instability Measurement System

For

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California 94035

ADVANCED COMMUNICATIONS INFORMATION MANAGEMENT

RESEARCH • DEVELOPMENT • ENGINEERING

ADCOM

A TELEDYNE COMPANY

NASA CR 73,442

Final Report

CHARACTERIZATION OF CAUSES OF SIGNAL PHASE AND FREQUENCY INSTABILITY

By

Elie J. Baghdady and Southard Lippincott

January 8, 1970

Prepared Under

Contract NAS2-5337

Design, Develop, and Fabricate an Oscillator Instability Measurement System

For

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California 94035

By

Teledyne ADCOM
808 Memorial Drive
Cambridge, Massachusetts 02139

G-150F

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

1. INTRODUCTORY STATEMENT OF PURPOSE

The purpose of this final report, prepared under Contract NAS2-5337, is to present the results of the study phase of Phase I of the contract. The study is concerned with the identification and characterization of the causes of signal phase and frequency instability whose measurement is facilitated by the use of equipment called the Oscillator Instability Measurement System (OIMS). The study includes:

A. The necessary measurements to be made with the OIMS on the Pioneer E spacecraft (prior to launch) and at the NASA Deep Space Station receivers in order to allow separate identification of phase and frequency instability caused by the following phenomena:

    1. Spacecraft oscillator
    2. Atmospheric multipath
    3. Scatter by atmospheric turbulence
    4. Fluctuations in time-variant spatial mean of the index of refraction
    5. DSS receiver under strong and noisy signal conditions.

B. The development of procedures and analytical characterization necessary to separate the various phase and frequency instabilities determined from the OIMS measurements specified in A above.

2. SURVEY OF CAUSES AND ASSOCIATED CHARACTERISTICS OF PHASE AND FREQUENCY INSTABILITY OF A RECEIVED OSCILLATOR SIGNAL

The various causes of phase and frequency instability of a received oscillator signal are listed in Table 1. The special characteristics (or "signatures") of each of these causes determine their measurability and the ability to separate their effects on the measured parameters of the composite received signal. A brief review of these characteristics will now be made. The source and the receiver causes of instability are similar in character and they will be covered under one major heading in Sec. 3. The propagation disturbances are covered in Sec. 4.

Table 1

CAUSES OF PHASE AND FREQUENCY INSTABILITY

1. Source Causes (Oscillator)

   1.1 Internal Causes

   - Flicker Noise
   - Oscillator Loop Noise
   - Power Supply Ripple

   1.2 External Causes

   - Flicker Noise in Amplifiers and Frequency Processors (Mixers, Dividers, Multipliers, etc.)
   - Additive Noise in Output Stages
   - Spurious Radiation Pickup

2. Propagation Effects (Transmission Medium)

   - Variable Refractive Properties of Medium
   - Scattering by Atmospheric Turbulence
   - Atmospheric Multipath
   - Reflection Multipath

3. Receiver Causes

   - Front End Noise
   - Receiver Circuit Parameter Instabilities (essentially similar to those listed under Source Causes)

3. MODELS AND CHARACTERISTICS OF UNSTABLE OSCILLATIONS*

Having identified in Table 1 the intrinsic causes of the instabilities that determine the phase and frequency errors of reference oscillations, we now turn to the problem of mathematical modeling of unstable oscillations and the characterization of instabilities caused by various mechanisms.

There are two approaches to the modeling of oscillations with frequency instabilities. These approaches may be called the causal approach and the black-box approach.

In the causal approach, a functional structure of the oscillation source is postulated, and various causes of instability are identified and combined with an assumed ideal oscillation signal. The model is thus centered on postulated occurrences within a box, and the consequences of these occurrences on the properties of the output signal are then pursued. This approach is essential for an understanding of the reasons for the instability, especially in the search for developing ultra-stable sources.

In the black-box approach, the model is centered on the resultant disturbed oscillation at the output terminals of the box, the point where the oscillation is available to the user. Thus, a number of output signal models are postulated, their properties explored, and the actual source output is identified, their properties explored, and the actual source output is identified piecewise (for different frequency regions) with one (or perhaps more) of these models.

---

* The material in this section is adapted from the paper "Short-Term Frequency Stability: Characterization, Theory and Measurement," by E. J. Baghdady, et al., Proc. IEEE, September, 1965.

3.1 Causal Approach

In the causal approach to the modeling of unstable oscillations, two types of disturbances that give rise to frequency instability are distinguished; namely, additive noise and multiplicative noise. In loop-controlled oscillators, the reference signal input may also contain a non-negligible noise component added to the reference signal.

Additive noise models take account of noise added to the oscillator signal either in the oscillator loop or in the buffer amplifier following the oscillator, as illustrated in Fig.3.1. The spectrum of this type of noise occupies a frequency region that often extends widely around the oscillator frequency. The noise introduced into (or originating in) the oscillator loop will be called internal additive noise; the noise added in the amplifier following the oscillator will be called external additive noise.

In addition to the additive noise near the oscillator frequency, multiplicative low-frequency processes are also present that effectively frequency modulate the oscillator. Examples of such processes are current and voltage fluctuations (flicker noise), mechanical vibrations, temperature variations, etc. The deviation ratio of the frequency modulation caused by these processes is usually large. Consequently, the shape of the resulting r-f spectrum is essentially identical with the shape of the probability density function of the resultant baseband noise process.

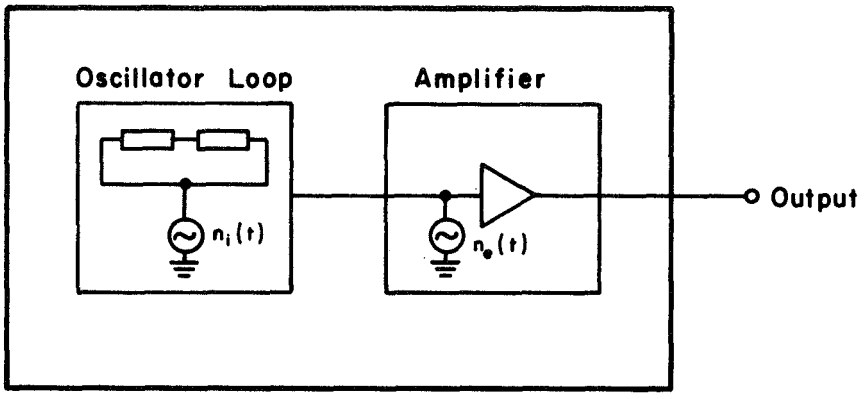

Fig.3.1 Block illustration of an oscillator with a buffer amplifier, showing additive noise sources.

$n_i(t)$ = internal noise source

$n_e(t)$ = external noise source

3.1.1 Additive-Noise Model: External Noise

For this model, it is assumed that the oscillator delivers a perfectly stable signal, to which relatively white gaussian noise is added. If the oscillator signal is expressed as

$$e_{osc}(t) = A \cos \omega_o t \tag{3.1}$$

the additive noise may be expressed in the form

$$n_e(t) = n_c(t) \cos \omega_o t - n_q(t) \sin \omega_o t \tag{3.2}$$

where $n_c(t)$ and $n_q(t)$ are white gaussian processes. The sum of the oscillator signal and the additive noise is

$$e_{out}(t) = [A + n_c(t)] \cos \omega_o t - n_q(t) \sin \omega_o t \tag{3.3}$$

Since we are dealing with very stable oscillators, it can be assumed that the rms value of the noise is much smaller than the rms value of the oscillator signal. Hence, only $n_q(t) \sin \omega_o t$ influences the phase of $e_{out}(t)$ significantly, and

$$e_{out}(t) \approx A[1 + n_c(t)/A] \cos \{\omega_o t + [n_q(t)/A]\} \tag{3.4}$$

Thus, a relatively weak externally added noise component results in envelope fluctuations, $n_c(t)/A$, and phase fluctuations, $\phi_e(t)$, given by

$$\phi_e(t) = n_q(t)/A \tag{3.5}$$

If, over the frequency range of interest, the mean-square one-side spectral density of the added noise is N_o volts^2/cps, then the two-side spectral density of $n_q(t)$ will also be N_o volts^2/cps and the phase perturbations of $e_{out}(t)$ will have a spectral density N_o/A^2 rad^2/cps. The added noise will

of course be modified by the r-f filtering in the buffer amplifier. This filtering restricts the bandwidth of the phase perturbations so that their spectrum is confined to a frequency region extending up to one-half of the amplifier r-f bandwidth. Thus, the mean-square spectral density $S_{\phi}(\omega)$ of the output signal phase fluctuations will be as illustrated in Fig. 3.2, where B_a cps denotes one-half of the amplifier bandwidth.

Since the instantaneous frequency fluctuations are given by the time derivative of the phase fluctuations, the mean-square spectral density $S_{\phi}^{\cdot}(\omega)$ of the frequency fluctuations is given by

$$S_{\phi}^{\cdot}(\omega) = \omega^2 S_{\phi}(\omega) \tag{3.6}$$

In the present case

$$S_{\phi}^{\cdot}(\omega) = \omega^2 N_o/A^2, \text{ for } |\omega| < 2\pi B_a \tag{3.7}$$

If the r-f amplifier frequency response beyond $\pm 2\pi B_a$ falls off faster than $1/\omega^2$, the mean-square spectral density of the frequency fluctuations will be as illustrated in Fig. 3.2.

It is clear that this type of noise in unstable oscillator outputs can be characterized by the following two gross parameters:

a) The relative noise density N_o/A^2.

b) The buffer amplifier bandwidth $2B_a$.

3.1.2 Additive-Noise Model: Internal Noise

In this model the additive noise enters into the oscillator loop as shown in Fig. 3.3. The noise $n_i(t)$ is assumed white in a frequency

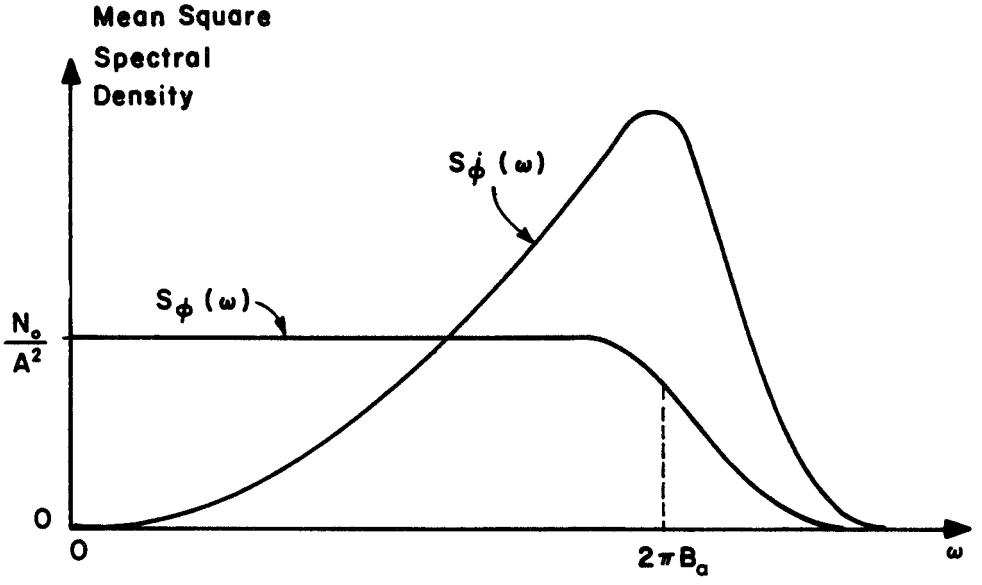

Fig. 3. 2 **Example of mean-square spectral density of phase fluctuations due to additive external noise.**

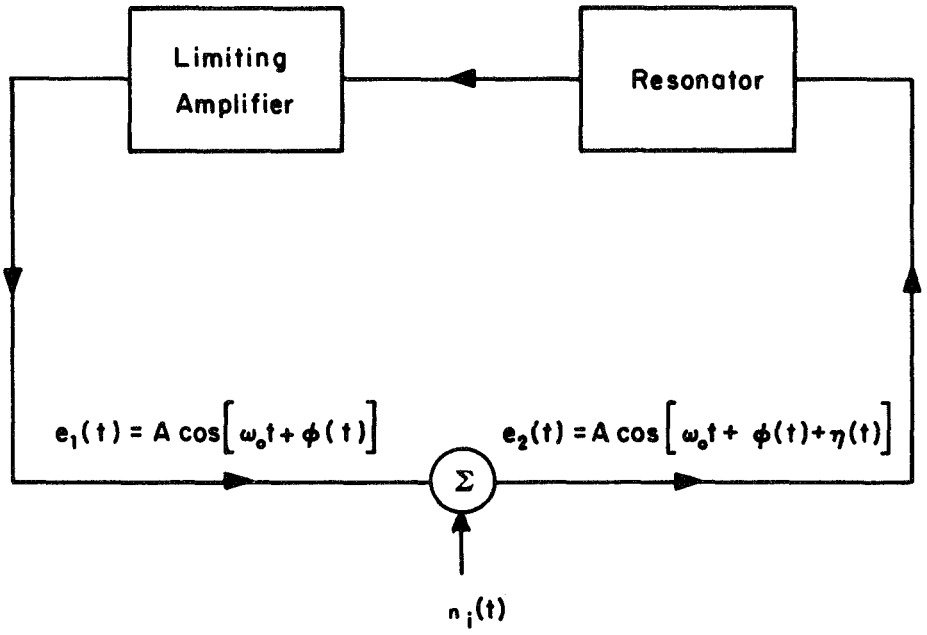

Fig. 3. 3 Block diagram of oscillator loop with internal additive noise.

region around the oscillator frequency. The additive noise $n_i(t)$ can be expressed in the form

$$n_i(t) = V_n(t)\cos[\omega_o t + \theta_n(t)] \tag{3.8}$$

where $V_n(t)$ and $\theta_n(t)$ are low-pass processes. Then the sum of the output of the limiter and the noise is

$$\begin{aligned} e_2(t) &= e_1(t) + n_i(t) \\ &= \{A + V_n(t)\cos[\phi(t) - \theta_n(t)]\}\cos[\omega_o t + \phi(t)] \\ &\quad + \{V_n(t)\sin[\phi(t) - \theta_n(t)]\}\sin[\omega_o t + \phi(t)] \end{aligned} \tag{3.9}$$

With $V_n(t) \ll A$ almost all of the time, the instantaneous phase of $e_2(t)$ is approximately $[\omega_o t + \phi(t) + \eta(t)]$, in which

$$\eta(t) \approx \frac{V_n(t)}{A}\sin[\phi(t) - \theta_n(t)] \tag{3.10}$$

Thus the effect of the additive noise is to introduce a fluctuating phase shift into the loop. Recognizing this, we can replace the adder in Fig.3-3 by an equivalent phase modulator as in Fig. 3-4.

When the phase shift $\eta(t)$ is equal to zero, the oscillator will oscillate at the frequency for which the phase shift through the resonator is zero (i.e., at its resonance frequency). However, when $\eta(t) \neq 0$, the introduction of the phase shift will change the oscillator frequency so that the phase shift in the resonator is equal to $-\eta(t)$ and the total phase shift around the loop is zero. Thus the variations of $\eta(t)$ will cause fluctuations of the oscillator frequency. In fact, Fig.3-4 represents the block diagram of a

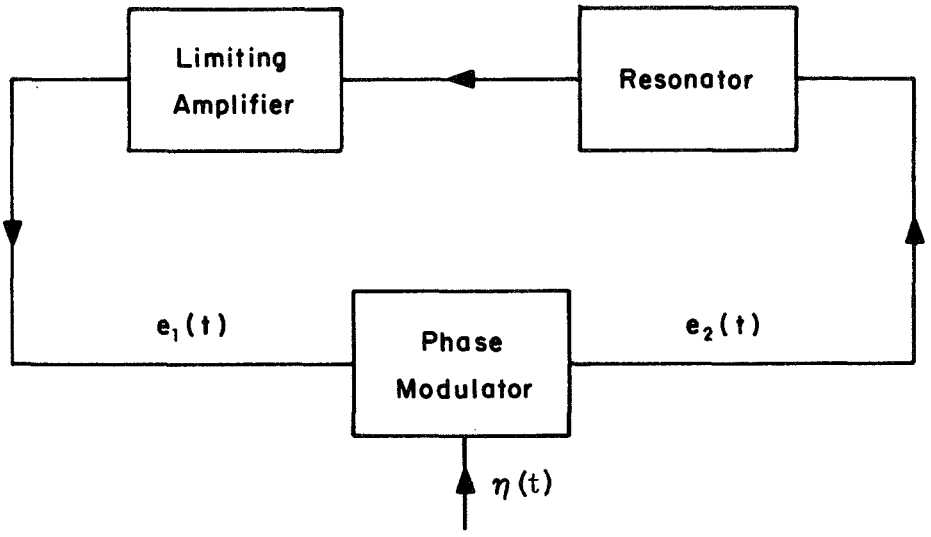

Fig. 3-4 Equivalent diagram of the oscillator in Fig. 3-1

familiar frequency-modulated oscillator. In this case, however, the control signal $V_n(t) \sin[\phi(t) - \theta_n(t)]$ is dependent on the phase $\phi(t)$ of the oscillator signal.

The frequency modulation sensitivity is determined by the phase slope (vs. frequency) of the resonator transfer function. Since the resulting frequency fluctuations will be small, only the part of the phase characteristic near the resonant frequency is of interest. In that region the phase characteristic can be assumed linear. For a resonator represented by a single tuned circuit with a half-power bandwidth of B_R cps, the phase shift $\phi(t)$ caused by a small frequency variation around the resonant frequency is approximately

$$\phi(t) \approx \dot{\phi}(t)/\pi B_R \tag{3.11}$$

Thus the phase condition for oscillations becomes

$$\eta(t) - (1/\pi B_R)\dot{\phi}(t) = 0 \tag{3.12}$$

whence

$$\dot{\phi}(t) = \pi B_R[V_n(t)/A] \sin[\phi(t) - \theta_n(t)] \tag{3.13}$$

This is a nonlinear differential equation with random input that seems very hard to solve. However some observations about the mean-square spectral density of $\dot{\phi}(t)$ may be made. If, for example, $\phi(t)$ was absent from the right-hand side of Eq.(3.13) then the mean-square spectral density of $\dot{\phi}(t)$ would be equal to the spectral density of $\pi B_R[V_n(t)/A] \sin\theta_n(t)$; i.e., it would be white with a mean-square spectral density equal to $(\pi B_R/A)^2 N_o$. The presence of $\phi(t)$ in the right-hand side will spread out the spectrum

with a downward slope away from $\omega = 0$. Aside from the tendency for peaking at $\omega = 0$ (because of the spectral convolutions associated with the nonlinearities inherent in (3.13)), the general level of the spectral density of $\dot{\phi}(t)$ will be lower than $(\pi B_R/A)^2 N_o$. Since it seems reasonable to assume that this spectral density is approximately uniform over the frequency range of interest, we can express it as $\alpha(\pi B_R/A)^2 N_o$, where α is a positive constant smaller than unity.

As in the case of the external noise it is clear that the amplifier following the oscillator will limit the spectrum of $\dot{\phi}(t)$ essentially to the frequencies up to B_a cps. The appearance of the mean-square spectral density of the phase and frequency fluctuations due to an internal noise source is shown in Fig. 3-5. In the region where $S_{\dot{\phi}}(f)$ is uniform, the spectrum of the phase is given by

$$S_{\phi}(f) = S_{\dot{\phi}}(f)/4\pi^2 f^2$$

The above type of oscillator instability can then be characterized by the following parameters:

a) The mean-square spectral density $\alpha(\pi B_R/A)^2 N_o$ of the oscillator frequency fluctuations.

b) The buffer amplifier bandwidth $2B_a$.

3.1.3 Multiplicative Noise Model

The frequency of an oscillator is always somewhat susceptible to variations in circuit parameters. An ultra-stable oscillator is, of course, designed so that the sensitivity to the parameter variations is very

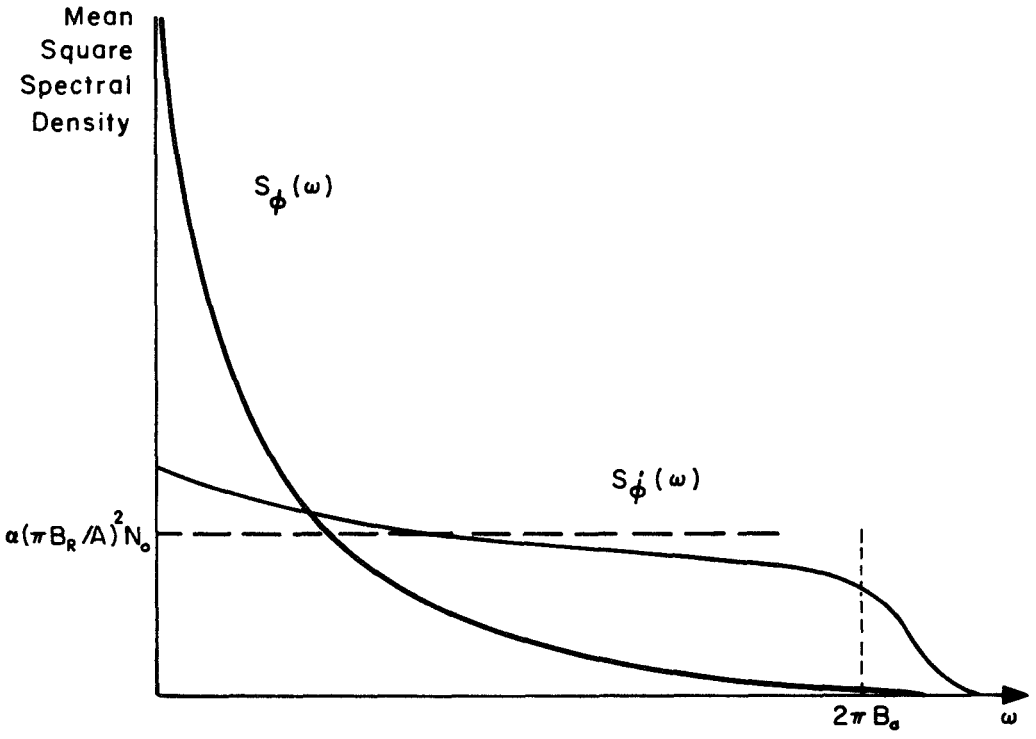

Fig. 3-5 Mean-square spectral densities of phase and frequency fluctuations due to internal noise source.

small. Considerable effort is also directed to keeping the parameters as constant as possible. Examples of phenomena that can cause parameter variations in an oscillator are temperature variations, voltage supply variations, current variations, magnetic field variations, mechanical vibrations, etc. These low-frequency processes essentially frequency modulate the oscillator. Some of them (as for example voltage supply variations due to hum or periodic vibrations) are periodic functions and will thus yield a discrete mean-square spectral density for the oscillator frequency or phase fluctuations. The other processes usually have continuous spectral densities that decrease monotonically with frequency.

The most common of the above processes are the low-frequency current fluctuations in the electronic tubes and transistors used in crystal oscillators. This type of fluctuation is called flicker noise. It has a spectral density of the form $S(\omega) = \beta/\omega$, where β is a constant. Since the energy of the flicker noise must be finite, this spectral density must ultimately become flat as $\omega \to 0$. Thus, the spectrum of the frequency fluctuations resulting from flicker noise can be characterized by the gross parameter β.

3.1.4 Loop-Controlled (APC and AFC) Oscillators[1]

There are two types of loop-controlled oscillators that are of wide interest, namely, automatic phase-controlled reference oscillators (APC) and automatic frequency-controlled reference oscillators (AFC). The performance of these systems in the presence of a strong reference signal is well-known: the instantaneous frequency noise caused by relatively white input additive gaussian noise is also gaussian with a mean-square spectral density that rises as the square of frequency and is ultimately bounded by the filtering effect of the driving filter in cascade with the low-pass equivalent closed-loop system function.

It is of interest here to consider the situation in which the input reference signal, if any, is weak compared with the noise present.

