Ionospheric Research

NASA Grant NGR 39-009-032

Scientific Report

on

"Enhancing Effective Instrument Resolution by an Analog Method"

by

M. F. Zabielski

September 30, 1966

Scientific Report No. 280

Ionosphere Research Laboratory

Submitted by: B. R. F. Kendall
B. R. F. Kendall, Associate Professor of Physics

Approved by: G. H. Waynick
A. H. Waynick, Professor of Electrical Engineering, Director, Ionosphere Research Laboratory

The Pennsylvania State University

College of Engineering

Department of Electrical Engineering

TABLE OF CONTENTS

			Page
Abstract			i
1.	Introduction		1
	1.1	Discussion of Resolution.	1
	1.2	Definition of Convolution Integral	1
	1.3	Conditions for Applying Integral to Describe Instrument Performance	2
	1.4	Application of Integral to Practical Spectrometers	3
	1.5	Previous Convolution Studies	4
	1.6	Definition of Deconvolution	4
	1.7	Previous Deconvolution Studies	5
2.	Statement of Problem		7
3.	Theory of Operation of the Convoluting Mode		8
4.	Theory of Operation of the Deconvoluting Mode		11
5.	Experimental Apparatus		16
	5.1	Description of Physical Structure of Computer	16
	5.2	Procedure for Constructing Shaped Electrode	16
	5.3	Auxiliary Memory	18
	5.4	Electronics of Convoluting Mode	18
	5.5	Electronics of Deconvoluting Mode	19
6.	Results		25
	6.1	Deconvolution Without Filter	25
	6.2	Deconvolution to a Delta Function Singlet	27
	6.3	Deconvolution to a Delta Function Doublet.	29
	6.4	Deconvolution to a Delta Function Triplet	29
	6.5	Deconvolution to a Square Function Doublet and Importance of Gain	33
7.	Conclusions and Summary		36
	7.1	Problems Encountered in Deconvolution	36
	7.2	Suggested Improvements to Computer	36
	7.3	Application to Practical Data	37
	7.4	Advantages of this Computer over Digital Methods	37
	7.5	Advantages of this Computer over Previous Analog Methods	38

Page

7.6 Implications of Further Development 39

Bibliography 41

ABSTRACT

An analog device which is capable of enhancing the effective resolution of certain analytical instruments is described. This enhancement is effected by the mathematical operation of deconvolution. An improvement in resolution by a factor of six has been achieved. This device is also capable of computing the convolution integral which is used to describe the effects of various instrument parameters on instrument performance. The theory of operation of this device is given and information sufficient for construction and operation is supplied. Examples of results obtained in several typical situations are presented.

Author

1. Introduction

1.1 Discussion of Resolution

When describing the performance of an instrument that has as its output a spectrum, it is convenient to define a term called resolution. The methods of defining resolution vary from one class of instruments to another because different physical quantities are being measured, and thus the line shapes vary in mathematical form. However, problems arise because there are inconsistencies in defining resolution within a given class of instruments. Rayleigh's criterion is frequently used for defining the resolution of an optical system. Physically, this corresponds to a 19 per cent drop in amplitude between two adjacent maxima of equal amplitude, but there are devices that can easily detect a 4-5 per cent drop in amplitude. As a result of this, there is disagreement in defining resolution. Consequently, it is always necessary to write the equation of resolution. However, even when the equation of resolution is given, it is still only a semi-quantitative description of an instrument's performance, for it attempts to give neither any reasons for the limitation in performance nor methods of treating the "unresolved" data.

1.2 Definition of Convolution Integral

It has been known since the last century that a more useful treatment of instrument performance exists in the Fourier convolution integral. Let us immediately write this equation

$$F(x') = \int_{-\infty}^{\infty} T(x)A(x'-x)dx \qquad (1)$$

where: $F(x')$ is the instrument's output spectrum which includes instrument aberrations and is called the output function;

$T(x)$ is the ideal spectrum that is free of instrument aberrations and is called the true function;

$A(x'-x)$ is a function that characterizes the aberrations introduced by the instrument on $T(x)$ and is called the apparatus function.

1.3 Conditions for Applying Integral to Describe Instrument Performance

Certain easily satisfied conditions are necessary for the use of this equation. The first condition is that the various peaks in the spectrum obey the law of superposition. The second is that the apparatus function be stable throughout an experiment. The third is that the peaks in the same spectral range have the same characteristic shapes when normalized.

It is of interest at this point to note the following. If we normalize the apparatus function

$$\int_{-\infty}^{\infty} A(x'-x)dx = 1 \qquad (2)$$

and employ the convolution identity found in Mikusinski (1959)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x'-x)T(x)dx'dx = \int_{-\infty}^{\infty} A(z)dz \int_{-\infty}^{\infty} T(x)dx \qquad (3)$$

then integrating equation (1) over x' yields

$$\int_{-\infty}^{\infty} F(x')dx' = \int_{-\infty}^{\infty} A(z)dz \int_{-\infty}^{\infty} T(x)dx \qquad (4)$$

i.e.

$$\int_{-\infty}^{\infty} F(x')dx' = \int_{-\infty}^{\infty} T(x)dx. \qquad (5)$$

This result is very satisfying from the viewpoint of energy in that the total energy in a spectrum is the same for $F(x')$ and $T(x)$, and hence, only a redistribution of energy occurs upon integration.

1.4 Application of Integral to Practical Spectrometers

There are two cases of particular interest in applying the convolution integral to practical spectrometers. The first is the case when $A(x'-x) = \delta(x'-x)$.* Carrying out the integration, we readily see that $F(x') = T(x')$. This case corresponds to a spectrometer that has infinite resolving power; thus, it is only of hypothetical interest. The second case is when $T(x) = \sum_i a_i\delta(x-b_i)$. This corresponds to a completely resolved spectrum that exists at the input. The result of the integration depends upon the width of the apparatus function and the distances between delta functions. If the distances between the delta functions are greater than the width of the base of the apparatus function, the integration yields a series of completely resolved peaks whose shapes are governed by $A(x'-x)$ and whose amplitudes are determined by the a_i coefficients. If the distances between the delta functions are less than the width of the base of the apparatus function, the integration yields a series of overlapping peaks

*$\delta(x'-x)$ is a Dirac delta function which is a mathematically improper function. It has the following properties:

(1) $\delta(x-b) = 0$ for $x \neq b$

(2) $\int f(x)\delta(x-b)dx = f(b)$ if the region of integration includes $x = b$, and is zero otherwise.

whose shapes are governed by $A(x'-x)$ and by the a_i coefficients.

To use the convolution integral in a practical situation one needs to know the apparatus function. This function can be determined by introducing a single component into the instrument which is mathematically representable by $T(x) = \delta(x-b)$. Again, carrying out the integration, we see $F(x') = A(x'-b)$. This method of determining $A(x'-x)$ is analogous to determining the impulsive response of electrical and mechanical systems, and thus $A(x'-x)$ can be viewed as characteristic of the response of the instrument.

1.5 Previous Convolution Studies

Because of the usefulness of the convolution integral in predicting the effect on the output function of various instrument parameters like finite slit widths, considerable effort has been spent on writing the causes of aberrations in terms of convolution transforms. The papers by Rautian (1958) and Duffieux (1960) are comprehensive reviews in this vein. The mathematical properties of convolutions which are used in these reviews are well documented in texts like that of Mikusinski (1959) and of Widder (1955). Descriptions of instrument aberrations have also been given in terms of information theory by King and Emslie (1951 and 1953).

1.6 Definition of Deconvolution

In addition to explaining the effects of aberrations on the true spectrum, this equation offers the opportunity to increase

the effective resolution of an instrument by solving the equation for T(x) knowing F(x') and A(x'-x). Solving for T(x) is known as deconvolution. Deconvolution is also known as 'unfolding', 'decomposition' and 'entschmierung'. The solution for T(x) by analytical methods is quite complicated and often tedious. However, the advent of the computer, both analog and digital, has stirred renewed interest in solving for T(x).

1.7 Previous Deconvolution Studies

Some of the deconvolution methods that are available are presented in the following publications. Purely analytical solutions have been demonstrated by Whittaker and Watson (1963), Fox and Goodwin (1953) and Sachenko (1961); however, practical application of these methods to generalized situations would require a digital computer. A method developed by Morrison, using an IBM 7074 computer, consists of taking the Fourier transformation of the output function and the apparatus function, followed by a point-by-point computation of the true function (1963). Similar methods have been developed by Mori and Doi (1964) and by Rollett and Higgs (1962). Iterative curve fitting techniques using the digital computer have been developed by Skarsgard (1961) and by Gardner (1960).

Manually adjusted analog systems using function generators and summers have been described by French et al (1954) and by Noble et al (1959). Noble's system has been developed by E. I. DuPont de Nemours and Company. Fully automated analog methods have been developed and another suggested by

Kendall (1961, 1962, and 1966).

2. Statement of Problem

The specific problem is the development of a new analog method of deconvolution. This thesis will describe an analog computer, based on distributed electrostatic fields, that is capable of both convolution and deconvolution and of automatically plotting the results. It will be shown that this method of convoluting and deconvoluting has distinct advantages over previous methods, both analog and digital.

3. Theory of Operation of the Convoluting Mode

In figure 1 the shaped electrode A moves along the x-axis over a surface T at a small distance d above it. The potentials on the surface T describe the true function T(x) and for any given value of x are uniform in the y direction. The shape of electrode A is given by the apparatus function $y = A(x'-x)$, where x' gives the instantaneous position of the electrode along the x axis.

The capacitor C is chosen to be sufficiently large so that the potential of electrode A is always very small compared with the potentials of surface T. The amplifier B amplifies the small variations in potential that appear across C. These represent the output function $F(x')$.

If electrode A is brought into the position shown from an initial position where all the potentials beneath the electrode were zero, there will be a redistribution of charge between A and C, which will produce a positive signal at the input of the operational amplifier if T(x) is described by positive potentials.

