The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma. Part. 2: Model identification

François Lefeuvre, L.R.O. Storey

CENTRE NATIONAL D'ETUDES DES TELECOMMUNICATIONS

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

CENTRE DE RECHERCHES EN PHYSIQUE DE L'ENVIRONNEMENT TERRESTRE ET PLANETAIRE

CRPE

NOTE TECHNIQUE
CRPE /41

The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma. II - Model identification.

by
F. LEFEUVRE and L.R.O. STOREY
CRPE/PCE
45045 - ORLEANS LA SOURCE - FRANCE

Mai 1977

CENTRE NATIONAL D'ETUDES
DES TELECOMMUNICATIONS

CENTRE NATIONAL DE
LA RECHERCHE SCIENTIFIQUE

CENTRE DE RECHERCHE EN PHYSIQUE DE
L'ENVIRONNEMENT TERRESTRE ET PLANETAIRE

NOTE TECHNIQUE CRPE/41

THE ANALYSIS OF 6-COMPONENT MEASUREMENTS OF A RANDOM ELECTROMAGNETIC WAVE FIELD IN A MAGNETOPLASMA. II - MODEL IDENTIFICATION

by

F. LEFEUVRE and L.R.O. STOREY

C.R.P.E./P.C.E.
45045 - ORLEANS LA SOURCE - FRANCE

Le Chef du Département P.C.E.

C. BEGHIN

Le Directeur

J. HIEBLOT

Mai 1977

CONTENTS

ABSTRACT

A method is proposed for the analysis of measurements of the six components of a random electromagnetic wave fields. This field, observed at a fixed point in a magnetoplasma, is assumed to be a Gaussian stationary and ergodic wide-band process. It is described statistically by the distribution of wave energy density with respect to the variables frequency and wave-normal direction. A parametric model of the wave distribution function is assumed. Using the principle of maximum likelihood, a method is developed for estimating the unknown parameters from data consisting of continuous measurements of the six field components over a finite time interval. The accuracy of the estimated parameters is evaluated. Finally, methods are proposed for testing the validity of the model, and for comparing the relative merits of different possible models when more than one exist.

1 - INTRODUCTION

The present paper deals with the problem of how to estimate the wave distribution function for a random electromagnetic wave field in a magnetoplasma, from measurements of the six components of the field at a fixed point. It is the second in a series of detailed papers on this and related topics.

In two short review papers that preceded this series (STOREY, 1971 ; STOREY & LEFEUVRE, 1974), we pointed out that a linear random field in a homogeneous magnetoplasma could be considered as the sum of fields due to a continuum of elementary plane waves, of different frequencies, propagated in different directions, without any mutual phase coherence. Such a field can only be described statistically. We chose to characterize it by a function that specifies the distribution of wave energy density with respect to frequency and the wave-normal direction ; this we named the wave distribution function (WDF). The second of the two papers referenced above forms an introduction to the present series.

In the first paper of this series (STOREY & LEFEUVRE, 1976), hereinafter referred to as paper I, we considered what we called the direct problem. Knowing the distribution function for the waves together with the values of the characteristic parameters of the plasma we sought to determine the statistical properties of the six field components. These properties are described by the 36 auto-covariance and cross-covariance functions of the six components, or by their auto-spectra and cross-spectra. We argued that it is preferable to work with the 36 spectra, which we arranged for convenience in a 6 x 6 spectral matrix $\mathbf{S}$. Then we showed that the elements of this matrix are related to the WDF by the following equation* :

$$(2.1) \qquad S_{ij}(\omega) = \frac{\pi}{2} \sum_{m} \int_{0}^{\pi} \int_{0}^{2\pi} a_{ijm}(\omega,\theta,\phi)\, F_{m}(\omega,\theta,\phi)\, d\theta\, d\phi$$

* All equations numbers in the present paper bear the prefix "2". Reference is made also to equations in paper I, which have numbers with the prefix "1".

Here ω is the wave angular frequency, while the subscripts i and j refer to the six field components. $S_{ij}(\omega)$ is either the mean auto-power spectrum of a single field component (if i = j), or the mean cross-power spectrum of two components (if i $\neq$ j). The coefficients $a_{ijm}(\omega,\theta,\phi)$ are related to the corresponding spectra for an elementary plane wave in the magneto-ionic mode m (ordinary or extraordinary), propagated in the direction specified, in spherical polar coordinates, by the polar angle θ and the azimuthal angle ϕ ; if the characteristics of the ambient plasma are known, then these coefficients can be calculated from the magneto-ionic theory. $F_m(\omega,\theta,\phi)$ is the distribution function for the waves in mode m. See paper I for fuller definitions of all those quantities. The equation (2.1), which enables the spectral matrix to be predicted when the distribution function is given (for instance, by the theory of the origin of the random wave field), represents the solution to the direct problem.

Here we are concerned with the corresponding inverse problem : namely, given data from which we can estimate the 36 elements of the spectral matrix, to what extent can the WDF be determined and how should this be done ? (Note that, for reasons given in paper I, only 33 of these elements are mutually independent).

At the start of the present work, it appeared that the solution to this problem would split into two consecutive parts : firstly, use of the 6-component field data to estimate the elements of the spectral matrix ; secondly, use of the matrix elements to estimate the distribution function (STOREY & LEFEUVRE, 1974). Now this is in fact true, but it does not represent the best way in which to derive the solution theoretically. The first part would be easy ; however, since the field data are stochastic, and exist only over a finite interval of time, the estimates of the 36 matrix elements are inevitably subject to statistical error. To develop the second part of the solution, it would be necessary to known the joint probability distribution of the 36 errors, which would be tiresome to derive and to manipulate. For this reason, it is preferable to by-pass the spectral matrix initially, and to treat the problem as one of estimating the WDF directly from the 6-component field data.

In general, an inverse problem such as this is not easy to solve. However, this particular problem is simplified if we recognize that the data concerning waves of different frequencies may be separated readily by spectral analysis. Then the problem is reduced to that of determining how the distribution varies as a function of direction at a fixed frequency, ω_o say.

Unfortunately even this simpler problem is still "improperly posed", in the sense that there are infinitely many wave distribution functions that correspond to a given set of experimental data. However, we shall assume in this paper that prior information is available about the form of the distribution function. More exactly, our assumption is that one or more theoretical models exist, each of which describes a possible distribution function completely when values are assigned to its characteristic parameters. Then the problem becomes clear-cut : if there is only one such model, it is that of estimating the parameters of this model, while if there are several models, we have the additional problem of choosing the best one. In either case, we must bear in mind that the true distribution function may not be describable by any of the models assumed.

We shall start with the case where there is only one model. Our treatment of it is illustrated by the flow chart of Fig. 1, in which the rectangles represent items of information and the arrows the operations that are performed on them. At the input we have three sets of information : firstly, measurements of the properties of the plasma at the point of observation ; secondly, certain prior information that enables us to construct a parametric model of the WDF ; thirdly, the set of field data. Then, using our knowledge of the plasma properties, we perform certain mathematical operations on the data in order to estimate the characteristic parameters of the model. Having done so, it is important to test the validity of the results. If they are unacceptable, then a new model must be devised ; this step is beyond the scope of the present paper. But if they are acceptable, then the procedure ends with this particular WDF as output.

In the case where there are several possible models, the procedure is only slightly more complex : first, the best set of parameters must be estimated for each model separately ; then the relative merits of the different models must be assessed ; finally the best set of results (model + parameter values) must be treated for validity. Fig. 2 illustrates the case with two models ; the extension to a larger number is obvious.

Problems that involve both parameter estimation and hypothesis testing (otherwise known as "state estimation") belong to the general field of identification theory (EYKHOFF, 1974). Even nowadays, the statistical foundations of this theory are the subject of controversy, between advocates of the concept of inverse probability based on Bayes" Theorem, and those of the concept of likelihood due to Fisher. Our treatment is based entirely on the likelihood concept, and uses the terminology of EDWARDS (1972).

This approach to the study of random wave fields has been taken previously in oceanography (for instance, by OLBERS et al., 1976), but not, so far as we known in plasma physics.

The general plan of this paper follows the data flow as sketched in Figs. 1 and 2. The section 2 describes the input information, comprising the plasma properties, the parametric model or models, and the V.L.F. field data. Section 3 describes how the field data are prepared, by successive filtering, demodulation, sampling, and analogue-to-digital conversion, for their subsequent treatment in a digital computer. Section 4, which is the heart of the paper, deals with the operations that must be performed on the input information in order to obtain the best estimates of the parameters of a given model : after a brief discussion of the relative merits of various estimation procedures (or "estimators"), leading to the choice of the method of maximum likelihood, an account in given of how the likelihood function is calculated and maximized. Section 5 describes how the merits of the different models are compared, and the best of them tested for validity, again on the basis of maximum likelihood ; also expressions are derived for the accuracy of the parameter estimates. Finally, in section 6, we summarize our results, discuss their limitations, and make suggestions for further study.

2 - INPUT INFORMATION

2.1. PLASMA PROPERTIES

We assume that, at the point of observation, measurements are made of such properties of the ambient magnetoplasma as suffice to characterize it in cold-plasma theory. The list of the relevant properties depends on the frequency range of the wave fields that are being studied. At frequencies large compared to the lower hybrid frequency, where the ions do not influence the waves significantly, it suffices to measure the magnetic field vector and the electron density. At lower frequencies the ionic composition must be measured also, unless it happens to be known beforehand, for instance from previous measurements that have shown it to be constant, in time and in space, throughout some large region containing the point of observation.

For the purposes of the subsequent analysis, knowledge of these properties is not absolutely indispensable. It would be possible to treat some or all of them as additional unknown parameters of the model for the WDF, and thus to determine them from the analysis itself. Such procedures would be the analogues, for random wave fields, of the various well-known methods for determining local plasma properties from observations of coherent wave fields, natural or artificial (STOREY, 1959 ; AUBRY, 1968 ; SCARF et al., 1969). However, a general rule in problems of statistical parameter estimation is that the fewer the unknown parameters, the more accurately can they be estimated, so it is better if the plasma properties are measured independently ; here we shall assume that this has been done, and that the errors of these measurements are negligible.

Knowing the plasma properties, the weighting functions $a_{ijm}(\omega,\theta,\phi)$ in the set of equations (2.1) can be calculated from the formulate given in appendix B of paper I. Then the only unknown parameters in this equation are those of the model for the distribution function $F_m(\omega,\theta,\phi)$.

The properties of the plasma also constrain the choice of this model. They determine whether, at a given frequency ω, both magneto-ionic modes can be propagated, or only one of them, or neither. For a propagated mode, they determine whether propagation occurs over the full range of wave-normal $0 \leqslant \theta \leqslant \pi$, or only over part of this range.

This constraint is best understood by referring to the classical CMA diagram (STIX, 1962 ; ALLIS, BUCHSBAUM & BERS, 1963) which is drawn in Fig. 3 for the case where the plasma contains only one type of positive ion. For instance, under the conditions corresponding to region 8 of this diagram, only the ordinary mode is propagated. Moreover, there is a range of θ, centred on $\pi/2$, within which neither mode can be propagated : waves in the ordinary mode are propagated only when their normal directions lie inside a certain cone with its axis along the magnetic field. This is the so-called "Whistler mode", in which much magnetospheric radio noise is generated.

Summarizing, for values of the variables ω and θ such that, for the known plasma properties, one or other mode is not propagated, the corresponding part of the WDF is necessarily zero. However, this is the only constraint that the plasma properties impose : at values of ω and θ for which the mode m is propagated, the plasma properties place no restrictions - and yield no information - on the form of $F_m(\omega,\theta,\phi)$.

2.2. PARAMETRIC MODEL

In addition to the constraint imposed by the plasma properties, prior information for modelling the WDF is generally provided by some theory of the generation and propagation of the random waves.

The theory need not necessarily be detailed : a few very simple and general ideas about the origin of the waves may suffice for constructing a phenomenological model, in which the distribution is usually a simple function of a few parameters. For instance, one may know that the source of the waves is so remote that it subtends

only a small solid angle at the point of observation, in which case the observed waves are likely to form a fairly narrow beam. Accordingly onemight be inclined to model their angular distribution with a 2-dimensional Gaussian function, the characteristic parameters being its peak value, the direction of the peak as specified by the corresponding angles θ and ϕ, and also one or two width parameters depending on whether one assumes the cross-section of the beam to be circular or elliptical ; in the latter case, a further angle may be required to specify the orientation of the major axis. Generally these parameters are not known before the field measurements are made : even if the position of the source is known, the direction of the beam may not be, because the waves may have been refracted on their way to the point of observation ; again because of refraction and also of scattering, the beam width may be uncertain even if the size of the source is known. Thus, at any given frequency, between 4 and 6 parameters have to be determined from the field measurements in order to specify the model completely. A phenomenological model such as this might be appropriate for the field of Alfvén waves in the solar wind.

On the other hand, if a detailed theory exists concerning both the place and mode of origin of the waves, then this prior information can be used to construct a physical model, in which the distribution is liable to be a relatively complicated function of a somewhat larger number of parameters. One set of parameters might relate to the instability that produces the waves ; for instance, it might characterize the particle distribution function for an unstable population of energetic electrons. Another set might characterize the spatial distribution of cold plasma, which governs the propagation of the waves. Some of these various parameters may be known from direct measurements, but most of them are likely to be unknown and hence must be determined from the field data. Physical models of this type could perhaps be devised for magnetospheric phenomena such as plasmaspheric E.L.F. hiss or hight-latitude V.L.F. hiss.

Whatever thetype of model, the number of adjustable parameters cannot be allowed to exceed the number of independent elements

of the spectral matrix. For reasons explained in paper I, this number is 33 (assuming that the field measurements include all 6 components), rather than the 36 that one might expect. The number of parameters is likely to be much less in any reasonable theory.

2.3. FIELD DATA

The same assumptions about the field data will be made here as in paper I. They may be summarized as follows : the data are signals obtained by continuous measurement of the electric and magnetic components of the random wave field ; the measurements are undistorted, and free from noise of technological origin ; the coordinate axes Oxyz, along which the components are measured, are right-handed rectangular with Oz parallel to the steady magnetic field ; the point of observation is fixed with respect to the plasma so the wave frequencies are not Doppler-Shifted ; each signal is a stationary and ergodic Gaussian random process of zero mean value ; the statistics of the set of observed signals is described fully by the spectral matrix $\mathbf{S}$, which is related by (2.1) to the wave distribution function $F_m(\omega,\theta,\phi)$. This data set is assumed to be complete : that is, it comprises all three components of the wave electric field vector $\vec{E}$ and all three components of the wave magnetic field vector $\vec{H}$. In practice, one or more of these six components may be missing. However, only a trivial extension of the theory is needed to treat such incomplete data sets.

For the purposes of the present theoretical analysis, these data are regarded as the six components of a generalized electric field vector $\vec{\mathcal{E}}$, defined by (1.2) as follows :

$$(2.2) \qquad \mathcal{E}_{1,2,3} \equiv E_{x,y,z} \qquad \mathcal{E}_{4,5,6} \equiv Z_o H_{x,y,z}$$

where Z_o is the wave impedance of free space. The elements $S_{ij}(\omega)$ of the spectral matrix are the mean auto-power spectra ($i = j$) and cross-power spectra ($i \neq j$) of these six components.

For the same purposes, we suppose these ideal statistically-stationary data to be given us over a time interval of duration T, starting at the instant t = 0. But in practice T cannot often be identified with the total duration of the data, since experimental data are seldom statistically stationary over long periods. To reconcile theory and practice, T should be identified rather with the length of time during which the data statistics can effectively be regarded as stationary. It is determined by testing the data for stationarity. In general it varies with time, and momentarily may even become so short as to make parts of the data useless ; these parts must be eliminated. The remaining data are divided into sections of length T ; and then each section is analysed separately by the methods described below.

In describing these statistical methods, we are naturally led to distinguish between ensemble averages, which are used to define the data statistics, and time averages over the interval $0 \leqslant t \leqslant T$, which are used to estimate them. We denote ensemble averages by triangular brackets enclosing the expressions concerned, and time averages by horizontal bars placed over these expressions. The ensemble average (or "true", or "expected") statistics will be represented by symbols with no special features, while the corresponding estimates based on time averages will be represented by the same symbols with caret superscripts. Thus if, for some random function x(t) satisfying our assumptions, we were led to define the mean power w as the ensemble average $\langle x^2 \rangle$, our estimate of this quantity would be the time average $\hat{w} = \overline{x^2} = T^{-1} \int_0^T x^2 \, dt$.

3 - DATA REDUCTION

3.1. FOREWORD

As a first step towards the analysis properly so called, which is described in sections 4 and 5, it is convenient to replace each of the six continuous signals in the data set by an equivalent sequence of numbers, this for two reasons, the first theoretical and the second practical. Firstly, considerations of statistical probability are more easily applied to discrete sequences than to continuous functions. Secondly, the analysis is too complicated to be performed by means other than a digital computer, which can accept data only in discrete numerical form.

The initial reduction of the data from continuous to discrete form involves three operations : (1) filtering ; (2) demodulation ; (3) sampling. These are described below, as if they were to be done in the given order. Also, for purposes of illustration, various analogue means for performing them are described briefly. However, neither this order nor these means are unchangeable. For instance, one could very well begin by sampling the data and converting them to numbers, then perform the remaining two operations in the computer.

Strictly speaking, these operations are preliminary processing rather than data reduction, since they are reversible and thus involve no reduction in the information content of the data. However, it will appear in section 4 that the first step in the subsequent analysis is to use these data to estimate the elements of the spectral matrix, by time-averaging the instantaneous auto-powers and cross-powers of the filtered signals. These operations destroy redundant information, so they are true data reduction. Here we suppose that they are done in the computer, after sampling and numerization. However, they could also be done by analogue means, either with or without prior demodulation, and since these means have some practical importance, they will be described in outline.

3.2. FILTERING

The final object of our analysis is to enable us to determine the value of the distribution function $F_m(\omega,\theta,\phi)$ for either of the two wave modes and for any set of values $(\omega_o,\theta_o,\phi_o)$ of the wave variables. This - if it could be done - would involve identifying the contributions to the signal power spectra from waves in a given mode and with ω in the range $\omega_o \pm \Delta\omega/2$, with θ in the range $\theta_o \pm \Delta\theta/2$, and with ϕ in the range $\phi_o \pm \Delta\phi/2$, the frequency bandwidth $\Delta\omega$ and the angular widths $\Delta\phi$ being sufficiently small that F_m can be considered as constant and equal to its value at $(\omega_o,\theta_o,\phi_o)$ throughout the corresponding ranges. Of course this object is not fully attainable, for reasons already explained in section 1. But it is partly attainable, inasmuch as in one variable - the frequency - the resolution is limited only by the length of time T over which the data can be considered as statistically stationary : the contributions that the waves in the frequency range $\omega_o \pm \Delta\omega/2$ make to the data signals can be isolated by frequency analysis (i.e. by passing these signals through narrow-band filters), and the bandwidth $\Delta\omega$ can be made as small as one wishes, subject to the restriction $T\Delta\omega \gg 1$. Accordingly this is the first step in the data reduction ; of course, if the data were acquired in space with narrow-band receivers, this step would be superfluous.

We shall suppose that all six data signals are passed through identical, ideal, narrow-band filters. These have rectangular pass-bands of width $\Delta\omega$ centred on the frequency ω_o, within which they neither attenuate the signals nor shift them in phase. Thus, at positive frequencies, the transfer function $Y(\omega)$ of the filters is assumed to be unity for $\omega_o - \Delta\omega/2 < \omega < \omega_o + \Delta\omega/2$, and zero elsewhere. We assume $\Delta\omega \ll \omega_o$. Their noise bandwidth, as defined by (1.4), is $\Delta\omega$. The symbol $g_i(t)$ will denote the narrow-band signal obtained by passing the wide-band signal $\xi_i(t)$ through such a narrow-band filter.

