

United States Patent	5,410,492
Gross , et al.	April 25, 1995

PROCESSING DATA BASE INFORMATION HAVING NONWHITE NOISE

Abstract

A method and system for processing a set of data from an industrial process and/or a sensor. The method and system can include processing data from either real or calculated data related to an industrial process variable. One of the data sets can be an artificial signal data set generated by an autoregressive moving average technique. After obtaining two data sets associated with one physical variable, a difference function data set is obtained by determining the arithmetic difference between the two pairs of data sets over time. A frequency domain transformation is made of the difference function data set to obtain Fourier modes describing a composite function data set. A residual function data set is obtained by subtracting the composite function data set from the difference function data set and the residual function data set (free of nonwhite noise) is analyzed by a statistical probability ratio test to provide a validated database.

Inventors: **Gross; Kenneth C.** (Bolingbrook, IL); **Morreale; Patricia** (Park Ridge, IL)

Assignee: **ARCH Development Corporation** (Chicago, IL)

Appl. No.: **068712**

Filed: **May 27, 1993**

Current U.S. Class: **702/185; 376/215; 376/216; 376/217; 706/914**

Intern'l Class: **G21C 007/36; G06F 007/00**

Field of Search: **376/215,216,217 395/914,915 364/492,576,554,600**

Parent Case Text

This application is a continuation-in-part of U.S. Ser. No. 07/827,776, filed Jan. 29, 1992 now U.S. Pat. No. 5,223,207.

The United States Government has rights in this invention pursuant to Contract W-31-109-ENG-38 between the U.S. Department of Energy and the University of Chicago.

Description

The present invention is concerned generally with a system and method for reliably processing a database of information having nonwhite noise characteristics. More particularly, the invention is concerned with a system and method for removal of nonwhite noise elements, such as serially correlated noise, from an incoming stream of data or from an existing database. This method allows use of reliable data, or providing a validated data set, from an industrial process and/or sensors that monitor the process. Such a system further allows accumulation of characteristic faulty data arising from nonwhite noise, which can be useful in understanding the varieties and characteristics of noise that cause false alarm conditions.

Conventional parameter-surveillance schemes are sensitive only to gross changes in the mean value of a process, or to large steps or spikes that exceed some threshold limit check. These conventional methods can pass through data containing bad information characterized by large numbers of false alarms (if thresholds are set too close to normal operating levels) or a large number of missed (or delayed) alarms (if the thresholds are set too expansively). Such a system gives rise to a faulted database of information for researchers or for engineers analyzing the status of an industrial process. Moreover, most conventional methods cannot perceive the onset

United States Patent	5,410,492
Gross , et al.	April 25, 1995

of a process disturbance or sensor deviation, which gives rise to a signal below the threshold level for an alarm condition.

In another conventional monitoring method, the Sequential Probability Ratio Test (SPRT) has found wide application as a signal validation tool in the nuclear reactor industry. Two features of the SPRT technique make it attractive for parameter surveillance and fault detection: (1) early annunciation of the onset of a disturbance in noisy process variables, and (2) the SPRT technique has user-specifiable false-alarm and missed-alarm probabilities. One important drawback of the SPRT technique, which has limited its adaptation to a broader range of nuclear applications, is the fact that its mathematical formalism is founded upon an assumption that the signals it is monitoring are purely Gaussian, independent (white noise) random variables.

It is therefore an object of the invention to provide an improved method and system for processing database information having nonwhite noise present.

It is another object of the invention to provide a novel method and system for statistically filtering industrial process signals having virtually any form of noise signal.

It is a further object of the invention to provide an improved method and system for operating on an industrial process signal database to remove unwanted serially correlated noise signals.

It is still an additional object of the invention to provide a novel method and system utilizing a database formed from a pair of signals to generate a difference function to be analyzed for anomalous data.

It is still a further object of the invention to provide an improved method and system for collecting a reliable database from at least one sensor for providing a real signal characteristic of a process and a predicted sensor signal allowing formation of a difference signal between the predicted and real signal for subsequent analysis and accumulation of data free from nonwhite noise contamination.

It is also an object of the invention to provide a novel method and system wherein a difference function is formed from a database of two sensor signals, and/or pairs of signals and nonwhite noise is removed enabling collection of reliable data.

It is yet an additional object of the invention to provide an improved method and system utilizing variable pairs of sensors for accumulating reliable data both for sensor degradation and industrial processes.