Consider an element of T with a width dx and potential t(x). The field between t(x) and any part of A directly above it is

$$E_x = t(x)/d \tag{6}$$

assuming that A is very close to ground. The differential element of charge dq that is induced on A is given by Gauss' law, therefore

$$dq = \epsilon_o E_x a(x-x')dx \tag{7}$$

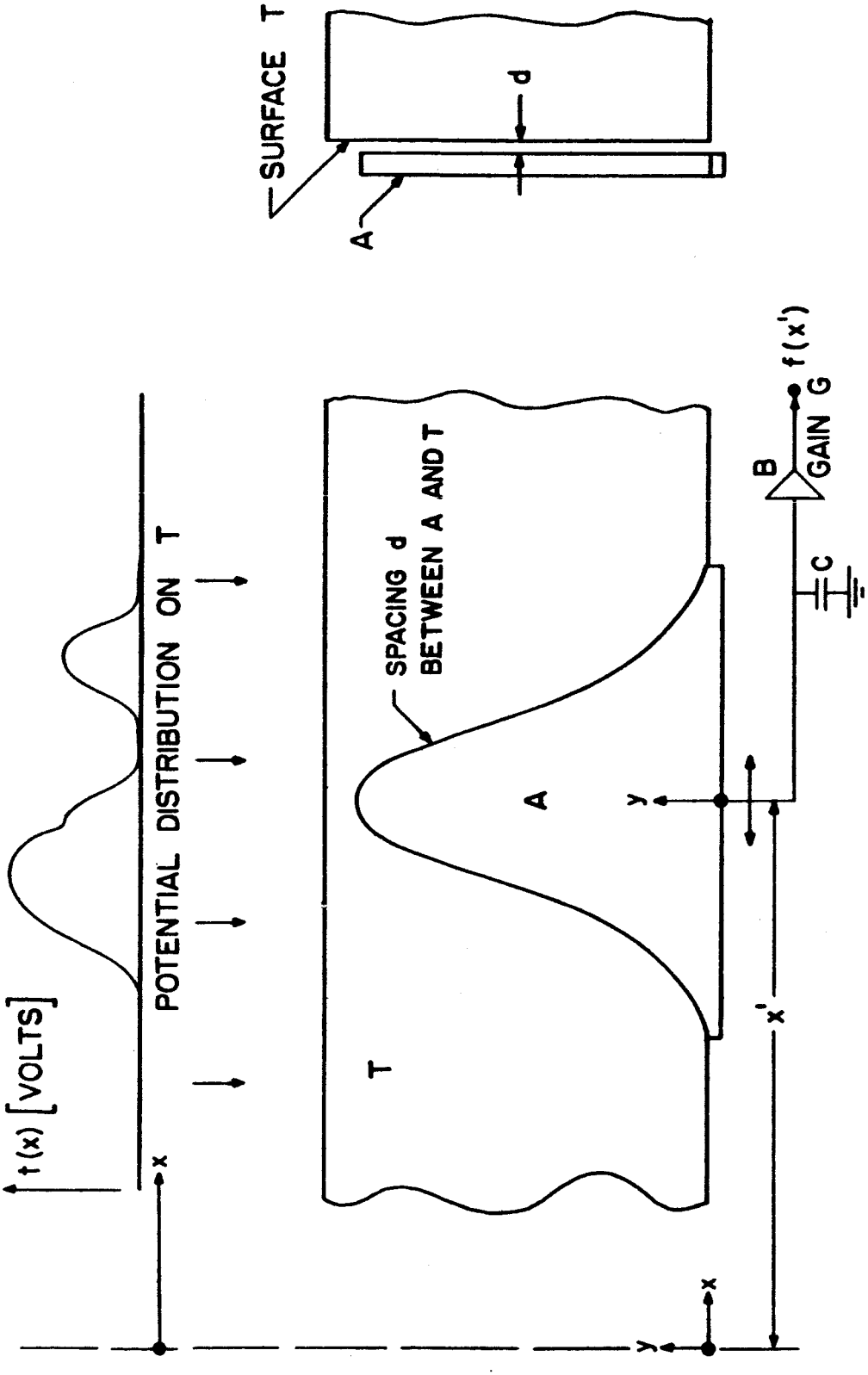


FIG.1 BASIC PRINCIPLE OF ANALOG SYSTEM FOR CONVOLUTION

where ϵ_o is the permittivity constant in m.k.s. units. To find the total induced charge on A, we integrate yielding

$$q(x') = \int_{-\infty}^{\infty} \epsilon_o t(x) a(x-x') dx/d \qquad (8)$$

Using the relation $q = cV$ and writing $V = f(x')$, we arrive at

$$f(x') = \frac{G\epsilon_o}{cd} \int_{-\infty}^{\infty} t(x) a(x-x') dx \qquad (9)$$

where c is the capacitance of C, assumed large compared with all other interelectrode capacities, and G is the gain of amplifier B.

Upon inspection, it is easily seen that equation (9) has the form of a convolution integral. Comparing it with equation (1), we see that if t(x) is made to correspond with the true function T(x), and a(x-x') to the apparatus function A(x'-x), then f(x') differs from F(x') only by a constant scaling factor. Care has to be taken so that the shapes representing A(x'-x) are oriented to allow for the reversed sense of the argument in a(x-x'). The complete range of values of F(x') are obtained for all values of x' under consideration by sweeping electrode A from one end of surface T to the other.

This section is an adaptation of the theory presented by Kendall (1966).

4. Theory of Operation of the Deconvoluting Mode

In figure 2 it can be seen that there are two capacitors connected to the input of amplifier B. The first capacitor consists of surface T and electrode A and shall be called C_A. The second is C_f which serves to isolate the amplifier B from the source of the voltage describing f(x'). Let us assume that the voltage V_f describing f(x') is always positive and that amplifier B has an inverted output. Since the input of amplifier B is isolated by the two capacitors, we may write

$$q_A + q_f = 0 \tag{10}$$

where q_A is the charge on C_A and q_f is the charge on C_f. Using the equation relating the charge and voltage on a capacitor, we may write

$$q_A = C_A(V_A - V_B) \tag{11}$$

$$q_f = C_f(V_f - V_B) \tag{12}$$

where V_B is the voltage that appears at the input of amplifier B. Placing equations (11) and (12) into equation (10), we get

$$C_A(V_A - V_B) + C_f(V_f - V_B) = 0 \tag{13}$$

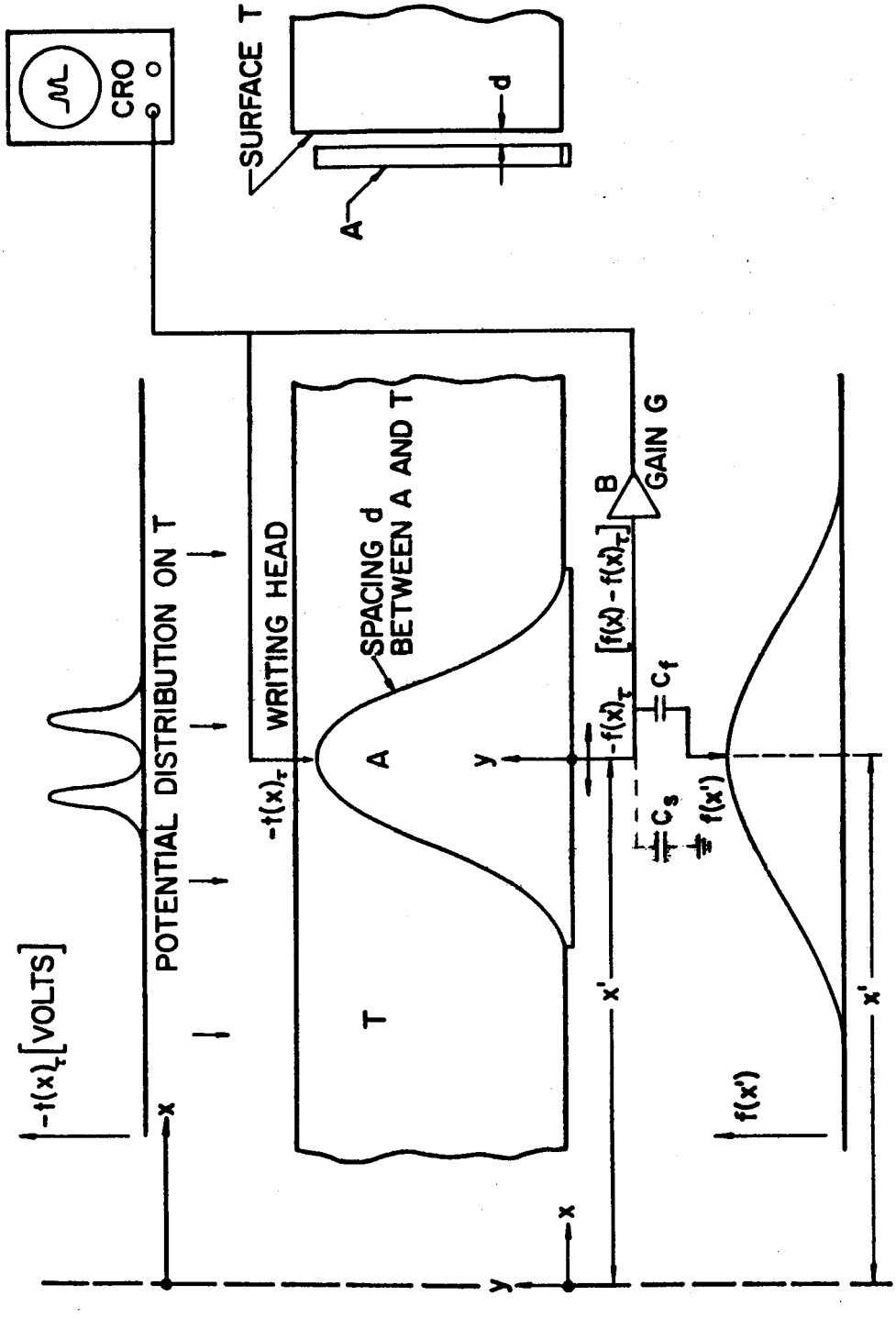


FIG.2 BASIC PRINCIPLE OF ANALOG SYSTEM FOR DECONVOLUTION

which upon solving for V_B becomes

$$V_B = \frac{C_A V_A + C_f V_f}{C_A + C_f} . \tag{14}$$

Returning to equation (11) and making the assumption that V_B is negligible compared to V_A, then

$$q_A = C_A V_A . \tag{15}$$

Placing this equation into (14) yields

$$V_B = \frac{q_A + C_f V_f}{C_A + C_f} . \tag{16}$$

To compute q_A, let us consider a differential element dx on surface T. Assuming again that the input of B is close to ground and applying Gauss' law while using the relation $E(x) = \frac{t(x)_T}{d}$ where $t(x)_T$ is a trial solution for t(x), then

$$dq_A = \frac{\epsilon_o t(x)_T a(x-x')dx}{d} . \tag{17}$$

But $dq_A = t(x)_T dC_A$, hence

$$dC_A = \frac{\epsilon_o a(x-x')dx}{d} \tag{18}$$

which upon integration yields

$$C_A = \int_{-\infty}^{\infty} \frac{\epsilon_o}{d} a(x-x')dx = \frac{\epsilon_o a}{d} \tag{19}$$

where a is just the area of electrode A. Reconsidering dq_A and

integrating, we get

$$q_A = \frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_\tau a(x-x')dx. \tag{20}$$