3.1.4.1 APC Oscillator

Consider the functional diagram of an APC loop shown in Fig. 3-6. Let

$$e_{in}(t) = [E_s + n_{c,\,if}(t)] \cos [\omega_{if} t + \psi(t)] - n_{q,\,if}(t) \sin [\omega_{if} t + \psi(t)] \qquad (3.14)$$

in which

E_s = constant amplitude of input reference signal

$\psi(t)$ = instantaneous phase fluctuations of input reference signal,

and $n_{c,\,if}(t)$ and $n_{q,\,if}(t)$ are instantaneous amplitudes of cophasal and quadrature components of the input noise. Also let the oscillation be represented by

$$e_{osc}(t) = E_{osc} \cos [\omega_{osc} t + \phi_{osc}(t)] \tag{3.15}$$

Under the assumption of stable feedback operation, and in the absence of spurious by-products in the output of the multiplier, we have, with $\omega_{if} = \omega_{osc}$,

$$\frac{1}{\mu\alpha} \dot{\phi}_{osc}(t) = \left\{ \left[1 + \frac{1}{E_s} n_{c,if}(t) \right] \sin \left[\psi(t) - \phi_{osc}(t) \right] + \frac{1}{E_s} n_{q,if}(t) \cos \left[\psi(t) - \phi_{osc}(t) \right] \right\} \boxtimes h_{\ell p}(t) \tag{3.16}$$

where $\boxtimes$ denotes convolution and $h_{\ell p}(t)$ is the impulse response of the low-pass filter. The existence of a feedback steady-state condition in which Eq.(3.16) gives an adequate description of loop performance requires that

$$|\psi(t) - \phi_{osc}(t)| \leq \pi/2 \tag{3.17}$$

If a perfect bandpass limiter operates on the sum of signal and noise before it is applied to the APC loop, Eq.(3.16) is replaced by

$$\frac{1}{\mu\alpha} \dot{\phi}_{osc}(t) = h_{\ell p}(t) \boxtimes \sin \left[\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t) \right] \tag{3.18}$$

where

$$\theta_{n,if}(t) = \tan^{-1} \frac{n_{q,if}(t)}{E_s + n_{c,if}(t)} \tag{3.19}$$

It is generally convenient to set

$$\phi_{osc}(t) = \phi_{osc,s}(t) + \phi_{osc,\ n,\ d}(t) \tag{3.20}$$

in which $\phi_{osc,\ n,\ d}(t)$ combines all the distortion and noise products in the phase of the oscillation and $\phi_{osc,s}(t)$ represents a pure signal term given by

$\psi(t) \boxtimes h_{eq}(t)$, $h_{eq}(t)$ being the impulse response of the low-pass linear equivalent strong-signal model of the closed-loop system.

When the input signal component is much weaker than the noise, Eq. (3.16) can be reduced to

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc,n,d}(t) \approx h_{\ell p}(t) \boxtimes \left\{\frac{1}{E_s} V_{n,if}(t) \sin\left[\theta_{n,if}(t) - \phi_{osc,n,d}(t)\right]\right\} \tag{3.21}$$

where
$$\theta_{n,if}(t) \approx \tan^{-1} n_{q,if}(t)/n_{c,if}(t) \tag{3.22}$$

and
$$V_{n,if}(t) \approx \sqrt{n^2_{c,if}(t) + n^2_{q,if}(t)} \tag{3.23}$$

Formally, Eq. (3.21) differs from Eq. (3.13) for the effect of an internal noise source only by the low-pass filtering effect $h_{\ell p}(t)$.

The introduction of a perfect bandpass limiter to operate upon the sum of signal and noise before the APC loop transforms Eq.

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc,n,d}(t) \approx h_{\ell p}(t) \boxtimes \sin\left[\theta_{n,if}(t) - \phi_{osc,n,d}(t)\right] \tag{3.24}$$

In order to understand the significance of Eqs. (3.21) and (3.24) we first note that in this discussion we are concerned primarily with the situations in which either

$$\phi_{osc,n,d}(t) \text{ fluctuates much more slowly than } \theta_{n,if}(t) \tag{3.25}$$

or
$$|\phi_{osc,n,d}(t)| \ll |\theta_{n,if}(t)| \leq \pi/2 \tag{3.26}$$

These conditions correspond to an APC loop whose low-pass transmittance

to the phase of the oscillation attenuates strongly all frequencies that are not much smaller than the bandwidth of the i-f system preceding the loop.

Thus, under either of conditions (3.25) and (3.26), Eqs. (3.21) and (3.24) can be approximated by

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc,n,d}(t) \approx h_{\ell p}(t) \boxtimes \left[\frac{1}{E_s} x_{q,if}(t)\right] \tag{3.27}$$

and

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc,n,d}(t) \approx h_{\ell p}(t) \boxtimes \sin\theta_{n,if}(t) \tag{3.28}$$

Integration yields

$$\phi_{osc,n,d}(t) \approx \frac{\mu\alpha}{E_s}\int^{t} h_{\ell p}(u) \boxtimes x_{q,if}(u)\,du \tag{3.29}$$

in the absence of prelimiting, and

$$\phi_{osc,n,d}(t) \approx \mu\alpha\int^{t} h_{\ell p}(u) \boxtimes \sin\theta_{n,if}(u)\,du \tag{3.30}$$

when the input signal plus noise is prelimited. Equations (3.29) and (3.30) show that irrespective of prelimiting or the lack of it, the APC loop treats the noise presented to it in an essentially linear manner with an equivalent linear filter system function given by

$$H_{o\ell}(s) = \mu\alpha H_{\ell p}(s)/s \tag{3.31}$$

as long as either of conditions (3.25) and (3.26) is satisfied. Note that $H_{o\ell}(s)$ is the "open-loop" transfer function of the APC loop.

Some remarks about the physical significance of the above results are in order. First, note that $H_{\ell p}(s)/s$ has a pole at the origin as long as $H_{\ell p}(s)$ is not allowed to have a zero there. The effect of this pole is to cause $|H_{o\ell}(j\omega)|^2$ to weight a nonzero noise density around $\omega = 0$ in a

hyperbolic manner. Thus, even though $|H_{\ell p}(j\omega)/\omega|$ cuts off more rapidly than $|H_{\ell p}(j\omega)|$ at the higher frequencies, the behavior near $\omega = 0$ does not allow the integral that yields the mean-squared value of the output noise to converge. The physical meaning of this is that although the mean value of $\phi_{osc,n,d}(t)$ will be zero, the unboundedness of the mean square indicates unbounded excursions that require a re-examination of condition (3.26). Physically, the oscillation beats continuously with the input noise. Heavy filtering of the higher-frequency noise beats by $H_{\ell p}(s)/s$ causes $\phi_{osc,n,d}(t)$ to be rather slow compared with $\theta_{n,if}(t)$, thus allowing the noise modulation of the oscillator frequency to have an essentially gaussian character and permitting the approximation of (3.21) by (3.27) and (3.24) by (3.28). The oscillator phase fluctuations wander out of step with the input noise fluctuations causing the loop to treat the noise essentially in an open-loop manner. If the input signal component is much weaker than the noise so that its contribution at the low-pass filter output is dominated by the noise, the loop will be unable to distinguish the signal from the equivalent of a weak noise component. No closed-loop control can therefore be established by the signal, and the oscillator frequency is modulated mainly by the gaussian noise as filtered by the low-pass filter $h_{\ell p}(t)$. The effect of this modulation of the oscillator frequency is known to cause the oscillator phase to stray in random-walk-like manner. But, even though the mean-square error in identifying the oscillator phase will diverge (directly, for random walk) with the length of the observation interval [because of the cumulative action of the integration in (3.29) and (3.30)], the mean-square error in identifying the

oscillator frequency will go to zero (inversely) with increasing length of the observation interval. Comparison of (3.30) and (3.29) also shows that prelimiting increases the noise level in the output.

3.1.4.2 AFC Oscillator

The functional diagram of an oscillator with an automatic frequency control loop is shown in Fig.3-7. We assume that the closed-loop system can lock properly to the desired signal in the absence of noise, and let $e_{in}(t)$ and $e_{osc}(t)$ be described by (3.14) and (3.15), respectively. The AFC loop is assumed to be driven with no prior amplitude limiting of signal plus noise. Under these conditions, we further assume that the mixer delivers in the nominal passband of the loop bandpass filter only the beat components

$$[E_s + x_{c,if}(t)] \cos [\omega_{bp} t + \psi(t) - \phi_{osc}(t)]$$

$$- x_{q,if}(t) \sin [\omega_{bp} t + \psi(t) - \phi_{osc}(t)]$$

in which $\omega_{bp} \equiv \omega_{if} - \omega_{osc}$ = nominal center frequency of the bandpass filter within the loop. If an amplitude-insensitive linear FM demodulator and a linear feedback from the discriminator output to the instantaneous frequency of the VFO are assumed, we can show that

$$\phi_{osc}(t) = k_d \beta_{fb} \left[\tan^{-1} \frac{A_q(t) \boxtimes h_{LP}(t)}{A_c(t) \boxtimes h_{LP}(t)} \right] \boxtimes h_{\ell p}(t) \qquad (3.32)$$

where

$$A_c(t) \equiv \left[1 + \frac{1}{E_s} x_{c,if}(t) \right] \cos \left[\psi(t) - \phi_{osc}(t) \right]$$

$$- \frac{1}{E_s} x_{q,if}(t) \sin \left[\psi(t) - \phi_{osc}(t) \right]$$

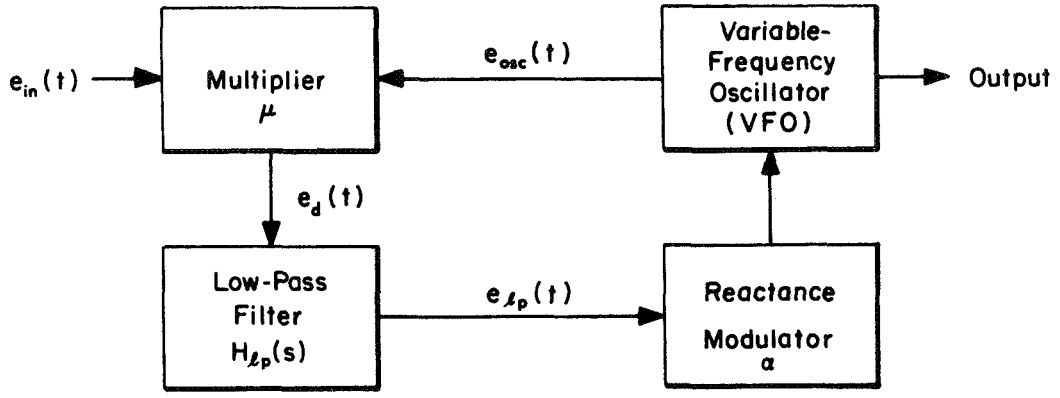

Fig. 3-6 Block diagram of APC oscillator.

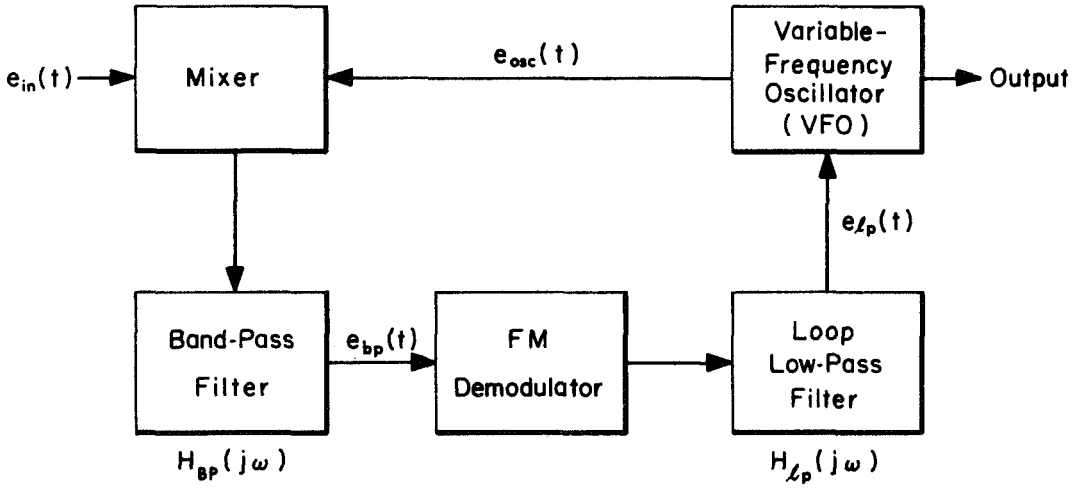

Fig. 3-7 Block diagram of AFC oscillator.

$$A_q(t) \equiv \left[1 + \frac{1}{E_s} x_{c,if}(t) \right] \sin \left[\psi(t) - \phi_{osc}(t) \right]$$

$$+ \frac{1}{E_s} x_{q,if}(t) \cos \left[\psi(t) - \phi_{osc}(t) \right]$$

and $h_{LP}(t)$ is the impulse response of the lowpass analog of the bandpass filter.

Now let the input reference signal be much weaker than the noise. Two cases are considered. In the first, we assume that $|\psi(t) - \phi_{osc}(t)| \ll 1$. We then have from Eq. (3.32)

$$\phi_{osc}(t) \approx k_d \beta_{fb} h_{\ell p}(t) \boxtimes$$

$$\tan^{-1} \frac{\{n_{q,if}(t) + n_{c,if}(t)\,[\psi(t) - \phi_{osc}(t)]\} \boxtimes h_{LP}(t)}{\{n_{c,if}(t) - n_{q,if}(t)\,[\psi(t) - \phi_{osc}(t)]\} \boxtimes h_{LP}(t)} \tag{3.33}$$

If we introduce some additional assumptions, this expression can be simplified still further. For example, if the loop low-pass filter bandwidth is relatively "sharply" restricted to a value below one-half of the bandpass filter bandwidth and if $k_d \beta_{fb}$ is not high enough to broaden the bandwidth of $\phi_{osc}(t)$ above one-half of the bandpass filter bandwidth, then

$$\phi_{osc}(t) \approx k_d \beta_{fb} h_{\ell p}(t) \boxtimes \tan^{-1} \left\{ \psi(t) - \phi_{osc}(t) + \frac{n_{q,if}(t) \boxtimes h_{LP}(t)}{n_{c,if}(t) \boxtimes h_{LP}(t)} \right\}$$

$$\approx k_d \beta_{fb} h_{\ell p}(t) \boxtimes [\psi(t) - \phi_{osc}(t)] + k_d \beta_{fb} h_{\ell p}(t) \boxtimes \phi_{n,LP}(t)$$

where

$$\phi_{n,LP}(t) = \tan^{-1} \frac{n_{q,if}(t) \boxtimes h_{LP}(t)}{n_{c,if}(t) \boxtimes h_{LP}(t)}$$

Rearrangement of terms yields (with $u_o(t)$ = unit impulse)

$$[u_o(t) + k_d\beta_{fb}h_{\ell p}(t)] \boxtimes \phi_{osc}(t) \approx k_d\beta_{fb}h_{\ell p}(t) \boxtimes [\psi(t) + \phi_{n,\,LP}(t)]$$

whence

$$\phi_{osc}(t) = h_{eq'}(t) \boxtimes [\psi(t) + \phi_{n,\,LP}(t)] \tag{3.34}$$

where $h_{eq'}(t)$ is the impulse response of a linear filter with transmittance

$$H_{eq'}(s) = \frac{k_d\beta_{fb}H_{\ell p}(s)}{1 + k_d\beta_{fb}H_{\ell p}(s)} \tag{3.35}$$

This system function differs from the more complete closed-loop system function associated with strong-signal models of AFC systems and frequency-compressive feedback demodulators only in that the transmission effects of the loop bandpass filter upon the frequency fluctuations of the input reference signal have been neglected.[1]

If instead of requiring $|\psi(t) - \phi_{osc}(t)| \ll 1$, we express $\phi_{osc}(t)$ as a sum [see Eq. (3.20)] of a pure signal-modulation component $\phi_{osc,\,s}(t)$, and noise and distortion component $\phi_{osc,\,n,\,d}(t)$ and require only that

$$|\psi(t) - \phi_{osc,\,s}(t)|_{max} \ll 1$$

then, Eq. (3.32) with the noise assumed much stronger than the signal leads to

$$\phi_{osc,\,s}(t) \approx \psi(t) \boxtimes h_{eq'}(t)$$

and

$$\phi_{osc,\,n,\,d}(t) \approx k_d\beta_{fb}h_{\ell p}(t) \boxtimes \tan^{-1}\frac{h_{LP}(t) \boxtimes \sin[\phi_{n,\,if}(t) - \phi_{osc,\,n,\,d}(t)]}{h_{LP}(t) \boxtimes \cos[\phi_{n,\,if}(t) - \phi_{osc,\,n,\,d}(t)]} \tag{3.36}$$

where

$$\phi_{n,\,if}(t) \equiv \tan^{-1} \frac{n_{q,\,if}(t)}{n_{c,\,if}(t)}$$

In each case, the phase fluctuations of the weak input phase reference are transmitted, filtered by $h_{eq'}(t)$, to the controlled oscillator phase without compression by the noise. However, in practice the FM noise impulses and the dips of the noise envelope below the drive threshold of the FM demodulator within the loop will prevent this FM demodulator from operating in the "amplitude-insensitive linear" manner assumed in the analysis. Moreover, the various factors that contribute to improper AFC tracking of signal plus noise will all be in evidence below the AFC noise threshold. Consequently, the comparatively severe disruptive noise effects may well completely mask the frequency fluctuations contributed by the input reference signal.

3.2 Black-Box Approach

In the black-box approach to the modeling of unstable oscillations, the source output is represented in various frequency ranges by signal models whose properties fit the properties of the source output over the specified portions of the spectrum. This approach is useful because it provides an alternate route to the characterization of source outputs with emphasis on the user needs. It also adds further insight into the mechanisms of instability in the signal structure, provides representations that are particularly suitable for analysis in subsequent parts of a system, by-passes the difficulties of dealing with nonlinear differential equations by assuming applicable signal representations or forms, and is naturally suited to the analysis of problems of spectral purity.

A common way to explain how an oscillation comes about in a properly designed loop without visible "provocation" is to rationalize that unavoidable random current or voltage fluctuations in various parts of the loop provide an excitation in a feedback loop with sharp selectivity (determined principally by the phase shift of a high-Q resonator) about the frequency of in-phase feedback. The regeneration around the loop builds up the feeble provocation to a strong signal whose amplitude is limited ultimately by automatic nonlinear gain control action or by actual saturation. This, together with the fact that the regenerative loop selectivity is of nonzero (albeit small) width about the frequency of in-phase feedback, suggests that the resulting signal may actually consist of a band of gaussian-like noise with or without a distinct sinewave at

the center of the band. In the immediate vicinity of the oscillation frequency, the oscillation signal may therefore be modeled by a band of amplitude-limited gaussian noise.

Other models for representing the signal over various frequency intervals are:

Sinewave plus gaussian noise,

Constant-amplitude carrier frequency-modulated by narrow-band gaussian noise,

Sinewave plus one or more discrete-frequency components, and

Combinations of the above.

Applicability of these various models depends upon the origin of the signal. For example, if the signal is a processed form of an originally extremely stable source, then frequency processing adds noise with a bandwidth extending over at least ten times the original source width of band. Therefore, sinewave plus gaussian noise would be a good model. If, however, the width of the source band is not too narrow relative to bandwidths of later processing circuits, then the appropriate model may be

(a) A band of noise, if passive filters only are used for tone isolation;

(b) A constant-amplitude signal frequency modulated by noise if a phase locked loop is used for "tone isolation."

We now discuss some of the above models to bring out their characteristic properties.

3.2.1 Amplitude-Limited Narrow-Band Gaussian Noise

If the center frequency is denoted ω_o rad/sec, bandpass amplitude-limited narrow-band gaussian noise may be expressed as

$$e_\ell(t) = A_\ell \cos[\omega_o t + \eta(t)] \tag{3.37}$$

where $\eta(t)$ is the instantaneous phase of a gaussian noise process and hence has a uniform distribution over the range $|\eta| \leq \pi$. The mean-square spectral density, $S_{\dot{\eta}}(\omega)$, of $\dot{\eta}(t)$ has been derived in a number of publications,[2,3,4,5] for rectangular as well as for gaussian shapes of the spectrum before limiting. Illustrative curves are sketched in Figs.3.8(a) and (b) ($\rho = 0$). It is interesting to observe the 1/f shape of $S_{\dot{\eta}}(\omega)$ over a narrow frequency range surrounding f = 0 and extending principally over a few times the bandwidth of the pre-limited process (defined by the resonator in the oscillator circuit).

The peak factor, $(\text{p.f.})_\eta$, of $\eta(t)$ is readily shown to be $\sqrt{3}$ whether η is considered to range over $|\eta| \leq \pi$ or over $0 \leq \eta \leq 2\pi$.

The r-f spectrum of this type of signal has also been studied[2,6] extensively, and is illustrated in Fig.(3.9).

3.2.2 Sinewave Plus Gaussian Noise

The properties of the resultant of sinewave plus gaussian noise are well known.[7,3] Of principal interest here are the following facts. First, under conditions of very high ratio of the mean-square value of the sinewave to the mean-square value of the gaussian noise,

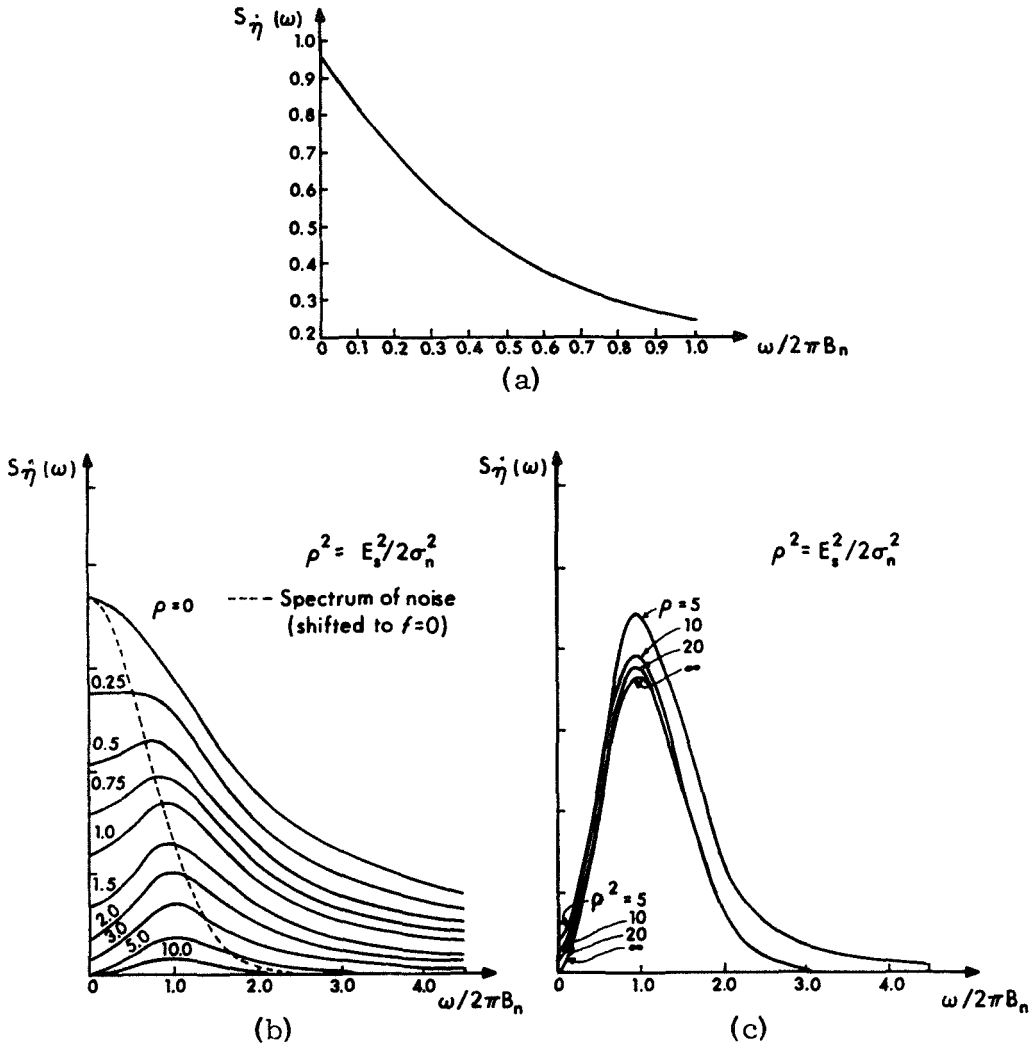

Fig. 3-8 Frequency Mean-Square Spectral Densities

(a) Mean-square spectral density of instantaneous frequency of narrowband gaussian noise with a rectangular input spectrum. (Adapted from Ref. 2)

(b) Mean-square spectral density of the frequency fluctuations, $\dot{\eta}(t)$, of the resultant of sine wave plus gaussian noise for different relative sine wave powers. (Adapted from Ref. 5)

(c) The same as (a) but for strong sine wave. (note change in scale). B_n is equal to half of the noise bandwidth before limiting. (Adapted from Ref. 5)

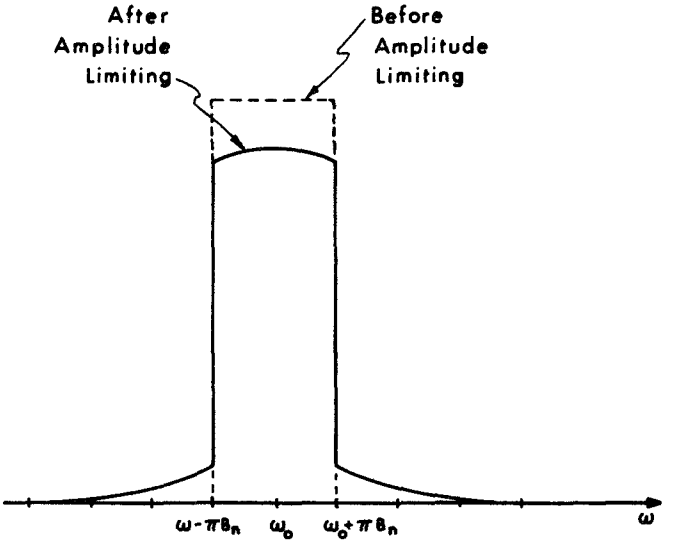

Fig. 3-9 Shape of r-f spectral density of bandpass limited gaussian noise for rectangular pre-limiting noise spectrum. (Adapted from Ref. 6)

the statistics of the phase and frequency fluctuations of the resultant signal are practically gaussian, and

the mean-square spectral density of the phase fluctuations practically differs only by a constant multiplier from the sum of the positive-frequency half and the negative-frequency half of the spectral density of the input noise, after both halves have been shifted down to $\omega = 0$.