Other objects, features and advantages of the present invention will be readily apparent from the following description of the preferred embodiments thereof, taken in conjunction with the accompanying drawings described below.

Brief Description of the Drawings

FIG. 1 illustrates the specified output of a pump's power output over time;

FIG. 2 shows a Fourier composite curve fit to the pump spectral output of FIG. 1;

FIG. 3 illustrates a residual function characteristic of the difference between FIGS. 1 and 2;

United States Patent	5,410,492
Gross , et al.	April 25, 1995

FIG. 4A shows a periodogram of the spectral data of FIG. 1, and FIG. 4B shows a periodogram of the residual function of FIG. 3;

FIG. 5A illustrates a noise histogram for the pump power output of FIG. 1, and FIG. 5B illustrates a noise histogram for the residual function of FIG. 3;

FIG. 6A shows an unmodified delayed neutron detector signal from a first sensor, and FIG. 6B is for a second neutron sensor. FIG. 6C shows a difference function characteristic of the difference between data in FIG. 6A and 6B; and FIG. 6D shows the data output from a SPRT analysis with alarm conditions indicated by the diamond symbols;

FIG. 7A illustrates an unmodified delayed neutron detector signal from a first sensor, and FIG. 7B is for a second neutron sensor; FIG. 7C shows a difference function for the difference between the data of FIG. 7A and 7B; and FIG. 7D shows the result of using the instant invention to modify the difference function to provide data free of serially correlated noise to the SPRT analysis to generate alarm information and with alarm conditions indicated by the diamond signals;

FIG. 8 illustrates a schematic functional flow diagram of a method of processing a data stream with FIG. 8A showing a first phase of a method of the invention and FIG. 8B showing the application of a method of the invention;

FIG. 9 illustrates schematically the implementation of data by a method of the invention using a DOS-type system;

FIG. 10 is similar to FIG. 9, but processes a plurality of SPRT data-analysis modules outputting data to the database; and

FIG. 11 illustrates an implementation of the method of the invention for a UNIX-type system.

Detailed Description of Preferred Embodiments

In a method of the invention, database information obtained from industrial process sensors can be used to accumulate reliable data and also to modify or terminate degrading or anomalous processes. The method of the invention can thus be applied to an incoming data stream to pre-validate the data or, applied to an existing database, can be analyzed to remove faulty data. The database sensor signals can therefore be used for research purposes as well as to manipulate input data to a statistical analysis technique, such as a process entitled Sequential Probability Ratio Test (SPRT). Details of this process and the inventions therein are disclosed in U.S. patent application Ser. No. 07/827,776, now U.S. Pat No. 5,223,207, which is incorporated by reference herein in its entirety. Also see the co-pending application incorporated by reference herein, filed contemporaneously and entitled, "System for Monitoring an Industrial Process and Determining Sensor" (Ser. No. 08/068,713). The procedures followed in a preferred method are shown generally in FIG. 8. In performing such a preferred analysis of the sensor signals, a dual transformation method is performed, insofar as it entails both a frequency-domain transformation of the original time-series data and a subsequent time-domain transformation of the resultant data. The data stream that passes through the dual frequency-domain, time-domain transformation is then processed with the SPRT procedure, which uses a log-likelihood

United States Patent	5,410,492
Gross , et al.	April 25, 1995

ratio test. A computer software appendix is also attached hereto covering the SPRT procedure and its implementation in the context of, and modified by, the instant invention.

In the preferred embodiment, successive data observations are performed on a discrete process Y, which represents a comparison of the stochastic components of physical processes monitored by a sensor, and most preferably pairs of sensors. In practice, the Y function is obtained by simply differencing the digitized data signals from two respective sensors. Let y_k represent a data sample from the process Y at time t. During normal operation with an undegraded physical system and with sensors that are functioning within specifications, the y_k should be normally distributed with mean of zero. Note that if the two data signals being compared do not have the same nominal mean values (due, for example, to differences in calibration), then the input signals will be pre-normalized to the same nominal mean values during initial operation.

In performing the monitoring of industrial processes, an incoming data stream can be validated or an existing database can be validated. The system's purpose is to declare data from a first system or a second system degraded if the drift in Y is sufficiently large that the sequence of observations appears to be distributed about a mean +M or -M, where M is our pre-assigned system-disturbance magnitude. We would like to devise a quantitative framework that enables us to decide between two hypotheses, namely:

H_1 : Y is drawn from a Gaussian probability distribution function (PDF) with mean M and variance σ^2 .