Placing (19) and (20) into (16)

$$V_B = \frac{\frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_\tau a(x-x')dx + C_f V_f}{\frac{\epsilon_o a}{d} + C_f}. \tag{21}$$

To extend the analysis to a practical situation, the stray capacity to ground C_S must be included, which upon introduction into (16) yields

$$V_B = \frac{\frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_\tau a(x-x')dx + C_f V_f}{\frac{\epsilon_o a}{d} + C_f + C_S}. \tag{22}$$

The voltage V_B could also be obtained by considering a Gaussian surface enclosing the sides of the capacitors that are directly connected to amplifier B. To continue, we write the equation for the output of amplifier B which is

$$t(x)_\tau = -GV_B = \frac{-G\left[\frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_\tau a(x-x')dx + C_f V_f\right]}{K} \tag{23}$$

where G is the gain of amplifier B and for convenience we have defined $K = \frac{\epsilon_o a}{d} + C_f + C_S$. Rewriting (23)

$$\frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_T a(x-x')dx + C_f V_f = \frac{t(x)_T K}{G}. \qquad (24)$$

If the gain is made very large compared to $t(x)_T K$ so that V_B is very small, then

$$\frac{\epsilon_o}{d} \int_{-\infty}^{\infty} t(x)_T a(x-x')dx + C_f V_f \cong 0 \qquad (25)$$

Therefore

$$V_f \cong \frac{\epsilon_o}{dC_f} \int_{-\infty}^{\infty} \left[-t(x)_T\right] a(x-x')dx \qquad (26)$$

However, if this is true then the voltages $t(x)_T$ must describe very closely the negative of the true function t(x) that gave rise to f(x'). Hence, if electrode A has been oriented to allow for the reversed sense of the argument in a(x-x'), we have deconvolution.

5. Experimental Apparatus

5.1 Description of Physical Structure of Computer

The same physical structure was used for both the convoluting and deconvoluting modes. The surface T consisted of 125 copper strips that were mounted on a plexiglass platen that had been machined to a tolerance of $\pm$ 0.001 in. These strips were insulated from each other and were connected to 0.01μf polystyrene-mylar capacitors. A plexiglass carriage equipped with wipers was made to roll on tracks that were also machined to $\pm$ 0.001 in. as shown in figure 3. This tolerance was necessitated by the $\frac{1}{d}$ relationship shown in equation (9) and the small value of d chosen to minimize fringing and to obtain a good signal to noise ratio. In order to maintain this tolerance under variations in climatic conditions, a system of braces and jacks was added.

The shaped electrode A was mounted on an insert that fitted into the frame of the carriage. This insert had a three point suspension system so that the electrode could be made parallel to T. The electrode was made from aluminum foil, and mounting to the insert was done with thin pressure sensitive double coated tape. Shaped electrode A was driven across T by the writing arm of the X-Y recorder.

5.2 Procedure for Constructing Shaped Electrode

The shape of the electrode was prepared by first making a tracing of the function $a(x-x')$. From this tracing a cardboard form of $a(x-x')$ was made. This form was placed over the aluminum foil that was already mounted on the insert. A razor's

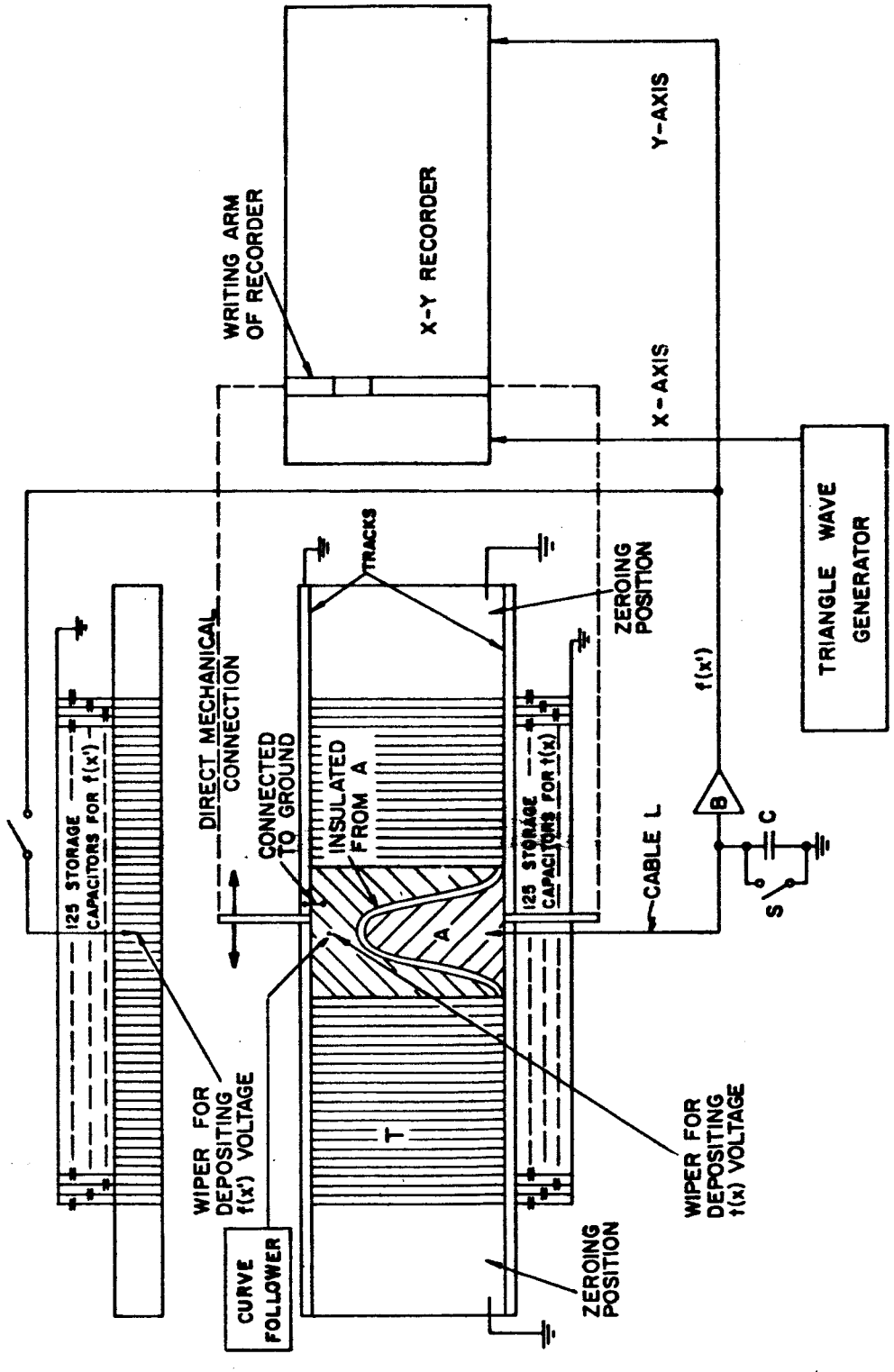


FIG. 3 EXPERIMENTAL APPARATUS FOR CONVOLUTION

edge was run along the edge of the form, thus producing the function on the foil. The cardboard form was moved slightly and a second line describing $a(x-x')$ was cut into the foil. A set of tweezers was used to remove the small strip of foil between the two functions. Resulting was the shaped electrode describing $a(x-x')$ that was immediately surrounded by the remaining portion of the foil which was grounded to minimize fringing as seen in figure 3.

5.3 Auxiliary Memory

In addition to the capacitive memory for $t(x)$, another 125 copper strips and capacitors of the same value and type were mounted on the platen. This served as an extra memory for $f(x')$. This is shown as detached from the platen in figure 3 to indicate that it was non-essential for the operation of the device in either mode of operation.

5.4 Electronics of Convoluting Mode

The electronics associated with the convoluting mode of operation are shown in figure 3. The copper strips and the electrode describing $a(x-x')$ corresponded respectively to surface T and electrode A mentioned in the section on theory. Voltages were introduced on T by direct contact to a power supply or automatically by connecting the output of the curve follower to the appropriate wiper on the cart. Changes in these voltages of less than 0.25 per cent were typical for storage times of ten hours. Polarization tests were also performed with voltages up to 350V with no detectable permanent polarization of the capacitors. Cable

L connected electrode A with amplifier B. This cable was a graphite coated coaxial cable where the graphite coating helped to reduce the noise introduced by the cable flexing as the carriage swept across surface T. Amplifier B was an electrometer with an input impedance of $10^{14}\Omega$. This high input impedance was necessary to minimize leakage. Since the capacity of the electrode A and associated cables was in the hundreds of picofarads range, a time constant in the ten thousand second range was possible which far exceeded the time for a sweep between the two zeroing positions. Capacitor C was a 1000 pf mylar capacitor. This capacity was large enough to keep the voltages on A near ground. The voltages on A seldom exceeded a maximum of 0.75 V which corresponded to several hundred volts on T; thus, the A voltages essentially did not alter voltage distributions on T. Switch S short-circuited capacitor C when electrode A was over the zeroing position to remove any residual charge on C. Switch S was specifically designed for operation in such a high impedance network. A shielded relay coil and arm were used to drive a plexiglass rod that made and broke the contacts of the shielded switch S. This method was necessary to eliminate the large switching transients associated with standard commercial relays. The switching transient associated with this arrangement was $\pm$ 2mV. Finally, the output of amplifier B was plotted on the X-Y recorder.