If the amplitude-limited resultant of the sinewave plus the gaussian component is expressed as in Eq.(3.37) and $e_\ell(t)$ is used to represent the output of an oscillator, the "initial" phase ϕ of the sinewave must be assumed to be uniformly distributed over $-\pi \leq \phi \leq \pi$, since all values of ϕ are equally likely. Then the unconditional probability density function $p(\eta)$ of the total phase will therefore also be uniform. However, the conditional probability density $P(\eta/\phi)$ of the total phase $\eta(t)$, given the phase ϕ of the sinewave, is shown in Fig. 3-10 for different relative strengths of the sinewave.

The mean-square spectral density $S_{\dot{\eta}}(\omega)$ of $\dot{\eta}(t)$ is shown in Figs. 3-8(b) and (c) for the case in which the noise spectral density before limiting has a gaussian shape. The shape of the input noise spectrum about the center frequency ω_o is superimposed as a dashed curve on the low-frequency curves in Fig. 3-8(b).

It is interesting to note from Fig. 3-8(b) that outside of the frequency range defined for each solid curve by the intersection with the dashed curve, the solid curve under consideration shows a $1/\omega$ dropoff with ω, whereas inside of this range the solid curve either flattens out or

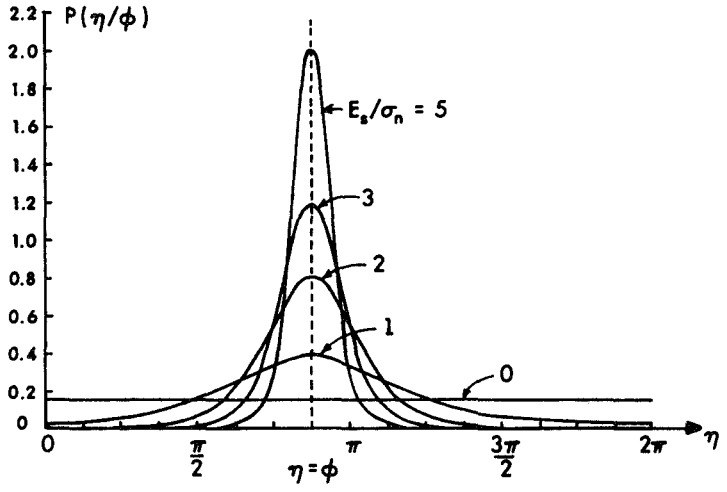

Fig. 3-10 Probability density distribution of the total phase η, conditioned on the phase ϕ of the sine wave, for various strengths of the sine wave. (Adapted from Ref. 5)

actually bends down as $\omega \to 0$. The mean-square spectral density will always have a nonzero value for $\omega = 0$ because of the impulsive component present in the instantaneous frequency of the resultant. (Only in the case of an infinitely strong sinewave is the spectral density strictly zero for zero frequency.) In view of the extremely narrow post-detection processing bandwidths normally encountered in applications of ultra-stable oscillations, no distinction between the impulsive component and the smooth component of the instantaneous frequency need be made.

For a nonzero spectral density for $\dot{\eta}(t)$ at $\omega = 0$, the mean-square spectral density of $\eta(t)$ is infinite at $\omega = 0$. This follows from the fact that the spectrum of the phase is obtained by dividing the frequency modulation spectrum by ω^2. The infinite spectral density for $\omega = 0$ results in an infinite rms value of the phase in any frequency region that includes $\omega = 0$. These properties are characteristic of a so-called random-walk process.

3.3 Combined Effect of Various Mechanisms

The various mechanisms of oscillator frequency instability modeled in the preceding sections may all be expected to be present in varying degrees in practical oscillators. The net effect of the combined mechanisms on the spectral density of the resulting instantaneous frequency fluctuations is illustrated in Fig. 3-11.

Of the various parts of the overall curve in Fig. 3-11, the part contained within $0 \leq \omega \leq 2\pi B_{osc}$ merits further explanation. By definition, B_{osc} is a measure of the "width" of the oscillation band, and is determined by the shape of the nominally "infinite-Q" selectivity curve for feedback around the oscillating loop. In an ultra-stable oscillator, B_{osc} will be an extremely small fraction of a cps, perhaps corresponding to periods of hours, days or longer. The scale of B_{osc} in Fig. 3-11 is greatly exaggerated in order to illustrate a guess about what the heretofore unmeasured part of the spectral density $S_{\dot{\phi}}(\omega)$ might look like. The suggested extensions of the curve of $S_{\dot{\phi}}(\omega)$ toward $\omega = 0$ are based on a model of the oscillation as a sinewave plus a band of gaussian noise arising from the various irregularities of the electronic phenomena associated with the loop components and shaped spectrally by the selectivity of the feedback under conditions of sustained oscillation. Two significant consequences of such a physically motivated model are:

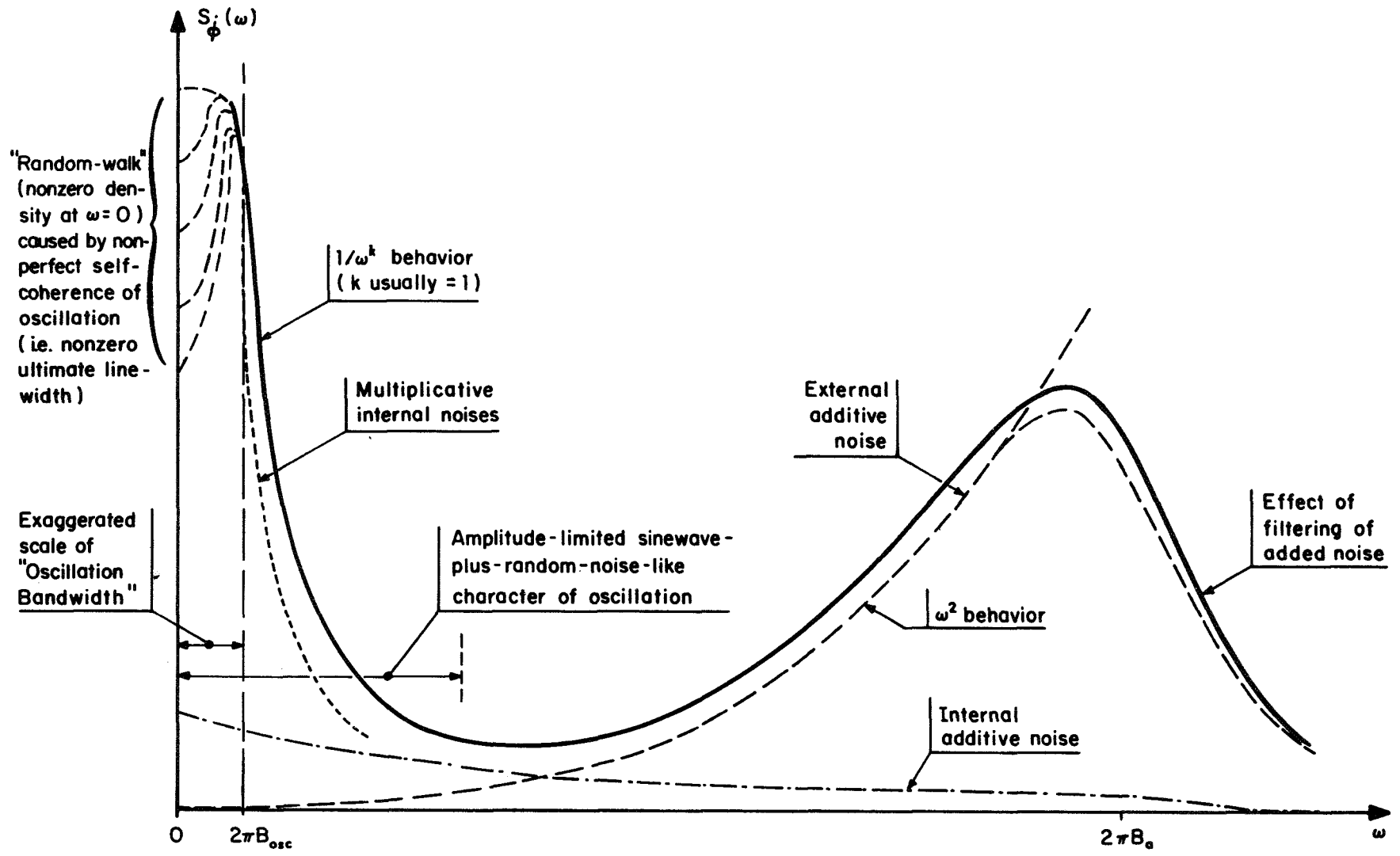

Fig. 3-11 Spectral density (solid curve) of instantaneous frequency fluctuations resulting from combination of various instability mechanisms. (Arbitrarily chosen relative levels for illustration only.)

a) The ultimate "linewidth" of the oscillation, as gauged by B_{osc}, is nonzero. This causes $S_{\phi}^{\cdot}(\omega)$ to have a nonzero value at $\omega = 0$, thus accounting for the inherent random-walk phenomenon in unlocked oscillators. $S_{\phi}^{\cdot}(0) = 0$ (and hence the random-walk effect is absent) only if the oscillation is perfectly self-coherent; i.e. its ultimate linewidth, B_{osc}, is zero.

b) The extremely slow fluctuations of the extremely narrow band of noise representing the ultimate oscillation signal are practically indistinguishable from the slow multiplicative noise phenomena (flicker noise, etc.). If one were to filter out the flicker and related noise effects within the loop by some high-pass filtering action, hyperbolic behavior of the spectral density of the frequency fluctuations immediately beyond $\omega = 2\pi B_{osc}$ should persist.

3.3.1 Instability Causes in Practical Circuits

Four principal types of exponent modulation effects produced by noise in oscillators may be distinguished. Any or all of them may be present in a particular oscillator. They may be grouped in one set of pairs by the type of noise and in another set of pairs by the region of the oscillator spectrum which is affected. The two types of noise are white noise (shot or thermal) and perturbation noise which is often referred to as flicker noise. The two regions of the oscillator's spectrum are within and without the resonator's effective half bandwidth or $f_m < f_o/2Q$ and $f_m > f_o/2Q$.

The effect produced by white noise for $f_m < f_o/2Q$ is a flat FM noise spectrum, the physical mechanism for which can be thought of as the voltage to frequency transference of the oscillator through its feedback network. The effect produced by white noise for $f_m > f_o/2Q$ is a flat ϕM noise spectrum. This assumes a single pole resonator, which is almost inariably the case for several reasons, among which is the maximization of the effective Q. This single pole applies an f^{-2} weighting to the FM noise spectral components outside the oscillator's half bandwidth, thus producing a flat ϕM spectrum. Alternatively one can assume that since the phase shift vs. frequency characteristic of the resonator is nearly zero outside its bandwidth then the voltage-to-frequency transference of the oscillator will be negligible in this region, therefore, flat white noise will appear directly as equally weighted flat AM and ϕM spectra. The magnitude of the flat ϕM mean square spectral density out of the oscillator itself is given by 2 FkT/P, where F is the actual operating noise figure of the active element of the oscillator and includes all noise components folded into the pass-band from harmonics of the fundamental frequency. The quantity kT is -174 dBm/Hz. P is the available power at the input of the active element of the oscillator. This flat ϕM spectrum extends out in frequency until it is attenuated by subsequent filtering.

The fundamental causes of flicker noise are not well understood. However, they appear to arise from random changes in physical parameters (such as charge density in a transitor) which manifest themselves as a rise in the effective noise figure of the device at low spectral frequencies. This rise in noise figure consistently obeys an f^{-1} law but it can vary widely in magnitude from one device to another. The effect produced by flicker noise for $f_m < f_o/2Q$ is an FM spectrum with an f^{-1} characteristic. This flicker FM spectrum has been shown to exist in crystal oscillators at spectral frequencies down to 3×10^{-7} Hz.[8] In high Q or low frequency oscillators where $f_o/2Q$ is small the flicker noise region of the oscillator's active elements can extend into the spectral frequency region beyond $f_o/2Q$ thereby giving rise to a flicker ϕM spectrum. Under these conditions, since flicker noise extends indefinitely to lower fluctuation frequencies, the oscillator will not exhibit a region of flat FM spectrum.

In addition to the exponent modulation effect produced by noise in an oscillator, there is an amplitude modulation noise spectrum. This appears as a flat AM spectrum for high spectral frequencies and is equal to the ϕM spectral level. For low fluctuation frequencies, the exponent modulation spectral density far exceeds the AM spectral density.

It is interesting to note that all physically realizable oscillators, with the single exception of the cesium beam resonator, have flicker FM noise spectra. The existence of flicker FM noise in the hydrogen maser is presently in question.[9]

There are two principal types of exponent modulation effects produced by noise in signal processing circuits other than oscillators. These are flat ϕM noise and flicker ϕM noise. There is also flat AM noise. These signal processing circuits can be divided into four classes:

1) amplifiers, linear and nonlinear;

2) frequency multipliers;

3) mixers (modulators or multipliers);

   a. frequency translators
   b. zero-frequency IF and phase detectors

4) frequency dividers.

The signal degradation that occurs in both linear and nonlinear amplifiers appears to be exclusively flat white noise. In linear amplifiers this noise is inversely proportional to the signal-to-noise ratio, and is equally divided between AM and ϕM. If N_o/S is the noise density to signal ratio then $2N_o/S$ is the ϕM spectral density. Where the amplifier noise figure and input signal power level are known, the ϕM density is given by $2FkT/P$. The AM modulation density is $200\ N_o/S = 200\ FkT/P$ percent. In nonlinear amplifiers the effects of saturation provide appreciable suppression of the AM components. According to FM theory, the capture effect reduces the angular modulation of the weaker signal by about 6 dB. Therefore, the ϕM spectral density added to a signal by a nonlinear amplifier should be about 6 dB lower than the equivalent linear amplifier, providing the operating noise figure is the same.

The signal degradation produced by well-designed frequency multipliers appears to be the same as in nonlinear amplifiers. However, the ϕM spectral density out of the multiplier is higher by N^2 (or $20 \log_{10} N$), where N is the multiplication factor. The AM spectral density is not increased by N. Therefore, the output of a multiplier for any appreciable N is exclusively ϕM. As the frequency multiplication factor N is increased, the performance of a frequency multiplier rapidly approaches the ideal, because the input phase noise $S\dot{\phi}(f_m)$ is scaled up by N^2 while the multiplier's internally generated noise is scaled up by smaller factors.

Thus, for large N, the scaled input noise predominates over all other sources of noise at the output of the multiplier.

A mixer, which for this discussion is defined as performing the function $e_o(t) = e_1(t) \cdot e_2(t)$, has two regions of interest. The first is where $e_o(t)$ does not exist at or near dc; that is, the mixer is used in the "heterodyne mode" as a frequency translator to a nonzero reference frequency as a frequency translator to a nonzero reference frequency. For this condition the performance is similar to that of an amplifier. Where one of the signals into the mixer is large compared with the other, the standard treatment for noise figure applies and the spectral density of ϕM noise is $2\,FkT/P$ where F is the overall cascade noise figure. There is an equal amount of AM noise. Where the amplitudes of the input signals are of comparable magnitude, the performance is that of a nonlinear amplifier. In both cases the noise spectra (ϕM and AM) will be flat. The second region of interest is where the output spectrum is centered about 0 Hz; that is, where the mixer is operated in the "homodyne" or zero-frequency IF condition, or as a phase detector. For this condition the inherent flicker noise of the diodes or other active elements produces the predominant spectral distribution. This effect is well known, for example, in doppler radars where the effective noise figure in the range of a few Hz to several kHz is several orders of magnitude above that which is obtained using the same mixer as a frequency converter to an IF in the MHz region. The same problem exists in phase detectors when very high signal-to-noise ratios are desired. Typical values for excess ϕM and AM noise density at 1 Hz are 20 to 70 dB above kT/P. This is indeed flicker noise and obeys an f^{-1} law.

The signal degradation produced by frequency dividers comes from two sources, flicker ϕM and broadband white noise (AM and ϕM). The spectral density level of angular modulation out of an ideal

frequency divider should be that of the input divided by N^2 where N is the division ratio. However, the two sources of degradation in the divider tend to mask this improvement for high purity signals. The broadband white noise level is usually determined by the last divider or the first amplifier following it. There appears to be a considerable amount of flicker ϕM noise generated by both regenerative analog dividers and digital dividers. In digital dividers it is clear that the time jitter of the state transition in response to a zero crossing of the input signal is due to the same internal device fluctuations that produce flicker noise in oscillators. The resulting flicker spectrum of ϕM confirms this. The same effect appears to occur in regenerative analog dividers. In general, as the division ratio N is increased, the noise generated within the divider will increase with respect to the scaled down input phase noise.

In all frequency synthesis systems some of these signal processing circuits are used. Since flicker ϕM noise is a principal component of both phase detectors and frequency dividers, it is not surprising that the output of many synthesizers has a large flicker ϕM noise level.

3.3.2 Equations of Composite Spectra

For many applications, especially those where the signal is either multiplied in frequency or used for timing or ranging purposes, the effects of angular modulation fluctuations far outweigh those of amplitude modulation. Therefore, we will concentrate on the former.

We have shown that the known physical processes produce only four distinct types of mean square spectral density, namely, flicker FM, flat FM flicker ϕM and flat ϕM (See Fig. 3-12 for illustrations). Therefore, it is possible to write equations for the spectral density of the FM and ϕM fluctuations as a function of spectral frequency f_m raised to the appropriate exponent. These are summarized in Table 2.

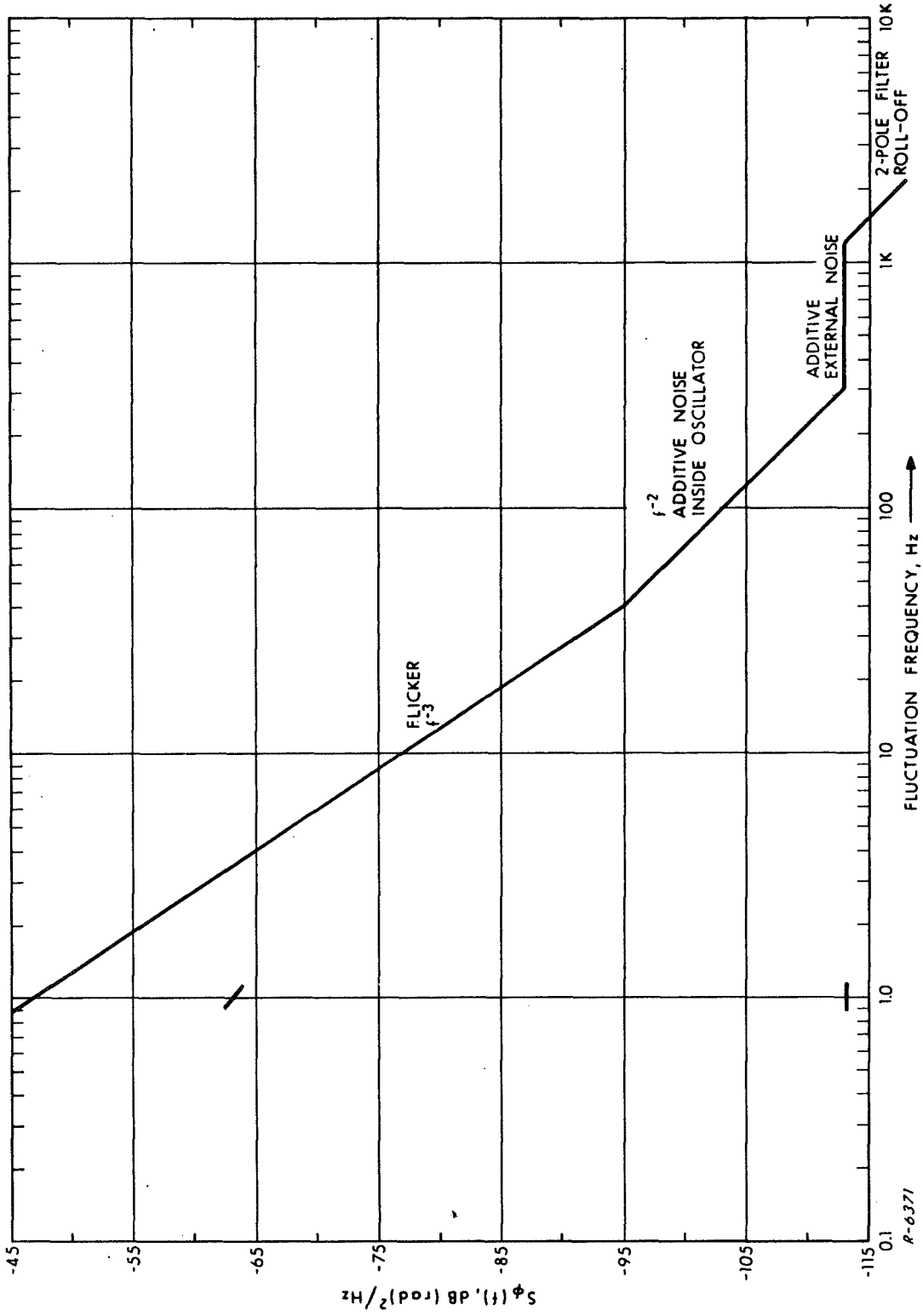

Fig. 3.12(a) Spectral Density Function of Pioneer E Local Oscillator Measured at the 114.6 MHz Output (Measured 7/31/69)

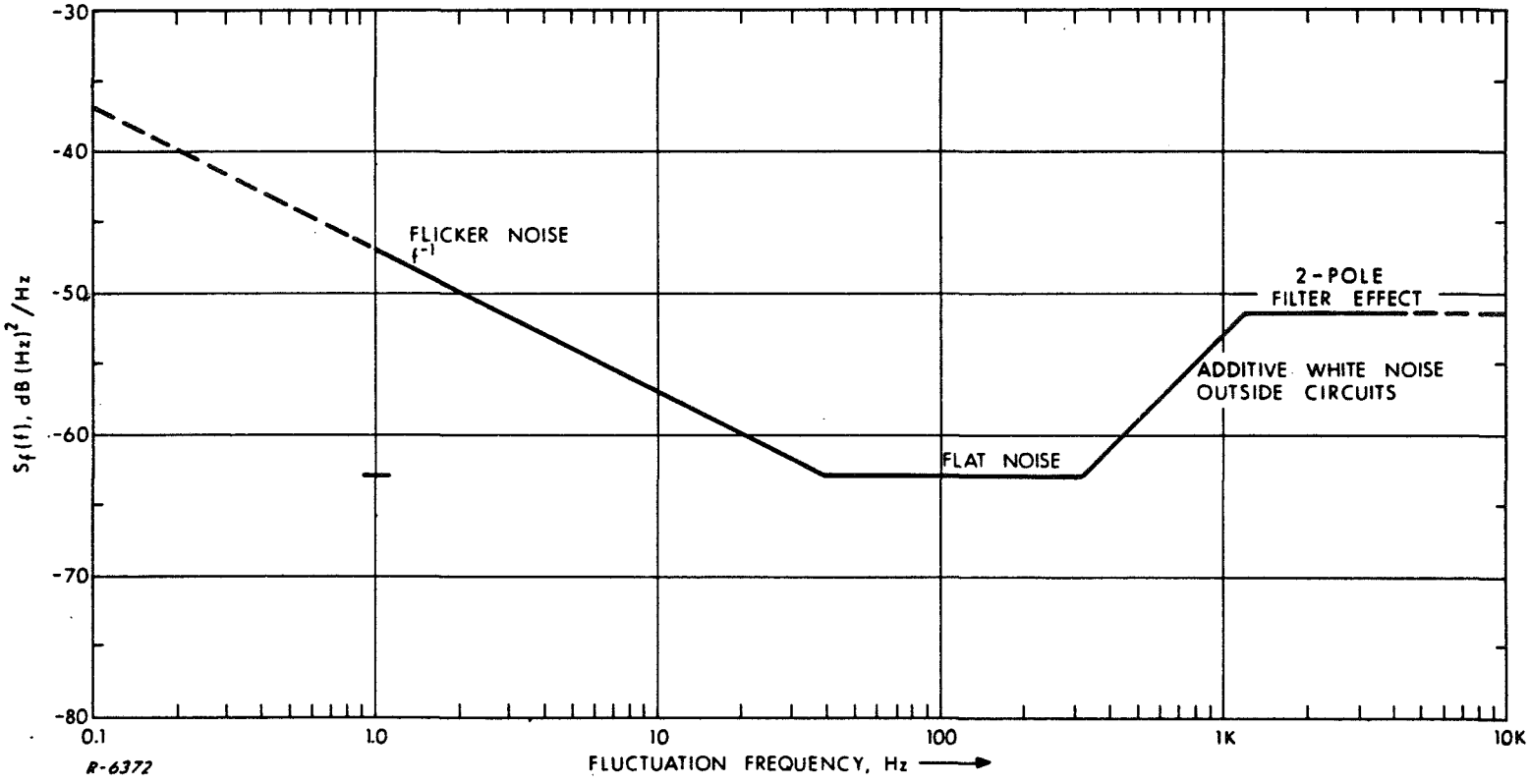

Fig. 3.12(b) Spectral Density of Frequency Fluctuations of Pioneer E Local Oscillator Measured at the 114.6 MHz Output (Measured 7/31/69)

Table 2

CLASSIFICATION AND DESCRIPTION OF OSCILLATOR INSTABILITY CHARACTERISTICS

FM Fluctuation Spectra				ϕM Fluctuation Spectra	
Description of Instability		Exponent of Power Spectral Density		Description of Instability	
		f_m^y	f_m^x		
Produced by the basic oscillator circuit and within its noise bandwidth.	Flicker Noise	-1	-3		
	Thermal Noise	0	-2		
		+1	-1	Flicker Noise	Produced in circuits external to the basic oscillator; such as amplifiers, multipliers, dividers, and mixers.
		+2	0	Thermal Noise	

Equations of Power Spectral Density

$$S_f(f_m) = Af_m^{-1} + B + Cf_m + Df_m^2 \ (\text{Hz})^2/\text{Hz} \qquad S_\phi(f_m) = Af_m^{-3} + Bf_m^{-2} + Cf_m^{-1} + D \ (\text{radians})^2/\text{Hz}$$

The constants A, B, C, and D have the same numerical values in both equations.

$$S_f(f_m) = Af_m^{-1} + B + Cf_m + Df_m^2 \qquad \text{Hz}^2/\text{Hz for FM}$$

and

$$S_\phi(f_m) = Af_m^{-3} + Bf_m^{-2} + Cf_m^{-1} + D \qquad \text{radians}^2/\text{Hz for PM}$$

where the coefficients implicitly include the dimensions:

$$A = \text{Hz}^3/\text{Hz}$$

$$B = \text{Hz}^2/\text{Hz}$$

$$C = \text{Hz}/\text{Hz}$$

$$D = 1/\text{Hz}$$

of mean square spectral density. As an example consider the case where the complete fluctuation spectrum consists of white phase noise which is passed through a rectangular filter of bandwidth $2F_1$. The mean square spectral density of this signal is then

$$S_\phi(f_m) = D \text{ rad}^2/\text{Hz from 0 to } F_1 \text{ Hz}$$

and the total rms phase deviation over the whole bandwidth is

$$\phi = \sqrt{DF_1} \text{ radians rms}$$

The numerical values of the corresponding coefficients in both equations are identical. This comes about because the conversion from S_ϕ to S_ϕ; adds a factor of $(2\pi)^2$ and the conversion from S_ϕ; to S_f subtracts the same factor. Alternatively, from FM theory: $\Delta f/f_m$ = Mod index = $\Delta\phi$ in radians, thus $S_f/f_m^2 = S_\phi$ in Hz^{-1} and $\text{radians}^2/\text{Hz}$, respectively. This correspondence of numerical values greatly enhances the ease of conversion from FM to ϕM spectral density and vice versa.