Now let us consider how the statistics of the filtered signals are related to those of the original data signals. On the new assumption that $F_m(\omega,\theta,\phi)$ is uniform over the frequency range

accepted by the filter, for both wave modes and for all values of θ and ϕ, it follows that $S_{ij}(\omega)$ also is uniform over this range, so - at positive frequencies - the cross-spectrum of the pair of filtered signals $g_i(t)$ and $g_j(t)$ is equal to $S_{ij}(\omega_o)$ for $\omega_o - \Delta\omega/2 < \omega < \omega_o + \Delta\omega/2$, and to zero elsewhere. The corresponding cross-covariance function, which is defined as the ensemble average $< g_i(t)\ g_i(t-\tau) >$ and which we shall call $r_{ij}(\tau)$, is the inverse Fourier transform of this spectrum, namely

$$(2.3) \qquad r_{ij}(\tau) = \frac{\Delta\omega}{\pi} \text{ sinc } \left(\frac{\tau\Delta\omega}{2\pi} \right) \{ \left[R_e S_{ij}(\omega_o) \right] \cos \omega_o \tau - \left[I_m\ S_{ij}(\omega_o) \right] \sin \omega_o \tau \}$$

where the symbols R_e and I_m denote the real and imaginary parts. The sinc function is

$$(2.4) \qquad \text{Sinc } u = \frac{\sin \pi u}{\pi u}$$

Evidently the cross-covariance function of the two signals $g_i(t)$ and $g_j(t)$ is determined completely by its values at two points, such as $\tau = 0$ and $\tau = \tau_o$ or $\tau = -\tau_o$, where $\omega_o \tau_o = \pi/2$, since

$$(2.5) \qquad r_{ij}(o) = \frac{\Delta\omega}{\pi} R_e \left[S_{ij}(\omega_o) \right]$$

$$(2.6) \qquad r_{ij}(-\tau_o) = - r_{ij}(\tau_o) = \frac{\Delta\omega}{\pi} I_m \left[S_{ij}(\omega_o) \right]$$

In (2.6), we have ignored the variation of the sinc function over the quarter-period at the frequency ω_o. Similarly the auto-covariance function of any one signal $g_i(t)$ is determined completely by its value for $\tau = 0$.

The results (2.4)-(2.6) concerning the statistic of the narrow-band real signals $g_i(t)$ can be expressed more elegantly in terms of the corresponding complex analytic signals $y_i(t)$. If $C_{ij} = \langle y_i(t)\, y_j^*(t) \rangle$ is the complex cross-covariance function of $y_i(t)$ and $y_j(t)$, then the following results hold :

$$r_{ij}(\tau) = \frac{1}{2} R_e\left[C_{ij}(\tau)\right] \tag{2.7}$$

$$C_{ij}(\tau) = \frac{2\Delta\omega}{\pi} S_{ij}(\omega_o)\ \mathrm{Sinc}\left(\frac{\tau\Delta\omega}{2\pi}\right) \exp\left(i\omega_o\tau\right) \tag{2.8}$$

At zero time-shift,

$$C_{ij}(0) = \frac{2\Delta\omega}{\pi} S_{ij}(\omega_o) = C_{ij}, \text{ say.} \tag{2.9}$$

Evidently all the information in the spectral matrix at the frequency ω_o is contained in the matrix $\mathbf{C}$ whose elements C_{ij} are the covariances at zero time-shift. $\mathbf{C}$ is known as the covariance matrix of the narrow band signals. The expression (2.9) is a special case of the more general results (1.5).

The facts set out above form the bases of a standard analogue method for estimating the complex covariances $C_{ij}(0)$, and hence the spectral matrix elements $(S_{ij}(\omega_o)$, from data of finite duration (MEANS, 1972 ; HARKER & ILIC, 1974). It involves taking time averages over the interval T. The fractional error of the resulting estimates is of the order of $(T\,\Delta\omega)^{-1/2}$, so the condition for them to be reliable is $T\Delta\omega >> 1$. This method is illustrated in fig. 4 ; following the convention announced in § 2.3, we denote such estimates by the use of caret superscripts.

3.3. DEMODULATION

It is well known that, for narrow-band signals, the analytic signal $y(t)$ that corresponds to a real signal $g(t)$ of mean frequency ω_o, can be regarded as a complex exponential of constant frequency ω_o, modulated by a complex envelope, $z(t)$ say (HELLSTROM, 1968). We shall write the latter in terms of its real and imaginary parts as

$$z(t) = x(t) - iy(t) \tag{2.10}$$

Thus

$$g_i(t) = R_e\left[y_i(t)\right] = R_e\left[z_i(t) \exp(i\omega_o t)\right] \tag{2.11}$$

$$= x_i(t) \cos \omega_o t + y_i(t) \sin \omega_o t \tag{2.12}$$

A minus sign has been included in (2.10) - though this is not customary - so as to avoid one in (2.12) and in various later equations. Given the instant corresponding to the origin of time, the two signals $x_i(t)$ and $y_i(t)$ contain all the information conveyed by $g_i(t)$. Like the latter, they are Gaussian random variables with zero mean value, but they are easier to deal with because they vary relatively slowly : their mean frequency is only $\Delta\omega/4$. Hence the second step in the data reduction consists of extracting these two components of the complex envelope, i.e. of "demodulating" each of the six narrow-band signals $g_i(t)$.

Fig. 5 illustrates a well-known analogue method of demodulation. First the narrow-band data signal is multiplied by the sinusoidal signals $\cos \omega_o t$ and $\sin \omega_o t$. Next, the product signals enter ideal low-pass filters which cut their spectra off at some frequency between $\Delta\omega/2$ and $\omega_o - \Delta\omega/2$, typically around $\omega_o/2$, without affecting their components at lower frequencies. Finally, the outputs from these two filters are $\frac{1}{2} x_i(t)$ and $\frac{1}{2} y_i(t)$ which - apart

from the numerical factor - are the required demodulated signals.

Instead of fixing the data bandwidth $\Delta\omega$ by means of the band-pass filters, as described in § 3.2, it is possible to do so with the low-pass filters, as shown in fig. 6. The wide-band data signals $\xi_i(t)$ are multiplied directly by the auxiliary signals $\cos \omega_o t$ and $\sin \omega_o t$, then the products are smoothed by low-pass filters with cut-off frequency $\Delta\omega/2$; we suppose that these filters have ideal rectangular low-pass characteristics. The procedure as shown requires the multipliers to be perfect, for otherwise components of the data signals around zero frequency or around the harmonics of ω_o may break through into the output. To avoid this risk, it is usual to pass each data signal, before demodulation, through a relatively wide band-pass filter, covering typically the band $0.5\omega_o - 1.5\omega_o$. The operations are best performed in this order when $\Delta\omega << \omega_o$, as we have assumed, since a low-pass filter cutting off at $\Delta\omega/2$ is easier to construct than a band-pass filter of bandwidth $\Delta\omega$ at the frequency ω_o.

In the present instance, where the six pairs of low-frequency signals $x_i(t)$ and $y_i(t)$ are to be used only for calculating their auto-covariances and cross-covariances, the phases of the auxiliary signals $\cos(\omega_o t)$ and $\sin(\omega_o t)$ need not be known so long as they are the same in all six channels. Likewise, it does not matter if the various band-pass and low-pass filters introduce phase-shifts, provided that these also are the same in all channels.

Having discussed alternative methods of demodulation, let us now examine the statistics of the demodulated signals, beginning with the auto-power spectra and auto-covariances. Again, we quote the spectra only at positive frequencies ; since the demodulated signal are purely real, the negative-frequency parts of their spectra are the complex conjugates of the positive-frequency parts.

The auto-spectra of $x_i(t)$ and of $y_i(t)$ are identical. They are rectangular, extending from $\omega = 0$ to $\omega = \Delta\omega/2$ with the constant value $2\, S_{ij}(\omega_o)$, and vanishing at all higher frequencies.

An inverse Fourier transform gives the corresponding covariance functions, which are of the form

$$(2.13) \qquad \langle x_i(t)\, x_i(t-\tau)\rangle = \langle x_i^2(t)\rangle \operatorname{sinc}\left(\frac{\tau\Delta\omega}{2\pi}\right)$$

Their common value at zero time-shift is

$$(2.14) \qquad \langle x_i^2(t)\rangle = \langle y_i^2(t)\rangle = \frac{\Delta\omega}{\pi}\, S_{ii}(\omega_o)$$

which is, of course, the mean power of each of these two signals.

The cross-spectrum of $x_i(t)$ with $y_i(t)$ is zero at all frequencies, so their mean cross-power is zero also :

$$(2.15) \qquad \langle x_i(t)\, y_i(t)\rangle = 0$$

These statements are true in the present context, but they would not be so in general, if the band-pass filters in the arrangement of fig. 5 were allowed to have an arbitrary response. They are true only if the function $|Y(\omega)|$ is symmetric about ω_o, which is the case for the ideal filters that we have assumed. In the arrangement of fig. 6, the required symmetry is achieved automatically, whatever the response of the low-pass filters may be. This is another advantage of the latter arrangement.

Finally, we need the cross-spectrum of the signal $x_i(t)$ or $y_i(t)$, with the signal $x_j(t)$ or $y_j(t)$. The cross-spectrum of the pair $x_i(t)$ and $x_j(t)$, which is equal to that of the pair $y_i(t)$ and $y_j(t)$, is again rectangular, extending from zero frequency to $\Delta\omega/2$, with the constant value $2\, R_e[S_{ij}(\omega_o)]$. The total mean cross-powers are

$$(2.16a) \qquad \langle x_i(t)\, x_j(t) \rangle = \langle y_i(t)\, y_j(t) \rangle = \frac{\Delta\omega}{\pi} R_e \left[S_{ij}(\omega_o)\right]$$

and, using (2.9), we have

$$(2.16b) \qquad \frac{\Delta\omega}{\pi} R_e\left[S_{ij}(\omega_o)\right] = \frac{1}{2} R_e\left[C_{ij}\right] = \alpha_{ij}, \text{ say.}$$

The cross-spectrum of the pair $x_i(t)$ and $y_j(t)$, which is equal to minus the cross-spectrum of $y_i(t)$ and $x_j(t)$, is of similar form, with the value $2\, I_m\left[S_{ij}(\omega_o)\right]$, and the cross-powers are

$$(2.17a) \qquad \langle x_i(t)\, y_j(t) \rangle = - \langle y_i(t)\, x_j(t) \rangle = \frac{\Delta\omega}{\pi} I_m\left[S_{ij}(\omega_o)\right]$$

and we also have

$$(2.17b) \qquad \frac{\Delta\omega}{\pi} I_m\left[S_{ij}(\omega_o)\right] = \frac{1}{2} I_m\left[C_{ij}\right] = \beta_{ij}, \text{ say.}$$

Note that $\alpha_{ji} = \alpha_{ij}$, but that $\beta_{ji} = -\beta_{ij}$. The various expressions quoted above for the auto-powers and cross-powers, in terms of the spectral matrix elements, are summarized in table 1.

They can be summarized more elegantly, in terms of the complex covariances of the complex envelopes, simply by writing

$$(2.18) \qquad \langle z_i(t)\, z_j^*(t) \rangle = C_{ij}$$

This equation follows from the previous ones because

$$(2.19) \qquad \langle z_i(t)\, z_j^*(t) \rangle = \langle x_i(t)\, x_j(t) + y_i(t)\, y_j(t) \rangle + i \langle x_i(t)\, y_j(t) - y_i(t)\, x_j(t) \rangle$$

These results form the bases for another analogue method for estimating the covariances ; it is illustrated in fig. 7A for the auto-covariances, and in fig. 7B for the cross-covariances. Note that each covariance can be estimated in two ways, that both ways are implemented, and that the two results are combined ; the cross-products of the low-frequency signals are added or subtracted as appropriate, i.e. as on the right-hand side of eq. (2.19). The reason for doing this is that their errors are uncorrelated, so adding them reduces the mean fractional error by a factor of $2^{1/2}$. Then, apart from the fact that the outputs are larger by a factor of 2, this method is equivalent in all respects to that of fig. 4.

We now introduce a new symbol to represent the set of 12 filtered and demodulated signals. We write

$$v_i(t) \equiv \begin{cases} x_{(i+1)/2}\ (t) & \text{if } i \text{ is odd} \\ y_{i/2}\ (t) & \text{if } i \text{ is even} \end{cases} \tag{2.20}$$

where the subcript i runs from 1 to 12. In ascending order, the 12 quantities v_i are x_1, y_1, x_2, y_2..... x_6, y_6. The signals $x_i(t)$ and $y_i(t)$ are now regarded as components of a 12 dimensional <u>data vector</u> $\vec{v}(t)$. This move is analogous to the introduction, in paper I, of the 6-dimensional generalized electric field vector $\vec{E}$ (see also equation 2.2) ; as before, it is made solely to simplify the algebra, and has no physical significance.

Moreover, we introduce a 12 x 12 <u>data matrix</u> $\mathbf{V}(t)$, the elements of which are products of the components of the data vector, and we define a matrix $\mathbf{\Lambda}$ such that

$$\Lambda_{ij} = \langle V_{ij}(t) \rangle = \langle V_i(t)\ V_j(t) \rangle \tag{2.21}$$

These covariances specify the statistical properties of the $v_i(t)$. Their values are given in table 2, using the symbols defined by equations (2.16) and (2.17) and listed in table 1. There are 144 of them, but only 33 are mutually independent.

As an alternative to the purely real 12-component data vector $\vec{\mathbf{v}}(t)$, we could use the 6-component complex data vector $\vec{z}(t)$, the components of which are the complex envelopes of the six narrow-band signals. Equally, we can define a complex data matrix $\mathbf{Z}(t)$, such that

$$Z_{ij}(t) = z_i(t)\, z_j^*(t) \tag{2.22}$$

The statistics of these quantities are specified by the complex covariance matrix $\mathbf{C}$, in accordance with (2.18).

In later sections, for clarity, we prefer to introduce new concepts in terms of the real data vector $\vec{v}(t)$ and matrix $\mathbf{V}(t)$; also we use the data in these forms in the numerical analysis. But, for conciseness, we shall often rewrite the analytic results in terms of the complex data vector $\vec{z}(t)$ and matrix $\mathbf{Z}(t)$.

3.4. SAMPLING

The final step in the data reduction is to sample the 12 low-frequency signals $v_i(t)$. Since each of them is limited to the band $|\omega| < \Delta\omega/2$, it can be replaced, without loss of information, by a sequence of samples taken from it regular intervals, provided that the sampling frequency $f_s = \omega_s/2\pi$ is greater or equal to the Shannon frequency $\Delta\omega/2\pi$. The sampling process consists of replacing the continuous signal $v_i(t)$ by the infinite sequence of discrete samples $v_i(t_n)$, where

$$t_n = n/f_s \tag{2.23}$$

and n takes all real integer values. The signal $v_i(t)$ can be reconstructed from the sequence $v_i(t_n)$ by means of the interpolation formula

$$(2.24) \qquad v_i(t) = \sum_{-\infty}^{\infty} v_i(tn)\ \mathrm{Sinc}\ [f_s(t - t_n)]$$

which demonstrates their equivalence.

The statements above apply to signals that are not limited in time, but those that we encounter in practice are of finite duration. We shall suppose that the interval over which the signals are observed is of duration $T = 2N/f_s$, and extends from the origin of time $t_o = 0$ to the instant $t_{2N} = T$. Including those from t_o and t_{2N}, altogether 2N + 1 samples of each signal are taken during this interval. For such signals, we shall use the interpolation formula

$$(2.25) \qquad v_i(t) \simeq \sum_{n=0}^{2N} v_i(t_n)\ \mathrm{sinc}\ [f_s(t - t_n)]$$

which is only approximate, as can be seen by comparing it with (2.24). The approximation is worst at the beginning and end of the sequence, and best in the middle. On the average, the errors are negligible so long as $N \gg 1$. Correspondingly, we shall use the following approximation for the <u>time average</u> auto-powers and cross-powers of the signals observed during the interval T :

$$(2.26) \qquad \overline{v_i(t)v_j(t)} \equiv T^{-1}\int_o^T v_i(t)v_j(t)dt \simeq (2N+1)^{-1}\sum_{n=0}^{2N} v_i(t_n)v_j(t_n)$$

The statistics of the samples are related in simple and evident ways to those of their parent signals : in particular, their ensemble average auto-powers and cross-powers are the qantities Λ_{ij} defined by (2.21) and listed in table 2. Moreover, the covariance

of pairs of samples of the same signal, separated by an interval Δn in their ordinal number n, is given by (2.13) with $\tau = \Delta n/f_s$; the cross-covariances of samples of two different signals vary with τ in the same way.

From these last results, it is easy to see that if the sample frequency f_s is equal to the Shannon frequency, then the set of values of the 12 signals $v_i(t)$ at any one sample time t_n is completely independant of the corresponding set at any other sample time t_m, where $m \neq n$. This state of affairs will simplify the theoretical analysis in section 4, so henceforth we shall assume that the signals $v_i(t)$ have been sampled at the Shannon frequency.

In practice, when the samples are taken, they are immediately converted into numbers so that they can be handled by a digital computer.

4 - PARAMETER ESTIMATION

4.1. CHOICE OF AN ESTIMATOR

Given a parametric model of the system under study, the problem of estimating its parameters is usually dealt with in two steps : firstly, we define what is sometimes called an "objective function" (BARD, 1974), that measures the extent to which the data agree (or disagree) with the predictions of the model ; secondly, we seek the values of the parameters that jointly maximize (or minimize) the objective function. This procedure constitutes the "estimator". Of course, the need remains to assess the validity of the model, both in the absolute sense and in comparison with any rival models, and also to determine the accuracy of the estimates of its parameters. The present sub-section is concerned only with the first step in the estimation procedure, namely the definition of an objective function.

The choice of the objective function depends on our prior knowledge about the statistical properties of the data. If we have no information at all on this subject, then our choice is arbitrary. For simplicity, it is usual to take the sum of the squares of the residuals, i.e. of the differences between the true values of the data and those predicted by the model. The minimization of this quantity constitutes the "least-squares estimator". On the other hand, if the probability density function of the data is known, then we prefer the likelihood function, which is proportional to the probability of obtaining the data when the parameters have a given set of values. This quantity has to be maximized, so we speak of the "maximum-likelihood estimator". Finally, if a prior probability can be assigned to any given set of parameter values, then we take as objective function the corresponding posterior probability, which - according to Bayes' theorem - is proportional to the product of the prior probability and of the likekihood function. The maximization of the posterior probability constitutes the "Bayes estimator".

The assumptions made in § 2.3 about the statistics of the field data place the present problem in the second category above, so we adopt the likelihood as our objective function. In § 4.2 below, we quote the general definition of this function, and derive the expression for it in our particular case. Then, in § 4.3, we explain the numerical method which we use to maximize it, and so to obtain the most likely set of parameter values for the given model. This will conclude section 4, but not the discussion of the likelihood function which has further uses : it also enables us to compare the relative merits of rival models, to test the validity of the best model, and to assess the accuracy of the parameter estimates. These topics will be discussed in section 5.

4.2. LIKELIHOOD AND SUPPORT FUNCTIONS

The concept of likelihood is related to that of probability (for a discrete variable), this being understood in the usual sense of the relative frequency of occurrence in a statistical ensemble. In the context of the analysis of stochastic experimental data (which momentarily we suppose to be discrete variables, though this is not so in the present case), the ensemble consists of all the possible sets of data that the experiment might have generated in the circunstances. The actual set of experimental data is a particular - but, we hope, representative - member of this ensemble. So as to be able to analyze the data, we assume that the circumstances of the experiment can be described by some physical model involving a finite number of paramaters. These, if the model is correct, suffice to specify the data statistics, i.e. the properties of the ensemble. The aim of the analysis is to estimate the set of parameters from the set of experimental data. Let H be the hypothesis that the parameters take a given set of values, and call the data set D. Let P(D/H) be the probability of obtaining the data D, granted the hypothesis H ; this probability is given by our statistical model. Then the <u>likelihood</u> L(H/D) of the hypothesis H, in the light of the data D and assuming this specific model, is defined as being proportional to P(D/H), the constant of proportionality

being arbitrary (EDWARDS, 1972).

When dealing - as we are here - with continuous rather than with discrete data, the concept of probability has to be replaced by that of probability density, and likelihood by the likelihood function. We write P(D/H) dD for the probability that the values of the data lie in the interval from D to D + dD, granted the hypothesis H ; the symbol P(D/H) now stands for the probability density. Then the likelihood function is defined as being proportional to P(D/H), with the constant of proportionality arbitrary.