H_2 : Y is drawn from a Gaussian PDF with mean 0 and variance σ^2 .

We will suppose that if H_1 or H_2 is true, we wish to decide for H_1 or H_2 with probability $(1-\beta)$ or $(1-\alpha)$, respectively, where α and β represent the error (misidentification) probabilities.

From the conventional, well-known theory of Wald, the test depends on the likelihood ratio l_n , where

$$l_n = \frac{\text{The probability of observed sequence } y_1, y_2 \dots, y_n \text{ given } H_1 \text{ true}}{\text{The probability of observed sequence } y_1, y_2 \dots, y_n \text{ given } H_2 \text{ true}} \quad (1)$$

After "n" observations have been made, the sequential probability ratio is just the product of the probability ratios for each step:

$$l_n = (PR_1) \cdot (PR_2) \cdot \dots \cdot (PR_n) \quad (2)$$

or

$$l_n = \prod_{i=1}^{i=n} \frac{f(y_i | H_1)}{f(y_i | H_2)} \quad (3)$$

where $f(y | H)$ is the distribution of the random variable y.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

Wald's theory operates as follows: Continue sampling as long as $A < I_n < B$. Stop sampling and decide H_1 as soon as $I_n \geq B$, and stop sampling and decide H_2 as soon as $I_n \leq A$. The acceptance thresholds are related to the error (misidentification) probabilities by the following expressions:

$$A = \frac{\beta}{1-\alpha}, \text{ and } B = \frac{1-\beta}{\alpha} \quad (4)$$

The (user specified) value of α is the probability of accepting H_1 when H_2 is true (false alarm probability). β is the probability of accepting H_2 when H_1 is true (missed alarm probability).

If we can assume that the random variable y_k is normally distributed, then the likelihood that H_1 is true (i.e., mean M , variance σ^2) is given by:

$$L(y_1, y_2 \dots, y_n | H_1) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{k=1}^n y_k^2 - 2 \sum_{k=1}^n y_k M + \sum_{k=1}^n M^2 \right) \right] \quad (5)$$

Similarly for H_2 (mean 0, variance σ^2):

$$L(y_1, y_2 \dots, y_n | H_2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{k=1}^n y_k^2 \right] \quad (6)$$

The ratio of (5) and (6) gives the likelihood ratio I_n

$$I_n = \exp \left[-\frac{-1}{2\sigma^2} \sum_{k=1}^n M(M - 2y_k) \right] \quad (7)$$

Combining (4) and (7), and taking natural logs gives

$$\ln \frac{\beta}{1-\alpha} < \frac{-1}{2\sigma^2} \sum_{k=1}^n M(M - 2y_k) < \ln \frac{1-\beta}{\alpha} \quad (8)$$

Our sequential sampling and decision strategy can be concisely represented as:

$$\text{If } I_n \leq \ln \frac{\beta}{1-\alpha}, \text{ Accept } H_2 \quad (9)$$

$$\text{If } \ln \frac{\beta}{1-\alpha} < I_n < \ln \frac{1-\beta}{\alpha}, \text{ Continue Sampling} \quad (10)$$

$$\text{And if } I_n \geq \ln \frac{1-\beta}{\alpha}, \text{ Accept } H_1 \quad (11)$$

United States Patent	5,410,492
Gross , et al.	April 25, 1995

Following Wald's sequential analysis, it is conventional that a decision test based on the log likelihood ratio has an optimal property; that is, for given probabilities α and β , there is no other procedure with at least as low error probabilities or expected risk and with shorter length average sampling time.