These equations account for the random or aperiodic fluctuations of frequency sources. To the spectral density thus given must be added any periodic modulations such as spectral lines produced by power frequency

components or spectral bulges (broadened lines) produced by external causes such as vibration or oven cycling. These can be accounted for separately from the purely random fluctuations.

If the frequency source has appreciable bandwidth limiting (such as an output filter) in the spectral frequency range of interest then the effect of this filter should be included. Thus a single-pole filter will add an additional f^{-2} slope to the spectrum of the highest spectral frequency components and a 2-pole filter will add f^{-4} slope. These additional slopes can be handled by using a subscript with the appropriate coefficient in the spectral density equation along with a notation of the applicable frequency range. For example, a spectrum whose highest spectral frequency range consists of flat ϕM noise is passed through a single-pole filter which in its skirt selectivity region produces an f^{-2} slope to the ϕM spectrum. This slope is associated with the coefficient B in the spectral density equation which is designated B_2 to differentiate it from the flat FM region B_1 at low spectral frequencies. Similarly, if a 2-pole filter is used in the same example the resulting ϕM spectral density with f^{-4} slope can be specified by $S_\phi(f_m) = Ef_m^{-4}$.

The numerical values of the coefficients in the spectral density equations thus provide a complete description of the random fluctuations in the frequency domain.

3.3.3 The Spectral Representation

There are three areas where the spectral representation of the fluctuations of a precision frequency source provides a beneficial approach to the understanding of the situation. These are 1) the area of theory and analyses, 2) the area of measurements, and 3) the area of specifications. In each of these areas there are significant differences between the approaches used in the time domain and those used in the frequency domain, and it will be useful to examine them.

Theory and Analysis

The spectral representations of the fluctuations of a frequency source are a complete description of all its properties except for drift which is readily handled separately. Because the spectrum describes the performance throughout the frequency domain, it allows a designer or user to apply standard circuit and system analysis techniques to ascertain the effects of various areas of the spectrum on the overall system performance. In contrast, the fractional frequency instability description does not allow these techniques to be used directly without first converting to the spectrum. This conversion process is subject to large errors and can be very ambiguous. For example, Allen[10] indicates that it is unlikely that one could distinguish between flat ϕM and flicker ϕM spectral shapes by the use of the usual time domain measurement techniques. In making frequency domain measurements, however, the ϕM fluctuations are obtained directly and their spectral analysis provides a clear meausre of the spectral shape and magnitude.

For applications involving frequency multiplication the spectral representation is most useful, first, because the spectral density level is directly increased by the square of the multiplication ratio; second, because the effect of bandlimiting filters on the total rms deviation must be assessed in the frequency domain, and, third, because the effects of spectral spreading or carrier reduction at very high multiplication ratios are readily apparent.

The effects of discrete spectral components arising from periodic modulations such as the power line frequency, gating periods, oven cycles and similar events are easily managed in the frequency domain but are difficult to handle in the time domain.

Measurements

When the spectral representation of the fluctuations of a frequency source is used, it is more convenient to perform measurements in the same information system, i.e., to extract directly the frequency fluctuations

and perform a spectral analysis. Measurements made in this manner provide a direct measure of the fundamental parameters of interest, namely, the spectrum, as opposed to the whole integrated and weighted fluctuation obtained by time domain measurements. In making spectral measurements any additive perturbations are readily apparent. Measurements of $S_f(f_m)$ have been made down to f_m of 3×10^{-7} Hz.[8]

Specifications

In addition to the reasons previously presented the user, the spectral representation of the frequency fluctuations is a powerful tool for the following reasons:

1) Direct comparisons of various types of oscillators can be made not only for those at the same frequency but for any frequency as the frequency ratio simply scales the whole spectral level.

2) Spectral shapes allow easier identification of problem areas.

3) The spectral representation has the capability of handling clearly all forms of fluctuations both periodic and random where the fractional frequency fluctuation form cannot distinguish between them.

4) Spectral regions can be covered from below 10^{-6} to above 10^{6} Hz.

Because the spectral representation covers such a wide frequency range, the areas of short, medium, and long term stability need definition in a manner which will encompass the needs of as many different applications as possible. Since these applications extend from deep space probes to millimeter wave communications and radar, the likelihood of satisfying all users is not great. However, with this in mind, we would like to suggest the following definitions:

Short Term	10^{6} to 10^{2} Hz,	1 μsec to 10 msec
Medium Term	10^{2} to 10^{-2} Hz,	10 msec to 100 sec
Long Term	10^{-2} to 10^{-6} Hz,	100 sec to 10^{6} sec

4. CHARACTERISTICS OF TROPOSPHERIC LINE-OF-SIGHT PROPAGATION EFFECTS

The various forms of transmission encountered in tropospheric, line-of-sight links are:

(a) Direct line-of-sight path, perturbed in amplitude by the time-and space-variant absorption properties and in phase by the corresponding time-and space-variant refractive fluctuations of the medium along the transmission path.

(b) Direct path plus reflection or scatter from ground surfaces or other neighboring obstacles or from inversion layers, and

(c) Atmospheric multipath, or splitting of the direct path into several paths of nearly the same properties by atmospheric stratifications along the direct path.

The characterization of each of these forms of propagation is essential to the determination of corresponding models that will ensure realistic laboratory synthesis, or simulation, of their effects on a transmitted signal.

In the present section, the essential characteristics of the above forms of propagation will be summarized on the basis of available data from field tests, as a preamble to the formulation of the mathematical representations needed for realistic laboratory synthesis of the effects of the propagation phenomena.

4.1 Characteristics of a Direct Line-of-Sight Path

The fluctuations in the densities of the constituents of the atmosphere (as functions of time and space variables) along a signal propagation path, give rise to fluctuating absorption and refraction effects. The absorption fluctuations are manifested in the amplitude of the signal, the variations of the refractive index cause ray bending, and hence path-length and angle of arrival changes from those of a straight line path

connecting the transmitting and receiving points. The time-variant nature of even the more gradual changes in the absorption characteristics and the refractive index gives rise to a scintillation, or "twinkling", of the received signal amplitude and phase, respectively. The scintillation is significant for microwaves, particularly insofar as it contributes the path amplitude and phase variabilities that figure significantly in the interference among a multiplicity of simultaneous paths.

Of principal interest in the analysis of multipath effects and of phase and frequency sensitive signal operations is the effect of the variability of the refractive index in the portion of the atmosphere traversed by a single propagation path upon the phase stability of the signal. The basic relationship between propagated signal phase along a single path and the refractive properties of the propagation medium stems from the fact that if $n(s,t)$ denotes the index of refraction along a path of length L, the transit time can be expressed as

$$\tau(t) = \frac{1}{c} \int_0^L n(s,t)\, ds$$

where s denotes distance along the path, the notation $n(s,t)$ expresses the dependence of the index of refraction upon both position along the path and time, and the variation of $n(s,t)$ with t is assumed to be sufficiently slow that during the transit time interval $n(s,t)$ does not change significantly. If we define

$$m_n(t) = \frac{1}{L} \int_0^L n(s,t)\, ds$$

then

$$\tau(t) = m_n(t)\, L/c$$

where $m_n(t)$ is recognized as the "spatial" mean value of $n(s,t)$ for the path of interest during the transit time interval centered about the time instant t. The transit-time phase shift of a sinusoidal carrier of wavelength λ can now be expressed as

$$\phi(t) = (L/\lambda)\, m_n(t) \qquad (4.1)$$

The received sinusoidal carrier will therefore portray random phase fluctuations, $\phi(t)$, with time if the atmospheric conditions along the path cause the spatial-mean index of refraction $m_n(t)$ to fluctuate randomly in time. The intensity of the phase fluctuations naturally varies inversely with the wavelength of the propagated carrier for a given path defined by a nominal length L and a time-variant spatial-mean index of refraction, $m_n(t)$.

The characteristics of the phase (as well as amplitude) instability imparted to a propagated radio signal can be determined from abundantly available data obtained experimentally on paths simulating the conditions for various applications and ranging from 700 m to 100 km. The best available empirical data are from experiments conducted, starting in 1954, by the Tropospheric Telecommunication Laboratory of the ESSA Institute for Telecommunication Sciences, over a variety of path configurations and climates, and for purposes including geodetic distance measurement, baseline synchronization and transmissions between ground-stations, aircraft and spacecraft. Two types of basic measurements were performed. In one set of measurements, two signals at slightly different frequencies are transmitted over the path in opposite directions. At each terminal, the signal received from the opposite terminal is heterodyned by means of a stable LO so that the resulting IF signals have identical frequencies except for the phase variations due to the propagation delay. One of the IF signals is returned to the recording terminal by an auxiliary microwave link and compared in phase with the other IF. In another set of measurements, attention is directed to the difference in phase shift over two propagation paths that are common at one terminal but diverge to separate distant terminals--as in an interferometer configuration. Fluctuations in this phase difference for the same carrier originating from the terminal common to the two divergent paths yields the "angle noise" introduced by the uncorrelated fluctuations in the refractive properties of the medium over the

separate paths. The conditions of the experiments have included the following ranges of parameters, studies in various combinations:

1. Radio frequency: 100 MHz to 35 GHz
2. Climate: Continental temperate to maritime tropical
3. Season: Throughout annual cycle
4. Path lengths: 700 m to 100 km
5. Path elevation angle: Horizontal to 9^{o}
6. Upper fluctuation response frequency: About 20 Hz
7. Continuous observation period: 2 to 15 days
8. Antenna sizes: 50 cm to 300 cm
9. Surface-reflection multipath: With and without

In general, the phase records show relatively large diurnal variations amounting to 20 to 30 parts per million of the nominal phase shift over the path. Superposed on these slow variations are smaller, more rapid fluctuations whose intensity often also shows a diurnal change, as illustrated in Fig. 4-1.

In a typical experiment, both λ and L were maintained as nearly constant as is feasible by using stable oscillators and antenna mounts. Thus, the observed variations in $\phi(t)$ can be assumed to reflect the time variations in the spatial average, $m_n(t)$, of the refractive index. Spectra of fluctuations are available from measurements of refractive index over a variety of conditions. Figure 4-2 is an example in which the rapid index fluctuations were obtained from microwave refractometers and the results were combined with values computed from meteorological observations over an 8-year period. The phase spectral density was measured directly over a 15-km path, but the length of sample limited the analysis to frequencies greater than 10^{-3} Hz.

In an experiment of special interest in bringing out the relative scales of dependence of amplitude and phase fluctuations upon the operating frequency for a single atmospheric path (with no attendant reflected or scattered paths), the terminals (Upolu Point, Hawaii, at an elevation of 30 m and Haleakala, Hawaii, at an elevation of 3040 m) had a geographical separation of about 65 miles. Signals at 9.6 (X-band) and

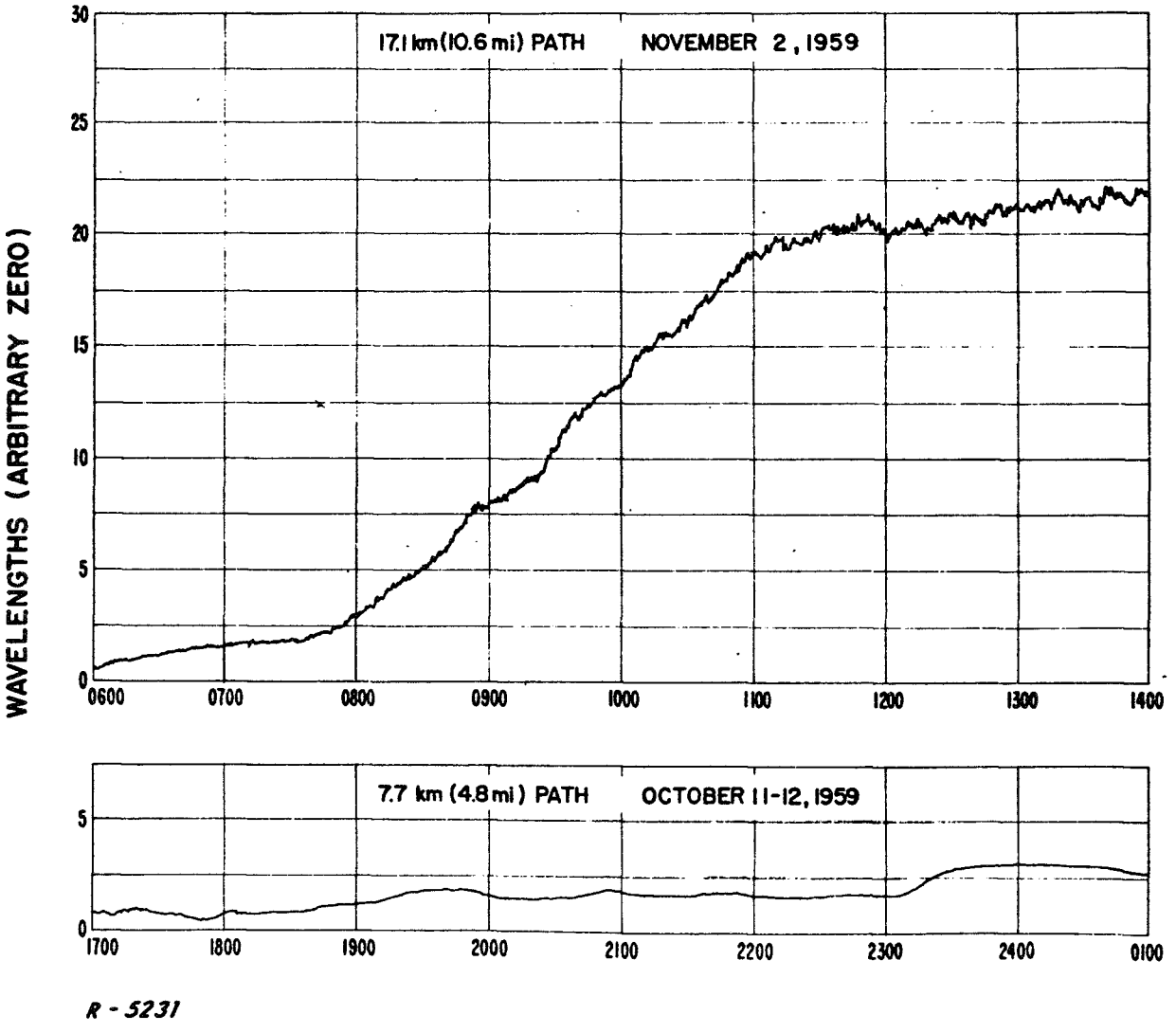

Fig. 4-1 Examples of Apparent Path Length Variations Observed at Cape Canaveral, Florida

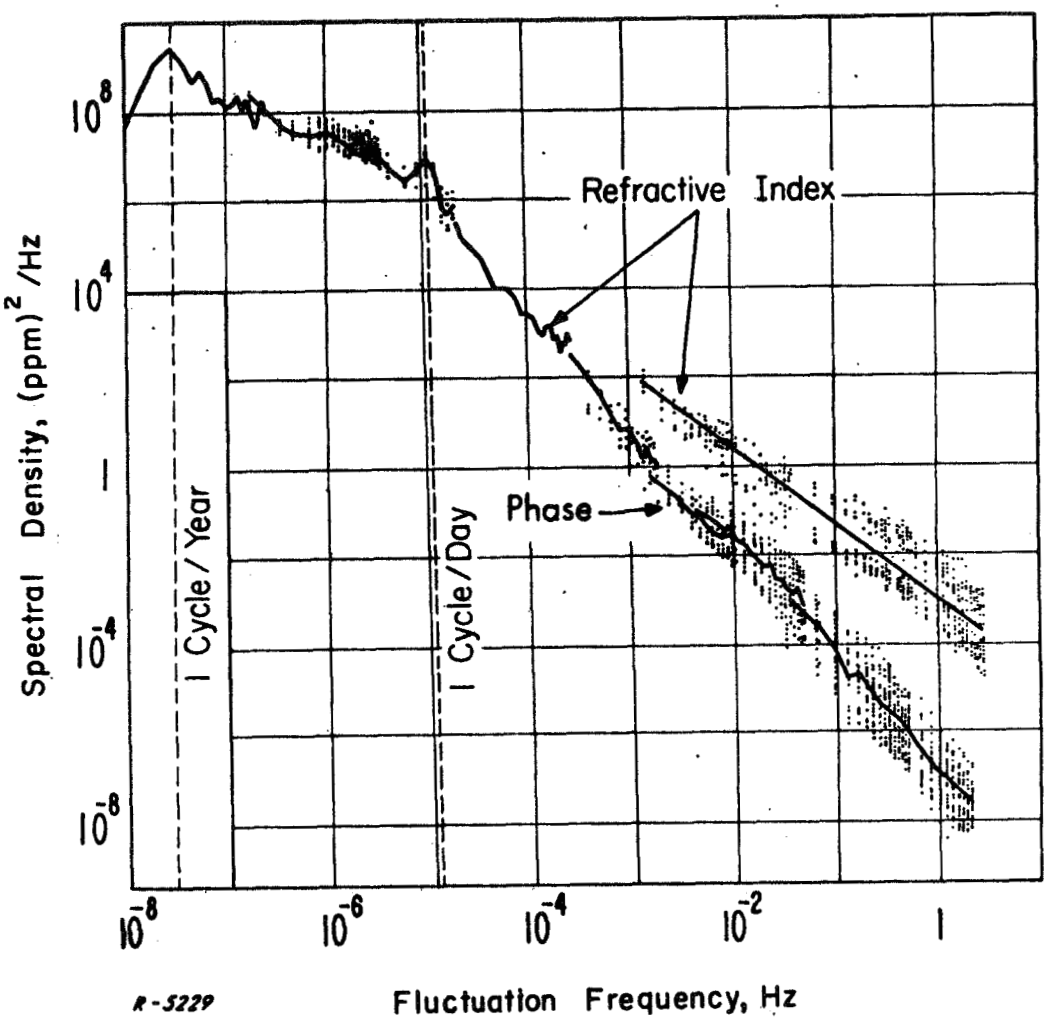

Fig. 4-2 Measured Spectral Density Functions of the Fluctuations in Refractive Index and Propagation Phase Shift of a Test Carrier. Log-log coordinates are used because of the wide range of densities and because of the form predicted by theory for certain spectral regions.

34.52 (K-band) were propagated simultaneously using the following antenna characteristics:

Antenna separation at Upolu Point	2.4 m
Antenna separation at Haleakala	3 m

(The relative positions of the X-band and K-band antennas at Haleakala were the reverse of those at Upolu Point, so that the two propagation paths intersected at approximately mid-path.)

Antenna type: Parabolic

Antenna polarization: Horizontal

Antenna diameters and computed half-power beamwidths:

At Haleakala, X-band	3.05 m, 0.7°
At Haleakala, K-band	1.83 m, 0.3°
At Upolu Point, X-band	1.83 m, 1.2°
At Upolu Point, K-band	1.83 m, 0.3°

Sample records of phase and amplitude fluctuations obtained in the above experiment are shown in Figs. 4-3 and 4-4. Examination of these figures shows that the "minute-to-minute" fluctuations appear to be quite well correlated at the two frequencies, but the more rapid fluctuations are not. The "minute-to-minute" (and longer-term) variations in X-band and K-band phase-of-arrival are illustrated in Fig. 4-5. These variations are expressed in terms of parts per million of the total phase change of a signal traversing the propagation path (which is also the parts per million variation in radio path length) and in terms of centimeters of radio path length. The plotted averages for the two variables are nearly equal in magnitude (standard deviations of 4.6 and 4.5 ppm for X and K band, respectively) and are very highly correlated (correlation coefficient: 0.9991).