We shall take this arbitrary constant as unity, in which case any expressions that we may derive for P(D/H) and for L(H/D) will be identical. The difference is that P(D/H) is a function of D, whereas L(H/D) is a function of H : when manipulating the probability density P(D/H) we consider the hypothesis H to be fixed and D to vary, while with the likelihood function L(H/D) the data are fixed and H is the variable.

But L(H/D), like P(D/H), still refers to an interval dD of the data set. This feature - among others - distinguishes the likelihood function from the posterior probability density, used in Bayesian estimation, which refers to an interval of the set of parameters of the model.

To apply these considerations to the problem in hand, we suppose that we have a parametric model $F_m(\omega_o,\theta,\phi;\vec{\psi})$ describing how the wave distribution function at the frequency ω_o depends on the angle variables θ and ϕ, and that this model is characterized by a number p of parameters $\psi_1,\psi_2,.....\psi_p$ which we denote collectively by the symbol $\vec{\psi}$: we group all the parameters together in this p-component parameter vector $\vec{\psi}$. Let H be the hypothesis that these parameters take the values $\hat{\psi}_1,\hat{\psi}_2.....\hat{\psi}_p$. This hypothesis H is to be assessed in the light of the data at our disposal, namely the 2N + 1 values of the 12-component data vector $\vec{v}(t_n)$.

First let us consider the probability density of individual values of this vector, i.e. the joint probability density of the 12 quantities $v_i(t_n)$ at a fixed t_n. It will be convenient to write the expression for this density in the customary matrix notation, which we now recapitulate. Bold-face capital letters denote square matrices ; thus we use the symbol $\mathbf{\Lambda}$ for the matrix whose elements are the covariances Λ_{ij} defined by (2.21). We use bold-face small letters to denote column vectors : thus $\mathbf{v}_n$ will be the column vector whose elements are the quantities $v_i(t_n)$. Now, according to our assumptions, each of these quantities is a random variable with a probability density obeying a normal (Gaussian) law. Hence the joint probability density of the 12 quantities $v_i(t_n)$ is a 12-variate normal distribution. In matrix notation, it is

$$P(\mathbf{v}_n/H) = \frac{1}{(2\pi)^6 |\mathbf{\Lambda}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{v}_n \mathbf{\Lambda}^{-1}\mathbf{v}_n\right) \tag{2.27}$$

where $|\mathbf{\Lambda}|$ is the determinant of the matrix $\mathbf{\Lambda}$, and $\mathbf{\Lambda}^{-1}$ is the inverse of this matrix. $\mathbf{v}'_n$ is a row vector, the transpose of the column vector $\mathbf{v}_n$ (KENDALL and STUART, 1969a).

The derivation of equation (2.27) assumes that $|\mathbf{\Lambda}|$ does not vanish. Actually this determinant would vanish in either of the following circunstances : firstly, if any linear relationships existed between the components $v_i(t_n)$ of the data vector (in other words, if any items of the data were redundant) ; secondly, if any component was zero. It should be understood that for $|\mathbf{\Lambda}|$ to vanish, these conditions would have to apply necessarily, for physical reasons, at all times, to all possible data sets : the fact that they might occur fortuitously in a particular data set, besides being very unlikely, has nothing to do with the question.

If $|\mathbf{\Lambda}|$ vanished for either of these reasons, then one or more components of the data vector would have to be eliminated, so as to obtain a linearly-independent set of non-zero data. The data vector $\vec{\mathbf{v}}$ and the matrix $\mathbf{\Lambda}$ would then be of lower order than before, so the theory would have to be modified accordingly, but this is easy. The determinant $|\mathbf{\Lambda}|$, of lower order, would be finite.

Moreover, $|\mathbf{\Lambda}|$ is always positive. Provided that its determinant is finite, any covariance matrix is positive definite (WILKS, 1962). This general result follows from the inequalities $\Lambda_{ij}^2 < \Lambda_{ii}\,\Lambda_{jj}$ that hold between the covariances. For any positive definite matrix, the determinant is positive.

After this discussion of the determinant $|\mathbf{\Lambda}|$, and now making the additional assumptions that the 12 components of the data vector are linearly independent and non-zero, we resume the derivation of the likelihood function.

The formula (2.27) for the probability density may be expressed in various other ways. For instance, the exponent may be written out in terms of the components of the data vector, using the fact that

$$\mathbf{v}_n' \mathbf{\Lambda}^{-1} \mathbf{v}_n = \sum_{i=1}^{12} \sum_{j=1}^{12} \frac{\text{cof}\ (\Lambda_{ij})}{|\mathbf{\Lambda}|}\ v_i(t_n) v_j(t_n) \tag{2.28}$$

where cof (Λ_{ij}) is the cofactor of Λ_{ij} in the determinant $|\mathbf{\Lambda}|$. Alternatively, the exponent can be expressed in terms of the data matrix $\mathbf{V}(t_n)$ defined in § 3.3. If we modify the notation slightly by setting $\mathbf{V}_n \equiv \mathbf{V}(t_n)$, we have

$$\mathbf{v}_n' \mathbf{\Lambda}^{-1} \mathbf{v}_n = \text{tr}\ (\mathbf{\Lambda}^{-1} \mathbf{V}_n) \tag{2.29}$$

where the symbol tr denotes the trace of the bracketed matrix product. Yet again, we can use the complex data vector and matrix. We let $\mathbf{z}_n$ be the column-vector representation of the vector $\vec{\mathbf{z}}(t_n)$ previously defined, set $\mathbf{Z}_n \equiv \mathbf{Z}(t_n)$, and exploit the following results which are proved in the appendix A :

$$t_r(\mathbf{\Lambda}^{-1} \mathbf{V}_n) = 2\, t_r(\mathbf{C}^{-1} \mathbf{Z}_n) \tag{2.30}$$

$$(2.31) \qquad |\mathbf{\Lambda}|^{1/2} = 2^{-6} |\mathbf{C}|$$

$\mathbf{C}$ is the complex covariance matrix. Then the formula for the probability density becomes

$$(2.32) \quad P(z_n/H) = \frac{1}{\pi^6 |\mathbf{C}|} \exp\left[-\mathbf{z}_n' \mathbf{C}^{-1} z_n\right] = \frac{1}{\pi^6 |C|} \exp\left[- t_r(\mathbf{C}^{-1} \mathbf{z}_n)\right]$$

Now let us derive the probability density of the complete data set D, comprising the 2N + 1 values of the data vector $\vec{v}(t_n)$, here called $\mathbf{v}_n$. As we saw in § 3.4, the values of this vector at different times t_n are wholly independent of one another, so the joint probability density of the complete set is simply the product of the individual probabilities :

$$(2.33) \quad P(D/H) = \left[\frac{1}{(2\pi)^6 |\mathbf{\Lambda}|^{1/2}}\right]^{2N+1} \exp\left[-\frac{1}{2} \sum_{n=o}^{2N} \mathbf{v}_n' \mathbf{\Lambda}^{-1} \mathbf{v}_n\right]$$

Proceeding as above, the sum in the exponent can be written out as follows :

$$(2.34) \quad \sum_{n=o}^{2N} \mathbf{v}_n' \mathbf{\Lambda}^{-1} \mathbf{v}_n = \sum_{i=1}^{12} \sum_{j=1}^{12} \frac{\mathrm{cof}(\Lambda_{ij})}{|\mathbf{\Lambda}|} \left[\sum_{n=o}^{2N} v_i(t_n)\, v_j(t_n)\right]$$

On the right-hand side of this equation, the sum in the square brackets can be rewritten as

$$(2.35) \quad \sum_{n=o}^{2N} v_i(t_n)\, v_j(t_n) = (2N + 1)\, \overline{v_i(t_n)\, v_j(t_n)}$$

where the second factor on the right-hand side is the average of the product concerned over the time interval T. This time average constitutes an estimate of the covariance Λ_{ij} defined by (2.21) ; in fact,

to the accuracy of the approximation (2.26), it is the maximum-likelihood estimate of Λ_{ij}, though we small not digress to prove this point. Recognizing it, and following the convention adopted at the end of § 2.3, we define

$$\hat{\Lambda}_{ij} \equiv (2N + 1)^{-1} \sum_{n=o}^{2N} v_i(t_n)\, v_j(t_n) \tag{2.36}$$

and we introduce the symbol $\hat{\boldsymbol{\Lambda}}$ for the 12 x 12 matrix of these quantities. (Warning. The caret superscript often symbolizes the adjoint of a matrix : this is not our usage). With the definition (2.26), the sum in the exponent on the right-hand side of (2.33) becomes

$$\sum_{n=o}^{2N} \mathbf{v}_n' \boldsymbol{\Lambda}^{-1} \mathbf{v}_n = (2N + 1)\, t_r(\boldsymbol{\Lambda}^{-1} \hat{\boldsymbol{\Lambda}}) \tag{2.37}$$

Hence our result for the probability density of the complete data set is

$$P(D/H) = \left[\frac{1}{(2\pi)^6 |\boldsymbol{\Lambda}|^{1/2}}\right]^{2N+1} \exp\left[-\left(N + \tfrac{1}{2}\right) t_r(\boldsymbol{\Lambda}^{-1} \hat{\boldsymbol{\Lambda}})\right] \tag{2.38}$$

As before, it is instructive to rewrite this result in terms of the complex data vector and matrix. It then becomes

$$P(D/H) = \left[\frac{1}{\pi^6 |\mathbf{C}|}\right]^{2N+1} \exp\left[-(2N+1)\, t_r(\mathbf{C}^{-1} \hat{\mathbf{C}})\right] \tag{2.39}$$

where $\hat{\mathbf{C}}$ is our estimate of the covariance matrix. The elements of $\hat{\mathbf{C}}$ are defined as follows :

$$\hat{C}_{ij} \equiv (2N + 1)^{-1} \sum_{n=o}^{2N} Z_i(t_n)\, Z_j^*(t_n) \tag{2.40}$$

By analogy with (2.9), we define our estimate $\hat{\mathbf{S}}$ of the spectral matrix $\mathbf{S}$ by setting

$$\hat{C}_{ij} \equiv \frac{2\Delta\omega}{\pi} \hat{S}_{ij} \tag{2.41}$$

Again, this is in fact the maximum-likelihood estimate. In these terms, our expression for the probability density takes the final form

$$P(D/H) = \left|\frac{1}{(2\Delta\omega)^6 |\mathbf{S}|}\right|^{2N+1} \exp\left[-(2N+1)\, t_r(\mathbf{S}^{-1}\hat{\mathbf{S}})\right] \tag{2.42}$$

In this expression, $\mathbf{S}$ is fixed while $\hat{\mathbf{S}}$ contains the variables derived from the 12 x (2N + 1) items of information $v_i(t_n)$ that comprise the data D.

The results (2.39) - (2.42) are the formal evidence for various facts that have been announced previously without proof. Firstly, (2.42) shows that the data enter into the expression for P(D/H) in reduced form, as the estimate $\hat{\mathbf{S}}$ of the spectral matrix $\mathbf{S}$. This confirms the statement, made in § 3.1, that the first step in the analysis of the 6-component random field data is one of data reduction, in which they are used to estimate the elements of the spectral matrix. Accordingly, from here on we shall refer to the 36 elements $\hat{\mathbf{S}}_{ij}$ of the matrix $\hat{\mathbf{S}}$ as the <u>reduced data</u>.

Secondly, the equations (2.39) and (2.40), together with (2.19), show that the process of data reduction involves combining the time-averaged products of the sampled low-frequency signals in pairs, in the ways suggested by (2.16a) and (2.17a) and illustrated in fig. 7.

Having derived the expression for the probability density in terms of the reduced data, the likelihood function follows immediately. As already explained, our expression for L(H/D) is identical

with (2.42), but now $\hat{\mathbf{S}}$ is fixed while $\mathbf{S}$ is variable, being a function of the p parameters ψ_k that characterize the model. The most likely set of values of these parameters is that which maximizes L.

Rather than maximize the likelihood function as such, it is more convenient to maximize the natural logarithm of L, which EDWARDS (1972) calls the support function. Apart from an additive term independent of $\vec{\psi}$, which we ignore, the support function is

$$(2.43) \quad l(\vec{\psi}) = -(2N+1)\left[\ln|\mathbf{S}| + t_r(\mathbf{S}^{-1}\hat{\mathbf{S}})\right]$$

where $\mathbf{S} = \mathbf{S}(\vec{\psi})$. This result will also be written as

$$(2.44) \quad l(\vec{\psi}) = -(2N+1)\,M(\vec{\psi})$$

$M(\vec{\psi})$ expresses the dependence of the support function on the model.

4.3 DERIVATION OF THE ESTIMATES

The most likely values of the parameters of a given model $F_m(\omega_o,\theta,\phi;\vec{\psi})$ are those that jointly maximize the support function. In discussing the process of maximization, it is instructive to begin with a restricted class of model - that of the linear models - and, within this class, to deal firstly with the special case of a 33-parameter model, i.e. one in which the number p of free parameters is equal to the number of independent elements of the spectral matrix. In so doing, we should emphasize that the models of which the identification forms the subject of the present paper, and which are supposed to have some physical basis, are most unlikely to be of the simple linear type : the linear models are of some use in solving the inverse problem, which will be the subject of the next paper in this series. In the present paper, nevertheless, the study of this simple class is helpful because it clarifies a number of general points.

A linear model is one of the form

$$F_m(\omega_o,\theta,\phi;\vec{\psi}) = \sum_{k=1}^{p} \psi_k(\omega_o,\theta,\phi)\ F_{mk}(\omega_o,\theta,\phi) \tag{2.45}$$

in which the wave distribution function $F_m(\vec{\psi})$ is the sum of p linearly independent known functions F_{mk}, weighted by the parameters ψ_k ; these weights are the same for the two magneto-ionic modes, denoted by the subcript m. For the random wave field described by this distribution function, the spectral matrix is of the form

$$\mathbf{S} = \sum_{k=1}^{p} \psi_k\ \mathbf{S}_k \tag{2.46}$$

where $\mathbf{S}_k$ is the spectral matrix corresponding to F_{mk} ; its elements are obtained by inserting F_{mk} into equation (2.1). Since the functions F_{mk} are linearly independent, so are the matrices $\mathbf{S}_k$. We first consider a linear model with p = 33.

The peculiarity of this model is that it can yield a spectral matrix $\mathbf{S}$ in exact agreement with the reduced data $\hat{\mathbf{S}}$. The matrix equation

$$\sum_{k=1}^{p} \psi_k\ \mathbf{S}_k = \hat{\mathbf{S}} \tag{2.47}$$

represents 33 independent linear equations involving 33 unknowns, so it has a unique solution which can be found by matrix inversion in the usual way. The corresponding value of the support function is

$$l_{max} = -\ (2N + 1)\ \left[\ln|\hat{\mathbf{S}}| + 6\right] \tag{2.48}$$

This result follows because $\hat{\mathbf{S}}$ is a 6 x 6 matrix, so $t_r\ (\hat{\mathbf{S}}^{-1}\hat{\mathbf{S}}\ ) = 6$.

We now justify our choice of the symbol 1_{max} by proving that this value of the support function is the maximum value, i.e. that

$$1(\vec{\psi}) \leqslant 1_{max} \tag{2.49}$$

for all $\vec{\psi}$, or in other words that

$$-\ 1_n|\mathbf{S}(\vec{\psi})|\ -\ t_r\ [\mathbf{S}^{-1}(\vec{\psi})\ \hat{\mathbf{S}}] \leqslant -\ 1_n|\hat{\mathbf{S}}|\ -\ 6 \tag{2.50}$$

Now, for any non-singular square matrices $\mathbf{A}$ and $\mathbf{B}$ of equal rank, we known that $t_r\ (\mathbf{AB}) = t_r\ (\mathbf{BA})$, and that $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$. Hence the inequality (2.43) is equivalent to

$$-\ 1_n|\hat{\mathbf{S}}\ \mathbf{S}^{-1}(\vec{\psi})|\ +\ t_r\ |\hat{\mathbf{S}}\ \mathbf{S}^{-1}(\vec{\psi})] > 6 \tag{2.51}$$

Moreover, let λ_k be the k'th eigenvalue of the matrix $\hat{\mathbf{S}}\ \mathbf{S}^{-1}(\vec{\psi})$. From certain well-known results that relate the trace and the determinant of any square matrix to its eigenvalues, we have that

$$t_r[\hat{\mathbf{S}}\ \mathbf{S}^{-1}(\vec{\psi})] = \sum_{k=1}^{6} \lambda_k \tag{2.52}$$

$$|\hat{\mathbf{S}}\ \mathbf{S}^{-1}(\vec{\psi})| = \prod_{k=1}^{6} \lambda_k \tag{2.53}$$

so (2.51) can be rewritten as

$$(2.54) \qquad \sum_{k=1}^{6} (\lambda_k - 1_n \lambda_k) \geq 6$$

Now, as shown in the appendix (relations (2.83) to (2.86)) the eigenvalue of the product of two hermitian positive definite matrices is real and positive and the following inequality must hold :

$$(2.55) \qquad \lambda_k - 1_n \lambda_k \geq 1$$

This implies that (2.54) is always satisfied, which in turn proves (2.49). Hence, for the 33-parameter linear model, the value of $\vec{\psi}$ that yields $\mathbf{S}(\vec{\psi}) = \hat{\mathbf{S}}$ is also the maximum-likelihood estimate of $\vec{\psi}$, as one might expect.

But the inequality (2.49) has more far-reaching consequences, because the proof that we have just given in no way involves the nature of the model. Therefore the value 1_{max}, which the support function attains when the theoretical spectral matrix is in exact agreement with the reduced experimental data, is a maximum maximorum or absolute maximum value, independant of the model. If, with a given model, we find a solution for $\vec{\psi}$ that yields such agreement, then we say that this solution is an optimum, in the sense that there cannot be any other solution with a greater likelihood.

Unfortunately this in itself does not guarantee that an optimum solution exists, nor, if one exists, that it is unique.

A trivial example of a model that leads to ambiguous optimum solutions is the linear model with $p > 33$. Here, since the number of unknowns exceeds the number of reduced data, there is no unique optimum solution. The matrix equation (2.47) yields p-33 relations between the ψ_k, defining a continuous range of parameter sets that are all optimum in our sense of the word. Obviously such models are of no use whatsoever.

An example of a model for which no optimum solution exists in general is the linear model with $p < 33$. In this case the matrix equation (2.47) generally has no solution, so we must seek the set of parameter values ($\vec{\psi}$, say) that maximizes the support function $1(\vec{\psi})$, while recognizing that $1(\vec{\psi})$ will generally be less than the value 1_{max} given by (2.48). Here the linearity of the model does not help : the problem is essentially the same as it would be in the case where the model is non-linear, to which we shall come shortly.

But, before doing so, let us just summarize the facts that consideration of the linear models has helped to clarify : they are that when seeking a set of parameters that maximizes the support function, we need first to know whether any such set exists, and if so, whether it is optimum and whether it is unique. These are general points, applying to models of all types.

Accordingly we now turn to the general case, including the more realistic non-linear models. Formally, the most likely set of values of the parameters is a solution of the simultaneous equations.

$$\frac{\partial\, 1(\vec{\psi})}{\partial\, \psi_k} = 0 \qquad (k = 1,2.....p) \tag{2.56}$$

Explicitly, the equations to be solved are

$$\frac{1}{|\mathbf{S}(\psi)|}\ \frac{\partial |\mathbf{S}(\vec{\psi})|}{\partial\, \psi_k} + t_r \{ \frac{\partial |\mathbf{S}^{-1}(\vec{\psi})|}{\partial\, \psi_k}\ \hat{\mathbf{S}}\} = 0 \tag{2.57}$$

with the subscript k running from 1 to p. The right solution is the one corresponding to the highest maximum of $1(\vec{\psi})$.

This set of p equations is non-linear, even when the model is linear, for although, in such a case, the matrix $S(\vec{\psi})$ is a linear function of the parameter ψ_k, this is not true of its determinant

nor of its inverse.

In view of their non-linearity, of their complexity due to the size of the matrix **S**, and of their number (when p is large), we like to avoid having to solve these equations. Instead, we search for the maxima of $1(\vec{\psi})$ directly, using the iterative method of POWELL (1964). This involves calculating $1(\vec{\psi})$ for sucessive values of $\vec{\psi}$, which are chosen in such a way that 1 usually increases from one value to the next. Provided that $1(\vec{\psi})$ is a well-behaved function, this iteration converges to one of its maxima.