A primary limitation that has heretofore precluded the applicability of Wald-type binary hypothesis tests for sensor and equipment surveillance strategies lies in the primary assumption upon which Wald's theory is predicated; i.e., that the original process Y is strictly "white" noise, independently-distributed random data. White noise is well known to be a signal that is uncorrelated. Such white noise can, for example, include Gaussian noise. It is, however, very rare to find physical process data associated with operating machinery or other industrial processes that are not contaminated with serially-correlated, deterministic noise components. Serially correlated noise components are conventionally known to be signal data whose successive time point values are dependent on one another. Noise components include, for example, auto-correlated (also known as serially correlated) noise and Markov dependent noise. Auto-correlated noise is a known form of noise wherein pairs of correlation coefficients describe the time series correlation of various data signal values along the time series of data. That is, the data U_1, U_2, \dots, U_n have correlation coefficients $(U_1, U_2), (U_2, U_3), \dots, (U_{n-1}, U_n)$ and likewise have correlation coefficients $(U_1, U_3), (U_2, U_4)$ etc. If these data are auto-correlated, at least some of the coefficients are non-zero. Markov dependent noise on the other hand is a very special form of correlation between past and future data signals. Rather, given the value of U_k , the values of $U_n, n>k$, do not depend on the values of U_j where $j<k$. This implies the correlation pairs (U_j, U_n) given the value U_k , are all zero. If, however, the present value is imprecise, then the correlation coefficients may be nonzero. This invention can overcome this limitation to conventional surveillance strategies by integrating the Wald sequential-test approach with a new dual transformation technique. This symbiotic combination of frequency-domain transformations and time-domain transformations produces a tractable solution to a particularly difficult problem that has plagued signal-processing specialists for many years.

In the preferred embodiment of the method shown in detail in FIG. 8, serially-correlated data signals from an industrial process can be rendered amenable to the SPRT testing methodology described hereinbefore. This is preferably done by performing a frequency-domain transformation of the original difference function Y . A particularly preferred method of such a frequency transformation is accomplished by generating a Fourier series using a set of highest "1" number of modes. Other procedures for rendering the data amenable to SPRT methods includes, for example, auto regressive techniques, which can accomplish substantially similar results, described herein for Fourier analysis. In the preferred approach of Fourier analysis to determine the "1" highest modes (see FIG. 8A):

$$Y_1 = \frac{\alpha_0}{2} + \sum_{m=1}^{\frac{N}{2}} (a_m \cos \omega_m t + b_m \sin \omega_m t) \quad (12)$$

where $a_0/2$ is the mean value of the series, a_m and b_m are the Fourier coefficients corresponding to the Fourier frequency ω_m , and N is the total number of observations. Using the Fourier coefficients, we next generate a composite function. X_t , using the values of the largest harmonics identified in the Fourier transformation of Y_t . The following numerical approximation to the Fourier transform is useful in determining the Fourier coefficients a_m and b_m . Let x_j be the

United States Patent	5,410,492
Gross , et al.	April 25, 1995

value of X_t at the j^{th} time increment. Then assuming 2π periodicity and letting $\omega_m = 2\pi m/N$, the approximation to the Fourier transform yields:

$$a_m = \frac{2}{N} \sum_{j=0}^{N-1} x_j \cos \omega_m j \quad b_m = \frac{2}{N} \sum_{j=0}^{N-1} x_j \sin \omega_m j \quad (13)$$

for $0 < m < N/2$. Furthermore, the power spectral density (PSD) function for the signal is given by 1_m where

$$I_m = N \frac{a_m^2 + b_m^2}{2} \quad (14)$$

To keep the signal bandwidth as narrow as possible without distorting the PSD, no spectral windows or smoothing are used in our implementation of the frequency-domain transformation. In analyzing the data from a pumping system of the EBR-II reactor of Argonne National Laboratory (West), the Fourier modes corresponding to the eight highest 1_m provide the amplitudes and frequencies contained in X_t . In our investigations for the particular pumping system data accumulated, the highest eight 1_m modes were found to give an accurate reconstruction of X_t while reducing most of the serial correlation for the physical variables studied. In other industrial processes, the analysis could result in more or fewer modes being needed to accurately construct the functional behavior of a composite curve. Therefore, the number of modes used is a variable that is iterated to minimize the degree of nonwhite noise for any given application. As noted in FIG. 8A a variety of noise tests are applied in order to remove serially correlated noise.

The reconstruction of X_t uses the general form of Eqn. (12), where the coefficients and frequencies employed are those associated with the eight highest PSD values. This yields a Fourier composite curve (see end of flowchart in FIG. 8A) with essentially the same correlation structure and the same mean as Y_t . Finally, we generate a discrete residual function R_t by differencing corresponding values of Y_t and X_t . This residual function, which is substantially devoid of serially correlated contamination, is then processed with the SPRT technique described hereinbefore.