Typical examplies of the power spectral densities of phase variations are shown in Figs.4-6 and 4-7. In each figure, the ordinate has been normalized with respect to the dependence upon wavelength. These figures show that from 0.01 Hz to 5 Hz the phase fluctuation

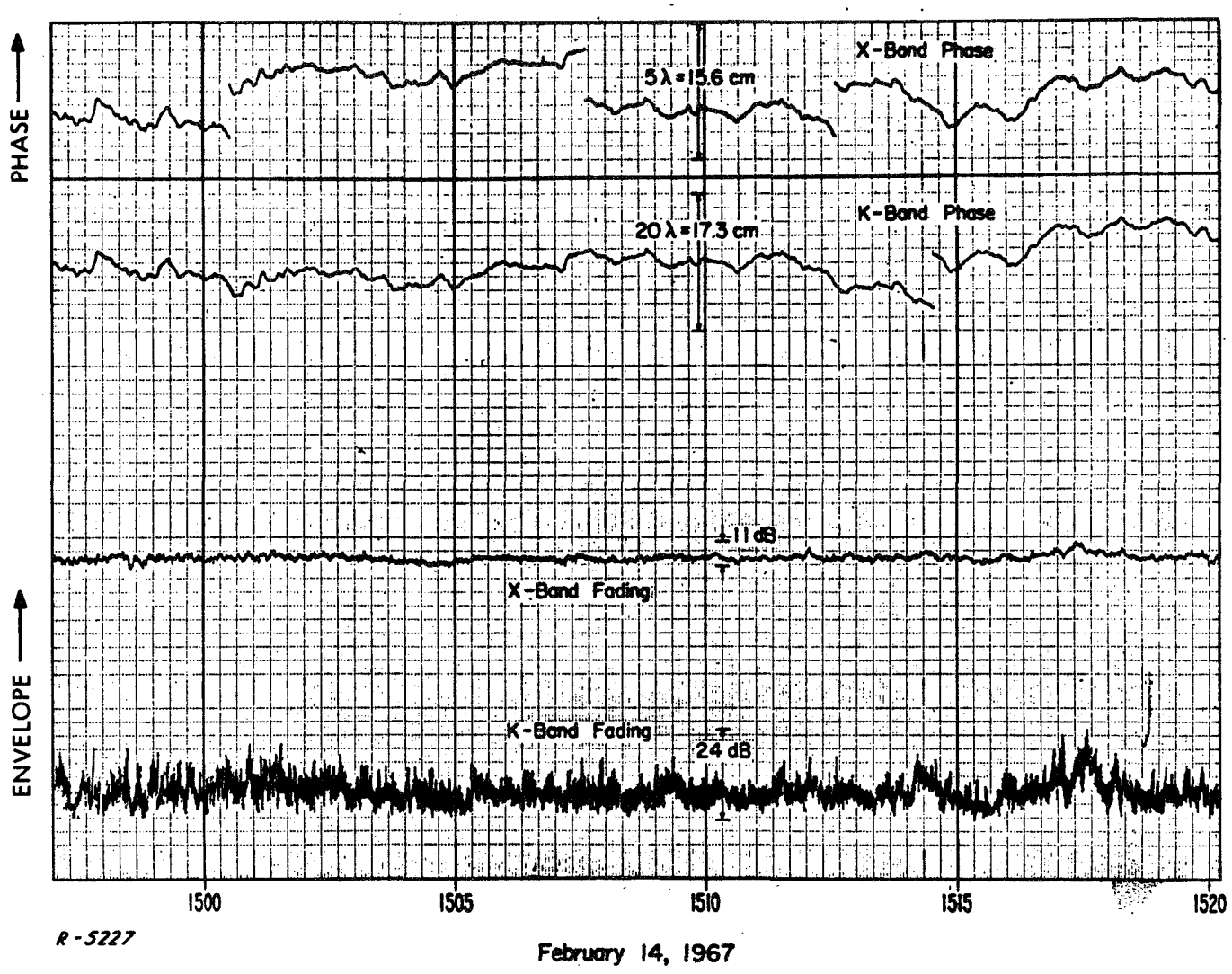

Fig. 4-3 Short-Term Phase and Envelope Fluctuations at X and K-bands 1457-1520, February 14, 1967. The scales of the two phase variables were adjusted to provide approximately equal sensitivity in terms of variations in the apparent or radio path length. (The frequent breaks in the phase records are scale shifts made automatically to keep the trace within bounds. The computer subsequently sensed these discontinuities and reconstructed the original variables.)

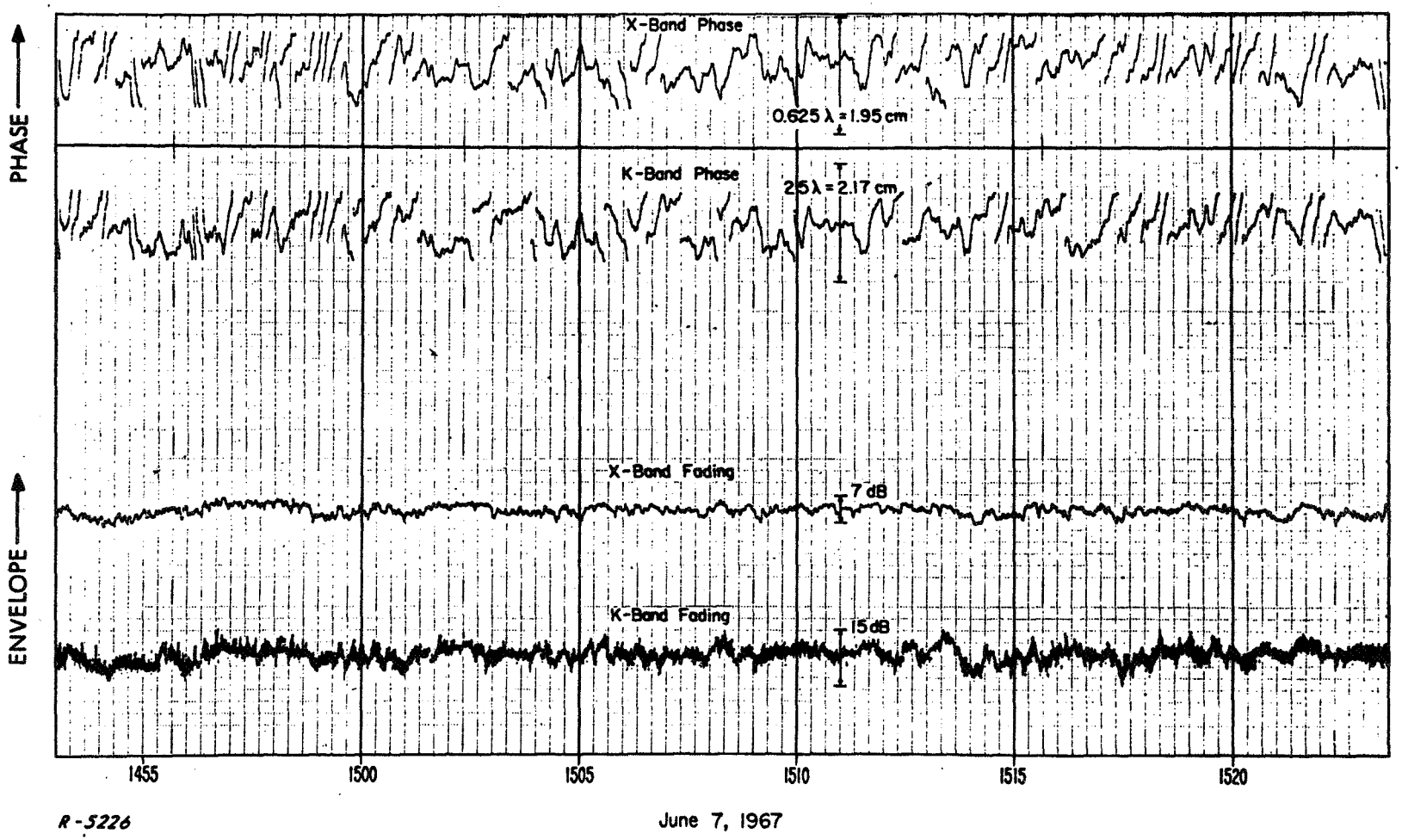

Fig. 4-4 Short-Term Phase and Envelope Fluctuations at X and K-Bands 1453-1523, June 7, 1967. (See Caption of Fig. 4-3)

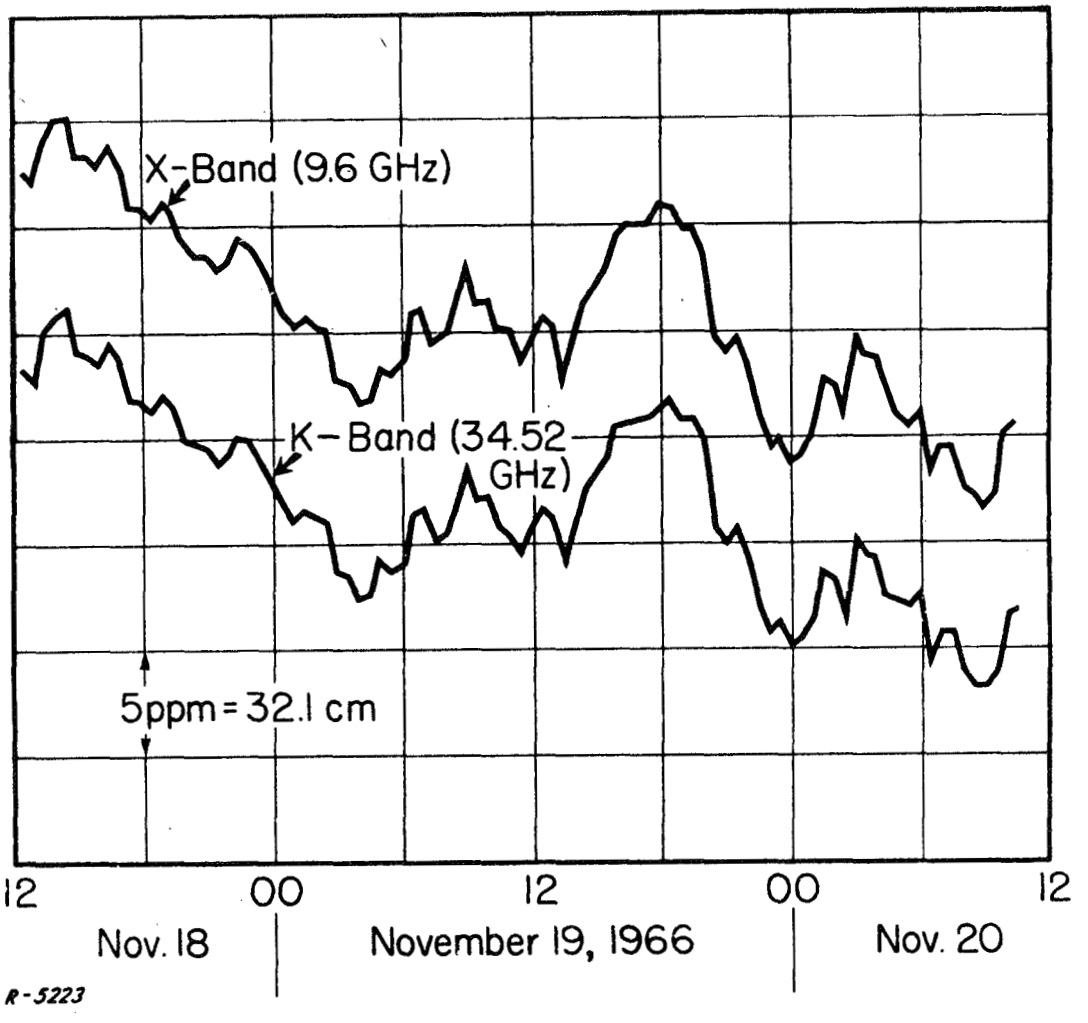

Fig. 4-5 Long-Term (Minute-to-Minute, or Longer) Average Fluctuations of X-Band and K-Band Phase. November 18-20, 1966.

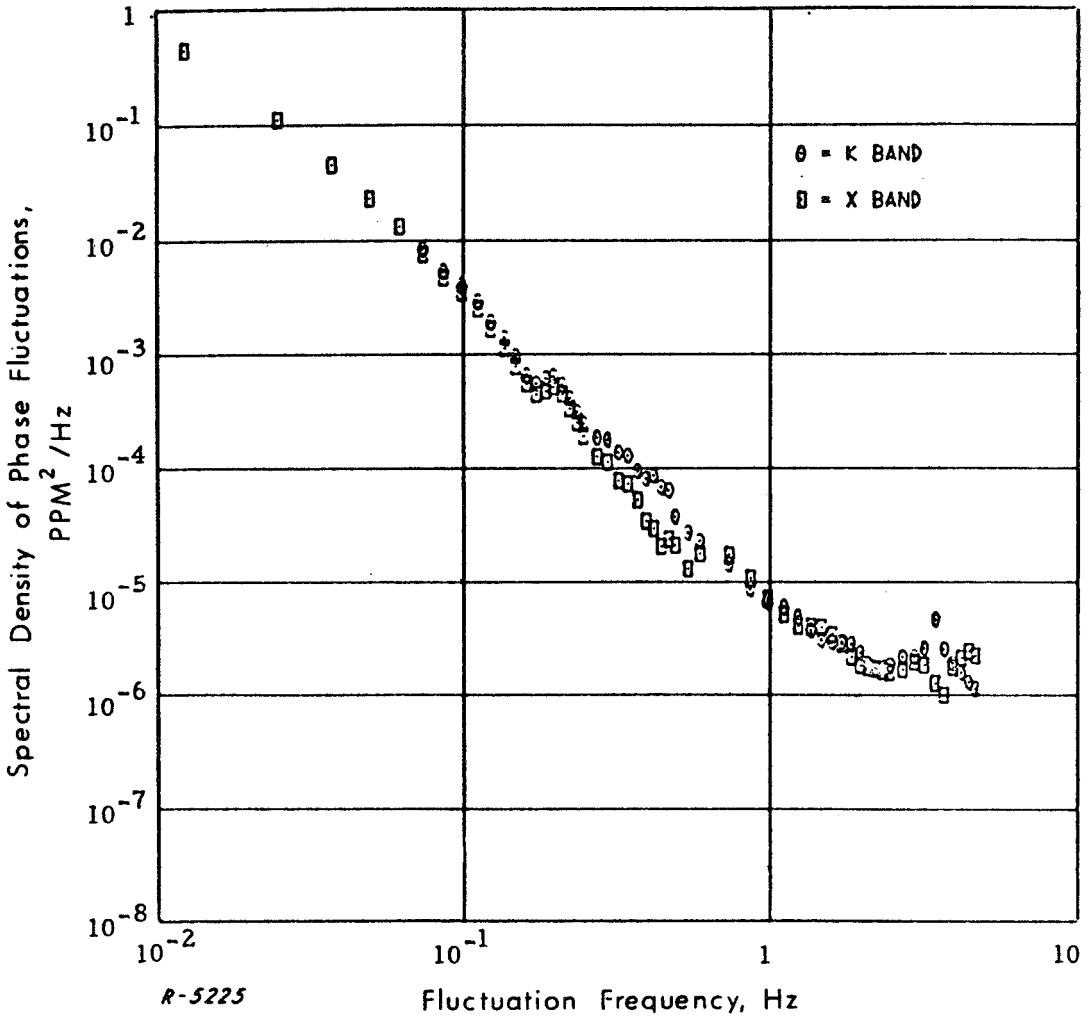

Fig. 4-6 Spectral Density of Phase Fluctuations at X and K-Bands 1457-1520, February 14, 1967.

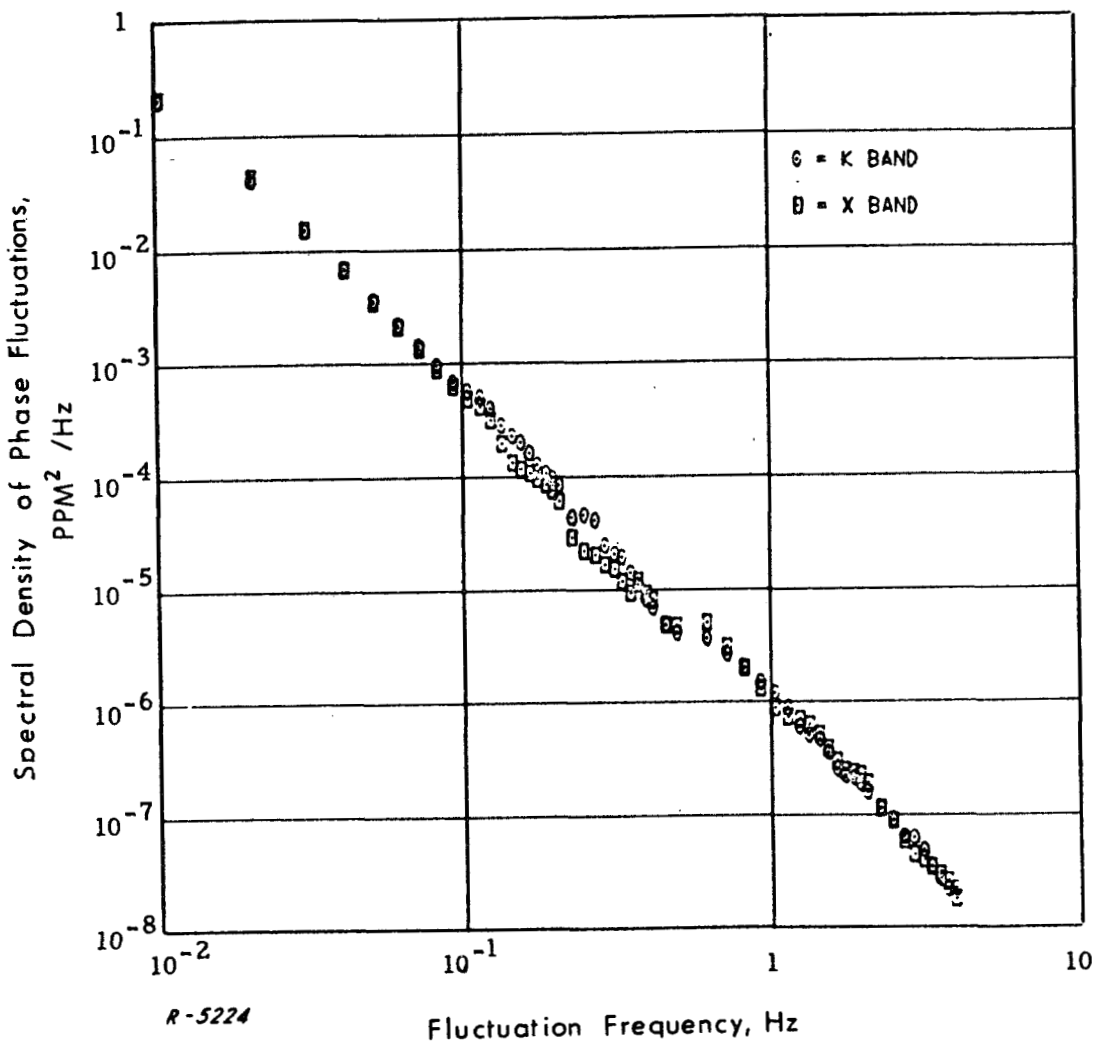

Fig. 4-7 Spectral Density of Phase Fluctuations at X and K-Bands 1453-1523, June 7, 1967.

spectral densities at X-band and K-band are nearly identical when each of the phase variables is expressed in terms of parts per million, and follow an $f^{-8/3}$ dependence on fluctuation frequency. The departure of the February spectra from the $f^{-8/3}$ dependence above about 1 Hz is the result of the relatively compressed recording scale used then and during most of the June run. The flattening of the spectrum in that region is caused by noise contributed by the recording and analysis processes. Hence, except for the June sample in Fig. 4-7, the analysis of the phase data discussed here does not extend beyond about 1 Hz.

The spectral density functions of the envelope fluctuations for the same periods are shown in Figs. 4-8 and 4-9. Note the following characteristics that are more or less typical of all the data examined:

(a) The general shape of the spectral density function of the envelope fluctuations is essentially the same at K-band as at X-band.

(b) The envelope-fluctuation spectral densities at K-band are consistently higher than at X-band, with wider separation of the two spectra at the higher fluctuation frequencies.

For a quantitative comparison of the spectra at X and K-band, the ratios of K-band to X-band spectral densities are plotted versus fluctuation frequency in Fig. 4-10 for February, and in Fig. 4-11 for June. These graphs show clearly the near equality between the X-band and K-band phase spectral densities at frequencies below 0.1 Hz. The average ratio of spectral densities over the decade from 0.1 to 1 Hz is also approximately equal to one.

The K-band to X-band spectral ratios for envelope fluctuations have an average of about 3 or 4 in the decade from 0.01 to 0.1 Hz, and rise sharply in the interval 0.2 to 0.4 Hz to an average of about 20. From Fig. 4-11 it can be seen that the latter ratio holds fairly constant from 0.4 to 5 Hz.

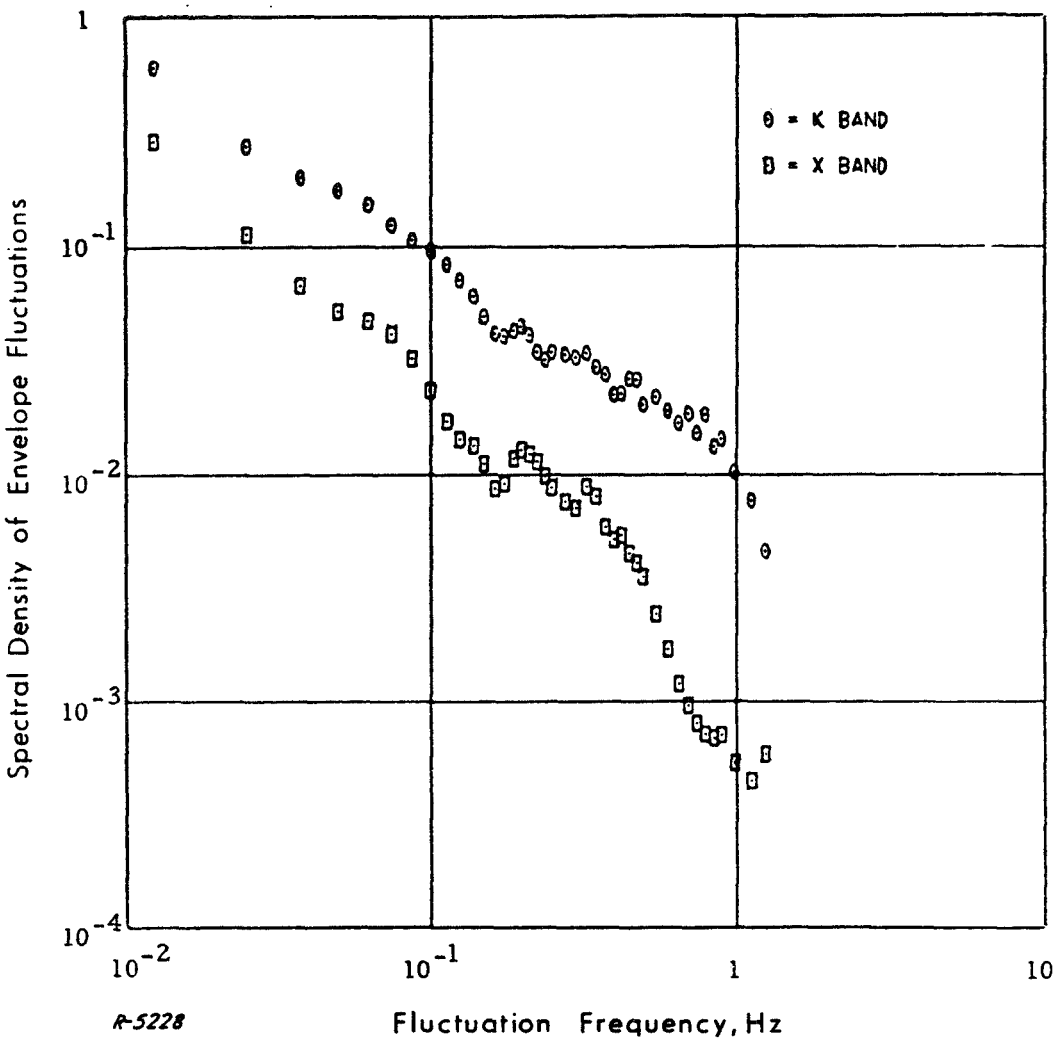

Fig. 4-8 Spectral Density Functions of Envelope Fluctuations at X and K-Bands 1457-1520, February 14, 1967. To permit comparison of spectral density levels at X-band and K-band, the data were expressed in terms of the ratio of the instantaneous voltage to the mean voltage. Hence the spectral density units are (voltage ratio)^2/Hz.

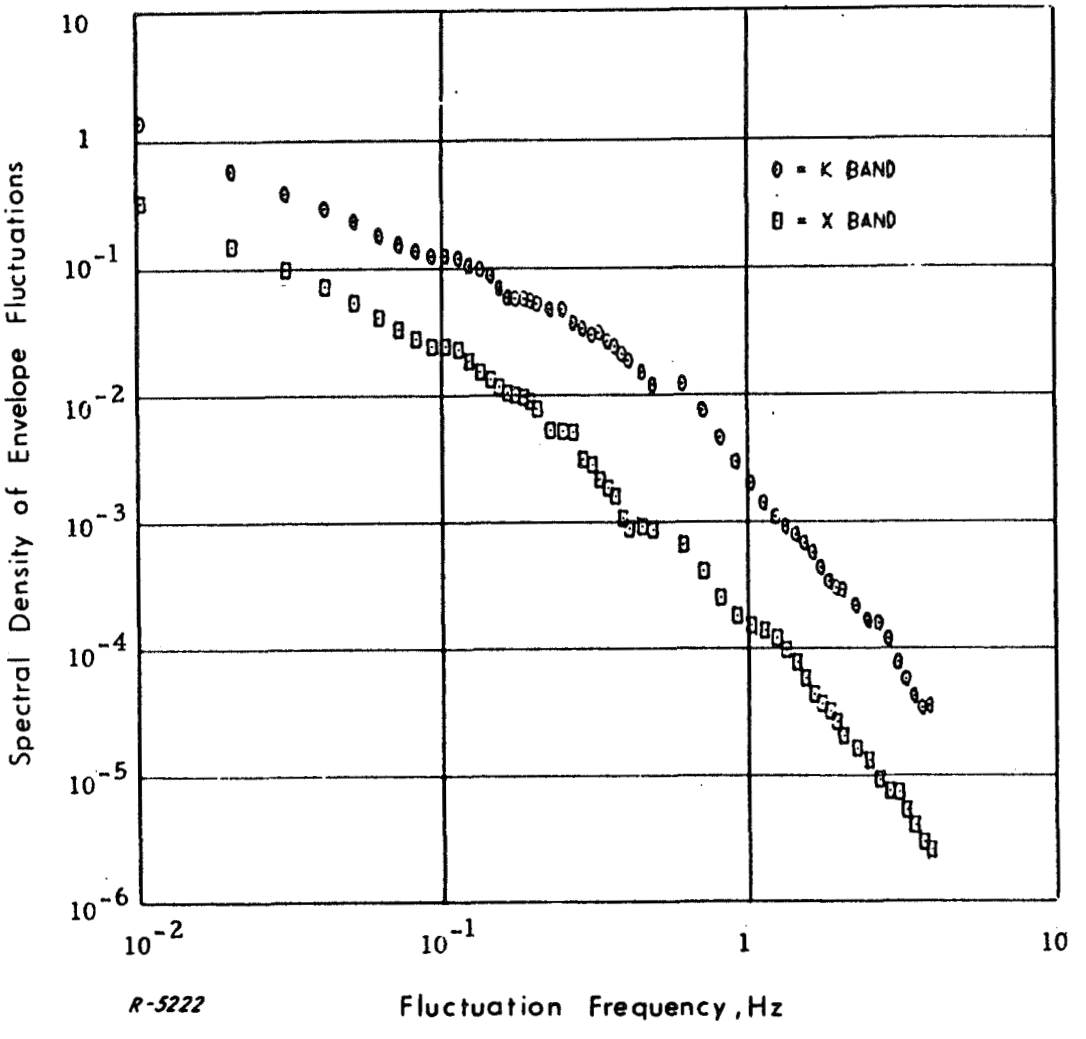

Fig. 4-9 Spectral Density Functions of Envelope Fluctuations at X and K-Bands 1452-1523, June 7, 1967 (see Caption of Fig. 4-8).