When a maximum is found, we have still to make sure that it is unique, or if not, that it is the highest maximum. Here again, the only way to proceed is to search numerically for others. We restart the iteration at diverse values of $\vec{\psi}$, all distant from the one that we have already found, in order to see whether it can converge to other maxima ; clearly this procedure is not infallible. If no further maxima are found, then we take our estimate of $\vec{\psi}$ from the one that we have already.

If, on the other hand, there are several maxima, our conduct depends on the heights of the secondary ones compared with that of the highest. If none are of comparable height, then we pick the highest. But if one or more are equal to the highest, then we recognize that several solutions are equally likely and that, in the absence of further information, we have no reason to prefer one to another. Finally, if there are one or more of only slightly lesser likelihood, then again we must retain them all in addition to the highest, while bearing in mind their relative likelihoods.

The bases for such judgements are explained more fully in the next section, which is concerned with how we assess the results of the analysis described above. The topics discussed include the relative merits of several different models, the degree of validity of the best model, and the accucary of the estimates of its parameters.

5 - ASSESSMENT OF THE RESULTS

5.1. COMPARISON OF SEVERAL MODELS

In the previous section, we assumed that our model of the wave distribution function was given, and we saw how to derive the most likely estimates of its parameters. Now let us assume instead that several alternative models can be envisaged for the WDF, and that we have no prior grounds for ranking them in any order of preference. Our problem is to assess the relative merits of the rival models, and eventually to select one as the best and reject the others. To solve this problem, we appeal once more to the likelihood concept. But note that whereas, in a parameter estimation problem, the hypothesis to be tested relates to sets of values of the parameters of a given model, which usually are continuous variables, in a model selection problem the hypothesis concerns the models themselves, which are discrete entities ; in other words, the latter is a problem of state estimation rather than of parameter estimation (EYKHOFF, 1974).

For the purpose of comparison of several models, we suppose that the parameter estimation process described in section 3 has already been performed for each of them, using the same data D in every case, so that for each we know the most likely set (or sets) of values of its parameters, together with the associated value of the likelihood function or of the support function. Let us label the rival models 1, 2, 3.....etc., and let the associated likelihoods be L_1, L_2, L_3...etc. and the supports l_1, l_2, l_3.....etc. Then the basis for assessing the relative merits of the different models is simply the ratios of their likelihoods, or - what is equivalent - the differences of their supports (EDWARDS, 1972).

How should these likelihood ratios be interpreted ? To simplify the question, let us suppose that there are only two models, and that $L_1 > L_2$. Then, by the definition of the likelihood function quoted in § 4.2, the ratio L_1/L_2 is the relative frequency with which

the two models would give rise to data in the interval from D to D + dD, if the experiment from which the actual data D originally came were to be repeated very many times in the same circunstances, i.e. with the same true WDF, whatever this may be. The fact that it would reproduce the data D this much more frequently is the reason for preferring the model 1.

As the ratio L_1/L_2 increases from unity, we are first inclined to say that models 1 and 2 are about equally likely, then that there is either a weak or a strong presumption in favour of model 1, and finally that model 2 must be rejected entirely. Of course, any such categoric interpretation of the values of a continuous variable is quite arbitrary, but nevertheless it has a certain practical convenience. In particular, the decision to eliminate unlikely models saves time, though occasionally it may be mistaken. The present authors define these categories by analogy with the usual practice of statisticians with regard to a random variable (x, say) that has a probability density given by the normal (Gaussian) law

$$p(x) \propto \exp\left[-(x-xm)^2/2\sigma^2\right] \tag{2.58}$$

Where x_m is the mean value and σ the standard deviation. They usually consider that a deviation of x from x_m in the range from o to σ is insignificant, that one in the range from σ to 2σ is weakly significant, that one between 2σ and 3σ is strongly significant, and finally that beyond 3σ the probability of a deviation occurring by chance is negligible. The values of the exponent in (2.58) corresponding to the limits of these ranges are 0, 1/2, 4/2, and 9/2 respectively. We define our categories by dividing the difference of support

$$l_1 - l_2 = l_n(L_1/L_2) \tag{2.59}$$

into the same ranges, as shown in table 3 ; the ranges of the likelihood ratio L_1/L_2 are then the same as those of the probability density ratio $p(x_m)/p(x)$ in the case of the Gaussian random variable.

(Incidentally, we employ similar criteria to distinguish between different maxima in the likelihood function for any single model. It will be recalled that this question arose at the end of § 4.3).

The steps described above correspond to the block labelled "Model Selection" in fig. 2, which illustrates the case where two models contend ; the extension to a greater number is obvious. Of course, these steps are unnecessary if there is only one model (fig. 1). In any case, the next step is to check the validity of the model selected.

5.2. VALIDITY OF THE BEST MODEL

Having thus picked the best model among those at our disposal, we must examine whether it is valid or not. A model is held to be valid if, firstly, it is physically realistic, and secondly, if it fits the data reasonably well.

We assume that the first point has been taken care of in the construction of the model, so that, for instance, the WDF is nowhere negative.

To examine the second point, we follow the same reasoning as in the previous sub-section. We compare the value $l(\vec{\psi})$ of the maximum of the support function for the best model to the optimum value l_{max}, as given by (2.41), which corresponds to perfect agreement between the model and the reduced data.

We interpret the difference $l_{max} - l(\vec{\psi})$ in accordance with table 3. If it exceeds our limit of 9/2, then this "best" model must be rejected, and no further progress can be made in the data analysis by model identification until some even better model has been found (see figs. 1 and 2).

(Incidentally, when analyzing data with a single model by the method described in section 4, we employ similar criteria to decide whether or not we have succeded in finding the highest maximum of the support function. If the apparent maximum value of this function turns out to be much less than l_{max}, yet there is no other reason for discrediting the model, then before rejecting it we first search carefully for another and higher maximum).

In practice, the selection of the best model and the test of its validity are made simultaneously, since both these steps involve a simple inspection of the support functions that issue from the parameter estimation.

5.3. ACCUCARY OF THE ESTIMATES

If the best model turns out to be valid, then there remains just one more step in the analysis, namely the assessment of the accuracy of the estimates of its parameters. The accucary of the parameter vector $\vec{\psi}$ determines in turn the accuracy with which the wave distribution function $F_m(\omega_o,\theta,\phi\ ;\ \vec{\psi})$ is known at any point (θ_o,ϕ_o).

First, the question of what is meant by the "accuracy" of $\vec{\psi}$ has to be clarified. This "accuracy" is not to be confused with that of the estimation process by means of which $\vec{\psi}$ is computed from the data. On our hypotheses, given the theoretical model on the one hand and the experimental data on the other, the likelihood and support functions are perfectly defined. The only errors that can arise in locating their common highest maximum are numerical, so with proper case they can be made as small as we wish. (For the purpose of discussion, we discount momentarily the possibility of a gross error due to a secondary maximum of the support function). In reality, the accuracy of which we are speaking is that of $\hat{\psi}$ as an estimate of the true value of the parameter vector. The possibility of $\hat{\psi}$ being inaccurate in this sense is witnessed by the fact that neighbouring values of $\vec{\psi}$ may be only slightly less likely. Evidently the assessment of the accuracy of $\hat{\psi}$ involves somehow specifying the size and shape of the region of parameter space within which $l(\vec{\psi})$

is comparable with $1(\hat{\psi})$.

In the most general case, where the support function is of intricate shape, the only sufficient statement of the situation would be a table or graph of this function throughout the region where its value does not fall short of the highest maximum by more than some limit, such as 9/2 (see § 5.1). This is particularly true if the support function exhibits non-negligible secondary maxima. Moreover, we must bear in mind that the model itself may be in error ; if $1(\hat{\psi})$ for one or more rival models lies within the stated range below $1(\hat{\psi})$ for the best model, then the parts of their support functions that lie in this range should be presented also. But clearly these are counsels of perfection, since it is impracticable to tabulate or graph the support function if the number of parameters is more than about 3, as it generally is.

Fortunately, a more concise assessment of the accuracy of the estimates is possible in the frequent situation where the support function has only one significant maximum (EDWARDS, 1972). To make the method clear, let us begin with the simple case illustrated in fig. 8, where the model involves just one parameter ψ. We wish to specify the accuracy of our estimate $\hat{\psi}$ of this parameter, corresponding to the maximum of the support function $1(\psi)$. Clearly, if the peak around $\hat{\psi}$ is well-defined, it will suffice for this purpose to specify the interval $\Delta\psi$ between the two points, one on either side of the maximum, where the values of the support function are less than the maximum value by some agreed amount m, for instance 1/2 (fig. 8A).

An analytic approximation for this quantity can be obtained by fitting a parabola to the curve of the support function in the neighbourhood of its maximum (fig. 8B). That is, we approximate the peak of the support function by the quadratic.

$$1(\psi) \simeq 1(\hat{\psi}) - (\psi - \hat{\psi})^2/2w^2 \qquad (2.60)$$

The width parameter w is determined by requiring that this approximate function have the same radius of curvature as the true support function at the maximum. Hence

$$w^2 = - \left(\frac{d^2 l}{d\psi^2} \right)^{-1} \qquad (\psi = \hat{\psi}) \tag{2.61}$$

In other words, the approximation (2.60) is just the power-series expansion of $l(\psi)$ around the point $\hat{\psi}$, limited to its first two terms. (Note that the representation of the support function by a parabola is equivalent to representating the likelihood function by a Gaussian curve, and that w is the standard deviation of this Gaussian). If the width $\Delta\psi$ of the peak of the support function is measured between points m units below the maximum, then it is approximately

$$\Delta\psi \simeq 2\sqrt{2m}\, w \tag{2.61}$$

This is a reasonable approximation provided that the peak is not too asymmetrical ; in the latter case, a good tactic is to improve the symmetry by a suitable non-linear transformation of the variable ψ (EDWARDS, 1972).

It is interesting to know how the accuracy depends on the bandwidth of the original data, and on the length of time during which they were recorded. These quantities enter the expression (2.43) for the support function via the factor 2N + 1, which is the number of discrete samples that are taken of each of the low-frequency signals. From § 3.4, we have

$$2N + 1 \simeq 2N = \frac{T\Delta\omega}{2\pi} = T\,\Delta f, \text{ say,} \tag{2.63}$$

where Δf is the frequency bandwidth corresponding to the pulsatance bandwidth $\Delta\omega$. Combining this result with (2.44) and (2.62), we find

$$w = (T\,\Delta f)^{-1/2} \left(\frac{d^2M}{d\psi^2} \right)^{-1/2} \tag{2.64}$$

The second derivative of M has no systematic dependence on T or Δf, so the error in the determination of ψ is inversely proportional to the square root of the product (Time x Bandwidth).

(Similary,in the tests discussed in § 5.1 and § 5.2 above, the details of the model affect only the function $M(\psi)$. The difference of support corresponding to a given difference of M increases in direct proportion to the product $T\,\Delta f$).

This method for assessing the accuracy of the estimates can be extended to the case where the model has any number p of parameters. As before, we approximate the peak of the support function by a power-series expansion, limited to the terms of first and second order. This quadratic form may be expressed conveniently in matrix notation :

$$l(\boldsymbol{\psi}) \simeq l(\hat{\boldsymbol{\psi}}) - \frac{1}{2} (\boldsymbol{\psi} - \hat{\boldsymbol{\psi}})' \, \mathbf{B} (\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}) \tag{2.65}$$

The vectors $\boldsymbol{\psi}$ and $\hat{\boldsymbol{\psi}}$ are those previously called $\vec{\psi}$ and $\hat{\vec{\psi}}$, and $\mathbf{B}$ is the matrix of the second derivative of $l(\boldsymbol{\psi})$ with respect to the parameters, taken at the maximum ; i.e. its elements are

$$B_{ij} \equiv \frac{\partial^2 l}{\partial \psi_i \, \partial \psi_j} \qquad (\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}) \tag{2.66}$$

To derive the accuracy of the estimates, we proceed by analogy with the 1-parameter case, and consider the size and shape of the boundary in $\boldsymbol{\psi}$ space on which

$$l(\hat{\boldsymbol{\psi}}) - l(\boldsymbol{\psi}) \simeq \frac{1}{2} \, \delta\boldsymbol{\psi}' \, \mathbf{B} \, \delta\boldsymbol{\psi} = m \tag{2.67}$$

where $\delta \boldsymbol{\psi} = \boldsymbol{\psi} - \hat{\boldsymbol{\psi}}$. The bounding surface defined by this equation is a p-dimensional ellipsoid. Our discussion of it will treat the general case where the number p of parameters is arbitrary, but will be illustrated by reference to the case with only two parameters, ψ_1 and ψ_2, in which the boundary is an ellipse and can be graphed.

The interpretation of the bounding surface in terms of accuracy of the estimates is relatively simple if the matrix B is diagonal, i.e. if the mixed second derivatives of l (those with $i \neq j$) all vanish. In this case, the principal axes of the ellipsoid are parallel to those of the parameters ψ_i. Knowing this, the shape of the ellipsoid is specified completely by the lengths of its axes, and these are directly interpretable as accuracies of the corresponding parameters, exactly as in the 1-dimensional case. For each parameter, the accuracy is

$$\Delta\psi_i \simeq 2\sqrt{2m}\, w_i \quad \text{with} \quad w_i = -\left(\frac{\partial^2 l}{\partial \psi^2_i}\right)^{-1} \tag{2.68}$$

For the 2-dimensional case, this situation is illustrated in fig. 9A.

If the matrix $\mathbf{B}$ is not diagonal, the principal axes of the ellipsoid are not parallel to the axes of the parameters. A 2-dimensional situation of this type is illustrated in part B of fig. 9. For each parameter, the range of variation corresponding to the projection of the ellipse onto the appropriate axis has been taken to be the same as in part A. But it is obvious that the simple statement of these ranges is not an adequate description of the accuracies of the estimates of the parameters, as it was in the previous case. When one inquires within what ranges their true values may possibly lie, around their most likely values, the parameters must not be considered separately but jointly. To describe their range of joint variation fully, all the elements of $\mathbf{B}$ must be specified.

Of course, it would be possible to make the matrix B diagonal by a suitable change of variables. With the present variables, the axes of the p-dimensional ellipsoid defined by Eq. (2.67) are parallel to the eigenvectors of B. The new variables would be chosen so that these axes became parallel to the coordinate axes in parameter space. Then the ranges of variation of these new parameters could each be specified by a single number, as in Eq. (2.68). The disadvantage of this procedure is that the new parameters usually have no physical significance.

6 - CONCLUSION

Subject to the hypotheses made in sections 1 and 2, a method has been developed in this paper for analyzing measurements of the 6 components of a random electromagnetic wave field in a magnetoplasma, given one or several parametric models of the wave distribution function.

This method enables us, on the basis of a finite set of data, to test the validity of the models (and hence the validity of the theories that were invoked to construct them), both relative to one another and relative to the eventual unknown true model. Thus it indicates which models and theories can be considered as being acceptable, and which must be rejected.

Obviously it may happen that the state of the theory does not allow us to propose any acceptable model, in which case we would be led to consider general models that assume no prior knowledge of the form of the WDF. Alternatively, we may wish to free ourselves of all preconceptions, and to seek some description of the entire range of acceptable models. These are different forms of the general inverse problem, which will be treated in the next paper of this series.

Meanwhile, however, we should make it clear that such general models cannot replace models of the type considered in the present paper, based on phenomenological or on physical knowledge, this for the following reasons :

- known constraints such as the positivity of the WDF appear naturally in the models considered here, whereas they have to be imposed artificially on the general models ;

- parameters with physical significance are less liable to be poorly determined by the data than are arbitrarily-chosen parameters ;

- models with only a few parameters can be identified more precisely than the general models, which involve many parameters ;

- at some stage in the reasoning, physical considerations must be introduced, in any case, to pick a representative model from the infinite set of possible ones.

In fact, these two appraches to the problem of the analysis of multi-component field data are complementary rather than competitive.

The present approach, through model identification, needs to be developed further. Firstly, the theory should be extended to include the case where the data are degraded by additive noise. Secondly, it is necessary to relax the assumption, made in § 3.2, that the narrow-band filters through which the data are passed have ideal rectangular pass-bands. Finally, our method needs to be adapted to the study of random electromagnetic wave fields in situations where some of the assumptions made in this paper (and in paper I) are not applicable. For instance, for studying hydromagnetic (Alfven) waves in the solar wind, account should be taken of the Doppler effect due to the motion of the point of observation relative to the solar-wind plasma, and use should be made of measurements of the 3 components of the plasma velocity vector as well as of the 6 electromagnetic field components. For these last purposes, the theory presented in paper I also needs further development.

REFERENCES

ALLIS, W.P., BUCHSBAUM, S.J. and A. BERS, 1963, "Waves in Anisotropic Plasmas", M.I.T. Press, Boston.

AUBRY, M.P., 1968, "Some results of the FR-1 satellite experiment on the VLF wave field in the zone close to the transmitter", J. Atmos. Terr. Phys. 30, 1161-1182.

BARD, Y., 1974, "Nonlinear Parameter Estimation", Academic Press, New York.

EDWARDS, A.W.F., 1972, "Likelihood", Cambridge Univ. Press, London.

DYKHOFF, P., 1974, "System Identification", Wiley, New York.

HARKER, K.J. and D.B. ILIC, 1974, "Measurement of plasma wave spectral density from the cross-power density spectrum", Rev. Sci. Instrum. 45, 1315-1324.

HELLSTROM. C.W., 1968, "Statistical Theory of Signal Detection" (2nd. ed.), Pergamon Press, Lodnon.

KENDALL, M.G. and A. STUART, 1969 a, "The Advanced Theory of Statistics. Volume 1 : Distribution Theory" (3rd edition), Griffin, London.

KENDALL, M.G. and A. STUART, 1969 b, "The Advanced Theory of Statistics. Volume 2 : Inference and Relationship" (3rd edition), Griffin, London.

MEANS, J.D., 1972, "The use of the three-dimensional covariance matrix in analyzing the polarization properties of plane waves", J. Geophys. Res. 77, 5551-5559.

OLBERS, D.J., MULLER, P., and J. WILLEBRAND, 1976, "Inverse technique analysis of a large data set", Phys. Earth and Planet. Interiors, 12, 248-252.

POWELL, M.S.D., 1964, "An efficient method for finding the minimum of a function of several variables without calculating derivatives", Comput. J. 7, 155-162.

SCARF, F.L., KENNEL, C.F., FREDRICKS, R.W., GREEN, I.M., and G.M. CROOK, "AC fields and wave particle interactions" in "Particles and Fields in the Magnetophere" (Ed : B.M. Mc CORMAC), pp. 275-291, Reidel, Dordrecht, Holland.

STIX, T.H., 1962, "The theory of Plasma Waves", Mc Graw-Hill, New York.

STOREY, L.R.O., 1959, "A method for measuring local electron density from an artificial satellite", J. Res. Nat. Bur. Standards, 63D, 325-340.

STOREY, L.R.O., 1971, "Electric field experiments : - alternating fields", in "The ESRO Geostationary Magnetospheric Satellite", European Space Research Organization, Neuilly-sur-Seine, Report n° SP-60, 267-279.

STOREY, L.R.O. and F. LEFEUVRE, 1974, "Theory for the interpretation of measurements of a random electromagnetic wave field in space", in "Space Research XIV" (Eds : M.J. RYCROFT and R.D. REASENBERG), Akademic-Verlag, Berlin.

STOREY, L.R.O. and F. LEFEUVRE, 1976, "The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma. 1 - The direct problem", Note Technique CRPE/12.

WILKS, S.S., 1962, "Mathematical Statistics", Wiley, New York.

TABLE 1

Mean power	Symbol	Expression
$< x_i^2 (t) >$	α_{ii}	$\frac{\Delta\omega}{\pi} S_{ii} (\omega_o)$
$< y_i^2 (t) >$	α_{ii}	$\frac{\Delta\omega}{\pi} S_{ii} (\omega_o)$
$< x_i (t) y_j (t) >$	α_{ij}	$\frac{\Delta\omega}{\pi} R_e[S_{ij}(\omega_o)]$
$< y_i (t) y_j (t) >$	α_{ij}	$\frac{\Delta\omega}{\pi} R_e[S_{ij}(\omega_o)]$
$< x_i (t) y_j (t) >$	β_{ii}	0
$< x_i (t) y_j (t) >$	β_{ij}	$\frac{\Delta\omega}{\pi} I_m[S_{ij}(\omega_o)]$

Auto-powers and cross-powers of the filtered and demodulated signals, expressed in terms of the spectral matrix elements.