FIGS. 9 and 10 schematically illustrate general implementation of the methods of processing EBR-II data; in particular FIGS. 9 and 10 are implemented on a DOS-type system that does not allow concurrent processing. FIG. 11 illustrates a detailed implementation on a UNIX type system that allows multiple concurrent processing.

In a specific example application of the above referenced methodology, certain data variables were monitored from the Argonne National Laboratory (West) reactor EBR-II. In particular, EBR-II reactor coolant pumps (RCPs) and delayed neutron (DN) monitoring systems were tested continuously to demonstrate the power and utility of the invention. The RCP and DN systems were chosen for initial application of the approach because SPRT-based techniques have already been under development for both the systems. All data used in this investigation were recorded during full-power, steady state operation at EBR-II. The data have been digitized at a two-per-second sampling rate using 2^{14} (16,384) observations for each signal of interest. FIGS. 1-3 illustrate data associated with the preferred spectral filtering approach as applied to the EBR-II primary pump power signal, which measures the power (in kW) needed to operate

United States Patent	5,410,492
Gross , et al.	April 25, 1995

the pump. The basic procedure of FIG. 8 was then followed in the analysis. FIG. 1 shows 136 minutes of the original signal as it was digitized at the 2-Hz sampling rate. FIG. 2 shows a Fourier composite constructed from the eight most prominent harmonics identified in the original signal. The residual function, obtained by subtracting the Fourier composite curve from the raw data, is shown in FIG. 3. Periodograms of the raw signal and the residual function have been computed and are plotted in FIG. 4. Note the presence of eight depressions in the periodogram of the residual function in FIG. 4B, corresponding to the most prominent periodicities in the original, unfiltered data. Histograms computed from the raw signal and the residual function are plotted in FIG. 5. For each histogram shown we have superimposed a Gaussian curve (solid line) computed from a purely Gaussian distribution having the same mean and variance. Comparison of FIG. 5A and 5B provide a clear demonstration of the effectiveness of the spectral filtering in reducing asymmetry in the histogram. Quantitatively, this decreased asymmetry is reflected in a decrease in the skewness (or third moment of the noise) from 0.15 (raw signal) to 0.10 (residual function).

It should be noted here that selective spectral filtering, which we have designed to reduce the consequences of serial correlation in our sequential testing scheme, does not require that the degree of non-normality in the data will also be reduced. For many of the signals we have investigated at EBR-II, the reduction in serial correlation is, however, accompanied by a reduction in the absolute value of the skewness for the residual function.

To quantitatively evaluate the improvement in whiteness effected by the spectral filtering method, we employ the conventional Fisher Kappa white noise test. For each time series we compute the Fisher Kappa statistic from the defining equation

$$\kappa = \left[\frac{1}{N} \sum_{k=1}^N I(\omega_k) \right]^{-1} I(L) \quad (15)$$

where $I(\omega_k)$ is the PSD function (see Eq. 14) at discrete frequencies ω_k , and $I(L)$ signifies the largest PSD ordinate identified in the stationary time series.

The Kappa statistic is the ratio of the largest PSD ordinate for the signal to the average ordinate for a PSD computed from a signal contaminated with pure white noise. For EBR-II, the power signal for the pump used in the present example has a κ of 1940 and 68.7 for the raw signal and the residual function, respectively. Thus, we can say that the spectral filtering procedure has reduced the degree of non-whiteness in the signal by a factor of 28. Strictly speaking, the residual function is still not a pure white noise process. The 95% critical value for Kappa for a time series with 2^{14} observations is 12.6. This means that only for computed Kappa statistics lower than 12.6 could we accept the null hypothesis that the signal is contaminated by pure white noise. The fact that our residual function is not purely white is reasonable on a physical basis because the complex interplay of mechanisms that influence the stochastic components of a physical process would not be expected to have a purely white correlation structure. The important point, however, is that the reduction in non-whiteness effected by the spectral filtering procedure using only the highest eight harmonics in the raw signal has been found to preserve the pre-specified false alarm and missed alarm probabilities in the SPRT sequential testing procedure (see below). Table I summarizes the computed Fisher Kappa statistics for thirteen EBR-II plant signals, which are used in the subject surveillance systems. In every case the table shows a substantial improvement in signal whiteness.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