5.5 Electronics of Deconvoluting Mode

The electronics of the deconvoluting mode are shown in the block diagram of figure 4. Surface T, electrode A, cable L,

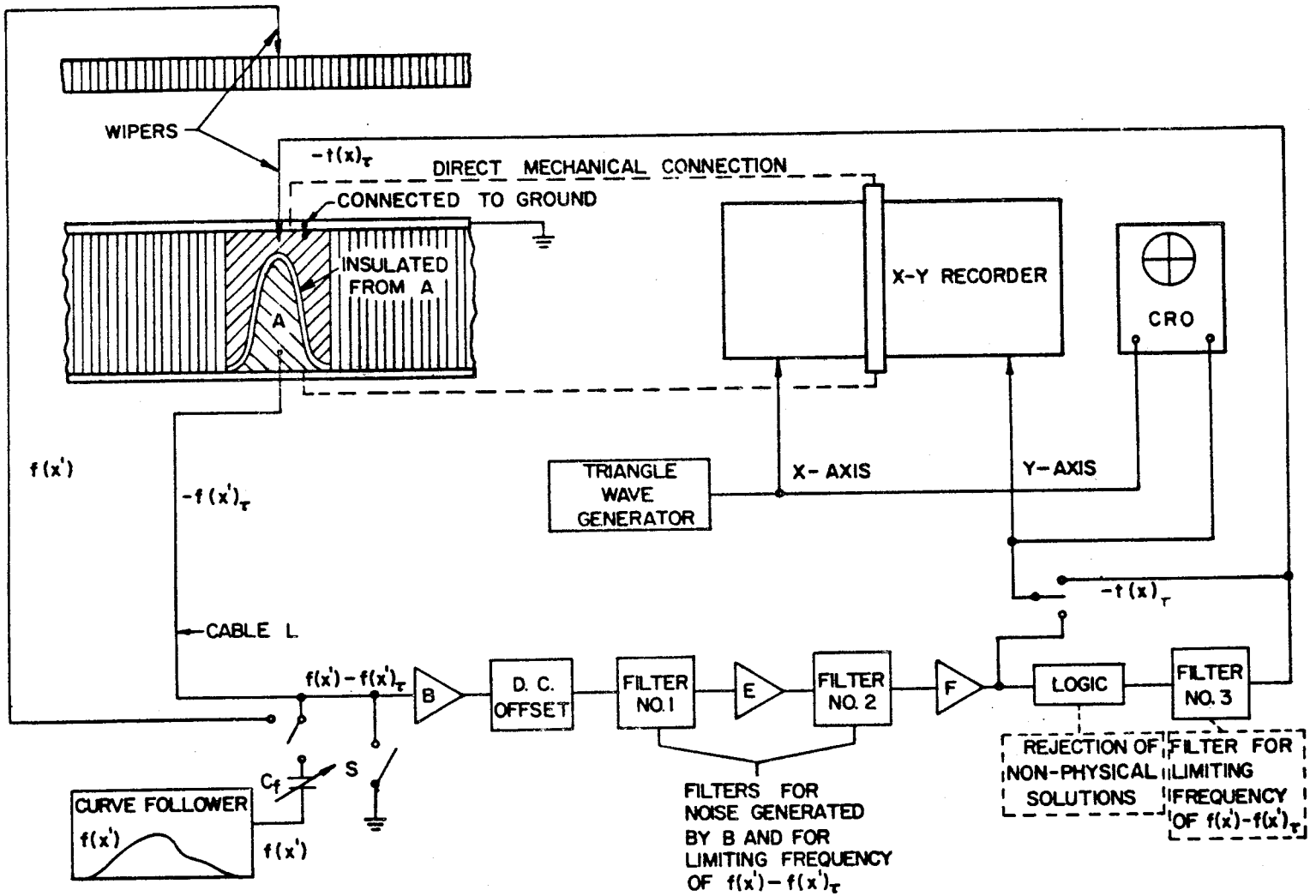


FIG. 4 EXPERIMENTAL APPARATUS FOR DECONVOLUTION

switch S and amplifier B have been described previously. The dc offset was introduced into the feedback loop to allow easy compensation for the small dc level of amplifier B. Filters 1 and 2 were necessary to eliminate high frequency noise generated by amplifier B, to reduce sixty-cycle noise, and to limit the response of the system. Amplifiers E and F were added to provide the large gain required by the theory. Amplifier B was run with a gain of 300 when the system was near equilibrium, while amplifiers E and F were run at gains of 10 and 2 respectively, giving a net amplifier gain of 6000. The electronic details of the dc offset, filters 1 and 2, and amplifiers E and F are shown in figure 5. The logic was introduced to prevent any positive voltages from being deposited on T. This corresponded to rejecting nonphysical solutions since most spectra, excluding derivative spectra, are such that they lie on only one side of the base line. Filter 3 was constructed to eliminate frequencies that were low enough to pass filters 1 and 2 but were still able to prevent $t(x)_T$ from being smooth. However, the smoothing of filter 3 did not remove the important structure of $t(x)_T$ as will be seen in the results. The voltage V_f describing $f(x')$ was generated by an optical line follower viewing plots of $f(x')$. The maximum output voltage of the line follower was 6V but an additional potentiometer was included to control the maximum voltage appearing across C_f. A capacitor was placed in this curve follower network to smooth V_f. Finally, the capacitor C_f was a trimmer capacitor ranging in value from 80-150 pf. This

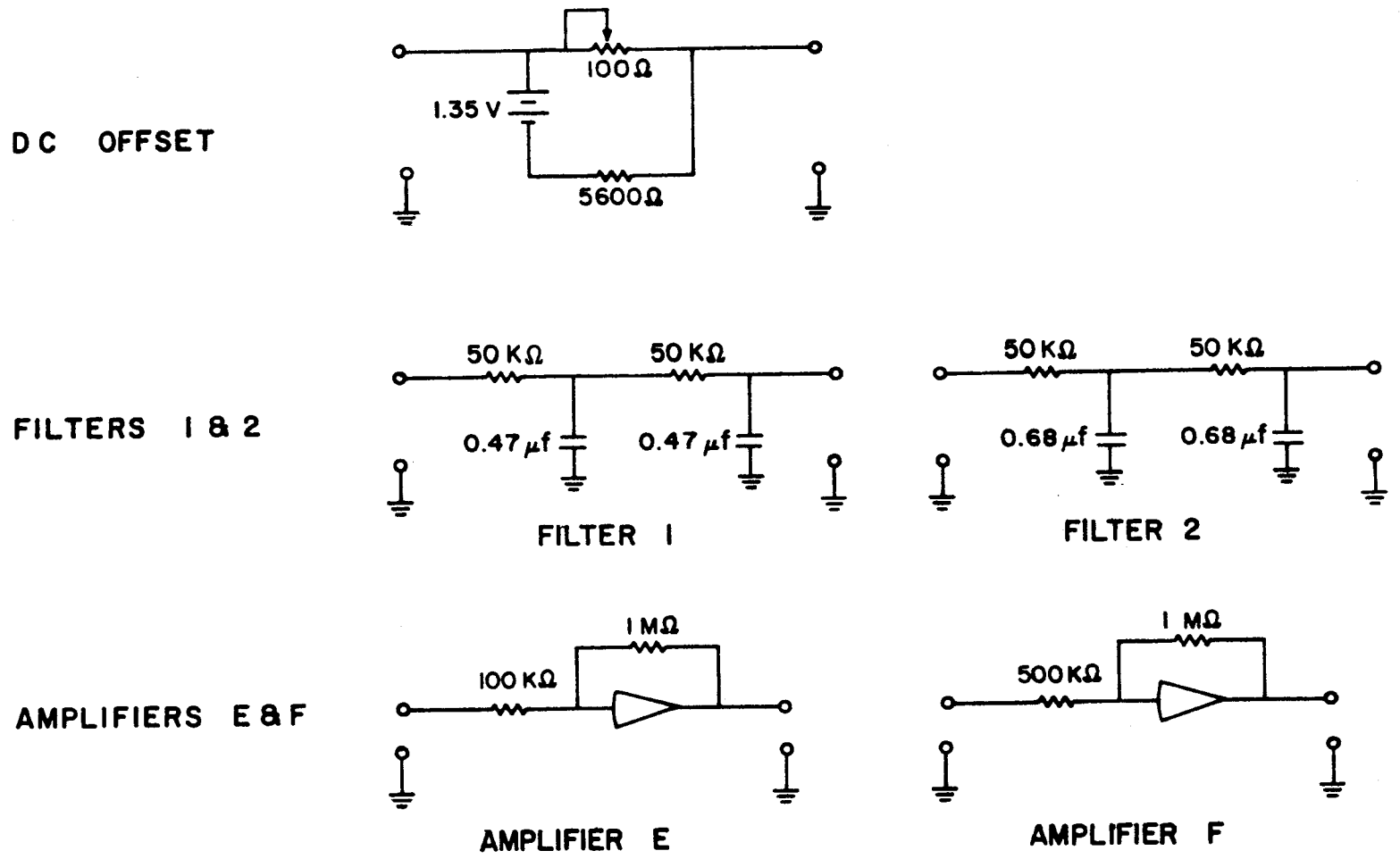


FIG. 5 ELECTRONIC DETAILS OF DC OFFSET, FILTERS 1&2, AND AMPLIFIERS E & F

capacitor gave added control over the magnitude of the voltage appearing at the input of amplifier B. The exact electronic details of the logic, filter 3, and the curve follower network are shown in figure 6.