TABLE 2

i \ j	1	2	3	4		11	12
1	α_{11}	0	α_{12}	β_{12}		α_{16}	β_{16}
2	0	α_{11}	β_{21}	α_{21}		β_{61}	α_{61}
3	α_{12}	β_{21}	α_{22}	0			
4	β_{12}	α_{21}	0	α_{22}			
11	α_{16}	β_{61}				α_{66}	0
12	β_{16}	α_{61}				0	α_{66}

Values of the ensemble mean signal powers $< v_i(t)\ v_j(t) >$

TABLE 3

$l_1 - l_2$	l_1/l_2	Interpretation
0 - 0.5	1 - 1.7	Both models equally likely
0.5 - 2.0	1.7 - 7.4	Model 1 slightly preferable
2.0 - 4.5	7.4 - 90	Model 1 highly preferable
> 4.5	> 90	Model 2 rejected

Authors interpretation of support differences on likelihood ratios.

ILLUSTRATIONS

1 - Data flow chart for the case with only one model.

2 - Data flow chart for the case with two models.

3 - CMA diagram for a plasma with only one type of positive ion.
Π_e and Π_i are the angular plasma frequency for electron and ion,
Ω_e and Ω_i are the angular gyrofrequency for electron and ion.

4 - Analogue method for estimating the elements of the covariance matrix from the narrow-band filtered signals, without demodulation.

5 - Analogue method for demodulating the narrow-band filtered signals.

6 - Analogue method for demodulating wide-band signals, and subsequently narrowing their bandwidth.

7 - Analogue method for estimating the elements of the covariance matrix from the narrow-band filtered signals, after demodulation : (A) auto-covariances ; (B) cross-covariances.

8 - Illustrating how the accuracy of the parameter estimate is assessed, in the case of a model with only one parameter : (A) definition of the width of the peak of the support function ; (B) parabola fitted to peak, yielding and analytic approximation for the width.

9 - Cross-section of the peak of the support functions, in the case of a model with two parameters : (A) when the matrix B is diagonal ; (B) in general.

APPENDIX A : MATRIX ANALYSIS

Let $\mathbf{V}_n$ be the 12 x 12 data matrix defined in section 3 and $\mathbf{Z}_n$ its associated 6 x 6 complex matrix. The matrix $\mathbf{V}_n$ can be written as shown in table 2 :

$$(2.69)\quad \mathbf{V}_n = \begin{bmatrix} \alpha_{11} & 0 & \alpha_{12} & \beta_{12} \cdots & \alpha_{16} & \beta_{16} \\ 0 & \alpha_{11} & -\beta_{12} & \alpha_{12} \cdots & -\beta_{16} & \alpha_{16} \\ & & & & & \\ \alpha_{16} & -\beta_{16} & & \cdots & \alpha_{66} & 0 \\ \beta_{16} & \alpha_{16} & & \cdots & 0 & \alpha_{66} \end{bmatrix}$$

Also the matrix $\mathbf{Z}_n$ can be written as follows :

$$(2.70)\quad \mathbf{Z}_n = \begin{bmatrix} Z_{11} & Z_{12} \cdots & Z_{16} \\ Z_{12}^{*} & Z_{22} \cdots & Z_{26} \\ & & \\ Z_{16}^{*} & Z_{26}^{*} \cdots & Z_{66} \end{bmatrix}$$

with

$$Z_{ij} = 2\,\alpha_{ij} + 2i\,\beta_{ij}$$

For the sake of simplicity, the subscript n has been omitted in the matrix elements.

In the same way the 12 x 12 covariance matrix Λ, also defined in section 3, can be written

$$(2.71)\quad \Lambda = \begin{bmatrix} a_{11} & 0 & a_{12} & b_{12} \cdots\cdots & a_{16} & b_{16} \\ 0 & a_{11} & -b_{12} & a_{12} \cdots\cdots & -b_{16} & a_{16} \\ & & & & & \\ a_{16} & -b_{16} & \cdots\cdots & & a_{66} & 0 \\ b_{16} & a_{16} & \cdots\cdots & & 0 & a_{66} \end{bmatrix}$$

and its associated 6 x 6 complex matrix $\mathbf{C}$ is

$$(2.72)\quad \mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \cdots\cdots & C_{16} \\ C_{12}^{*} & C_{22} \cdots\cdots & C_{26} \\ & & \\ C_{16}^{*} & C_{26}^{*} \cdots\cdots & C_{66} \end{bmatrix}$$

with

$$C_{ij} = 2\, a_{ij} + 2i\, b_{ij}$$

We want to show that

$$(2.73)\qquad \mathrm{tr}\,(\Lambda^{-1} \mathbf{V}_n) = 2\ \mathrm{tr}\,(\mathbf{C}^{-1} \mathbf{Z}_n)$$

and

$$(2.74)\qquad |\Lambda|^{1/2} = 2^{-6}\, |\mathbf{C}|$$

Let us start with the relation (2.74). We need first of all to demonstrate that any eigenvalue of $\mathbf{C}$ is proportional to a double eigenvalue of Λ.

Now, if λ is an eigenvalue of $\mathbf{C}$, we have

$$\mathbf{C}\vec{U} = \lambda\vec{U} \tag{2.75}$$

with $\vec{U}$ the complex vector whose compontents U_i are

$$U_i = \delta_i + i\,\gamma_i$$

In the same way, if λ is a double eigenvalue of , we have

$$\Lambda\vec{W} = \lambda\vec{W} \tag{2.76}$$

Two eigenvectors $\vec{W}_1$, $\vec{W}_2$, linearly independants of one another, correspond to the same eigenvalue λ. Many choices of eigenvectors are possible. Intuitively we see that, to compare (2.75) and (2.76), we can take as eigenvectors

$$\vec{W}_1 = \begin{bmatrix} \delta_1 \\ -\gamma_1 \\ \delta_2 \\ -\gamma_2 \\ \\ -\delta_6 \\ \gamma_6 \end{bmatrix} \quad ; \quad \vec{W}_2 = \begin{bmatrix} \gamma_1 \\ \delta_1 \\ \gamma_2 \\ \delta_2 \\ \\ \gamma_6 \\ \delta_6 \end{bmatrix}$$

For simplicity, let us adopt the block matrices notation. The relation (2.76) can be rewritten

$$\Lambda_B \vec{W}_B = \mathbf{G}\vec{W}_B \tag{2.77}$$

with the following definitions :

$$\Lambda_B = \begin{bmatrix} B_{11} & B_{12} \cdots\cdots\cdots\cdots B_{16} \\ B_{21} & B_{22} \cdots\cdots\cdots\cdots B_{26} \\ \\ B_{61} & B_{62} \cdots\cdots\cdots\cdots B_{66} \end{bmatrix}$$

$$B_{ij}\ (i < j) = \begin{bmatrix} a_{ij} & b_{ij} \\ -b_{ij} & a_{ij} \end{bmatrix} ; \quad B_{ii} = \begin{bmatrix} a_{ii} & 0 \\ 0 & a_{ii} \end{bmatrix} ; \quad B_{ij}\ (i > j) = \begin{bmatrix} a_{ij} & -b_{ij} \\ b_{ij} & a_{ij} \end{bmatrix}$$

$$\vec{W}_B = \begin{bmatrix} W_{B_1} \\ W_{B_2} \\ \\ W_{B_6} \end{bmatrix} \qquad ; \qquad W_{Bj} = \begin{bmatrix} \delta_j & \gamma_j \\ -\gamma_j & \delta_j \end{bmatrix}$$

$$G = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

Let us develop the relation (2.75) :

$$(2.78.a) \quad \sum_{j=1}^{k-1} (a_{kj}\delta_j + b_{jk}\gamma_j) + a_{kk}\delta_k + \sum_{j=k+1}^{6} (a_{kj}\delta_j - b_{kj}\gamma_j) = \frac{\lambda}{2}\,\delta_k$$

$$(2.78.b) \quad \sum_{j=1}^{k-1} (a_{jk}\gamma_j - b_{jk}\delta_j) + a_{kk}\gamma_k + \sum_{j=k+1}^{6} (b_{kj}\delta_j + a_{kj}\gamma_j) = \frac{\lambda}{2}\,\gamma_k$$

In the same way, we develop the relation (2.77) ;

$$(2.79) \quad \sum_{j=1}^{k-1} B_{kj}\, W_{Bj} + B_{kk}\, W_{Bk} + \sum_{j=k+1}^{6} B_{kj}\, W_{Bj} = G\, W_{Bk}$$

so

$$(2.80)\quad \sum_{j=1}^{k-1} \begin{bmatrix} a_{kj}\delta_j + b_{kj}\gamma_j & a_{kj}\gamma_j - b_{kj}\delta_j \\ b_{kj}\delta_j - a_{kj}\gamma_j & b_{kj}\gamma_j + a_{kj}\delta_j \end{bmatrix} + \begin{bmatrix} a_{kk}\delta_k & a_{kk}\gamma_k \\ -a_{kk}\gamma_k & a_{kk}\delta_k \end{bmatrix}$$

$$\sum_{j=k+1}^{6} \begin{bmatrix} a_{kj}\delta_j - b_{kj}\gamma_j & a_{kj}\gamma_j + b_{kj}\delta_j \\ -b_{kj}\delta_j - a_{kj}\gamma_j & -b_{kj}\gamma_j + a_{kj}\delta_j \end{bmatrix} = \lambda \begin{bmatrix} \delta_k & \gamma_k \\ -\gamma_k & \delta_k \end{bmatrix}$$

We derive the following four equations :

$$(2.81.a)\quad \sum_{j=1}^{k-1} (a_{kj}\delta_j + b_{kj}\gamma_j) + a_{kk}\delta_k + \sum_{j=k+1}^{6} (a_{kj}\delta_j - b_{kj}\gamma_j) = \lambda\delta_k$$

$$(2.81.b)\quad \sum_{j=1}^{k-1} (a_{kj}\gamma_j - b_{kj}\delta_j) + a_{kk}\gamma_k + \sum_{j=k+1}^{6} (a_{kj}\gamma_j + b_{kj}\delta_j) = \lambda\gamma_k$$

$$(2.81.c)\quad \sum_{j=1}^{k-1} (a_{kj}\gamma_j - b_{kj}\delta_j) + a_{kk}\gamma_k + \sum_{j=k+1}^{6} (b_{kj}\delta_j + a_{kj}\gamma_j) = \lambda\gamma_k$$

$$(2.81.d)\quad \sum_{j=1}^{k-1} (b_{kj}\gamma_j + a_{kj}\delta_j) + a_{kk}\delta_k + \sum_{j=k+1}^{6} (-b_{kj}\gamma_j + a_{kj}\delta_j) = \lambda\delta_k$$

Remembering that $a_{kj} = a_{jk}$ and $b_{kj} = b_{jk}$, we see that, apart from a multiplicative factor of 2, the relations (2.81.a) and (2.81.d) are equivalent to the relation(2.78.b). It follows that, if 2λ is an eigenvalue of $\mathbf{C}$, then λ is a double eigenvalue of $\mathbf{\Lambda}$.

Now, by definition,

$$|\mathbf{C}| = \prod_{i=1}^{6} 2\lambda_i = 2^6 \prod_{i=1}^{6} \lambda_i$$

and

$$|\mathbf{\Lambda}| = \prod_{i=1}^{12} \lambda_i = \prod_{i=1}^{6} \lambda_i^2$$

Therefore,

$$|\mathbf{\Lambda}|^{1/2} = 2^{-6} \; |\mathbf{C}|$$

This is the result required.

To demonstrate the relation (2.73), let us show that any eigenvalue of the matrix $(\mathbf{C}^{-1} \; \mathbf{Z}_n)$ is a double eigenvalue of the matrix $(\mathbf{\Lambda}^{-1} \; \mathbf{V}_N)$. Let us start by calculating the eigenvalue of $(\mathbf{C}^{-1} \; \mathbf{Z}_n)$. If $\vec{U}$ is an eigenvector with components $U_j = \delta_j + i \; \gamma_j$, we have

$$\mathbf{C}^{-1} \; \mathbf{Z}_n \; \vec{U} = \lambda \; \vec{U} \tag{2.82}$$

Since $\mathbf{C}$ and $\mathbf{Z}_n$ are Hermitian, positive definite and invertible, it follows that

$$\lambda = \frac{(\vec{U})^T \; \mathbf{Z}_n \; \vec{U}}{(\vec{U}^*)^T \; \mathbf{C} \; \vec{U}} \tag{2.83}$$

Now

$$(\vec{U}^*)^T \; \mathbf{C} \; \vec{U} = \sum_{k=1}^{6} \left\{ \sum_{j=1}^{k-1} u_k u_j^* \; C_{jk} + |U_k|^2 \; C_{kk} + \sum_{j=k+1}^{6} u_k u_j^* \; C_{kj} \right\} \tag{2.84}$$

and, if we go one step further,

$$(2.85)\quad (\vec{U}^{*})^{T}\ \mathbf{C}\ \vec{U} = 2\sum_{k=1}^{6}\left\{(\delta_k^{\,2}+\gamma_k^{\,2})\,a_{kk} + 2\sum_{j=k+1}^{6} a_{kj}\,(\delta_k\delta_j+\gamma_k\gamma_j) + b_{kj}\,(\gamma_k\delta_j-\gamma_j\delta_k)\right\}$$

Thus it appears that $(\vec{U}^{*})^{T}\ \mathbf{C}\ \vec{U}$ is purely real. If we do the same for $(\vec{U}^{*})^{T}\ \mathbf{Z}_n\ \vec{U}$, we find the following real value for λ ;

$$(2.86)\quad \lambda = \frac{\sum_{k=1}^{6}\left\{(\delta_k^{\,2}+\gamma_k^{\,2})\,\alpha_{kk} + 2\sum_{j=k+1}^{6}\left[\alpha_{kj}(\delta_k\delta_j+\gamma_k\gamma_j) + \beta_{kj}\,(\gamma_k\delta_j-\gamma_j\delta_k)\right]\right\}}{\sum_{k=1}^{6}\left\{(\delta_k^{\,2}+\gamma_k^{\,2})\,a_{kk} + 2\sum_{j=k+1}^{6}\left[a_{kj}(\delta_k\delta_j+\gamma_k\gamma_j) + b_{kj}\,(\gamma_k\delta_j-\gamma_j\delta_k)\right]\right\}}$$

Now let us try to find an analogous relation for the eigenvalue of $\mathbf{\Lambda}^{-1}\ \mathbf{V}_n$. Assuming that the eigenvalue of $\mathbf{\Lambda}^{-1}\ \mathbf{V}_n$ are double ones and using the block matrix notation again, we have

$$(2.87)\quad (\mathbf{\Lambda}_B^{-1}\ (\mathbf{V}_n)_B)\ \vec{W}_B = \mathbf{G}\ \vec{W}_B$$

with $\mathbf{\Lambda}_B$, $\mathbf{G}$ and $\vec{W}_B$ defined as above and with

$$(\mathbf{V}_n)_B = \begin{bmatrix} D_{11} & D_{12}\ldots\ldots\ldots D_{16} \\ D_{21} & D_{22}\ldots\ldots\ldots D_{26} \\ & \\ D_{61} & D_{62}\ldots\ldots\ldots D_{66} \end{bmatrix}$$

$$D_{ij}(i<j) = \begin{bmatrix} \alpha_{ij} & \beta_{ij} \\ -\beta_{ij} & \alpha_{ij} \end{bmatrix};\quad D_{ii} = \begin{bmatrix} \alpha_{ii} & 0 \\ 0 & \alpha_{ii} \end{bmatrix};\quad D_{ij}(i>j) = \begin{bmatrix} \alpha_{ij} & -\beta_{ij} \\ \beta_{ij} & \alpha_{ij} \end{bmatrix}$$

Because the matrices $\mathbf{\Lambda}_B$ and $(\mathbf{V}_n)_B$ are invertible, we can writte :

$$(2.88)\quad \mathbf{G}\ \vec{W}^{T}\mathbf{\Lambda}_B\ \vec{W} = \vec{W}_B^{\,T}\ (\mathbf{V}_n)_B\ \vec{W}$$

or,

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}$$

if m_{ij} and n_{ij} are the elements of the matrices $\vec{W}^T \mathbf{\Lambda}_B \vec{W}$ and $\vec{W}_B^T (\mathbf{V}_n)_B \vec{W}$.

Now, by doing some more algebra we obtain

$$m_{11} = m_{22} = \sum_{k=1}^{6} \left\{ \alpha_{kk} (\delta_k^2 + \gamma_k^2) + 2 \sum_{j=k+1}^{6} \left[\alpha_{kj} (\delta_k \delta_j + \gamma_k \gamma_j) + \beta_{kj} (\gamma_k \delta_j - \gamma_j \delta_k) \right] \right\}$$

$$n_{11} = n_{22} = \sum_{k=1}^{6} \left\{ a_{kk} (\delta_k^2 + \gamma_k^2) + 2 \sum_{j=k+1}^{6} \left[a_{kj} (\delta_k \delta_j + \gamma_k \gamma_j) + b_{kj} (\gamma_k \delta_j - \gamma_j \delta_k) \right] \right\}$$

$$m_{12} = m_{21} = n_{12} = n_{21} = 0$$

It follows immediatly that the double eigenvalues of $\mathbf{\Lambda}^{-1} \mathbf{V}_n$ are equal to the eigenvalue of $\mathbf{C}^{-1} \mathbf{Z}_n$ given by (2.86).

The trace of a matrix being equal to the sum of the eigenvalues, we have

$$\text{tr} (\mathbf{\Lambda}^{-1} \mathbf{V}_n) = 2 \text{ tr} (\mathbf{C}^{-1} \mathbf{Z}_n)$$

which is the result (2.73) that we set out to prove.

ACKNOWLEDGEMENTS

The authors are most grateful to Professor D.A. Gurnett for suggesting this problem, in a private communication to L.R.O. Storey in 1965. Our thanks are also due to our colleague C. Delannoy for help in the numerical analysis. This work was sponsored by the Centre National d'Etudes des Télécommunications under the contract n° CIR 75-32.