The complete SPRT technique integrates the spectral decomposition and filtering process steps described hereinbefore with the known SPRT binary hypothesis procedure. The process can be illustratively demonstrated by application of the SPRT technique to two redundant delayed neutron detectors (designated DND A and DND B) whose signals were archived during long-term normal (i.e., undegraded) operation with a steady DN source in EBR-II. For demonstration purposes a SPRT was designed with a false alarm rate, α , of 0.01. Although this value is higher than we would designate for a production surveillance system, it gives a reasonable frequency of false alarms so that asymptotic values of α can be obtained with only tens of thousands of discrete observations. According to the theory of the SPRT technique, it can be easily proved that for pure white noise (such as Gaussian), independently distributed processes, α provides an upper bound to the probability (per observation interval) of obtaining a false alarm--i.e., obtaining a "data disturbance" annunciation when, in fact, the signals under surveillance are undegraded.

FIGS. 6 and 7 illustrate sequences of SPRT results for raw DND signals and for spectrally-whitened DND signals, respectively. In FIGS. 6A and 6B, and 7A and 7B, respectively, are shown the DN signals from detectors DND-A and DND-B. The steady-state values of the signals have been normalized to zero.

TABLE 1 – Effectiveness of Spectral Filtering for Measured Plant Signals

Plant Variable I.D.	Fisher Kappa Test Statistic (N = 16,384)	
	Raw Signal	Residual Function
Pump 1 Power	1940	68.7
Pump 2 Power	366	52.2
Pump 1 Speed	181	25.6
Pump 2 Speed	299	30.9
Pump 1 Radial Vibr (top)	123	67.7
Pump 2 Radial Vibr (top)	155	65.4
Pump 1 Radial Vibr (bottom)	1520	290.0
Pump 2 Radial Vibr (bottom)	1694	80.1
DN Monitor A	96	39.4
DN Monitor B	81	44.9
DN Detector 1	86	36.0
DN Detector 2	149	44.1
DN Detector 3	13	8.2

Normalization to adjust for differences in calibration factor or viewing geometry for redundant sensors does not affect the operability of the SPRT. FIGS. 6C and 7C show point-wise differences of signals DND-A and DND-B. It is this difference function that is input to the SPRT technique. Output from the SPRT method is shown for a 250-second segment in FIGS. 6D and 7D.

Interpretation of the SPRT output in FIGS. 6D and 7D is as follows: When the SPRT index reaches a lower threshold, A , one can conclude with a 99% confidence factor that there is no

United States Patent	5,410,492
Gross , et al.	April 25, 1995

degradation in the sensors. For this demonstration, A is equal to 4.60, which corresponds to false-alarm and missed-alarm probabilities of 0.01. As FIGS. 6D and 7D illustrate, each time the SPRT output data reaches A, it is reset to zero and the surveillance continues.

If the SPRT index drifts in the positive direction and exceeds a positive threshold, B, of +4.60, then it can be concluded with a 99% confidence factor that there is degradation in at least one of the sensors. Any triggers of the positive threshold are signified with diamond symbols in FIGS. 6D and 7D. In this case, since we can certify that the detectors were functioning properly during the time period our signals were being archived, any triggers of the positive threshold are false alarms.

If we extend sufficiently the surveillance experiment illustrated in FIG. 6D, we can get an asymptotic estimate of the false alarm probability α . We have performed this exercise using 1000-observation windows, tracking the frequency of false alarm trips in each window, then repeating the procedure for a total of sixteen independent windows to get an estimate of the variance on this procedure for evaluating α . The resulting false-alarm frequency for the raw, unfiltered, signals is $\alpha = 0.07330$ with a variance of 0.000075. The very small variance shows that there would be only a negligible improvement in our estimate by extending the experiment to longer data streams. This value of α is significantly higher than the design value of $\alpha = 0.01$, and illustrates the danger of blindly applying a SPRT test technique to signals that may be contaminated by excessive serial correlation.

The data output shown in FIG. 7D employs the complete SPRT technique shown schematically in FIG. 8. When we repeat the foregoing exercise using 16 independent 1000-observation windows, we obtain an asymptotic cumulative false-alarm frequency of 0.009142 with a variance of 0.000036. This is less than (i.e., more conservative than) the design value of $\alpha = 0.01$, as desired.

It will be recalled from the description hereinbefore regarding one preferred embodiment, that we have used the eight most prominent harmonics in the spectral filtration stage of the SPRT technique. By repeating the foregoing empirical procedure for evaluating the asymptotic values of α , we have found that eight modes are sufficient for the input variables shown in Table I. Furthermore, by simulating subtle degradation in individual signals, we have found that the presence of serial correlation in raw signals gives rise to excessive missed-alarm probabilities as well. In this case spectral whitening is equally effective in ensuring that pre-specified missed-alarm probabilities are not exceeded using the SPRT technique.