In figure 4, it can be seen that the oscilloscope and recorder displayed the trial solution as the voltages on T were modified. The output of amplifier F was also displayed. This output contained nonphysical results before they were rejected by the logic. This rejection in turn caused a further modification of the trial solution because of the closed loop nature of the system. Viewing the magnitude of these nonphysical results gave a measure of the mismatch between $f(x')$ and $f(x')_T$. It should be noted at this time that the method of successive modifications of the voltages on the $t(x)_T$ memory and the viewing of the nonphysical results before the logic are common to the computers of Kendall (1961 and 1962). Moreover, the rejection of nonphysical results was included in the programs of Mori and Doi (1964).

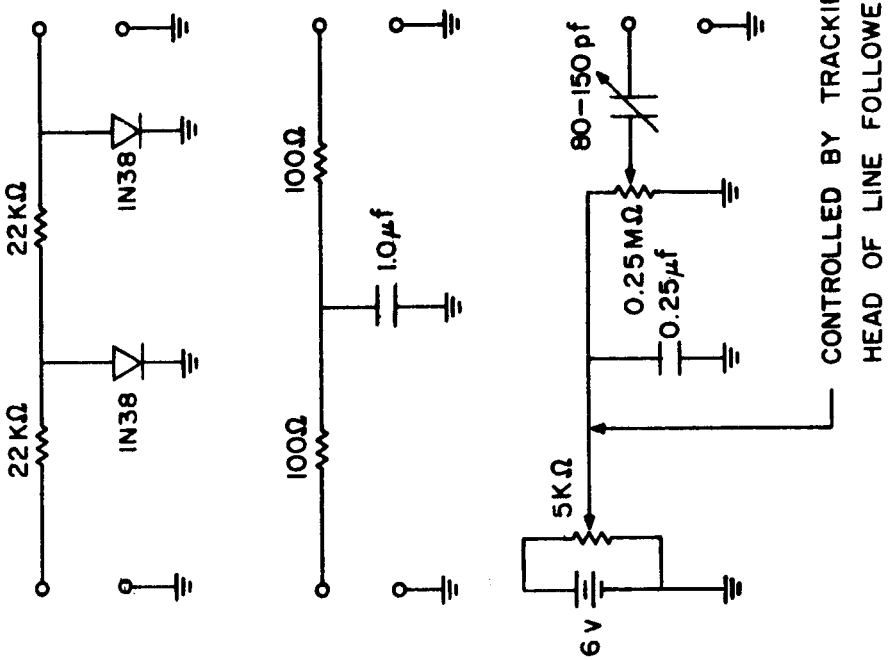


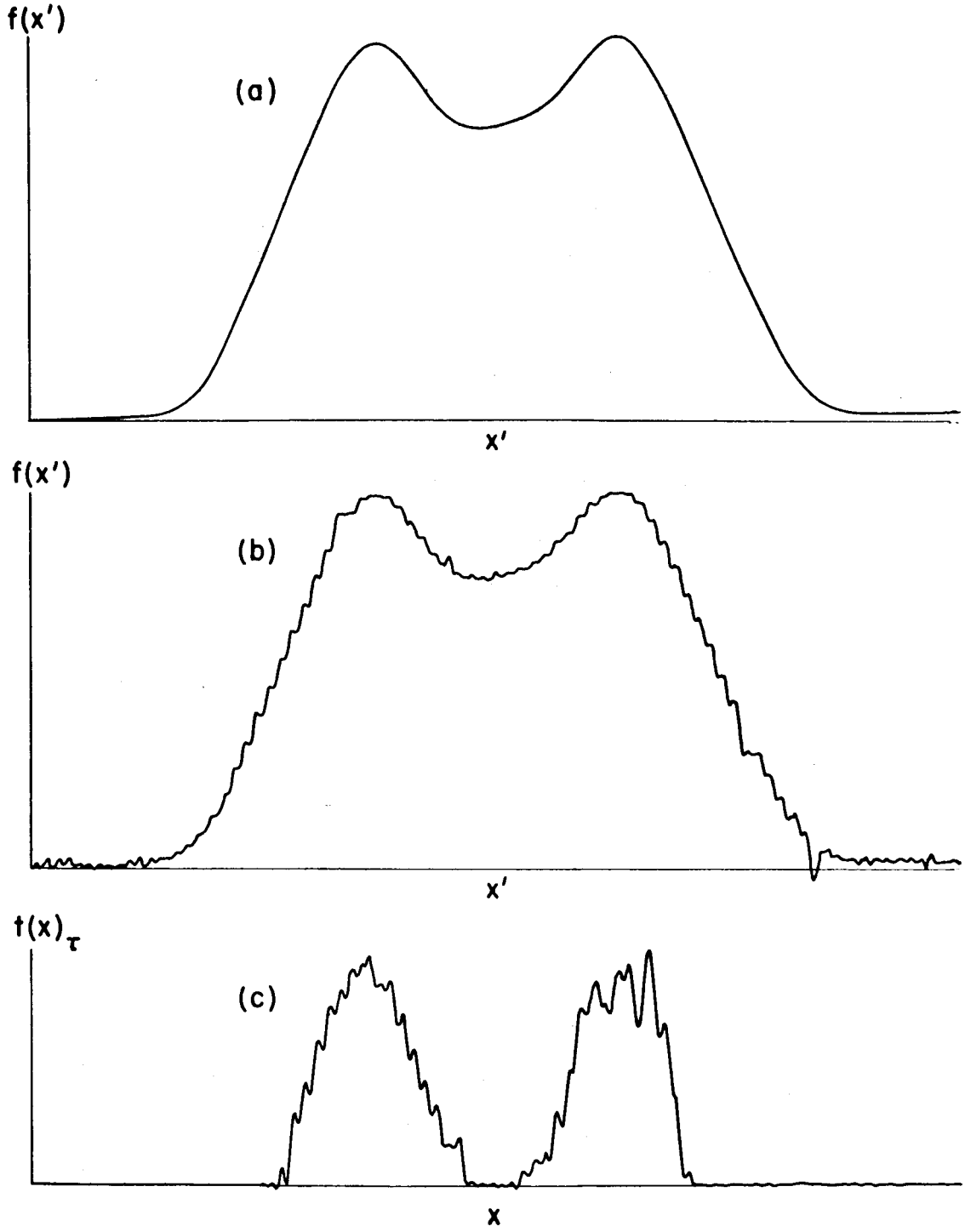
FIG. 6 ELECTRONIC DETAILS OF LOGIC, FILTER 3, AND CURVE FOLLOWER

6. Results

6.1 Deconvolution Without Filter

To illustrate the importance of filter 3, figure 7(c) shows data that was obtained before its introduction into the system. Curve 7(a) is the convolution f(x') in which a(x-x') is an isosceles triangle of arbitrary width b at the base and t(x) is two approximately isosceles triangles of width 0.246 b at the base and a distance 0.556 b between their apexes. This convolution was obtained electronically. The slight asymmetry is due to the poor regulation of the power supply that was used to introduce t(x) on the strips; thus, the triangles were not exactly isosceles. It can be seen, also, that f(x') does not exactly return to the base line. This is due to the flexing of cable L and some minor leakage. Curve 7(b) is a plot of the convolution as stored in the discrete auxiliary memory described in the section on the experimental apparatus. Curve 7(c) is the deconvolution of 7(b). The jagged structure of 7(c) can be principally attributed to the insufficient response suppression of filters 1 and 2. A secondary cause of this structure was the discrete nature of the auxiliary memory; however, even when the curve follower was used as the memory for f(x'), this structure was still observed. The attempt to produce the triangular structure of t(x) is apparent in 7(c).

The remaining deconvolution data that shall be presented were obtained with filter 3 in the system and with the curve follower serving as the memory for f(x').



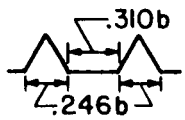
(a) CONVOLUTION OF $a(x-x')$ =  AND $t(x)$ = 
(b) f(x') STORED BY AUXILIARY MEMORY
(c) DECONVOLUTION t(x) OF f(x') WITHOUT FILTER 3

FIGURE 7

6.2 Deconvolution to a Delta Function Singlet

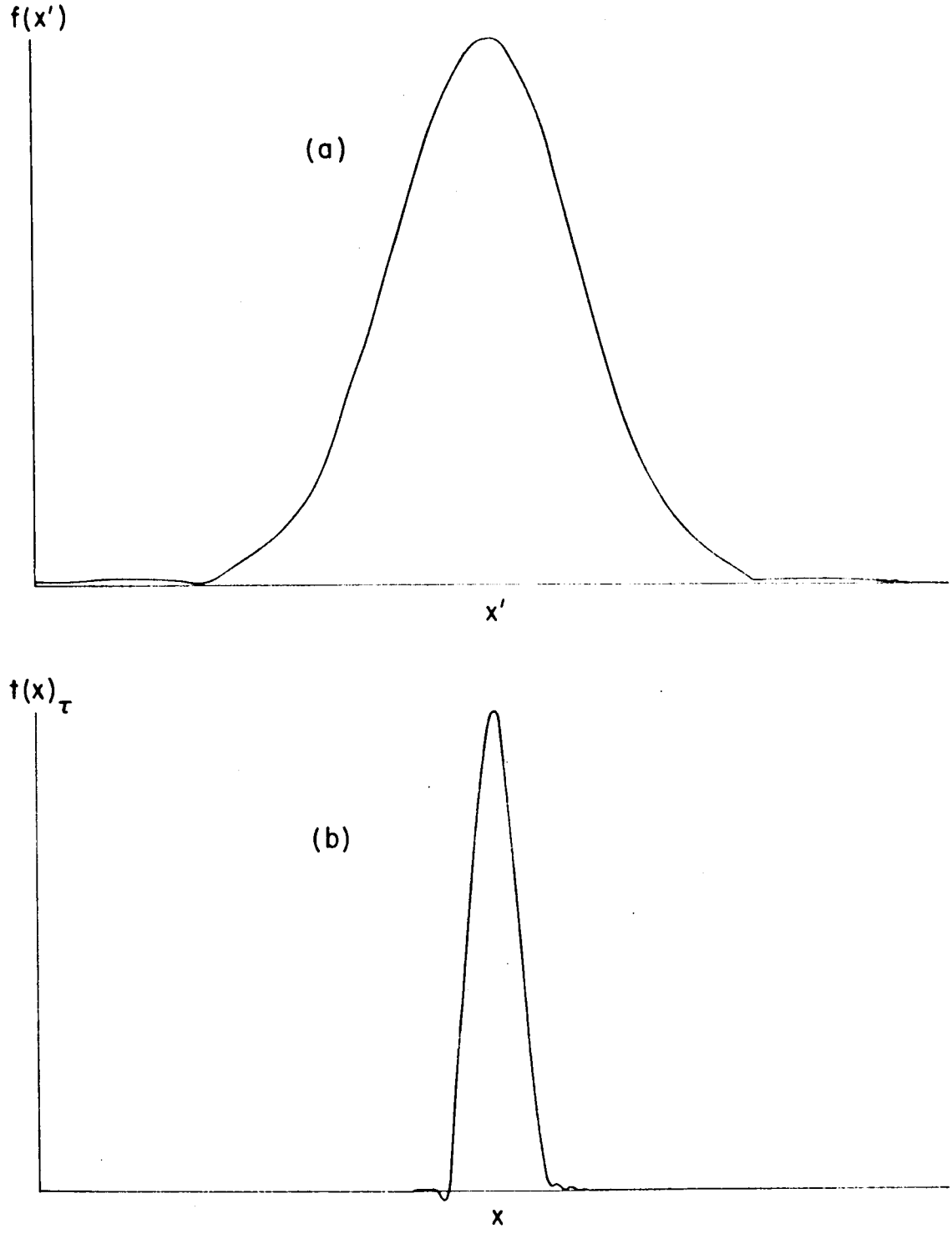
Curve 8(a) is the convolution f(x') in which