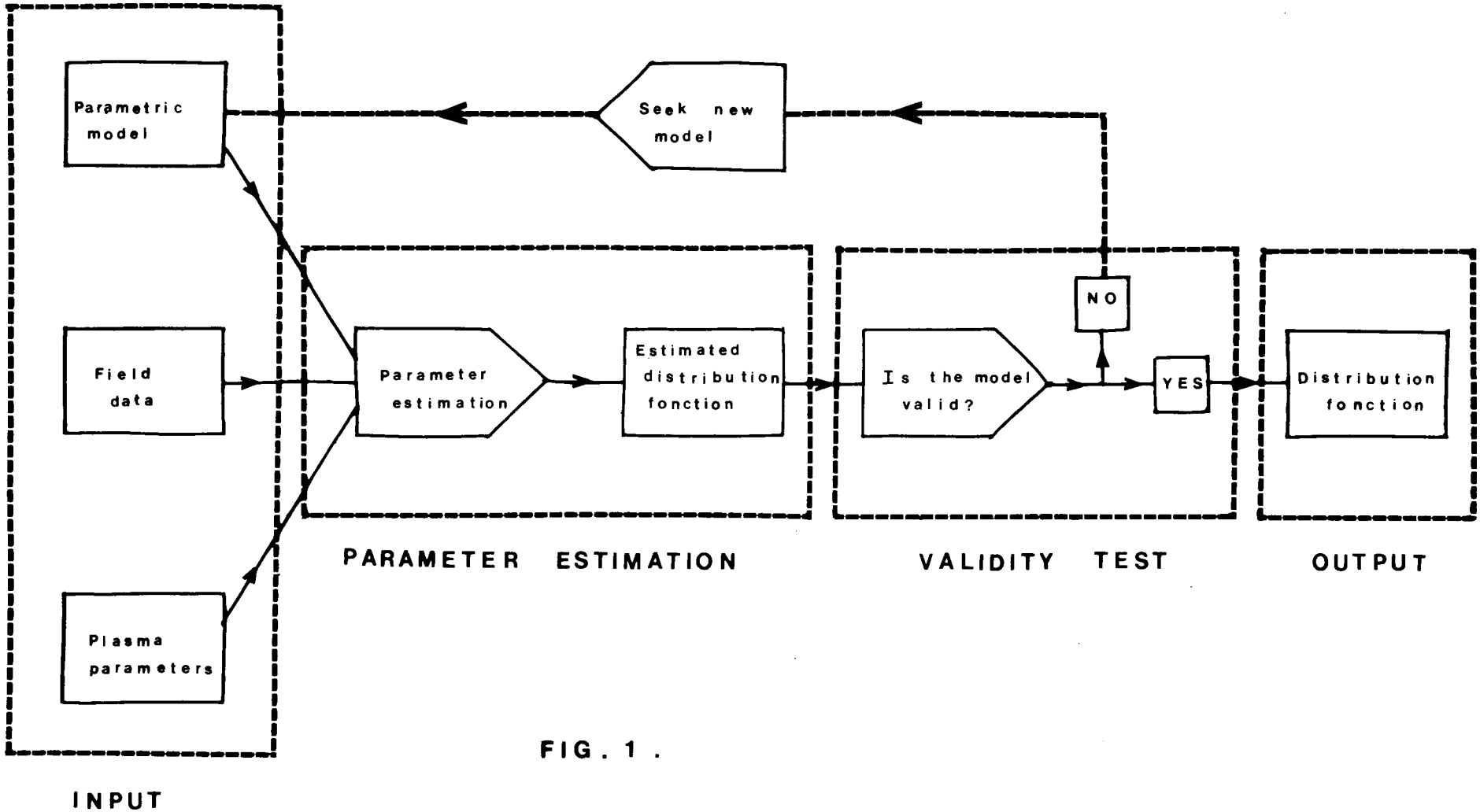
Parametric model
Seek new model
Field data
Parameter estimation
Estimated distribution fonction
Is the model valid?
NO
YES
Distribution fonction
Plasma parameters
PARAMETER ESTIMATION
VALIDITY TEST
OUTPUT
INPUT

FIG. 1.

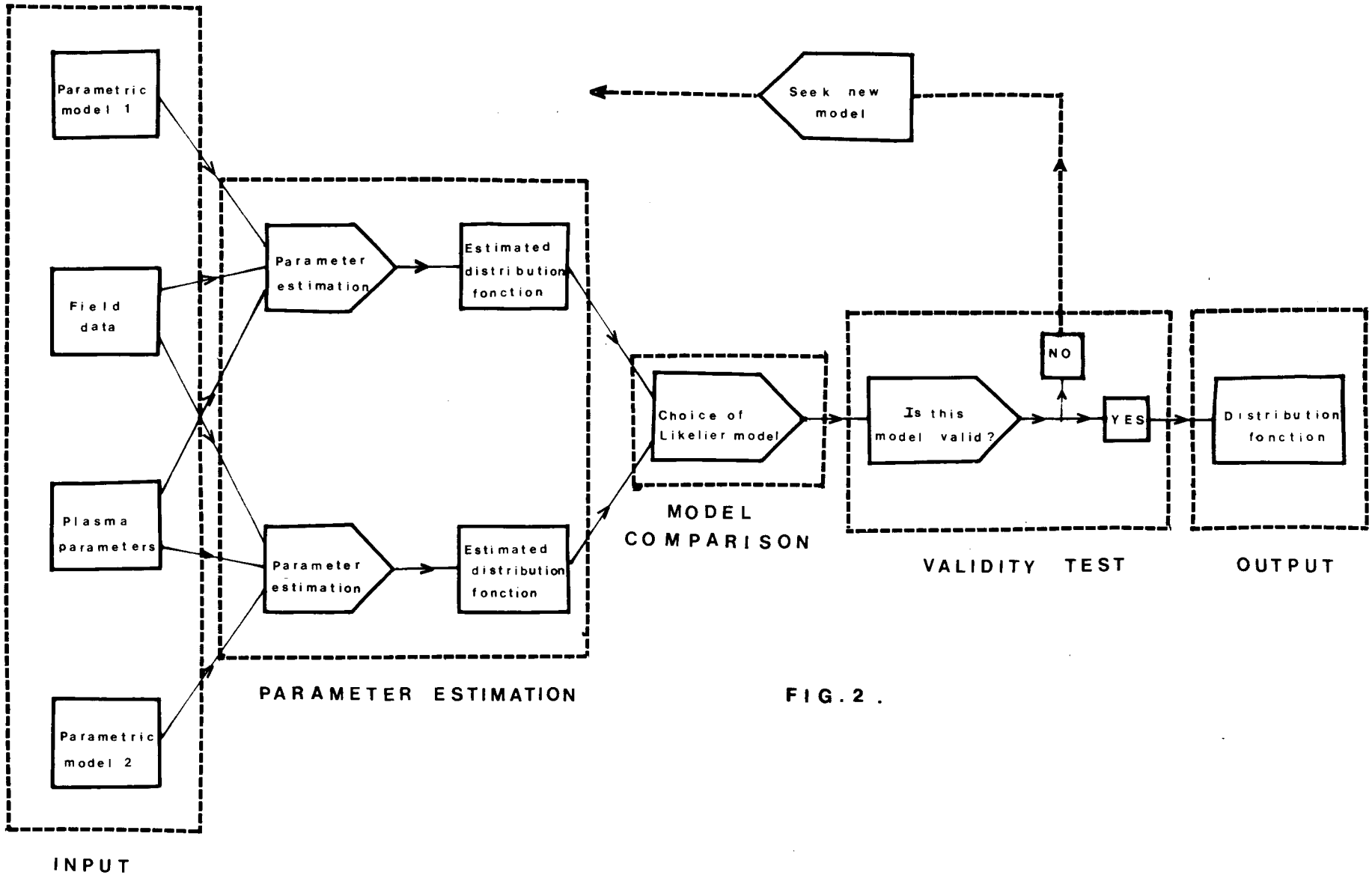
Parametric model 1
Field data
Plasma parameters
Parametric model 2
INPUT
Parameter estimation
Parameter estimation
Estimated distribution fonction
Estimated distribution fonction
PARAMETER ESTIMATION
Choice of Likelier model
MODEL COMPARISON
Seek new model
Is this model valid ?
NO
YES
VALIDITY TEST
Distribution fonction
OUTPUT

FIG. 2.

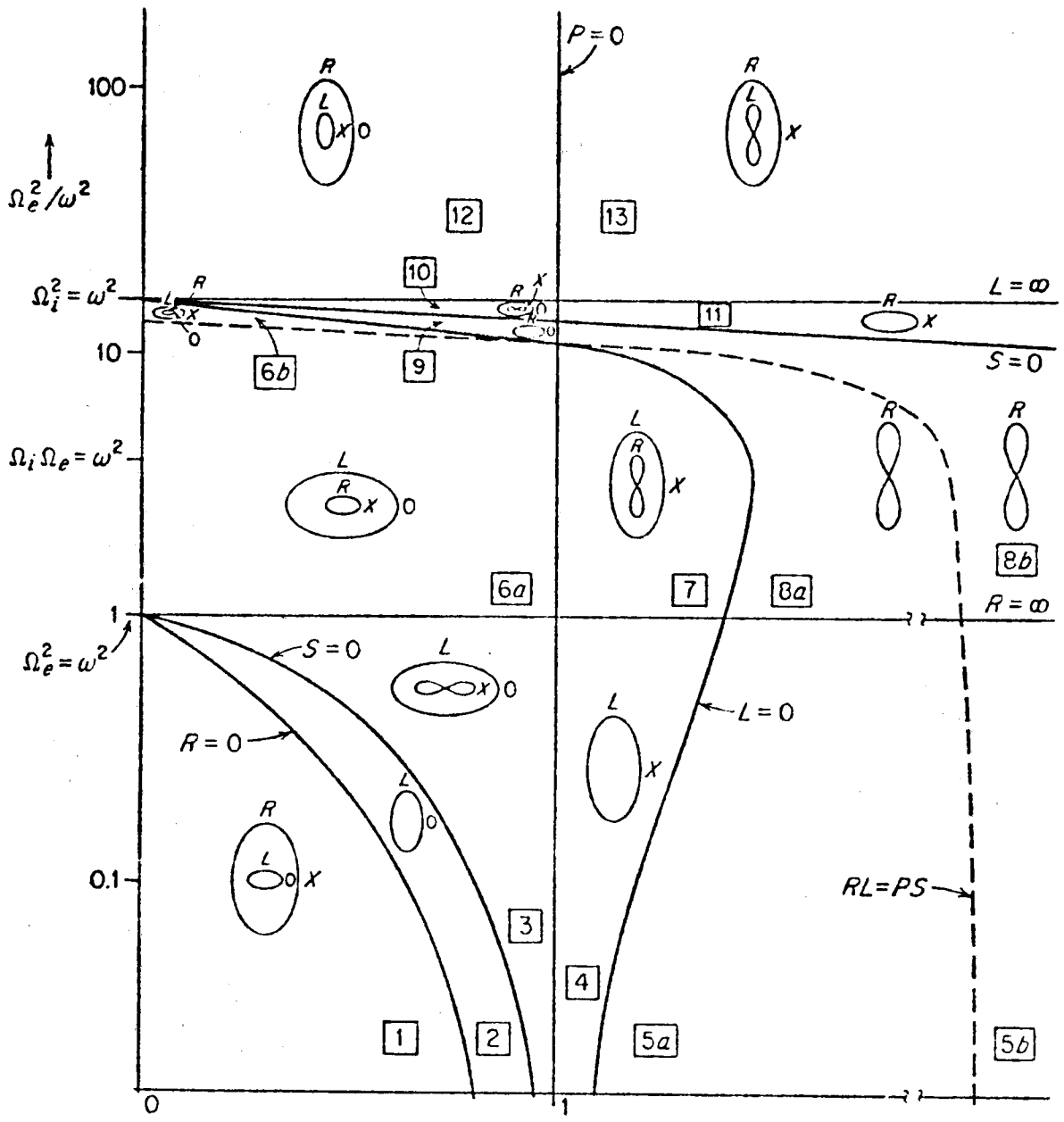
P = 0
100
Ω_e^2/ω^2
$\Omega_i^2 = \omega^2$
10
$\Omega_i \Omega_e = \omega^2$
1
$\Omega_e^2 = \omega^2$
0.1
0
1
$(\Pi_i^2 + \Pi_e^2)/\omega^2$
L = ∞
S = 0
R = ∞
R = 0
L = 0
RL = PS
12
13
10
11
6b
9
6a
7
8a
8b
5a
5b
4
3
2
1

FIG.3

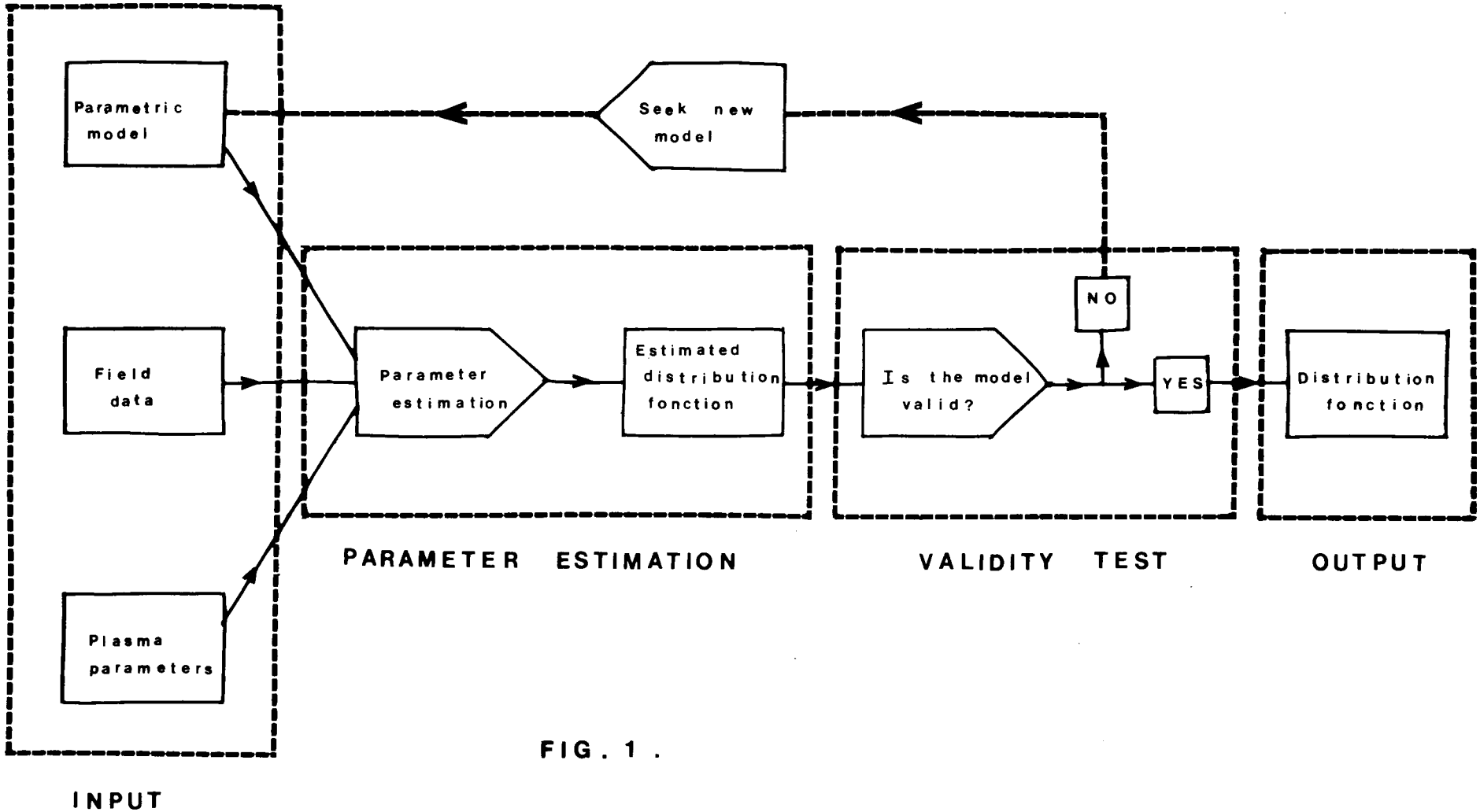
Parametric model
Seek new model
Field data
Parameter estimation
Estimated distribution fonction
Is the model valid?
NO
YES
Distribution fonction
Plasma parameters
PARAMETER ESTIMATION
VALIDITY TEST
OUTPUT
INPUT

FIG. 1.

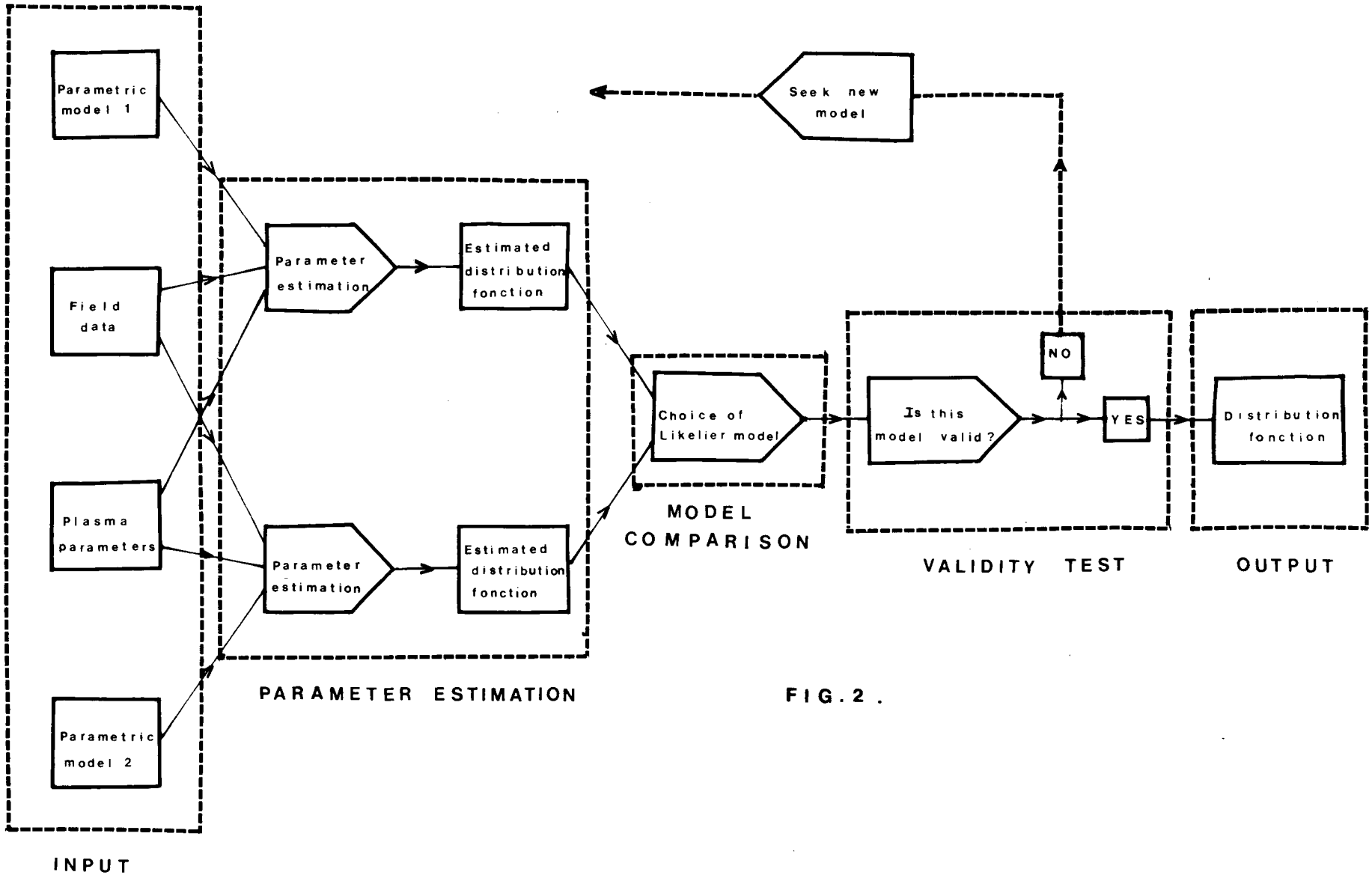
Parametric model 1
Field data
Plasma parameters
Parametric model 2
INPUT
Parameter estimation
Parameter estimation
Estimated distribution fonction
Estimated distribution fonction
PARAMETER ESTIMATION
Choice of Likelier model
MODEL COMPARISON
Seek new model
Is this model valid ?
NO
YES
VALIDITY TEST
Distribution fonction
OUTPUT

FIG. 2.