In another different form of the invention, it is not necessary to have two sensor signals to form a difference function. One sensor can provide a real signal characteristic of an ongoing process and a record artificial signal can be generated to allow formation of a difference function. Techniques such as an auto regressive moving average (ARMA) methodology can be used to provide the appropriate signal, such as a DC level signal, a cyclic signal or other predictable signal. Such an ARMA method is a well-known procedure for generating artificial signal values, and this method can even be used to learn the particular cyclic nature of a process being monitored enabling construction of the artificial signal.

The two signals, one a real sensor signal and the other an artificial signal, can thus be used in the same manner as described hereinbefore for two real sensor signals. The difference function

United States Patent	5,410,492
Gross , et al.	April 25, 1995

Y is then formed, transformations performed and a residual function is determined which is free of serially correlated noise.

Fourier techniques are very effective in achieving a whitened signal for analysis, but there are other means to achieve substantially the same results using a different analytical methodology. For example, filtration of serial correlation can be accomplished by using the autoregressive moving average (ARMA) method. This ARMA technique estimates the specific correlation structure existing between sensor points of an industrial process and utilizes this correlation estimate to effectively filter the data sample being evaluated.

A technique has therefore been devised to produce reliable databases free from false alarm information, which integrates frequency-domain filtering with sequential testing methodology. This method provides a solution to a problem that is endemic to industrial signal evaluation surveillance. For example, it is not uncommon for sensors to become degraded during service, and faulty signals will be generated. Such faulty signals can give rise to false alarms and result in accumulation of a faulted database intended for use in the future as a reliable source of information. The subject invention particularly allows identification of valid or corrupted data and consequent use of reliable database information in analyzing industrial systems. For example, this invention would be particularly useful in meteorological systems, aeronautical design systems, automotive simulation systems or any system wherein experimental data can be used for modeling or design projects. Further, one can evaluate ongoing slow degradation that evolves over a long time period (gradual de-calibration bias in a sensor, appearance of a new radiation source in the presence of a noisy background signal, wear out or buildup of a radial rub in rotating machinery, etc.). The system thus can alert a researcher or operator of the incipience or onset of a disturbance long before it would be apparent to visual inspection of strip chart or CRT signal traces, and well before conventional threshold limit checks would be tripped. This permits the researcher to either have a reliable database for research use or to actively use the database to terminate, modify or avoid events that might otherwise lie outside technical specification guidelines or availability goals. Thus, in many cases a user of such a database can anticipate or schedule corrective actions (sensor replacement or recalibration; component adjustment, alignment, or rebalancing; etc.) to be performed during a scheduled system outage.

Another important feature of the technique that distinguishes it from conventional methods is the built-in quantitative false-alarm and missed-alarm probabilities. This is quite important in the context of high-risk industrial processes and applications. The invention makes it possible to apply formal reliability analysis methods to an overall system comprising a network of interacting SPRT modules that are simultaneously monitoring a variety of plan variables. This amenability to formal reliability analysis methodology will, for example, greatly enhance the process of granting approval for nuclear-plant applications of the invention, a system that can potentially save a utility millions of dollars per year per reactor.

While preferred embodiments of the invention have been shown and described, it will be clear to those skilled in the art that various changes and modifications can be made without departing from the invention in its broader aspects as set forth in the claims provided hereinafter.

Claims

What is claimed is:

United States Patent	5,410,492
Gross , et al.	April 25, 1995

1. A method of processing at least one database obtained from at least one of an industrial process and at least a first and second sensor for determining fault conditions of data characteristic of a process, comprising the steps of:

collecting data from at least a first and second sensor to redundantly accumulate data from at least one physical variable of the industrial process to provide a first data signal from said first sensor and a second data signal from said second sensor, each said data signal being characteristic of the one physical variable;

obtaining a difference function of the data characteristic of the arithmetic difference pair wise between said first data signal and said second data signal at each of a plurality of different times of sensing the one physical variable;

obtaining a frequency domain transformation of said first difference function to procure Fourier coefficients corresponding to Fourier frequencies;

generating a composite function over a time domain using the Fourier coefficients;

obtaining a residual function over time by determining the arithmetic difference between the difference function and the composite function, the residual function having reduced serially correlated noise;

operating on the residual function using computer means for performing a statistical analysis technique to determine whether alarm condition data is present in at least one of the data base from the industrial process and from the at least first and second sensor, the residual function including white noise characteristics of an uncorrelated function of reduced skewness relative to the difference function and input to the statistical analysis technique; and

producing said at least one database having data with identified alarm conditions associated therewith, allowing separation of said data with identified alarm conditions.