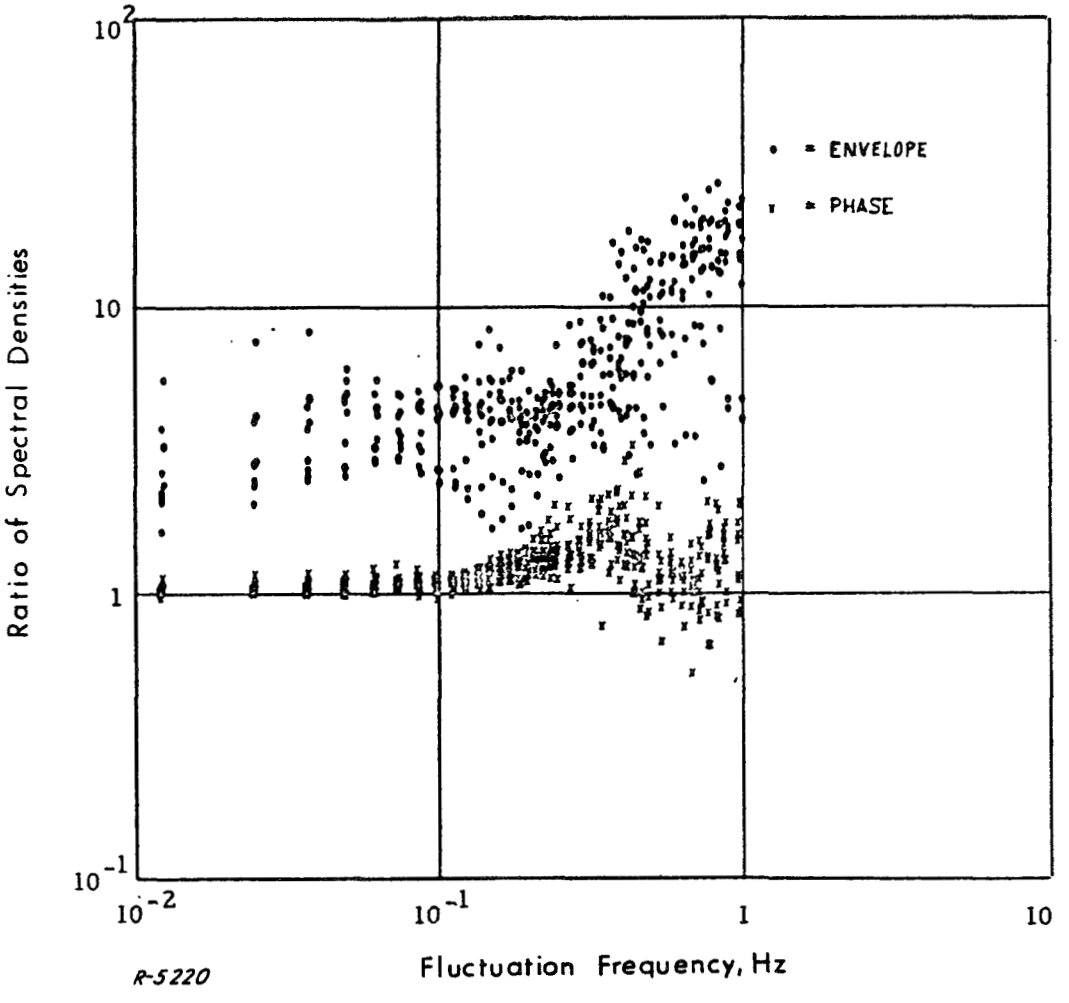

Fig. 4-10 K to X-Band Spectral Density Ratios, February 14-15, 1967.

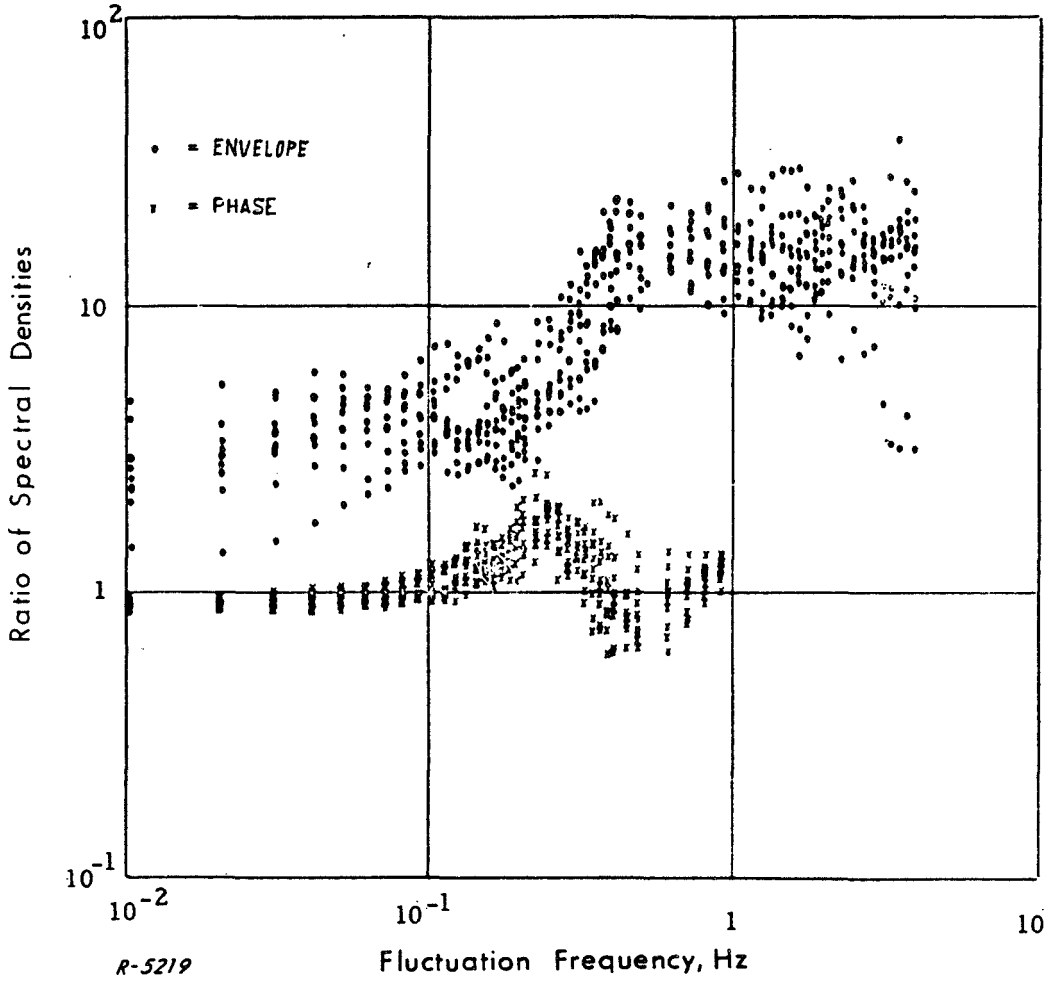

Fig. 4-11 K to X-Band Spectral Density Ratios, June 6-9, 1967.

The variations of phase difference over paths extending from a common transmitting antenna to two separated receiving antennas differ qualitatively from those of the individual paths. The two spaced receiving antennas act as a spatial filter to reduce the contributions of structural features large compared with the path separation. Thus, for a spacing of 5 m, index variations that are highly correlated over distances of 5 m or more would tend to affect each path in the same way and thus have little effect on the phase difference. Features that are small compared with the path separation would tend to produce uncorrelated phase variations in the two paths with the result that the variance of the difference would be the sum of the separate variances.

The spatially more extensive variabilities in the traversed atmosphere tend to be associated with the lower-frequency fluctuation components. Thus, the spectral density of the phase difference fluctuations is expected to be lower than that of the phase fluctuations on a single path for the lower fluctuation frequencies, and to approach asymptotically the sum of the separate path variances at the higher frequencies. This behavior is partially illustrated in Fig. 4-12, where spectral densities of the phase difference for 5- and 790-m separations are compared with the phase spectral density for one of the two paths.

For the 790-m case, the refractive index fluctuations that account for the frequency range shown were apparently uncorrelated between the paths. If the spectral densities extended to lower frequencies, we would expect the upper curve to cross over and drop below the single-path spectral density. Since the two paths for the 790-m case were statistically similar, the spectral density function of their phase difference can be expected, and actually turns out, to be about twice the value for the single path.

In the interpretation of the physical causes of the amplitude and phase fluctuations discussed above, it is important to note that scattering of microwaves by gaseous particles and water droplets in the atmosphere

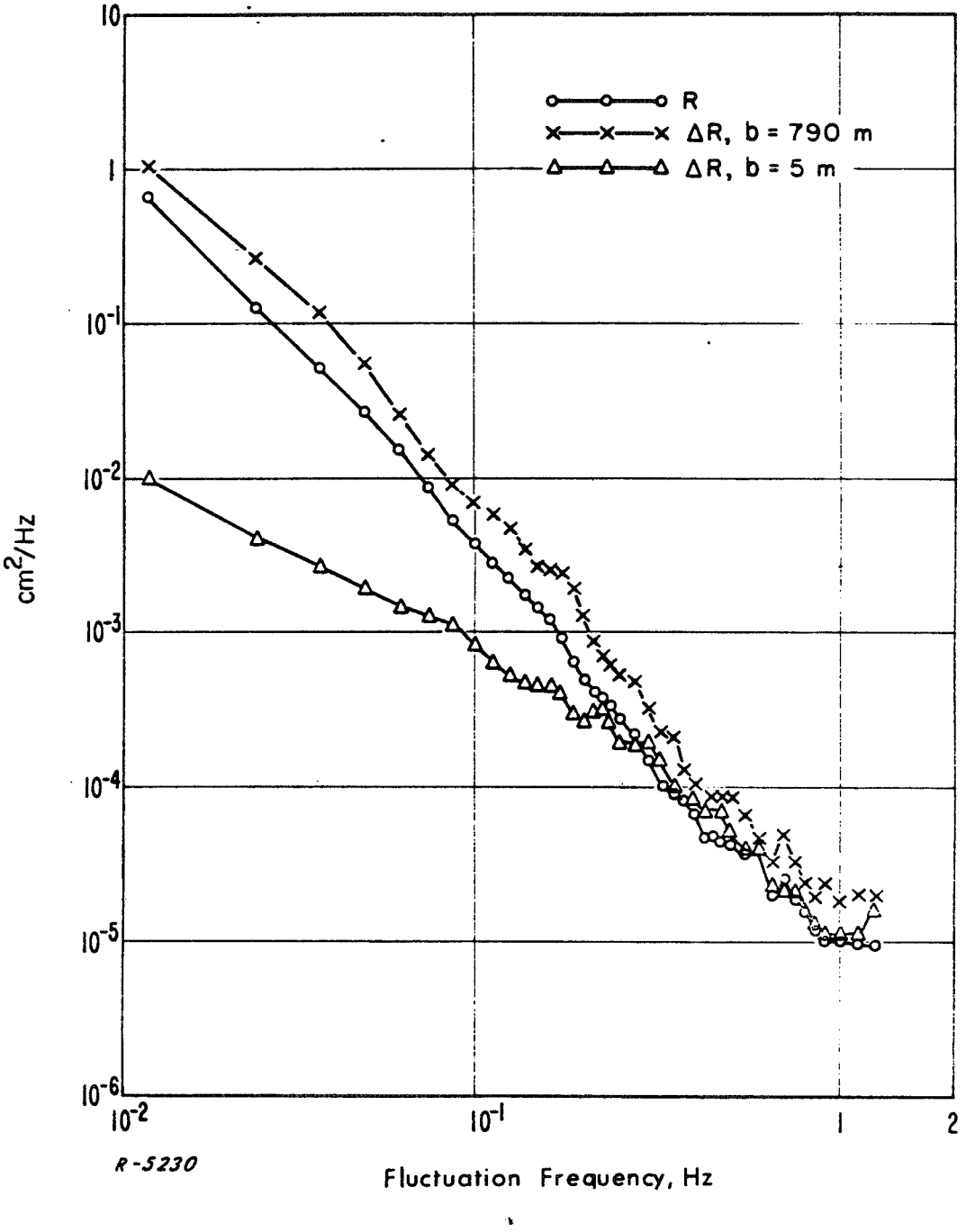

Fig. 4-12 Comparison of Fluctuation Spectral Density Functions for the Phase Difference for 5 m and 790 m Separations with the Density Function for One of the Two Paths.

is negligible. Clouds consisting of water droplets and ice crystals affect the transmission of electromagnetic waves through absorption, scattering, and thermal noise emission. However, below 100 GHz, cloud water droplets are very small compared with a wavelength, and the separations between droplets are such that multiple scattering can be neglected. The absorption coefficient of small water particles varies directly with frequency, whereas the scattering coefficient varies as the fourth power of frequency. Consequently, at the lower frequencies (centimetric and millimetric waves) absorption tends to dominate while particle scattering becomes important at the higher frequencies, becoming predominant at optical frequencies. Below 100 GHz, the scattering cross-section of a water droplet is orders of magnitude smaller than the absorption cross-section. Indeed, scattering by atmospheric particles can be neglected for wavelengths in excess of 1 micrometer.

4.2 Characteristics of Multipath in Tropospheric, Line-of-Sight Propagation

The various propagation mechanisms that account for the larger-scale disturbances in ground-to-ground tropospheric line-of-sight transmission are illustrated in Fig. 4.13.

Depending upon the mechanism that enables the transmitted signal to arrive at the receiving site via more than one path, it is possible to distinguish between

(1) Reflection-multipath fading, and

(2) Atmospheric-multipath fading.

4.2.1 Reflection-Multipath Fading

In reflection-multipath, illustrated in Fig. 4.13(a) and (b), the signal arrives at the receiver via at least two paths:

(a) The direct optical path, and
(b) An indirect path made possible by;

(i) Ground-surface reflection,

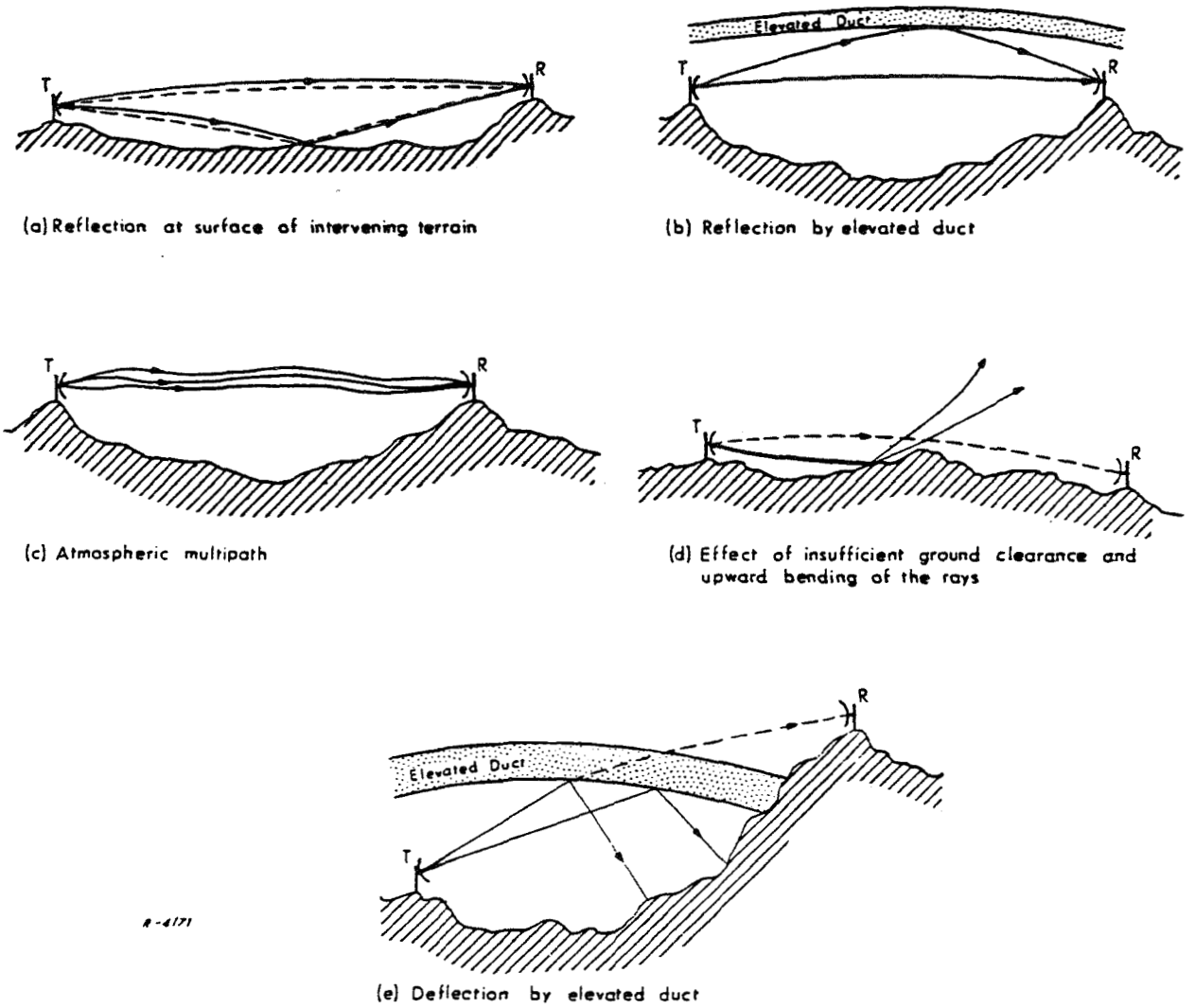

Fig. 4-13 Multipath and Signal Deflection Mechanisms in Tropospheric Line-of-Sight Transmission

(ii) Water-surface reflection,

(iii) Reflection by an inversion layer or duct that lies above both transmitting and receiving antennas, and

(iv) Reflection from spurious objects (such as aircraft, etc.) of transient duration in the intervening space.

Ground-surface and water-surface reflections happen all the time, day and night, and their effects upon the received signal strength can be correlated with the condition of the reflecting surface as well as with changes in the atmospheric conditions. Duct reflection disappears when the layer that causes it vanishes.

4.2.1.1 Ground-and Water-Surface Reflection

Ground-and water-surface reflections may range from specular when the surface is smooth, sometimes resulting in a reflected wave that is almost as strong as the direct wave, to diffuse when the surface is very rough, thus resulting in a reflected wave that is scattered from a widely spread area with a consequent delay spread and a doppler spread if either, or both, of the terminals is in motion and/or if the surface scatterers are not stationary. Specular reflections result in severe fading when the reflected wave arrives in phase opposition to the direct wave. The relative phases of the direct and reflected paths change continually because of changing path lengths caused by the refractive-index fluctuations along the traversed paths, and because of doppler differences if the reflecting surface features are in motion or/and if either (or both) of the terminals are in motion.

The effect of ground reflections can be reduced by using sufficient antenna directivity in order to favor the direct ray and suppress the reflected ray.

For separations between terminals such that grazing angles in excess of one or two degrees are experienced at the reflecting

surfaces or objects, the reflection coefficient depends critically upon the reflecting surface conductivity and dielectric properties, the surface roughness, the polarization of the wave, and the grazing angle. When the grazing angle is small -- less than one degree -- as for ground-to-ground transmissions over paths in excess of 10 miles, the effects of factors other than surface physical roughness are negligible.

The effects of physical features of the reflecting surface on the characteristics of the reflected ray depend on the wavelength of the radiation and the grazing angle ψ_2. The criterion used to determine if a surface with protrusions could be considered smooth or rough is the Rayleigh criterion, which states that a wrinkled surface can be considered smooth relative to an incident wave if

$$h_o \sin \psi_2 < \lambda/8$$

where

h_o = height of surface protrusion

ψ_2 = grazing angle of incident wave

λ = wavelength of the incident radiation

For example, at $\psi_2 = 5^\circ$, surfaces whose obstacles do not exceed about 8 cm may be considered smooth for waves at 5 GHz, while the figure for waves at 30 GHz is approximately 1.5 cm. This shows that reflections from ripply ground or water surfaces are most likely to result in diffuse rather than specular secondary paths at operating frequencies in excess of a few GHz.

The overall effect of roughness is to reduce the effective reflection coefficient ρ to the range $0.1 < \rho < 0.3$ resulting in a reduction of fade depths, as well as to introduce some further fluctuation if the roughness is time-variant (such as for sea-wave motion).

For reflection from a body of water, the surface roughness will not generally be sufficient to result in a reduction of the reflection coefficient to a low level, except perhaps in rough weather with exceptionally high winds. The reflection is therefore often essentially specular and very bad fades can always be expected, especially with

horizontally polarized waves. But with vertically polarized waves the reflection coefficient, though very nearly equal to unity for very small grazing angles, drops to about 0.7 for grazing angles near 1^{o}, which may lead to significantly shallower fades (especially at the lower microwave frequencies) than is experienced with horizontally polarized waves.

4.2.1.2 Duct-Reflection Paths

A duct, or an inversion layer, is a horizontal layer of air whose properties are such that the modified index of refraction decreases rapidly with height through it. Under normal conditions, the modified index of refraction increases with height through the atmosphere. An elevated duct is bounded above and below by layers of normal air, with a sharp discontinuity in the derivative of the index of refraction with respect to height in penetrating the duct from below and from above. Rays incident upon either side of the duct at low grazing angles will be reflected (see Fig. 4.13(b))essentially specularly with an efficiency that increases very rapidly as the grazing angle goes to zero. In fact, the reflection coefficient of the layer is effectively unity for grazing angles less than about one-half of a degree and the reflection coefficient declines very sharply at high grazing angles. The name "duct" for such layers stems from the phenomenon of ducting (or waveguide effect) experienced when radio waves that enter the layer at appropriate angles of incidence are trapped within the layer and travel within it over long distances. The duct acts as a waveguide (with imperfect walls). An estimate for the wave-lengths best sustained for a given duct height d is

$$\lambda \leq \lambda_{max} = 0.014\, d^{3/2}$$

where λ is in centimeters and d is in feet. For example, $\lambda_{max} = 1$ cm for d = 20 ft and $\lambda_{max} = 10$ cm for d = 80 ft. The ducting phenomenon is important mainly for microwave frequencies because the depths of atmospheric ducts are generally of the order of 100 ft. The above limit on trappable wavelength is only a nominal cutoff. Radiations of longer

wavelength may also be trapped but they will suffer greater attenuation. Only partial trapping may occur at wavelengths of shorter than centimetric values.

For ground ducts, if the transmitting and receiving antennas are within the duct, attenuation can be as low as 5 dB per 100 miles for the trapped mode calculated by the formula given above. However, the attenuation at the interfaces may be very high.

In the case of an elevated duct the radiation may suffer an attenuation at the interface of the walls of the duct. Elevated ducts are usually encountered at elevations of 1,000 to 5,000 feet. Small changes in the elevation or structure of the layer can cause large changes in the phase of the duct-reflected ray relative to the direct ray at the receiver.