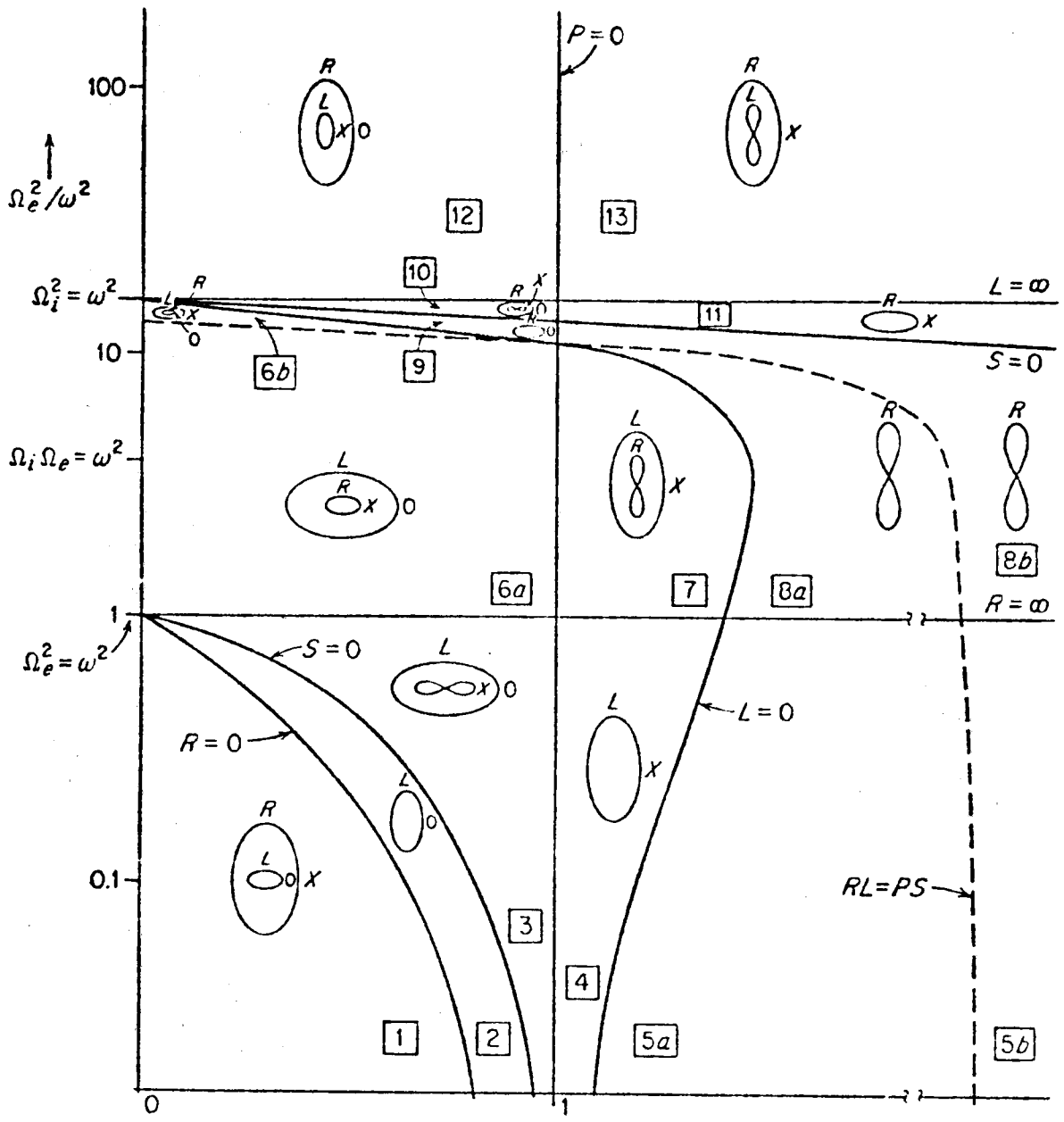
P = 0
100
Ω_e^2/ω^2
$\Omega_i^2 = \omega^2$
10
$\Omega_i \Omega_e = \omega^2$
1
$\Omega_e^2 = \omega^2$
0.1
0
1
$(\Pi_i^2 + \Pi_e^2)/\omega^2$
L = ∞
S = 0
R = ∞
R = 0
L = 0
RL = PS
12
13
10
11
6b
9
6a
7
8a
8b
5a
5b
4
3
2
1

FIG.3

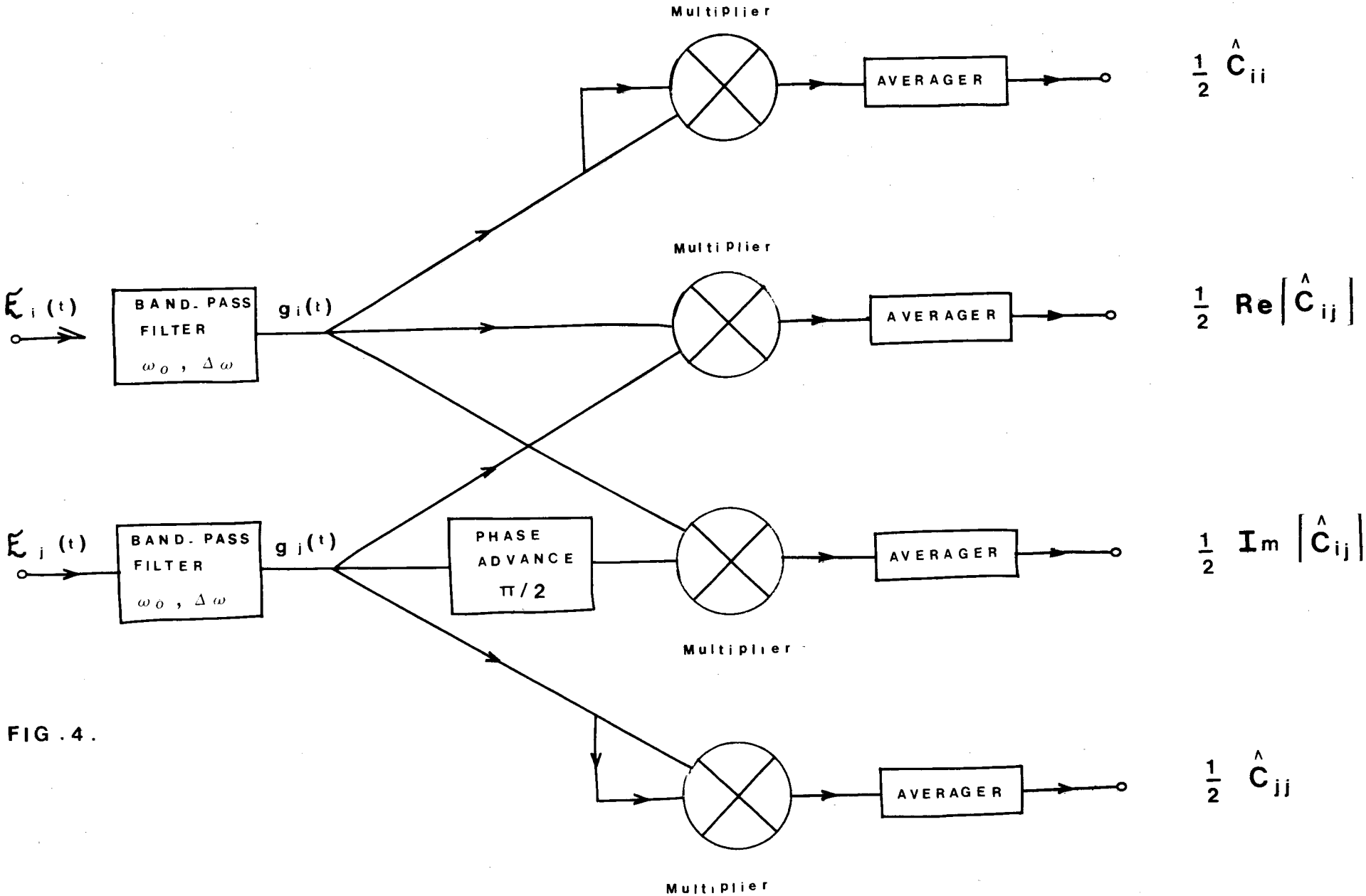
Multiplier
AVERAGER
$\frac{1}{2}\hat{C}_{ii}$
Multiplier
$\xi_i(t)$
BAND-PASS FILTER
ω_o , $\Delta\omega$
$g_i(t)$
AVERAGER
$\frac{1}{2}\,Re\left[\hat{C}_{ij}\right]$
$\xi_j(t)$
BAND-PASS FILTER
ω_o , $\Delta\omega$
$g_j(t)$
PHASE ADVANCE
$\pi/2$
AVERAGER
$\frac{1}{2}\,Im\left[\hat{C}_{ij}\right]$
Multiplier
AVERAGER
$\frac{1}{2}\,\hat{C}_{jj}$
Multiplier

FIG . 4 .

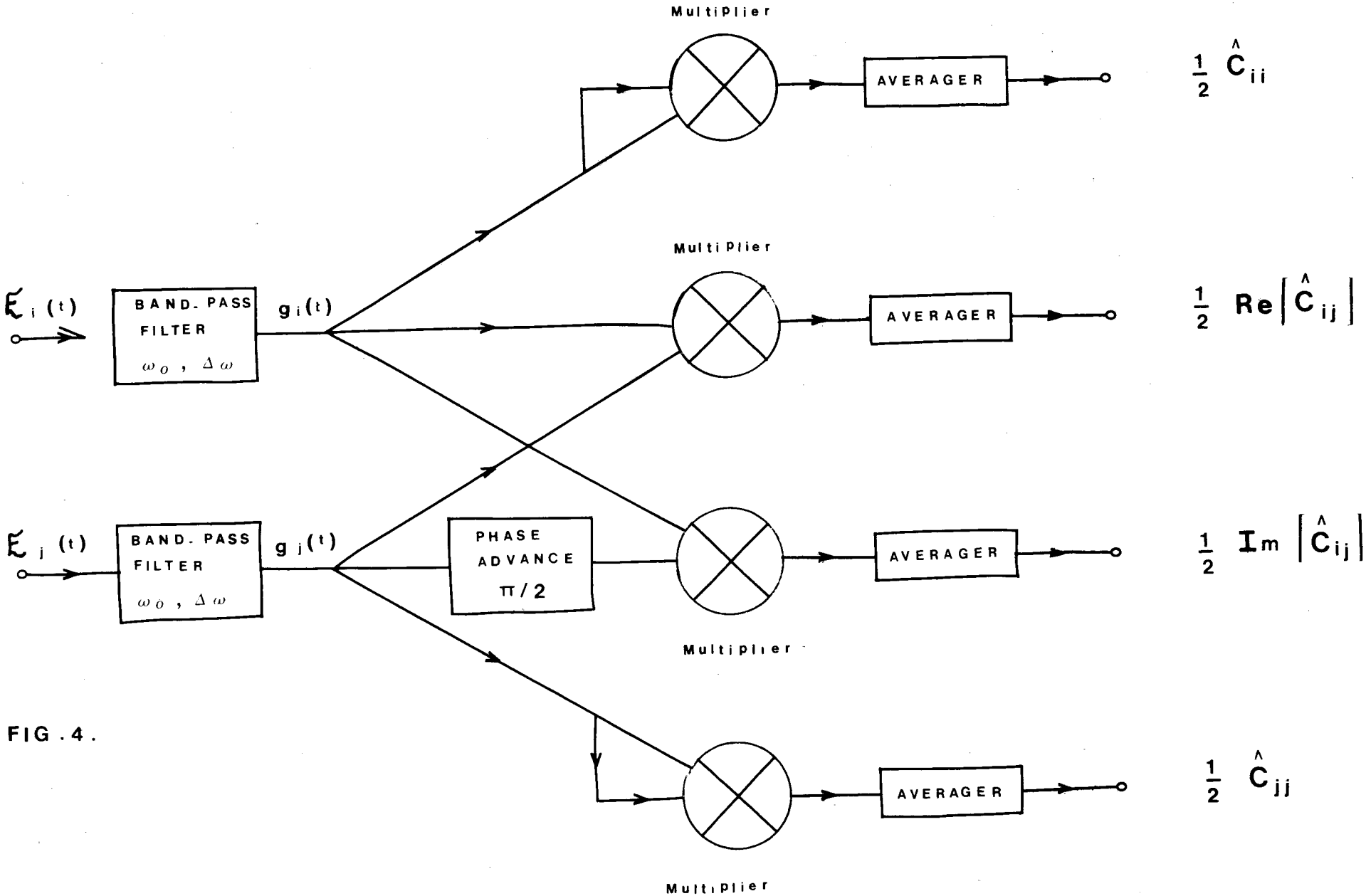
Multiplier
AVERAGER
$\frac{1}{2}\hat{C}_{ii}$
Multiplier
$\xi_i(t)$
BAND-PASS FILTER
ω_o , $\Delta\omega$
$g_i(t)$
AVERAGER
$\frac{1}{2}\,Re\left[\hat{C}_{ij}\right]$
$\xi_j(t)$
BAND-PASS FILTER
ω_o , $\Delta\omega$
$g_j(t)$
PHASE ADVANCE
$\pi/2$
AVERAGER
$\frac{1}{2}\,Im\left[\hat{C}_{ij}\right]$
Multiplier
AVERAGER
$\frac{1}{2}\,\hat{C}_{jj}$
Multiplier

FIG . 4 .

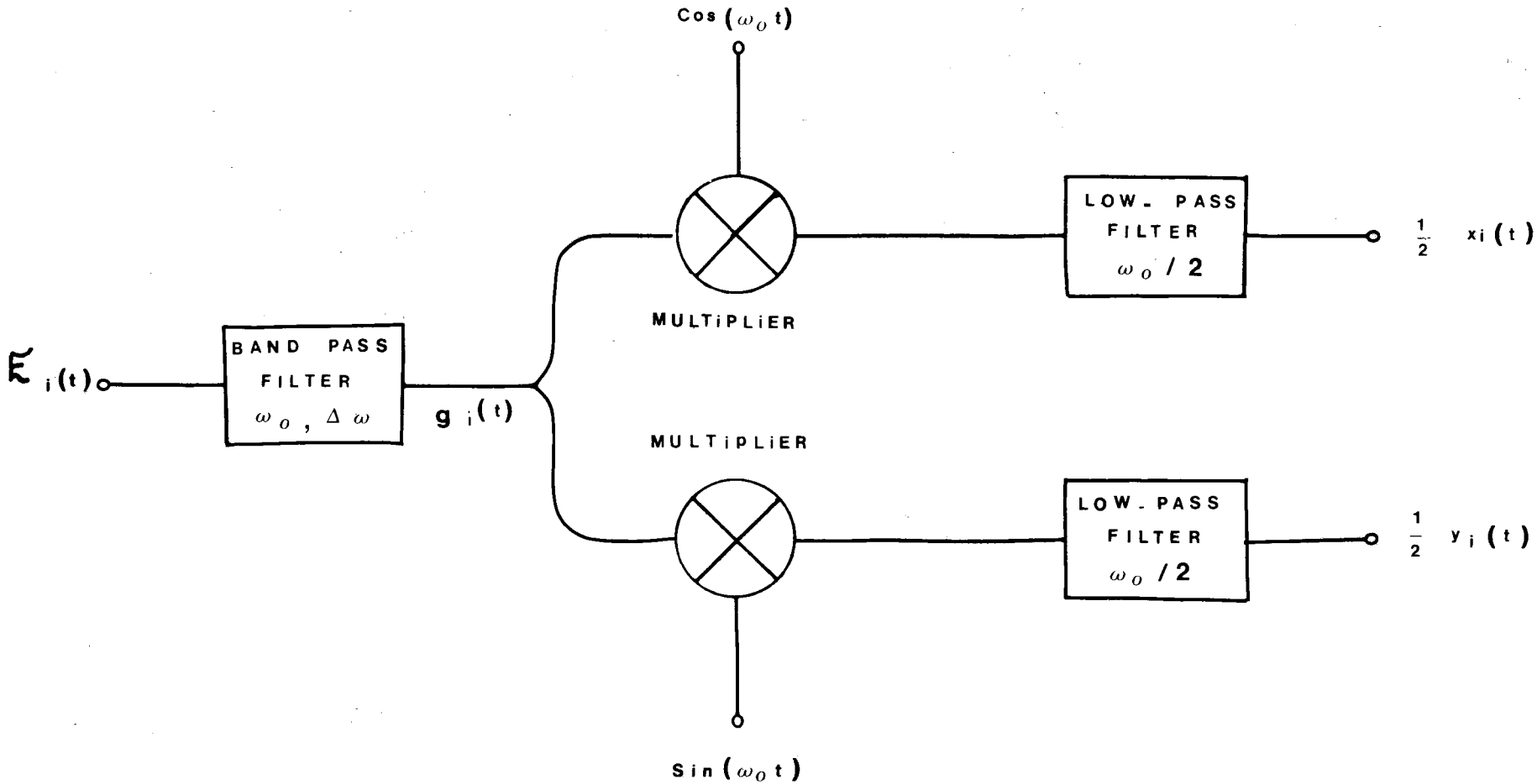
Cos (ω_o t)
MULTiPLiER
LOW - PASS FILTER ω_o / 2
$\frac{1}{2}$ x_i (t)
$\tilde{E}_i$(t)
BAND PASS FILTER ω_o , $\Delta\omega$
g_i(t)
MULTiPLiER
LOW - PASS FILTER ω_o / 2
$\frac{1}{2}$ y_i (t)
Sin (ω_o t)

Fig. 5

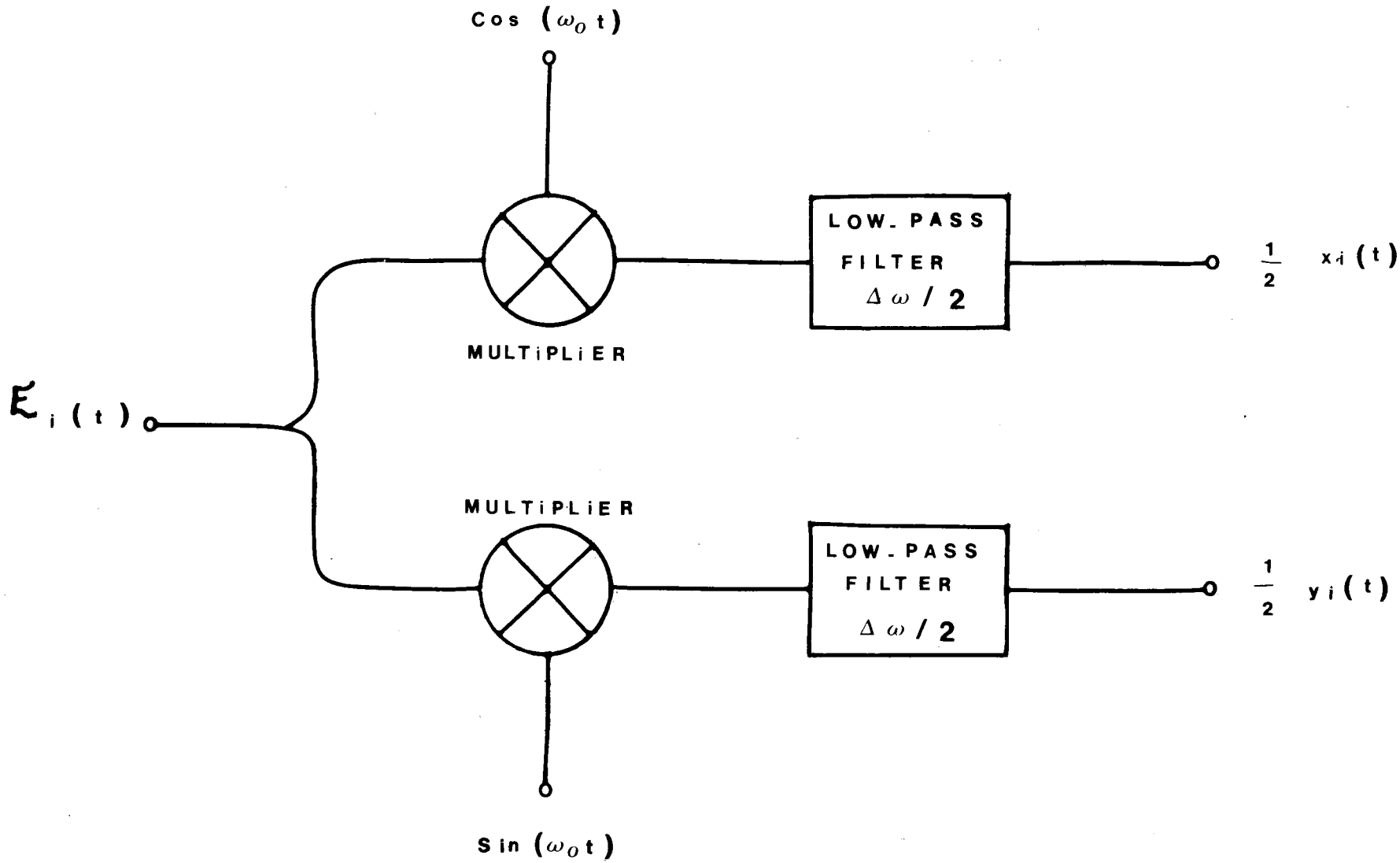
Cos (ω_0 t)
MULTIPLIER
LOW-PASS FILTER $\Delta\omega$ / 2
$\frac{1}{2}$ x_i (t)
E_i (t)
MULTIPLIER
LOW-PASS FILTER $\Delta\omega$ / 2
$\frac{1}{2}$ y_i (t)
Sin (ω_0 t)

FIG. 6.