$$a(x-x') = \frac{e^{\frac{-(x-x')^2}{2}}}{\sqrt{2\pi}}$$

and $t(x) = \delta(x)$. As is apparent, $a(x-x')$ is a normalized Gaussian function of $\sigma = 1$ where σ is the standard deviation. Applying the rule for integration involving Dirac delta functions, it is obvious that

$$f(x') = \frac{e^{\frac{-x'^2}{2}}}{\sqrt{2\pi}}.$$

The error between this electronically computed integration and the hand calculated integration is 0.52 per cent with respect to the width at half height. Again, the failure of f(x') to return to the base line can be attributed to cable flexing and leakage. Cable L was oriented to reduce the stresses introduced by flexing. Considerable effort was spent in minimizing leakage. Cable lengths were reduced to a minimum and extra insulation was added where necessary and physically possible. Curve 8(b) represents the deconvolution back to the "delta function". This represents an enhancement in resolution by a factor of approximately 6; that is, the width at half-height of 8(b) is a sixth of the width at half-height of 8(a). The system could not return completely to the delta function, which is equivalent to having only one strip charged, because the response of the system has been deliberately limited for reasons already stated.



(a) CONVOLUTION OF A GAUSSIAN AND A DELTA FUNCTION SINGLET
(b) DECONVOLUTION OF 8(a)

FIGURE 8

6.3 Deconvolution to a Delta Function Doublet

Curve 9(a) is the convolution f(x') of

$$a(x-x') = \frac{e^{\frac{-(x-x')^2}{2}}}{\sqrt{2\pi}}$$

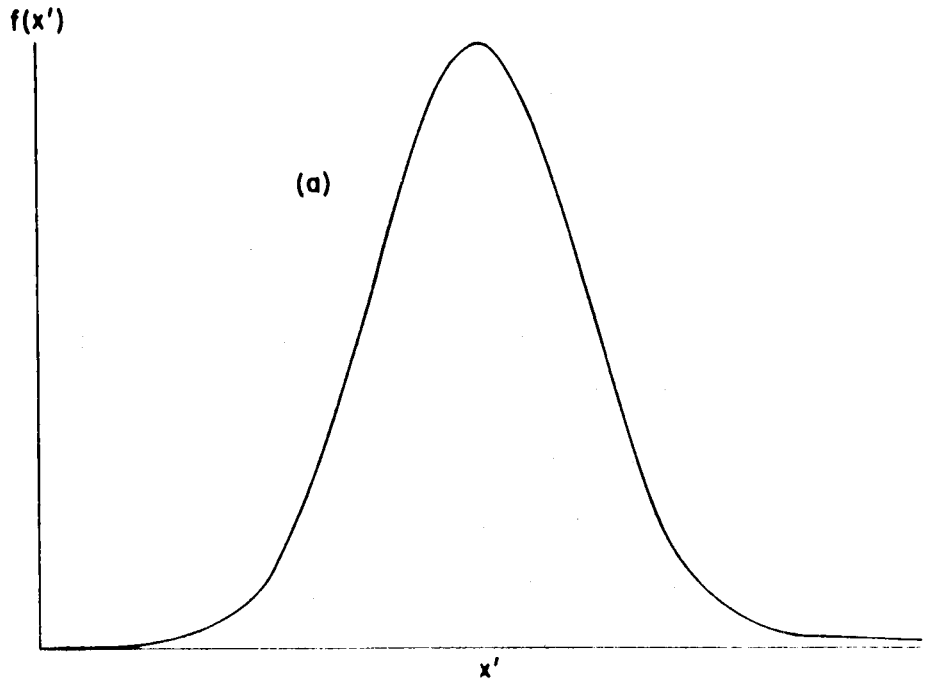
and $t(x) = a_1\delta(x + 0.267w) + a_2\delta(x - 0.267w)$ where the amplitude coefficients a_1 and a_2 are equal and their sum equals one and where $\underline{w}$ is the width at half-height of $a(x-x')$. More commonly this is known as a doublet with equal peak heights. This integration can be compared with figure 11(a) which was hand calculated. The error in width at half-height is 1.7 per cent due to causes already mentioned. Curve 9(b) represents the deconvolution of 9(a). The doublet structure is obtained but there is a small error in peak heights. Instead of amplitude coefficients being equal to 0.500, a_1 equals 0.470 and a_2 equals 0.530. This error can be attributed to the errors in the convolution 9(a), to the leakage and cable flexing, and to the ill-behaved mathematical nature of deconvolution.

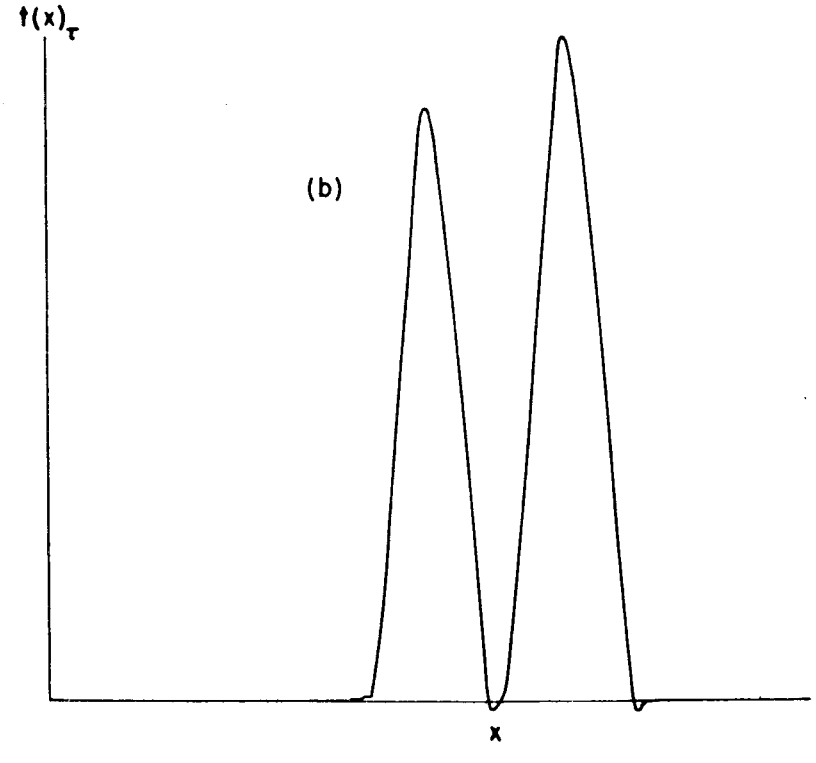
6.4 Deconvolution to a Delta Function Triplet