4.2.2. Atmospheric-Multipath

It is a well-established fact that rapid variations with height in the refractive index of the atmosphere, caused by atmospheric stratifications which usually prevail during windless summer nights, give rise to more than one radio path through the air between transmitter and receiver at microwave frequencies (see Fig. 4.13(c)). Experiments conducted with very short pulses over a 22-mile G/G path at 4,000 MHz have revealed the occasional appearance of two and sometimes three or more paths with path length differences ranging to a maximum observed value of 7 feet. The strength of the various paths and their pathlength differences were observed to change in a matter of minutes and at times none of them was identifiable as the "main" path. In another series of experiments using vertical angle of arrival measurements, it was observed that daytime transmissions were via a single path. Night time transmissions in these same experiments featured the appearance of two rays whose vertical angles of arrival differed by a maximum of about one-half of a degree and sometimes three rays 1/4 of a degree apart.

Path differences of up to 11 feet were also reported. In all of these experiments the roughness of the intervening ground and the presence of vegetation precluded the presence of ground-surface reflections of any consequence (the effective reflection coefficient was about 0.2).

In G/G links, the measured statistics of the envelope fluctuations of the signal under conditions of severe multipath resulting in sharp, deep fades are closely approximated by the Rayleigh distribution. This is brought out in Fig. 4.14. When fewer than 4 significant paths account adequately for the fading, the Rayleigh distribution will not provide a satisfactory approximation. Figure 4.14 presents distributions that may be used in predicting the performance for most cases with acceptable accuracy.

4.3 Mathematical Models of Tropospheric LOS Propagated Signals

The propagation data gathered to date from extensive, well-performed and well-analyzed field measurements, and abstracted in the preceding sections, enable us now to formulate the mathematical models of propagated signals, and hence the functional operations needed for laboratory synthesis of the propagation effects.

4.3.1 Synthesis of Direct LOS Path Propagation Effects

From the data presented in Section 2.1, it appears that if the signal

$$e_{in}(t) = \cos \omega_o t \tag{4.2}$$

is radiated by the transmitting antenna, the signal received over a direct line-of-sight path can be expressed as

$$e_{dp}(t) = V(t) \cos [\omega_o t + \phi_o(t)] \tag{4.3}$$

where V(t) represents the fluctuating envelope and $\phi_o(t)$ the phase fluctuations. Note that we can write

$$\phi_o(t) = \phi_{fast}(t) + \phi_{slow}(t) \tag{4.4}$$

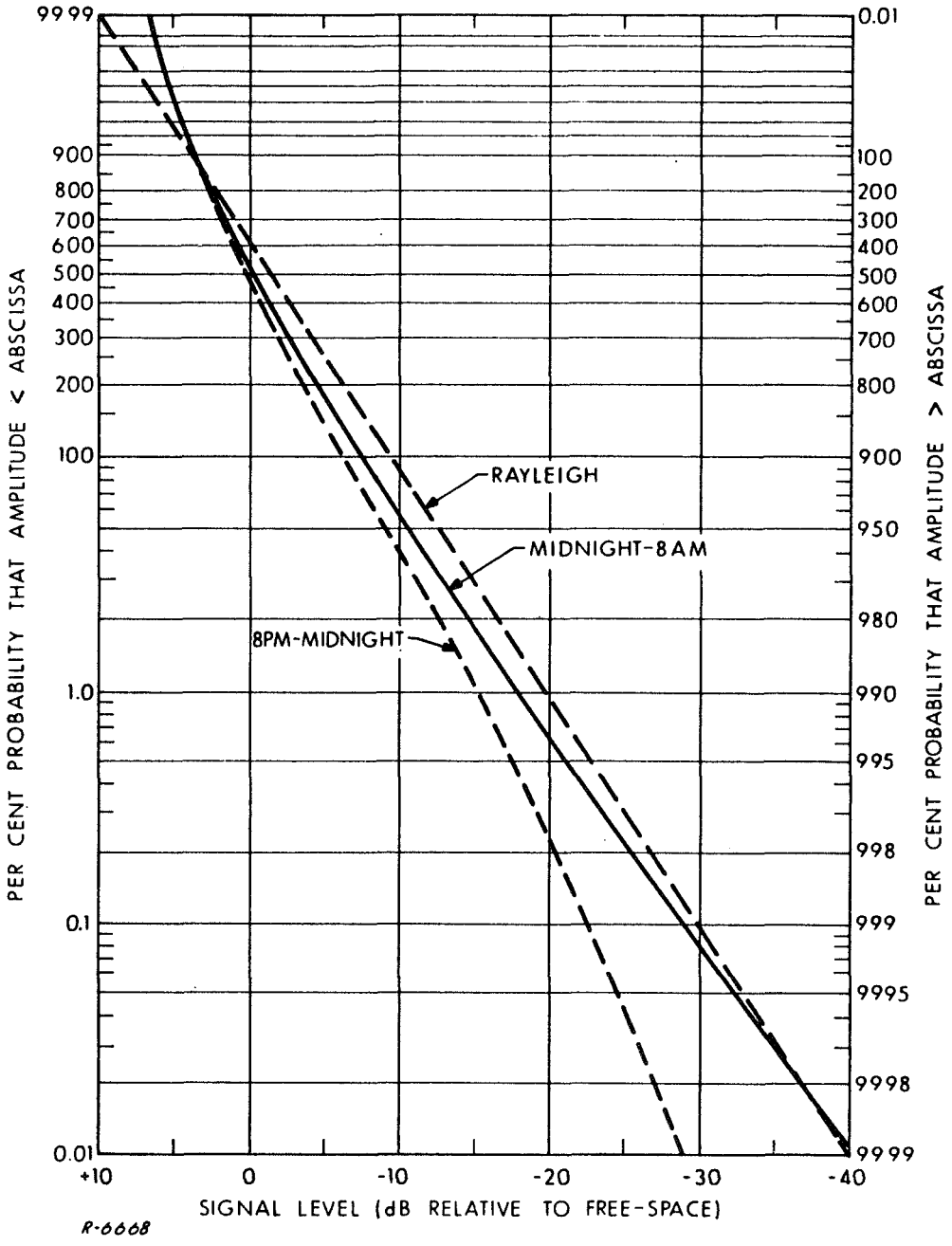

Fig. 4-14 Typical Envelope Fading Statistics for Microwave Line-of-Sight Transmission

where $\phi_{fast}(t)$ represents the relatively rapid "instant-to-instant" phase fluctuations, and $\phi_{slow}(t)$ represents the relatively slowly fluctuating average value (over a few seconds) of the phase. From Figs. 4-3 and 4-4, it is apparent that

$$\overline{\phi^2_{fast}(t)} \ll 1 \tag{4.5}$$

but that $\phi_{slow}(t)$ may actually amount to relatively large drifts (up to a few times 2π or more depending upon the path length and the operating frequency) in a period of minutes. For two different operating frequencies f_{o1} and f_{o2} we would expect from Eq. (4.1) that if $\phi_{ok}(t)$, k = 1 or 2, represents the total phase fluctuations in Eq. (4.3), then

$$\phi_{o2}(t) = (f_2/f_1)\phi_{o1}(t) \tag{4.6}$$

The mean squared value of the phase fluctuations contributed by a particular range of fluctuation frequencies can be determined as the area between the spectral density curves and the portion of the horizontal axis covering the specified range of frequencies in Figs. 4-6 and 4-7. For the range from 10^{-2} Hz to 10 Hz, the results are 5.1×10^{-3} $(\text{ppm})^2$ for Fig. 4-6, and 1.7×10^{-3} $(\text{ppm})^2$ for Fig. 4-7. For a 30 km path and f_o = 10 GHz, the rms phase fluctuations are about 26^o for Fig. 4-6 and 15^o for Fig. 4-7.

The probability distribution law that governs the instantaneous values of the phase fluctuations can be surmized as follows. We first associate the faster fluctuations represented by $\phi_{fast}(t)$ with the random fluctuations of incremental contributions from successive spatial regions traversed by the signal over the propagation path. The slower drifts represented by $\phi_{slow}(t)$ are then associated with the relatively slow variations in the delay characteristics of larger spatial regions containing the path. If we further assume that the more rapidly variable incremental delays that make up $\phi_{fast}(t)$ are independent for the successive parts of

the traversed path, and that they are all comparable in importance, then the central limit theorem suggests that the instantaneous values of $\phi_{fast}(t)$ obey a gaussian probability density function restricted practically to within an interval of width 2π. The total phase fluctuation $\phi_o(t)$ can therefore be treated as a gaussian process added to a slowly fluctuating median represented by $\phi_{slow}(t)$. For typical G/G path lengths (nominally 30 km) the total mean squared value of $\phi_o(t)$ is small compared with unity, and the statistics of the $\phi_o(t)$ process can be considered gaussian for intervals of many minutes.

The intensity of the envelope fluctuation process carried in V(t) can also be estimated from the records in Figs. 4-3 and 4-4. Specifically, for the X-band signal (9.6 GHz), Fig. 4-3 suggests a dynamic range of about 11 dB for the envelope fluctuations, with the fluctuations falling within a range of about 6 dB most of the time; the corresponding numbers from Fig. 4-4 are about 7 dB and 4 dB, respectively. The mean squared values of the envelope fluctuations can be determined as the areas between the spectral density curves and the horizontal axis in Figs. 4-8 and 4-9. The numbers for the various cases are listed in Table 3.

The probability distribution law governing the envelope fluctuations in V(t) can be surmized from the phenomenological cause of the fluctuations. Thus, if the phase fluctuations are correctly attributed to the time variations of the spatial-mean refractive index, $m_n(t)$, the envelope fluctuations must be attributed to the attendant time variations of the spatial-mean atmospheric absorption over the traversed path. The total attenuation sustained at a given instant of time may be considered as the result of a large number of independently variable elemental attenuations sustained in cascade along the traversed path. If the component increments of attenuation introduced over various parts of the propagation path are expressed in dB, the total attenuation in dB will be the sum of a large number of independent random components, the properties of each

Table 3

CHARACTERISTICS OF ENVELOPE FLUCTUATIONS

		Dynamic Range (in dB)		Estimated Nominal Peak Swing a_{max}*	Mean Square (10^{-2} to 10Hz) in $(ppm)^2$	Ratio of Mean Squares K/X		Peak Factor (peak/rms)
		Peak-to-Peak	Most Values Within			Feb.	June	
X-Band (9.6 GHz)	Feb.	11	6	0.56	1.2×10^{-2}	5.82	5.86	5.1
	June	7	4	0.38	0.87×10^{-2}			4.1
K-Band (34.52GHz)	Feb.	24	16	0.88	7.0×10^{-2}			3.3
	June	15	9	0.70	5.1×10^{-2}			3.1

* $a_{max} = \frac{k-1}{k+1}$, k = antilog [(peak-to-peak, in dB) / 20]

of which becomes lost in the properties of the total. From the central limit theorem, this concept of the composition of the total attenuation along the path suggests that the statistics of the envelope fluctuations should be log-normal; i.e., the fluctuations expressed in dB should be characterized by a gaussian probability distribution function. Specifically, if V_t represents the value of V(t) at a particular instant of time, t, and if

$$\alpha_t = \ln V_t \tag{23}$$

then α_t is a random variable with a gaussian probability density function

$$p_\alpha(A) = \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-(A-m_\alpha)^2 / 2\sigma_\alpha^2} \qquad -\infty < A < \infty \tag{24}$$

where m_α is the mean value of ln V, and σ_α is its variance. Accordingly, it is readily shown that

$$p_V(v) = \frac{1}{v\,\sigma_\alpha \sqrt{2\pi}} e^{-(\ln v - m_\alpha)^2 / 2\sigma_\alpha^2}, \quad v \geq 0 \tag{25}$$

$$= 0, \text{ otherwise}$$

4.3.2 Coherence Bandwidth and Fading Rate Estimates

It is of interest at this point to introduce two important parameters, review their significance, and present estimates of them for LOS propagation.

In general, the coherence bandwidth of a propagation medium, as seen through the output terminals of the receiving antenna, may be defined as the maximum frequency range over which a desired degree of correlation (or phase coherence for some applications) is maintained among the transmission fluctuations of component sine waves. Thus all frequency components of a signal would fluctuate practically in step if they were contained within the coherence bandwidth, and the channel differential propagation effects upon link performance would then be considered negligible, or within tolerable bounds.

For most specific communication link applications, however, the definition of coherence bandwidth should be in terms of allowable degradation of performance; e. g., in terms of

1. Intermodulation distortion for FDM/FM
2. Intersymbol "cross-talk"
3. Average probability of error
4. rms value of instantaneous phase difference of responses to two sinewaves (carriers or subcarriers) of different frequencies.

In the type of space links of interest in the present study there is considerable emphasis on maximum usable data rate. Accordingly, coherence bandwidth should be defined on the basis of an allowable

degradation in performance

a. due to intersymbol interference, and

b. due to reduction in the mutual coherence between a sync reference, or a carrier reference, and the signal components at the frequencies most removed from the reference.

Inasmuch as the limitation on coherence bandwidth can ultimately be traced to, or at least be considered to result from, interference among a multiplicity of propagation paths with different delays and randomly time-variant characteristics, the coherence bandwidth is inversely proportional to the delay spread (or "duration" of the impulse response) of the medium. The proportionality "constant" differs from application to application, and is critically dependent on whether the impulse response of the medium is a single-pulse-like waveform, or a combination of separate, distinct and time-spaced pulses. Computations of probability of error in binary signal reception and intermodulation distortion in the reception of FDM/FM signals over channels characterized by "single pulse" type impulse responses have shown that for reasonable engineering criteria of allowable performance degradation, the coherence bandwidth will be roughly one-tenth or less of the reciprocal of the difference between the times of arrival of the earliest and the latest significant paths.

The fading rate (fading bandwidth, frequency smear or doppler spread) parameter is a measure of the bandwidth of the received waveform (the channel response) when the input to the channel is a pure sine wave. Thus, this parameter measures the nominal width of the dispersion in frequency experienced by each individual frequency component in the transmitted signal; it is a measure of the nominal rate at which the fluctuations in the channel, and hence the received signal perturbations, occur. When the test signal at the channel input

is a pure sine wave, the fading rate of the channel response to this test signal is completely determined by the relative motions of the various multipath-generating physical entities (reflectors, scatterers, boundary surfaces, etc.) and by the changes (if any) in the relative velocities of propagation over the various paths.

For microwave line-of-sight propagation between terminals at least one of which may be in motion, the coherence bandwidth and the fading rate depend on the hop length, the conditions of the intervening terrain and atmosphere, the antenna directivity characteristics and orientations, and the motions of either terminal and of the causes of the multipath. Thus, these gross transmission parameters are themselves random variables that, strictly speaking, can only be specified in statistical terms.

Fading rates attributable to the atmospheric refractive effects, in the absence of multipath, can be estimated from the spectral density plots in Figs. 4-6 and 4-7 for the phase fluctuations, and Figs. 4-8 and 4-9 for the envelope fluctuations. Note that the phase spectral density plots are in terms of parts per million, and that rescaling of the vertical axis for particular link lengths and operating frequencies is necessary prior to estimating the fluctuation rate in terms of the fluctuation frequency beyond which the phase-fluctuation spectral density would be considered negligible. The explicit formula for converting from parts per million (ppm) to phase change $\Delta\phi$ in radians is

$$\Delta\phi = 10^{-6}(\text{ppm})\, 2\pi f_o (L/c)$$

where f_o is the operating frequency in Hz, L is the nominal path length and c is the velocity of electromagnetic propagation in the medium. Consequently, conversion of the vertical scale in Figs. 4-6 and 4-7 from $(\text{ppm})^2/\text{Hz}$ to $(\text{rad})^2/\text{Hz}$ can be accomplished by multiplying the present ordinate readings in those figures by

$$4 \times 10^{-11} \quad f_o^{\,2}(L/c)^2$$

For example, if $f_o = 2.3$ GHz and L = 300 km, then the multiplication factor is 2.116×10^2.

4.4 Measurement of Channel Transmission Characteristics and Parameters

The multipath structure, delay differences, properties of signals received over individual paths, and the various gross transmission parameters of a channel can all be determined experimentally by means of appropriately chosen test signals. The two basic test signals for propagation media (as for all linear media) are the impulse in time (or very short pulse) and the impulse in frequency (or the sinusoidal time function). In principle, either type of signal is sufficient for a complete characterization of linear channels. In practice, the choice of test signal design is based largely on considerations of implementation and data handling requirements. For example, the short pulse is often the natural choice for isolating directly the various mechanisms of propagation leading to different paths and for bringing out the delay spread characteristics of the channel. The sinusoidal carrier is useful both for analysis of multipath structure and quite naturally for exploring directly the statistical properties of amplitude and phase fluctuations of a propagated signal that can be attributed to the channel variations. Channel measurements for characterizing a channel in terms of the performance characteristics of particular communication techniques require special types of modulated signals, as we shall elaborate later on in this discussion.

4.4.1 Basic Test Signals and Related Objectives for Channel Characterization Measurements

A number of basic test signals of special importance in channel characterization measurements will now be listed along with the channel

characteristics that are naturally brought out by each.

a. Single short pulse for bringing out:

   (1) Impulse response and its gross characteristics, such as general appearance, duration or multipath spread;

   (2) Multipath structure;

   (3) Delay characteristics of individual paths (arrival times and their distributions);

   (4) Amplitude distribution of isolated paths (when these are discernible) or of the envelope of the resultant impulse response, including such gross features as envelope peak.

b. Single sinusoid to determine:

   (1) Statistics of the envelope of the received signal; and

   (2) Statistics of the phase and frequency of the received signal.

   For (1) and (2):

      (a) Parameters of short-time averages (first-order statistics);

      (b) Parameters of short-time joint distributions (mainly second order, although higher order would be important to check a gaussian model), particularly autocorrelation or power spectral density (which brings out the fading bandwidth).

c. Two short pulses separated in time

   (1) At same frequency to determine:

      (a) Self-decorrelation (time constant, etc.); and

      (b) Other characteristics of the dependence of the impulse response upon the time of application of the impulse.

(2) At different frequencies to determine decorrelation as a function of frequency separation.

d. Two sinusoids at different frequencies to determine:

(1) Spaced-frequency correlation (which brings out the coherence bandwidth and diversity bandwidth).

(2) Spaced-frequency, spaced-time correlation.

4.4.2 Measurement of Multipath Structure and Characteristics

A. Impulse Response Tests: If very narrow, widely spaced pulses of RF are radiated, the receiver will receive more than one pulse, in response to each radiated pulse, if multipath is present. Each received replica of the same transmitted pulse will correspond to one path, and the paths can be resolved if their delay differences exceed the duration of the transmitted pulse, and if the receiver bandwidth exceeds a few times the reciprocal of the pulse duration. In this way, resolvable paths can be studied, and their relative amplitudes and delay differences measured directly.

B. Sinusoidal Carrier Tests: Two types of CW signal tests can be considered for the measurement of multipath characteristics. These can be listed as

(a) Single-frequency tests, and

(b) Multiple-frequency tests, by means of

(i) a frequency-swept carrier, or

(ii) a carrier amplitude-modulated by a number of tones, suitably spaced in frequency.

Single-frequency tests can be performed directly on a radiated carrier. The statistical characterization of fading by means of a single-frequency test signal has already been discussed in Section 4.3.2. We now discuss the use of such a signal for measuring the characteristics of specular multipath. For this purpose, consider the situation in which the signal is received via two specular paths with amplitude ratio, $\underline{a}$, and delay difference, t_d, so that the resultant carrier component can be expressed as

$$E_s \cos(\omega_c t + \phi_c) + aE_s \cos[\omega_c(t - t_d) + \phi_c]$$

$$= E_s \sqrt{1 + a^2 + 2a\cos\omega_c t_d} \cos\left[\omega_c t + \phi_c - \tan^{-1} \frac{a \sin\omega_c t_d}{1 + a\cos\omega_c t_d}\right]$$

The envelope function

$$\sqrt{1 + a^2 + 2a\cos\omega_c t_d}$$

fluctuates with $\omega_c t_d$ for each fixed value of $\underline{a}$. The peak-to-peak excursion of the envelope is

$$E_s[V(t)|_{max} - V(t)|_{min}] = 2aE_s$$

The envelope drops from a maximum to the next minimum through a change of π in $\omega_c t_d$. Such a change in $\omega_c t_d$ may be effected either by changing the location of the receiving antenna in order to change t_d, or by changing the operating test frequency ω_c. For a nominal frequency of 125 MHz, t_d need only change 4×10^{-9} sec (or 4 nanoseconds) to bring about the change from an envelope maximum to a minimum. Since 1 nanosecond of propagation time corresponds to approximately 1 foot of path-length difference, such changes in path length difference may be expected to result from

displacements of the receiver position within a volume in which $\underline{a}$ and E_s may be expected to be nearly constant or independent of receiver changes of position for operating frequencies of 100 MHz or more. In this way, the amplitude ratio, $\underline{a}$, of the two paths can be measured from the difference between a maximum and the next minimum of the envelope.

If more than one specular path is actually present, we may again expect that the fluctuations of the envelope over a traversed volume in which the signal level and the relative amplitude ratios would remain nearly constant can be used to determine the relative path amplitudes by numerically fitting the observed envelope curve to a curve obtained by combining a number of paths. This may be done expediently by assuming a suspected number of paths, writing the general expression for the envelope and generating a number of simultaneous equations equal to the number of unknowns by numerical substitutions from the measured envelope curve.

Note that the desired change of π in $\omega_c t_d$ can also be effected either by sweeping ω_c over a suitable range or by introducing other components at suitably spaced frequencies by means of amplitude modulation by discrete tones. This would allow both $\underline{a}$ and t_d to be determined from the readings on the changing envelope of the frequency-swept carrier, or the simultaneous readings on the envelopes of the separate components at the different frequencies. This method can be extended to the case of more than two paths by the technique of setting up systems of simultaneous equations, as before, or by the choice of modulation tones that could be

adjusted to cancel out the contribution of particular specular paths to the resultant signal.

C. Simultaneous Pulsed and Sinusoidal Carrier Tests: The pulsed and sinusoidal carrier tests can be performed simultaneously without any significant mutual interference. The results of the two types of tests can be used to check one another, or can be combined to improve the quality of the measurement results, when the paths can be resolved with the pulse tests. The pulse measurements also yield directly the number of significant paths present, their delay differences, their amplitude ratios, and any observable delay spread for individual paths. This information can be employed directly to guide the data reduction and computational tasks of the frequency-envelope measurements. The sinusoidal carrier measurements can also be used to resolve the multipath structure even when the pulse equipment employed is incapable of resolving these paths. Finally, the single-carrier technique is particularly suited to the investigation of scatter-type, or diffuse, paths as previously explained in Section 4.3.2.

4.4.3 Measurement of Gross Channel Parameters

We shall now state briefly how the various gross channel parameters defined in Section 4.3.2 can be measured.

A. Measurement of Coherence Bandwidth

General Interpretations of Coherence Bandwidth. - For some applications the coherence bandwidth should be defined in terms of properties of the cross-correlation of corresponding aspects of the

responses to two sinewaves (carriers or subcarriers) of different frequencies; e.g., cross-correlation of

1. Total or "complex" responses
2. Envelopes of responses
3. Instantaneous phases of responses
4. Instantaneous frequencies of responses.

For most applications, however, the definition of coherence bandwidth should be in terms of allowable degradation of performance; e.g., in terms of

1. Intermodulation distortion for FDM/FM
2. Intersymbol "cross-talk"
3. Average probability of error
4. rms value of instantaneous phase difference of responses to two sinewaves (carriers or subcarriers) of different frequencies.

General-Purpose Measurement. - Apply a sinusoidal carrier with a constant amplitude at the transmitting end and sweep its frequency at an f_m Hz (1 or a few kHz) rate with a peak frequency deviation of ΔF Hz that assumes well-calibrated values. For each value of ΔF, detect the envelope of the received signal and measure the f_m kHz component in it. A back-to-back measurement should first be made to establish the residual FM-to-AM conversion in the equipment itself. Any increase over this value when the signal traverses the medium will result from the channel frequency selectivity. The value of $2\Delta F$ that results in some specified increase in the f_m level in the signal envelope is a measure of the coherence bandwidth of the transmission medium.

Analog Transmission. - Use FDM/FM with frequency deviation of $\pm\Delta F$ Hz. Fill the subchannel space with white noise leaving a slot in a desired space. Measure the intermodulation products within the slot in a back-to-back arrangement to determine the equipment contribution, as a function of ΔF. Transmit the signal over the channel and measure the increase in intermodulation products as a function of ΔF.

From the resulting curve of intermodulation distortion vs. ΔF and the relation of bandwidth occupied by the signal to ΔF, the coherence bandwidth can be defined as the bandwidth corresponding to the ΔF that raises the percentage intermodulation distortion by a specified amount.

Digital Transmission. - Repeat the steps of the preceding measurement with bit streams for digital transmission and use a specified average bit error probability as the performance criterion. The bandwidth occupied by the signal when a specified degradation is encountered in the average bit error probability is the coherence bandwidth of the medium.