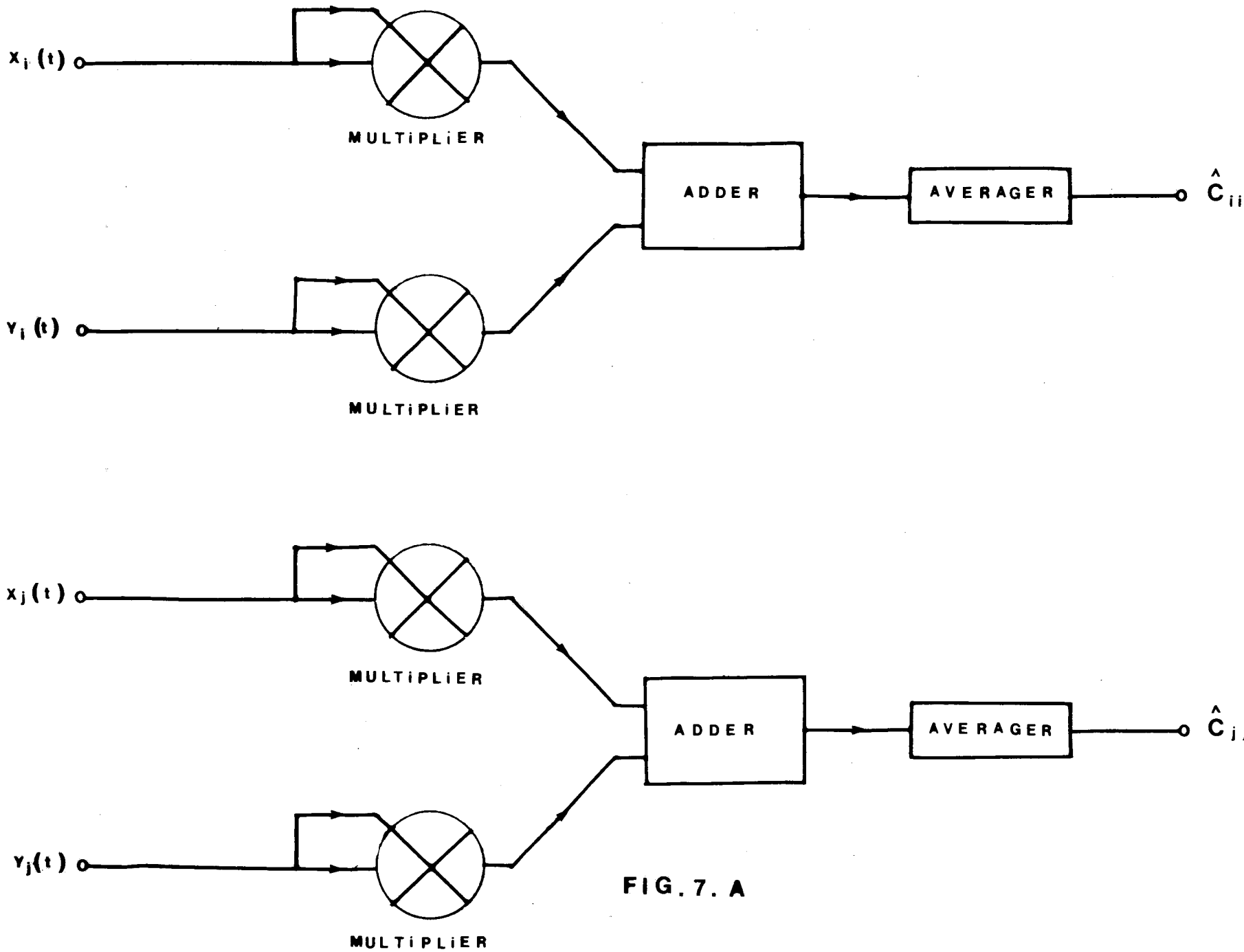
$x_i(t)$
MULTIPLIER
ADDER
AVERAGER
$\hat{C}_{ii}$
$y_i(t)$
MULTIPLIER
$x_j(t)$
MULTIPLIER
ADDER
AVERAGER
$\hat{C}_{jj}$
$y_j(t)$
MULTIPLIER

FIG. 7. A

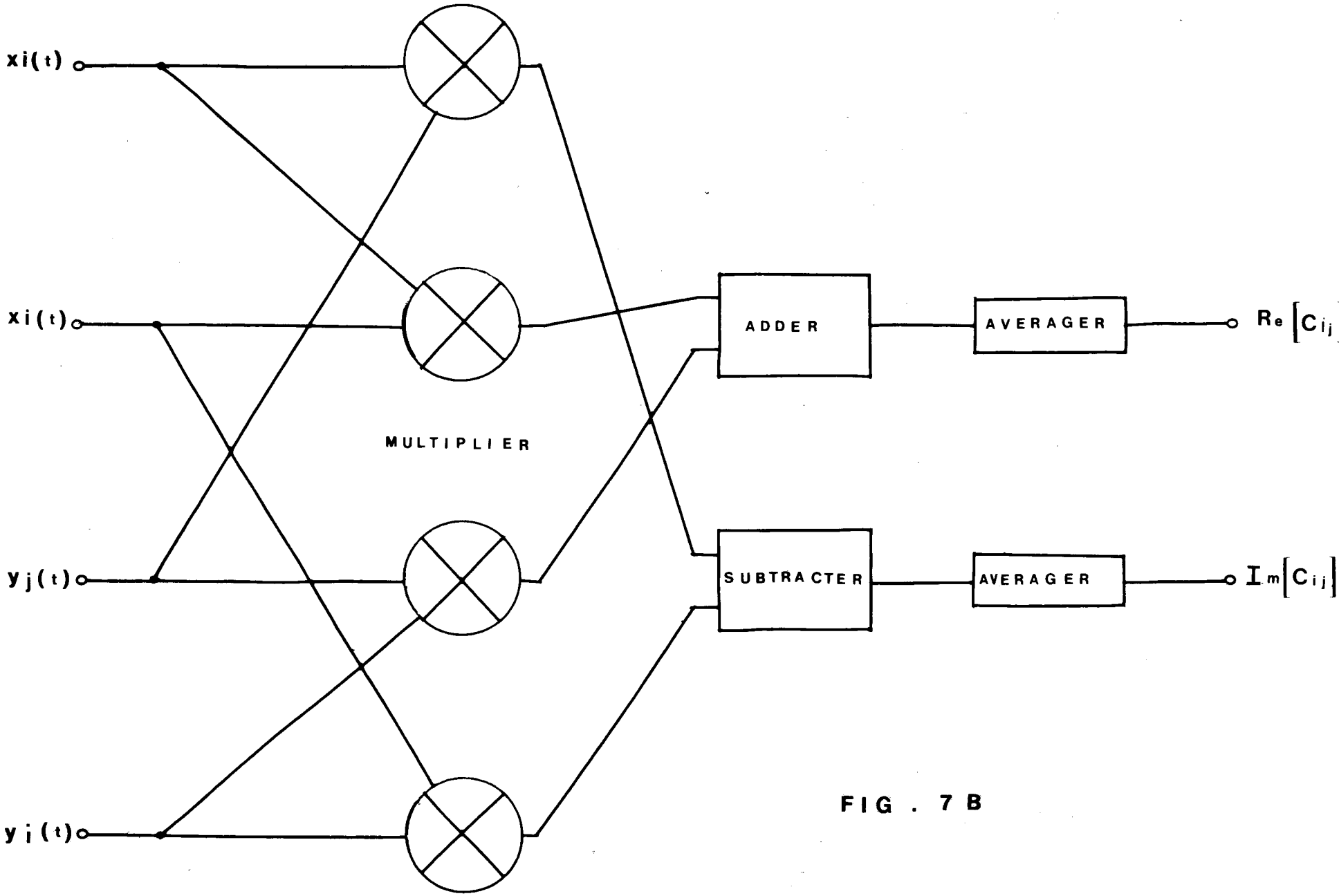
xi(t)
xi(t)
yj(t)
yi(t)
MULTIPLIER
ADDER
SUBTRACTER
AVERAGER
AVERAGER
Re[Cij]
Im[Cij]

FIG. 7B

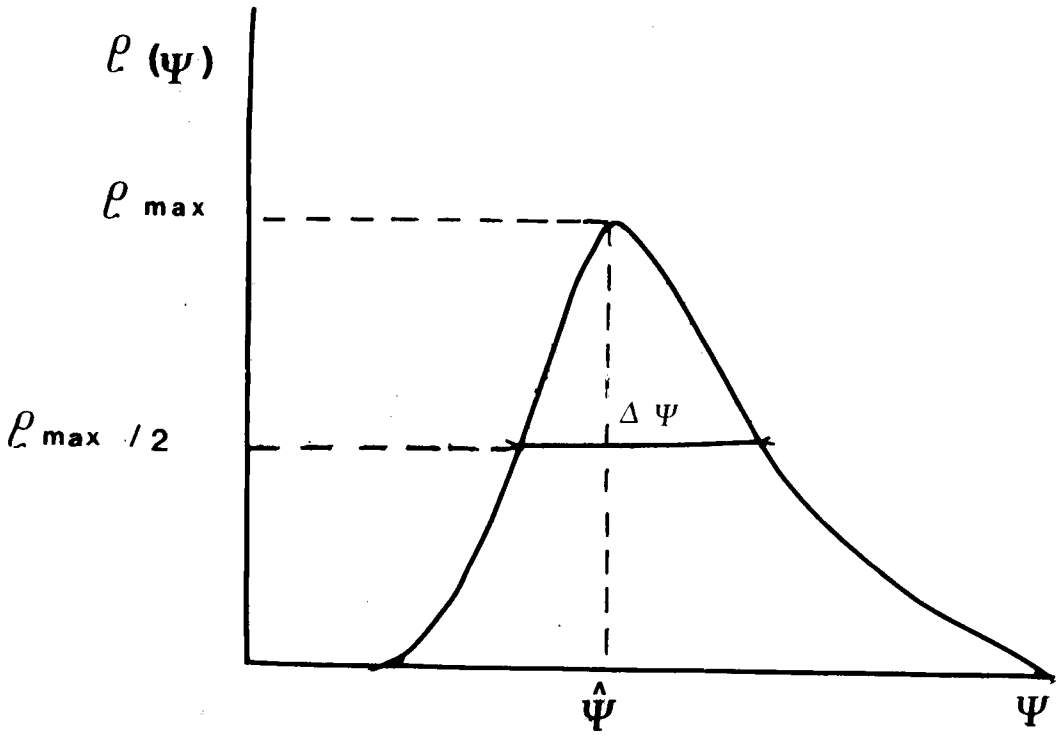

FIG 8. A

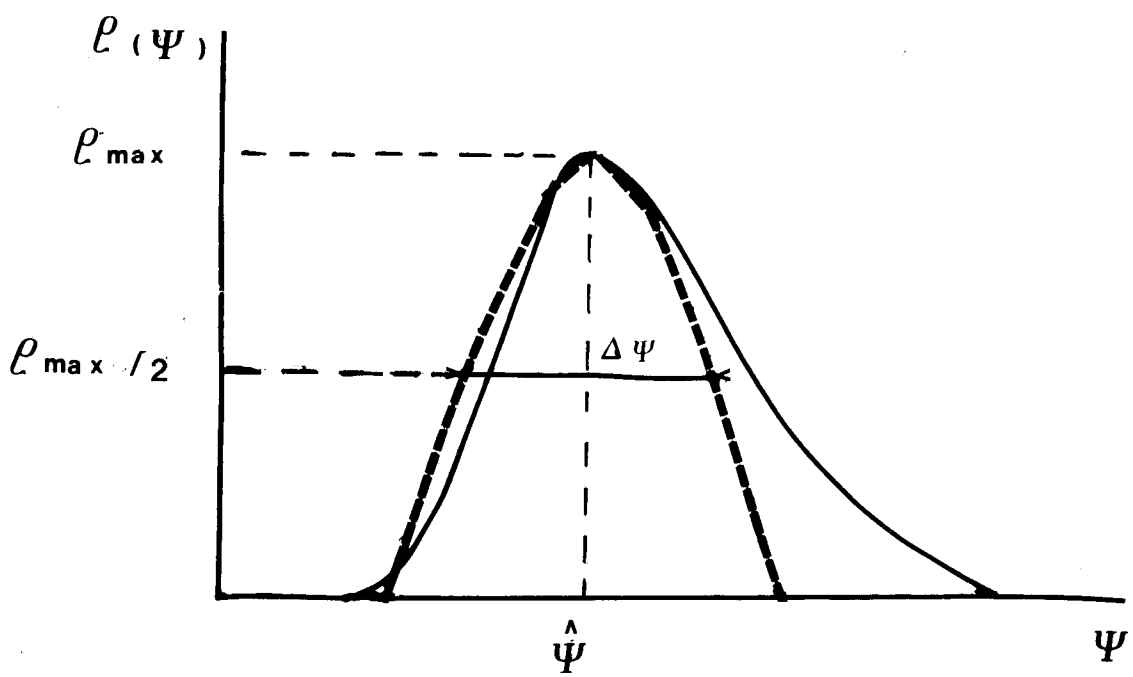

FIG 8. B

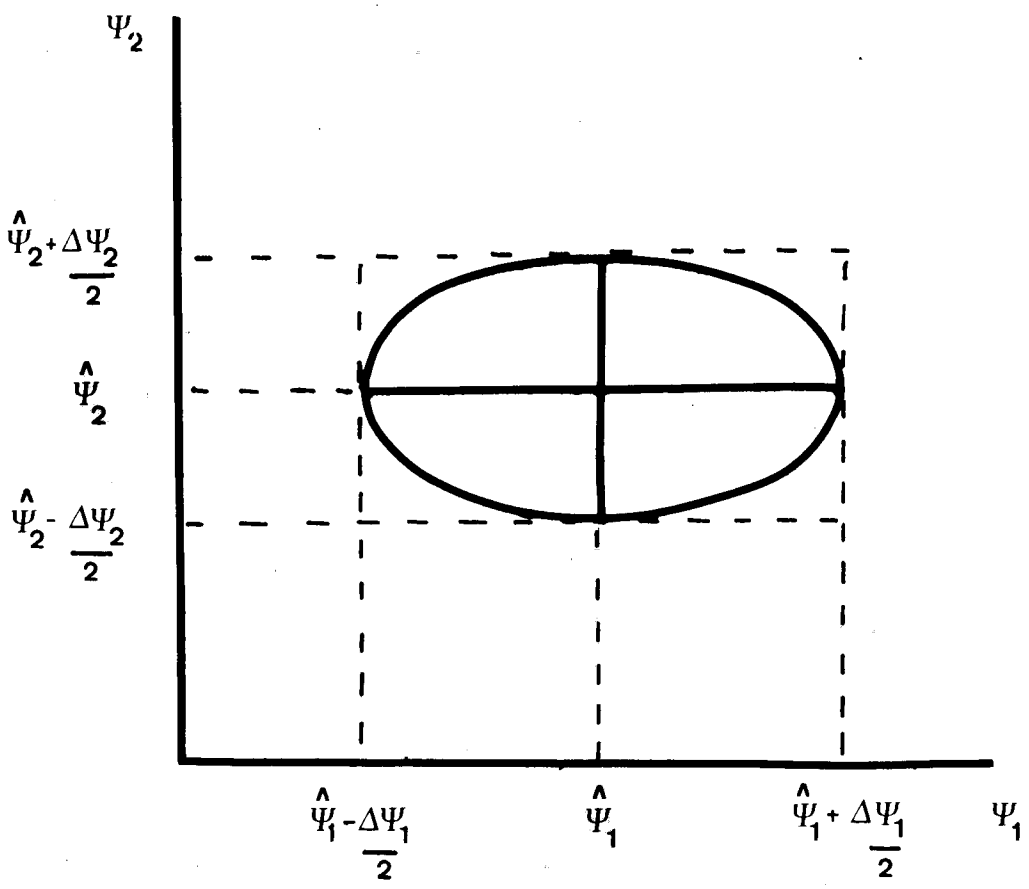
Ψ_2
$\hat{\Psi}_2 + \frac{\Delta\Psi_2}{2}$
$\hat{\Psi}_2$
$\hat{\Psi}_2 - \frac{\Delta\Psi_2}{2}$
$\hat{\Psi}_1 - \frac{\Delta\Psi_1}{2}$
$\hat{\Psi}_1$
$\hat{\Psi}_1 + \frac{\Delta\Psi_1}{2}$
Ψ_1

FIG 9.A

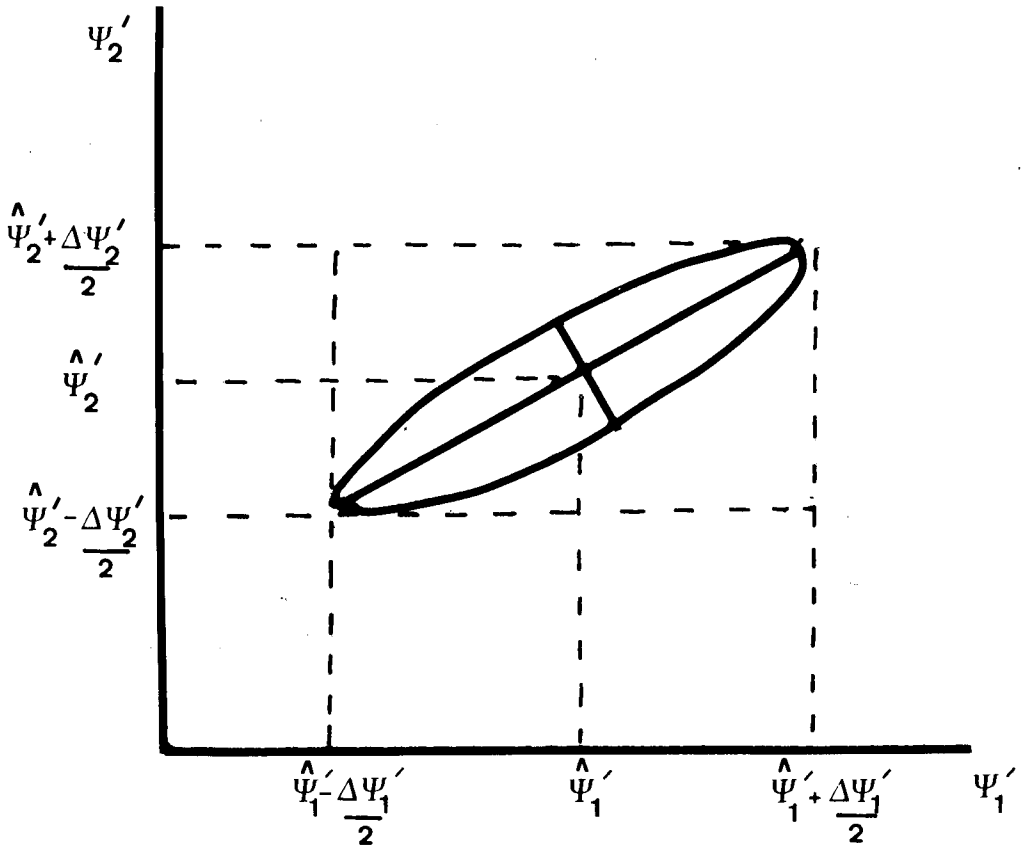
Ψ_2'
$\hat{\Psi}_2' + \frac{\Delta\Psi_2'}{2}$
$\hat{\Psi}_2'$
$\hat{\Psi}_2' - \frac{\Delta\Psi_2'}{2}$
$\hat{\Psi}_1' - \frac{\Delta\Psi_1'}{2}$
$\hat{\Psi}_1'$
$\hat{\Psi}_1' + \frac{\Delta\Psi_1'}{2}$
Ψ_1'

FIG 9.B

CRPE
Centre de Recherches
en Physique de l'Environnement
terrestre et planétaire

Avenue de la Recherche scientifique
45045 ORLEANS CEDEX

Département PCE
Physique et Chimie
de l'Environnement

Avenue de la Recherche scientifique
45045 ORLEANS CEDEX

Département ETE
Etudes par Télédétection
de l'Environnement

CNET - 38-40 rue du général Leclerc
92131 ISSY-LES-MOULINEAUX

CNET imp 75/1163