Curve 10(a) is the convolution of

$$a(x-x') = \frac{e^{\frac{-(x-x')^2}{2}}}{\sqrt{2\pi}}$$

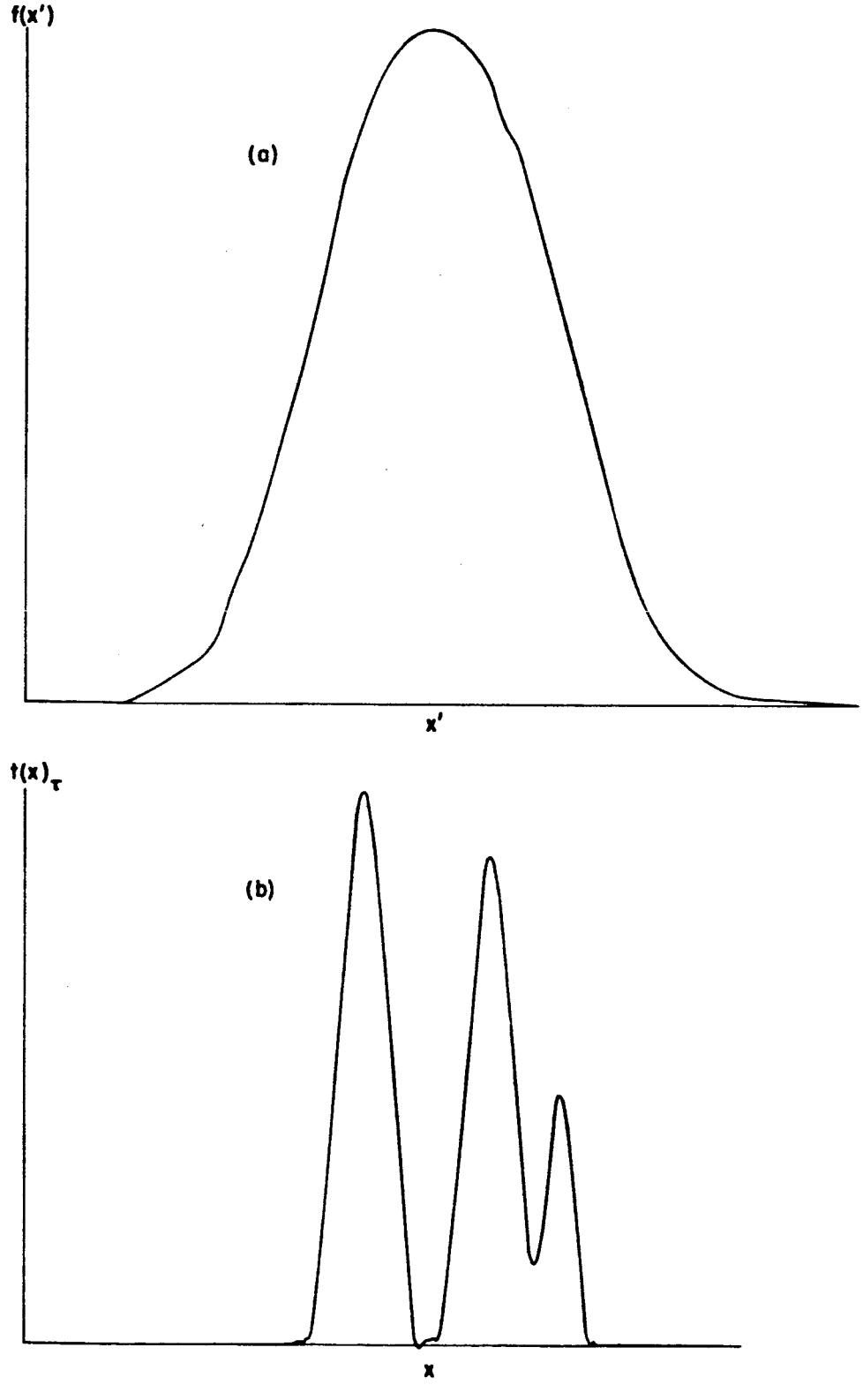
and $t(x) = a_1\delta(x + 0.533w) + a_2\delta(x) + a_3\delta(x - 0.533w)$ where the amplitude coefficients a_1, a_2, and a_3 are equal and their sum equals one and where $\underline{w}$ is the width at half-height of $a(x-x')$. This is simply known as a triplet with equal peak heights. The error in width at half-height of 10(a) compared to the hand





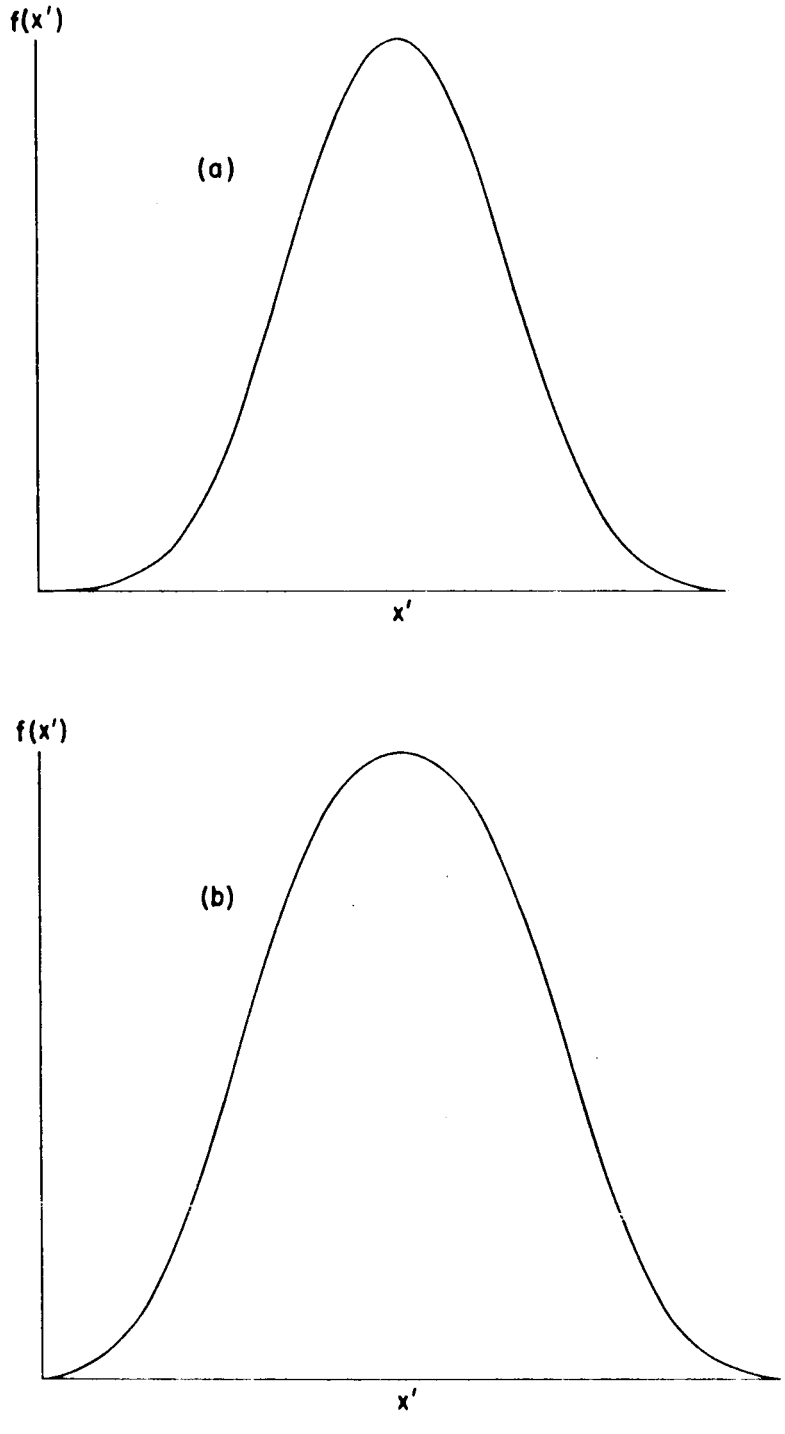
(a) CONVOLUTION OF A GAUSSIAN AND A DELTA FUNCTION DOUBLET
(b) DECONVOLUTION OF 9(a)

FIGURE 9



(a) CONVOLUTION OF A GAUSSIAN AND A DELTA FUNCTION TRIPLET
(b) DECONVOLUTION OF 10(a)

FIGURE 10



(a) HAND COMPUTED CONVOLUTION OF A GAUSSIAN AND A DELTA FUNCTION DOUBLET
(b) HAND COMPUTED CONVOLUTION OF A GAUSSIAN AND A DELTA FUNCTION TRIPLET
FIGURE 11

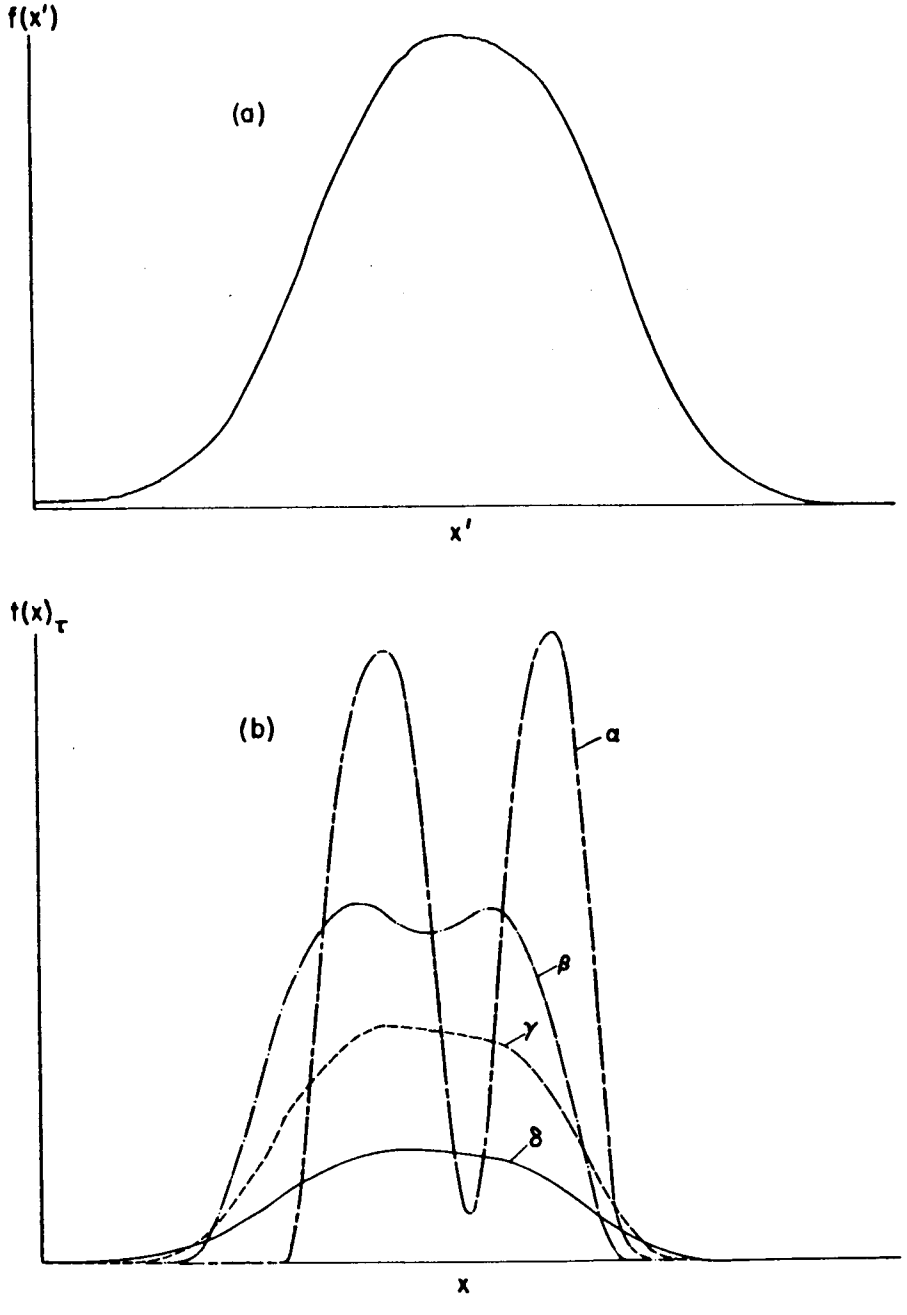
calculated 11(b) is 2.00 per cent. Curve 10(b) is the deconvolution of 10(a). The triplet structure is obtained but there are errors in peak heights. Instead of all the amplitude coefficients being equal to 0.333, a_1 equals 0.424, a_2 equals 0.372 and a_3 equals 0.194. Again, the sources of error are the same as those stated in the previous paragraph.