B. Measurement of Fading Rate or Fading Bandwidth

One measure of the fading rate is given by the number of times per sec that the carrier envelope crosses its median level with a positive or a negative slope. The determination of this number may be made either from a recording of the envelope of the channel response to a constant-amplitude input sinusoidal carrier, or it can be implemented electronically by means of a heavy clipper followed by a differentiator, a polarity selector and a pulse counter, cascaded in the stated order.

Another measure of fading rate is the width of the received carrier envelope spectral density. This spectral density is determined by measuring the envelope correlation coefficient

$$\rho(\tau) = \frac{\overline{E(t)\,E(t+\tau)} - \overline{E(t)}^2}{\overline{E^2(t)}}$$

The Fourier transform of $\rho(\tau)$ yields the desired fluctuating-envelope spectral density function

$$S_{c,\,env.}(f) = \mathcal{F}\{\rho(\tau)\}$$

The fading rate is a measure of the bandwidth of $S_{c,\,env.}(f)$.

As an illustration, empirical results for a troposcatter medium show that a fit of the form

$$\rho(\tau) = e^{-(\tau/\tau_o)^n}, \quad n = c_o/\tau_o + c_1$$

can sometimes be achieved, where

τ_o = correlation time for a drop in the value of $\rho(\tau)$ to 1/e, and

c_o and c_1 are parameters that depend on the carrier frequency.

Thus, the general procedure for fading rate as a measure of fading spectral width can be summarized by the following steps:

(1) Measure $\rho(\tau)$.

(2) Fit the empirical points into a suitable curve.

(3) Fourier transform the empirical curve to obtain $S_{c,\,env.}(f)$.

(4) Obtain the desired measure of spectral width.

C. Measurements on $H(j\omega, t)$

The amplitude characteristic $|H(j\omega, t)|$ can be determined for short intervals (milliseconds) by transmitting a constant-amplitude test signal with a repetitive linear frequency sweep covering a desired frequency range. The envelope of the received signal will give a very close approximation to $|H(j\omega, t)|$ for intervals of one period of the repetitive frequency sweep if this period is sufficiently short.

Alternately, a set of discrete frequencies is used as the test signal, spaced by appropriate frequency increments and covering a desired frequency range. The envelopes of the individually observed components at the receiver reveal $|H(j\omega, t)|$ at the transmitted frequencies. The desired test signal may be generated by frequency modulating a carrier by a single tone, and adjusting the gains for the detected tone envelopes at the receiver in inverse proportion to the relative amplitudes of the spectral lines of interest. The same signal may also be generated locally at the receiver and used to shift each of the individual doppler-spread spectral lines in the received signals down to zero center frequency for determination of the spaced-frequency complex-envelope correlation function.

The phase stability of the medium may be measured by transmitting a test signal made up of one or several frequency-spaced carriers and returning this signal back from the receiving system by means of a coherent transponder in one of two forms:

(a) either on a subcarrier modulated on a carrier that is coherently related to the received-signal carrier by a fixed frequency ratio; or

(b) with a direct fixed frequency multiplier on the entire signal.

Comparison of the transmitted signal components with the returned signal components yields the one-way (for (a)) or the two-way (for (b)) phase instabilities of the medium at the frequencies of the signal components employed.

5. IDENTIFICATION AND SEPARATION OF CAUSES FROM MEASURED RECORDS OF INSTABILITY

The various causes of instability discussed in the preceding sections are generally characterized by partially or totally overlapping spectra. The basis for identifying and separating the effects of such causes lies in the facts that

a) The spectral signatures (or functional dependence of the spectral density functions) of the instabilities associated with the various causes portray differences that are to some extent distinguishable.

b) It may be possible in some cases to assume that each of the causes becomes predominant in its effects over all the others at an identifiable and separate stage of a mission.

Thus, measurements made while the spacecraft is on the launch pad should bring out functions $\phi(t)$, $\dot{\phi}(t)$, $S_{\phi}(f)$, $S_{\dot{\phi}}(f)$ and probability density functions of envelope and phase that are each characteristic of what should be attributed to the source instabilities.

The tropospheric propagation effects may be anticipated and bounded functionally and quantitatively from the empirical data covered in Section 4. The troposphere is estimated to have an effective vertical depth of about 10 km when the spacecraft is beyond the earth's atmosphere. The estimated path length through the troposphere depends, of course, upon the angle above the horizon and the range to the spacecraft before it leaves the earth's atmosphere. Direct measurement of the perturbations due to the troposphere during the ascent phase may be handicapped by the presence of rocket-exhaust plasma trailing the ascending vehicle. The deterministic parameters (spectral densities, probability densities, dynamic ranges, etc.) should be predicted, with the aid of the existing empirical data, from a knowledge of the position or trajectory information of the spacecraft.

The principal effect of the ionosphere at S band frequencies is upon the polarization of the propagated wave. The effect of fluctuations in polarization may be significant at the levels of perturbation of interest in a study of signal instabilities, when the electromagnetic wave is linearly polarized.

When the spacecraft is so far away that the received signal level relative to the additive noise density is not sufficiently high, the perturbations introduced by interference from the additive noise may well dominate all other effects.

Finally, fluctuations introduced by multipath effects are generally noticeably different from the relatively small, "grassy" looking effects of other phenomena.

5.1 General Measurements on Narrowband Signals

The measurable parameters that bring out the instability characteristics of a signal are:

(a) Instantaneous phase or frequency fluctuations

(b) Instantaneous envelope fluctuations

(c) Instantaneous amplitude of components obtained by resolving the signal cophasally and in quadrature with a reference stable carrier at some convenient frequency.

In situations where it is known a priori that the envelope of the signal from the original source should be constant, any observed envelope fluctuations can be attributed directly to the effects of external causes of instability, such as additive noise and propagation perturbations. In situations where it can be assumed that the effects under consideration are relatively small, a number of interesting observations can be made.

First, the sum of a constant-envelope signal and a relatively weak noise can be expressed in the form

$$E_s \cos [\omega_{osc} t + \phi_{osc}(t)] + n(t)$$

$$= E_s[1 + n_c(t)/E_s] \cos [\omega_{osc} t + \phi_{osc}(t)] - n_q(t) \sin [\omega_{osc} t + \phi_{osc}(t)]$$

$$\approx E_s[\ 1 + n_c(t)/E_s]\ \cos\ [\ \omega_{osc}t + \phi_{osc}(t)\ -\ n_q(t)/E_s] \qquad (5.1)$$

where n(t) has been resolved into two components, one instantaneously co-phasal with the signal with an instantaneous amplitude $n_c(t)$, the other instantaneously in quadrature with the signal with an instantaneously amplitude $n_q(t)$. Equation (5.1) shows that the envelope fluctuations under conditions of constant envelope signal and relatively weak noise, are approximately given by $n_c(t)$. Note that if $\phi_{osc}(t)$ is slow relative to the $n_c(t)$ and $n_q(t)$ fluctuation (i.e., of much narrower bandwidth), and if the spectrum of n(t) is completely to one side or the other of the frequency ω_c, then $n_c(t)$ and $n_q(t)$ are Hilbert transforms of each other. This signifies that under the conditions stated, the instantaneous envelope fluctuation waveform, normalized relative to the signal amplitude, is the Hilbert transform of the component of the instantaneous phase fluctuation waveform that is due to the additive disturbance. This fact provides an interesting method by which $\phi_{osc}(t)$ can be isolated from the total phase fluctuation of the resultant signal in Eq. (5.1).

More generally, if we write

$$n(t) = n_c(t)\ \cos\ \omega_c t\ -\ n_q(t)\ \sin\ \omega_c t \qquad (5.2)$$

and if the spectrum of n(t) falls on both sides of ω_c, let

$$n_c(t) = n_u(t) + n_\ell(t) \qquad (5.3)$$

where $n_u(t)$ is associated with the spectral components above ω_c, and $n_\ell(t)$ with those below ω_c. Then it can be shown that

$$n_q(t) = \hat{n}_u(t)\ -\ \hat{n}_\ell(t) \qquad (5.4)$$

where the "hat" denotes "Hilbert transform of." But

$$\hat{n}_c(t) = \hat{n}_u(t) + \hat{n}_\ell(t)$$

This shows that when the spectrum of n(t) includes components on both sides of ω_c, $\hat{n}_c(t) \neq n_q(t)$.

Suppose we now add a very stable local oscillation

$$e_{\ell o}(t) = E_{\ell o}\cos(\omega_{\ell o} + \phi_{\ell o}) \tag{5.5}$$

to the unstable received signal, $e_s(t)$, such that $\omega_{\ell o}$ is a frequency above the highest frequency of interest in $e_s(t)$, and $E_{\ell o}$ is much stronger than the amplitude of $e_s(t)$. The sum can be expressed as

$$\begin{aligned} e(t) &\equiv e_{\ell o}(t) + e_s(t) \\ &= [E_{\ell o} + s(t)]\cos(\omega_{\ell o} + \phi_{\ell o}) + \hat{s}(t)\sin(\omega_{\ell o} + \phi_{\ell o}) \\ &\approx [E_{\ell o} + s(t)]\cos[\omega_{\ell o} t + \phi_{\ell o} + \hat{s}(t)/E_{\ell o}] \end{aligned} \tag{5.6}$$

The OIMS can be used to measure $s(t)$ and $\hat{s}(t)/E_s$. This together with the fact that

$$\begin{aligned} e_s(t) &= s(t)\cos(\omega_{\ell o} t + \phi_{\ell o}) + \hat{s}(t)\sin(\omega_{\ell o} t + \phi_{\ell o}) \\ &= \sqrt{s^2(t) + \hat{s}^2(t)}\ \cos[\omega_{\ell o} + \phi_{\ell o} - \tan^{-1}\frac{\hat{s}(t)}{s(t)}] \end{aligned} \tag{5.7}$$

yields an alternate method for measuring the instability characteristics of $e_s(t)$.

One way to separate the effects of additive noise from those of the source and propagation instabilities is as follows. First, let the sum of oscillator signal and additive noise in Eq. (5.1) be amplitude limited and the result expressed as

$$e_r(t) = \cos[\omega_{osc} t + \phi_o + \phi_{osc}(t) + \phi_{add}(t)] \tag{5.8}$$

where ϕ_o is a random variable, $\phi_{osc}(t)$ is a sample function of the random process that represents the inherent phase instabilities of the source, and $\phi_{add}(t)$ is a sample function of the random process that represents the additive effects. Next, note that if the instantaneous frequency fluctuations, $\dot{\phi}_{osc}(t) + \dot{\phi}_{add}(t)$, are detected by means of an OIMS, then the following

assumptions can usually be made:

a) the spectral density function of $\dot{\phi}_{add}(t)$ goes to zero at f = 0 most often as f^2 ;

b) the spectral density function of $\dot{\phi}_{osc}(t)$ is nonzero at f = 0.

Thus, lowpass filtering the instantaneous frequency fluctuations with an essentially flat response below one, or a few, Hz and an abrupt cutoff above it, and the complementary highpass filtering should yield fluctuations passing the spectral density functions of $\dot{\phi}_{osc}(t)$ and $\dot{\phi}_{add}(t)$ respectively. Such filtering is best done digitally.

5.2 Measurement Techniques with the OIMS

This section describes the measurement techniques which are recommended for use with the Teledyne ADCOM Model G-150 Oscillator Instability Measurement System (OIMS). For specific details of the OIMS the reader is referred to the instruction manual for this equipment.

5.2.1 The OIMS Equipment

The OIMS was developed to fulfill the need for the real-time extraction of the instantaneous phase and frequency fluctuations of high-stability frequency sources. It provides simultaneously three outputs in the form of calibrated voltages that are proportional to the phase, frequency, and amplitude of the signal under test.

The basic operation of the OIMS is the quadrature and co-phasal comparison of two signals--the signal under test and the internally generated and processed reference signal. The quadrature comparison provides the phase-fluctuation voltage, high-pass filtered by the PLL operation of the OIMS, thereby eliminating large, low-frequency phase excursions and allowing far greater sensitivity for phase fluctuation rates above one Hertz. The derivative of the quadrature comparison provides the frequency-fluctuation voltage, low-pass filtered by the PLL operation of the OIMS,

thereby eliminating a large amount of high-frequency noise for frequency fluctuation rates above a few hundred Hertz. The co-phasal comparison provides the amplitude-fluctuation voltage.

The three output voltages of the OIMS, which are proportional to the fluctuations of the signal under test, also contain the fluctuations of the reference signal and its processing by the OIMS. For most situations, the fluctuations of the OIMS and its reference signal are so much smaller than those of the signal under test that they may be ignored. For those few situations where the fluctuations of the signal under test are comparable with the residuals of the OIMS, reference should be made to the calibration procedures in the instruction manual and also comparisons between similar units should be made.

The parameters of voltage and frequency for the signal under test are specified in detail in the instruction manual.

5.2.2 Spectrum Analysis of OIMS Measured Fluctuations

The use of spectrum analysis on the output signals of the OIMS is particularly appropriate since these signals are provided at calibrated levels and with known and controllable frequency response. Equally important is the fact that a properly presented spectrum can identify the instability characteristics of a signal under test, as shown in Sec. 3 of this report.

The most meaningful method of presentation for spectral power density curves is a log-log basis with equal-weight decibel scales for each axis. The horizontal or independent-variable axis will be a logarithmic frequency scale or a linear scale in $10 \log_{10} f$ where f is spectral frequency or fluctuation rate in Hertz. The vertical or dependent-variable axis will be a decible scale with the dimensions of dB (radians^2)/Hz for phase spectral power density, $S_\phi(f)$ or dB (Hz^2)/Hz for frequency spectral power density, $S_f(f) = \frac{1}{(2\pi)^2} S_\phi(f)$.

There are two basic classes of spectrum analyzers that need to be differentiated for this discussion. These are the constant-numerical-bandwidth (CNBW) and the constant-percentage-bandwidth (CPBW) analyzers. The CNBW class includes most digital and all fixed-filter analog spectrum analyzers. The CPBW class includes the tunable-filter, analog type spectrum analyzer. The reason for differentiating the two types is their response to continuous or broadband spectra as opposed to discrete or line spectra. The reaponse of a CNBW analyzer to a flat white noise spectrum for example is a constant reading as a function of frequency. The response of a CPBW analyzer to this same spectrum, however, is a reading which is directly proportional to the spectral frequency being measured. Each of these classes has its particular advantages and disadvantages for the types of spectra encountered in this field.

Some of the expected characteristics of spectra that will be encountered in oscillators and frequency-processing equipment in general and in spacecraft systems in particular have been discussed extensively in Sec. 3 of this report. Briefly, however, we expect to have both continuous and line spectra intermixed. The continuous spectra will have slopes of f^{-4}, f^{-3}, f^{-2}, f^{-1}, flat, f^{+1}, f^{+2} corresponding to -40, -30, -20, -10, 0, +10, and +20 dB per decade of spectral frequency. The line spectral components mixed in with the continuous spectra make analysis difficult since the line components require a fast response whereas the continuous spectrum requires appreciable post-detection averaging to reduce the fluctuations-thus causing a slow response.

The operating techniques that are peculiar to this type of signal will be discussed below. It is assumed that the reader has fully familiarized himself with the normal operation of the particular spectrum analyzer being used. For more detailed background material, several analyzer manufacturers publish handbooks or application notes on spectrum analysis which may be helpful in avoiding operation pitfalls.

The use of a CNBW analyzer for mixed continuous and line spectra requires great care in making the measurements. First it is assumed that the output presentation of the analyzer is in the desired log-log form; that is, amplitude in decibels as a function of frequency on a logarithmic scale. If this is not the case, then it should be so instrumented with suitable logarithmic converters on the linear outputs of the analyzer. Whenever a linear operation analyzer is converted to logarithmic output, due precautions must be taken to ensure that the dynamic range, linearity and offsets are not degrading the data. In order to separate the line components from the continuous spectrum, a narrow analyzer bandwidth is desired. On the other hand, in order to keep the peak-to-peak fluctuations of the continuous spectrum measurement down to a reasonable value the product of the analyzer's filter bandwidth and the post detection averaging time should be about 100, but in any event not less than 10. Therefore for narrow bandwidths such as 1 to 10 Hz the long required averaging times push the sweep time required to cover a certain number of bandwidths up to very long times. As a result, for spectrum analysis with a sweeping type CNBW analyzer in the 1 Hz to 10 kHz range, the following procedures are suggested.

1. Select the sweep range to cover two decades of frequency. (10 to 1000 Hz.)
2. Select a bandwidth comparable with the lowest spectral frequency being measured. (10 Hz.)
3. Select an averaging time to give a time-bandwidth product of 100. (10 sec.)
4. Select a sweep time slow enough so that line components will not be widened due to the long averaging time.
5. Allow one decade of overlap between adjacent spectra for adjustment of fit.

With these precautions the CNBW analyzer should give good results.

The use of the CPBW analyzer presents certain interesting problems while alleviating others. First, as mentioned earlier, the

bandwidth is a certain fixed fraction or percentage of the instantaneous analysis frequency. Therefore the measured value of a flat spectrum increases as 10 log f. This characteristic is a distinct advantage when analyzing spectra with f^{-1} slopes since the resulting curve is flat. It is a convenient statistical observation that a flat spectrum has the most accurately estimated slope. Since spectra with f^{-1} slopes appear frequently in measurements using the OIMS, the CPBW analyzer is definitely advantageous for these measurements. Another advantage of the CPBW analyzer is the faster sweep speeds which can be realized since when the frequency is swept at a constant rate in log f then the number of bandwidths swept per unit time remains constant, which is a desirable situation. One problem inherent in many CPBW analyzers is the wide skirt response of the filter, a condition that results in line components masking the adjacent continuous spectrum. Another problem is the variation in the peak-to-peak fluctuation for a continuous spectrum as a function of analysis frequency. This condition is a direct result of the continuously changing time-bandwidth product of the analyzer. A third problem is the difficulty of quickly converting the measured spectral level of a continuous spectrum into density. This problem is solved with the aid of Fig. 5.1. The decibel values shown in this figure when added algebraically to the measured spectral level at a particular frequency will give the spectral density at that frequency. For a typical CPBW analyzer with 1/3 and 1/10 octave bandwidths the noise bandwidths are 23% and 7.1% respectively. For these values the correction equations become:

$$6.4 - 10 \log f \text{ for } 1/3 \text{ octave and}$$

$$11.5 - 10 \log f \text{ for } 1/10 \text{ octave.}$$

With these equations or Fig. 5.1, the conversion of a continuous spectral measurement to spectral density can be readily accomplished.

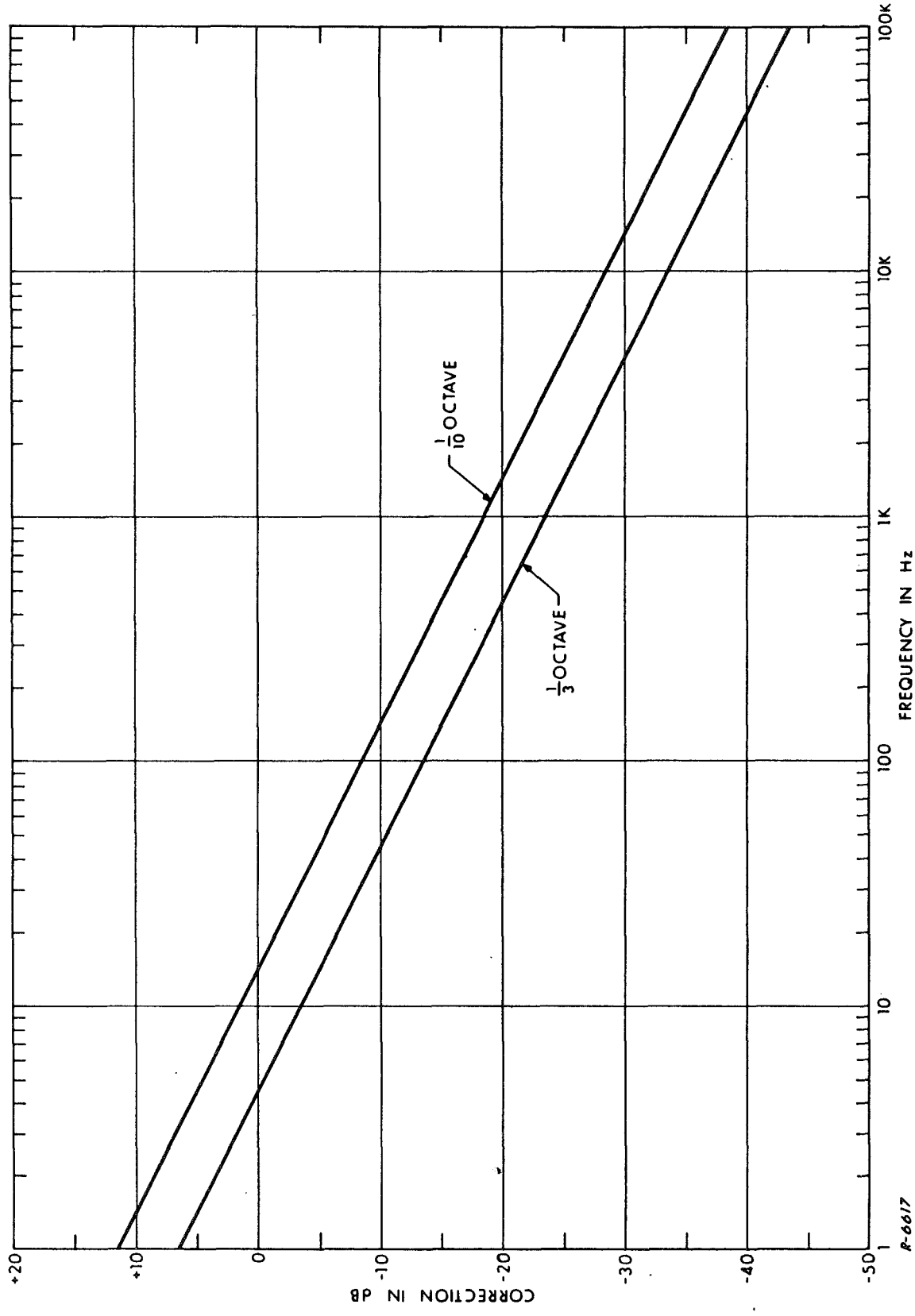

Fig. 5.1 DB Values for Converting Spectral Measurements with a CPBW Analyzer to True Power Spectral Density

For either class of spectrum analyzer there are certain precautions which, if carefully followed, will result in better quality results from the measurements. First, the OIMS ϕM and FM outputs are calibrated in dBV/radian and dBV/Hz respectively. These scale factors when subtracted from the analyzer's scale calibration in dBV will convert the measured data into dB rad or dB Hz respectively. Since there are several gain settings that can contribute to the final scale value, the settings of all contributing controls should be meticulously recorded for every measurement. Second, the data should be reduced by estimating the straight line slopes and combining measurement ranges immediately after the measurement is made if at all possible. If this is done much frustration will be avoided by spotting questionable data while it is still possible to re-do the measurement with revised settings or parameters in order to improve the accuracy of the measurement. Third, if the conditions allow it, measurements of both PM and FM spectra should be made covering overlapping frequency ranges and one set of reduced data should then be converted to the other. Note that:

$$S_{\phi}(f) \text{ in } \frac{(\text{rad})^2}{\text{Hz}} \times f^2 \text{ in } (\text{Hz})^2 = S_f(f) \text{ in } \frac{(\text{Hz})^2}{\text{Hz}}.$$

This conversion will cross check the data and add to the confidence level. If these precautions are followed, the difficult task of obtaining good quality spectral density measurements on signals that are difficult to measure will be considerably eased.

6. REFERENCES

1. E. J. Baghdady, "Theoretical Comparison of Exponent Demodulation by Phase-Lock and Frequency-Compressive Feedback Techniques," 1964 IEEE International Convention Record, Part 6, pp. 402-421.

2. M. C. Wang, in J. L. Lawson and G. E. Uhlenbeck, Threshold Signals, Vol. 18, Rad. Lab. Series, McGraw-Hill Book Company; New York, 1950.

3. S. O. Rice, "Statistical properties of a sinewave plus random noise," Bell Syst. Tech. Journ., vol. 27, pp. 109-157; January 1948.

4. F. L. H. M. Stumpers, "Theory of frequency-modulation noise," Proc. IRE, vol. 36, pp. 1081-1092; September 1948.

5. D. Middleton, Statistical Communication Theory, McGraw-Hill Book Company, New York; 1960.

6. R. Price, "A note on the envelope and phase-modulated components of narrow-band gaussian noise," IRE Trans. Inf. Th., pp. 9-13; September 1955.

7. D. O. North, "An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems," RCA Laboratories, Princeton, N. J., Tech. Rept. No. PTR-6C; June 25, 1943. Republished in Proc. IEEE, pp. 1016-1027; July 1963.

8. H. D. Tauzman, "Short Term Stability Measurements," IRE International Convention Record, vol. 9, Part 9, pp. 191-206, 1961.

9. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, Inc., New York, 1959.

10. L. S. Cutler, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards", presented at The Symposium on the Definition and Measurement of Short-Term Frequency Stability, Goddard Space Flight Center, Greenbelt, Md., Nov., 1964.