6.5 Deconvolution to a Square Function Doublet and Importance of Gain

Curve 12(a) is the convolution of

$$a(x-x') = \frac{e^{\frac{-(x-x')^2}{2}}}{\sqrt{2\pi}}$$

and $t(x) = a_1\left[H(x - 0.267w) - H(x - 0.533w)\right] + a_2\left[H(x - 1.066w) - H(x - 1.333w)\right]$ where $H(x)$ is a Heaviside step function, the coefficients a_1 and a_2 are equal with their sum equal to one, and $\underline{w}$ is the width at half height of $a(x-x')$. This true function is just two square functions of equal height and width and separated from each other by a distance equal to 0.533w. The reason for $f(x')$ not beginning at zero is that this computation was performed with the limits of integration reversed; that is, the shaped electrode was swept from right to left while in the previous computations the electrode was swept from left to right. The reasons for the small drift from the base line have been listed previously. Figure 12(b) is the deconvolution of 12(a). This figure illustrates the importance of sufficient loop gain. The amplifier gains that were used to generate curves δ and γ were 200 and 600 respectively. It is seen that these curves



(a) CONVOLUTION OF A GAUSSIAN AND A SQUARE FUNCTION DOUBLET
(b) DECONVOLUTION OF 12(a) WHERE $G_\alpha > G_\beta > G_\gamma > G_\delta$
FIGURE 12

follow quite closely $f(x')$. The amplifier gains that were used to generate curves β and α were 2000 and 6000 respectively. The doublet structure is apparent in curve β and the doublet is almost completely separated in curve α. The system could not produce square functions because of the filters but the error in peak heights is small. Instead of both amplitude coefficients being equal, a_1 equals 0.486 and a_2 equals 0.514.

7. Conclusions and Summary

7.1 Problems Encountered in Deconvolution

The ill-behaved mathematical nature of deconvolution has been demonstrated by Fox and Goodwin (1953). The dramatic and adverse effect of noise on deconvolution can be explained in view of this characteristic. The work of Rautian (1958) discusses this effect and the work of Morrison (1963) confirms it. With these facts in mind, the results that have been presented in the previous section are quite favorable despite some of the apparently large errors of the triplet deconvolution. Reiterating, these errors can be ascribed to the noise generated by the flexing of cable L, the leakage in the system and the errors of the convolutions being processed. All three of these sources of error can be viewed as noise.

7.2 Suggested Improvements to Computer

Fortunately, the elimination of these sources of error is possible. A system in which amplifier B is mounted on the carriage in which electrode A is inserted would remove the noise generated by the flexing of cable L. The cables that would be flexing in such a system would be connected to the curve follower and the output of amplifier B, which are low impedance sections of the circuitry. Amplifier B would have to meet the requirements of the theory and be small to facilitate mounting. This arrangement would reduce leakage by reducing the length of the cable in the input circuit of amplifier B. A further reduction in leakage could be obtained by mounting all cable connections in the input circuit on polystyrene or teflon blocks. With the

removal of these first two sources of error, some of the errors in the convolutions would in turn be removed.

7.3 Application to Practical Data

At this point the objection might be raised that the convolutions that were deconvoluted were generated by this device; thus, a circular argument seems to have been used to demonstrate that the computer functions properly. This objection can be dispelled by simply referring to the errors in width at half-height between the electronically computed convolutions and the hand calculated convolutions. The maximum error was only 2 per cent and the average error was 1.41 per cent. Because of these small errors between these two independently produced convolutions, it is safe to assume that data produced by analytical instruments can be processed.

7.4 Advantages of this Computer over Digital Methods

The advantages of this prototype deconvolution computer over the digital techniques are twofold. The first advantage is that of the time saved in processing a given spectrum. To process a spectrum with this computer, only shaped electrode A has to be made and aligned. Compared to the time spent in writing the digital programs, the time spent in performing these two operations is extremely small. This computer can be set up in a laboratory to process data immediately. Data cannot be processed immediately with the digital methods unless shared-time facilities are available which is usually not the case. The total time spent in performing a given computation is less than a minute which is,

in general, the same as or shorter than the digital times. The second advantage is that in order to obtain reasonably good deconvolution results smoothing techniques which consist of separate mathematical operations have to be introduced in the digital programs while this new computer does this smoothing automatically. Finally, the results of this computer are comparable to the results obtained by the digital methods.

7.5 Advantages of this Computer over Previous Analog Methods

The analog methods of French et al and Noble et al are not in the strictest sense capable of deconvolution. These devices synthesize the convolution by manually constructing peaks appropriately along the horizontal axis and adding together the ordinates of the overlapping portions of these peaks such that the ordinate of this sum at a given horizontal point equals the ordinate of the convolution at that point. The Model 310 Curve Resolver of the DuPont Company, as has been mentioned, is an adaptation of the Noble device. Curves 9(a) and 10(a) were processed by this DuPont device. The DuPont device correctly interpreted curve 9(a) as a doublet but the amplitude coefficients were in 40 per cent error while the computer described in this thesis yielded only a 6 per cent error in amplitude coefficients. The DuPont device incorrectly interpreted curve 10(a) as a doublet instead of a triplet while the computer described in this thesis correctly interpreted the triplet structure but with a rather large error in amplitude coefficients. In addition, the analyses of curves 9(a) and 10(a) by the DuPont device were

performed by the most experienced operator of this device in the DuPont organization. Moreover, because an operator manually constructs the solutions on this DuPont device, the solutions are subject to bias which is not the situation with the computer described in this thesis.

The previously automated computers of Kendall (1961 and 1962) were limited to dealing with situations in which the true function consisted of a series of delta functions at a limited number of positions. This new method is not restricted in this sense as demonstrated by the deconvolution involving the two square functions. The net amount of delta function data that can be processed at one time by the new method is approximately twice that of these earlier computers. The complexity of the electronics in these earlier computers far exceeds that of the new computer. The results of the doublet deconvolution are comparable to that of these earlier computers and it is thought that the suggested modifications can bring the triplet data into better agreement with the theoretical results.

7.6 Implications of Further Development

The further development of this computer is of great potential importance. Low resolution instruments which usually have the advantages of simple design, reliability, and compactness are not widely used. With the sixfold increase in effective resolution offered by this prototype computer, low resolution instruments could be used more widely and their advantages could be exploited. These advantages are of immediate importance in

the study of the upper atmosphere and space because of the expenditures in time and money associated with the firing of a rocket and the limited lift capabilities of present rockets.

The useful range of operation of moderate resolution devices can be extended with considerable savings in money compared to the cost of purchasing a high resolution device. The components of this computer are commercially available with the exception of the plexiglass platen with copper strips. The construction of this platen consumed a very small amount of material but a rather large amount of time. The time spent in this construction could be reduced by the methods employed in the construction of printed circuitry.

The resolution of high resolution instruments could be further increased by this computer thus allowing the detailed study of phenomena that are considered possible in theory but are beyond present experimental techniques. And finally, the discovery of totally new phenomena not known in theory would be made possible by the resolution enhancement offered by this computer to high resolution instruments.

BIBLIOGRAPHY

1. Duffieux, P. M., Revue Opt. Theor. Instrum., 39, 491, (1960).

2. Emslie, A. G. and King, G. W., J. Opt. Soc. Amer., 43, 658, (1953).

3. Fox, L. and Goodwin, E. J., Proc. Cambridge Phil. Soc., 245, 501, (1953).

4. French, C. S., Towner, G. H., Bellis, D. R., Cook, R. M., Fair, W. R., and Holt, W. W., Rev. Sci. Instrum., 25, 765, (1954).

5. Gardner, L. B., IRE Trans. Nuc. Sci., NS-7, 36, (1960).

6. Kendall, B.R.F., Rev. Sci. Instr., 32, 758, (1961).

7. Kendall, B.R.F., Rev. Sci. Instr., 33, 30, (1962).

8. Kendall, B.R.F., Jour. Sci. Instr., 39, 267, (1962).

9. Kendall, B.R.F., Jour. Sci. Instr., 43, 215, (1966).

10. King, G. W. and Emslie, A. G., J. Opt. Soc. Amer., 41, 405, (1951).

11. Mikusinski J., Operational Calculus, Pergamon, New York, (1959).

12. Mori, M. and Doi, K., Jap. J. Appl. Phys., 3, 112, (1964).

13. Morrison, J. D., J. Chem. Phys., 39, 200, (1963).

14. Noble, F. W., Hayes, J. E. and Eden, M., Proc. Inst. Radio Engrs., 47, 1952, (1959).

15. Rautian, S. G., Soviet Physics (Uspekhi), 1(66), 245, (1958).

16. Rollett, J. S. and Higgs, L. A., Proc. Phys. Soc., 79, 87, (1962).

17. Sachenko, V. P., Bull. Acad. Sci. USSR, 25, 1052, (1961).

18. Skarsgard, L. D., Johns, H. E. and Green, L. E. S., Radiation Research, 14, 261, (1961).

19. Whittaker, E. T. and Watson, G. N., A Course of Modern Analysis, Cambridge, New York, (1963).

20. Widder, D. V. and Hirschman, I., Convolution Transforms, Princeton Univ. Press, Princeton, (1955).