2. The method described is claim 1 wherein said computer means comprises an artificial intelligence system.
3. The method as defined in claim 1 wherein the residual function further comprises reduced Markov dependent noise.
4. The method as defined in claim 1 wherein the industrial process database comprises at least one of a chemical process database, a mechanical process database and an electrical process database.
5. The method as defined in claim 1 wherein the step of obtaining Fourier coefficients comprise iteratively determining the minimum number of Fourier harmonics able to generate the composite function.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

6. The method as defined in claim 1 further including at least one of the steps of identifying or removing data having alarm conditions associated therewith.
7. A method of accumulating a database with validated information from at least one of an industrial process and a sensor for determining fault conditions therein, comprising the steps of:
 - collecting data from at least one sensor to detect at least one physical variable of the industrial process to provide a real signal from said one sensor;
 - generating an artificial data signal characteristic of the one physical variable;
 - obtaining a difference function characteristic of the difference pair wise between said real signal and said artificial signal at each of a plurality of different times of sensing the one physical variable;
 - obtaining a frequency domain transformation of said difference function;
 - generating a composite function data base over a time domain;
 - obtaining a residual function over time by determining the difference between the difference function and the composite function database, the residual function including white noise characteristics of an uncorrelated signal of reduced skewness compared to the difference function; and
 - operating on the residual function using a computer means for performing a statistical analysis technique to determine whether alarm condition data is present in at least one of the industrial process and the at least one sensor, the residual function having said white noise characteristics input to the statistical analysis technique for providing the database with validated information.
8. The method as defined in claim 7 wherein the step of obtaining a frequency domain transformation comprises performing a Fourier transformation.
9. The method as deemed in claim 7 wherein the steps of obtaining a composite function over time comprises performing an autoregressive moving average analysis.
10. The method as deemed in claim 7 further including the step of determining a difference function for the artificial signal and the real signal, as well as a separate pair of real signals.
11. The method as defined in claim 7 wherein the residual function further comprises reduced Markov dependent noise.
12. The method as defined in claim 8 wherein the step of obtaining a frequency domain transformation comprises obtaining Fourier coefficients iteratively to determine the minimum number of Fourier harmonics able to generate the composite function database.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

13. A system for accumulating a database with validated signals from data obtained from at least one of an industrial process and a sensor for determining a fault condition therein, comprising:

first means for providing a first set of data characteristic of at least one physical variable of the industrial process;

second means for providing a second set of data calculatngly determined and related to the one physical variable of the industrial process;

third means for determining a difference function of data characteristic of the arithmetic difference pair wise between said first set of data and said second set of data at each of a plurality of different times of the one physical variable being sensed;

fourth means for obtaining a residual function of data over time by means for determining the arithmetic difference between the difference function data and the composite function data, the residual function data having reduced serially correlated noise; and

fifth means for operating on the residual function data including a computer means for performing a statistical analysis technique and for determining whether corrupted data is present in at least one of the data obtained from the industrial process and the sensor and with said third means, said fourth means, and said fifth means cooperatively providing the residual function data with white noise characteristics of an uncorrelated signal of reduced skewness relative to the difference function as an input to the statistical analysis technique to provide the database with validated signals.

14. The system as defined in claim 13 further including means for obtaining a frequency domain transformation of said difference function data.

15. The system as defined in claim 13 wherein said computer means comprises an artificial intelligence system.

16. The system as defined in claim 13 wherein said means for calculatngly generating a second set of data comprises computer means for executing a computer program.

17. The system as defined in claim 16 wherein the computer program includes an autoregressive moving average procedure.

18. The system as defined in claim 13 wherein the system includes at least one pair of sensor data sets.

19. The system as defined in claim 13 wherein said computer means executes a computer program including a statistical probability ratio test on the residual function data.

20. The system as defined in claim 13 further including means for at least one of identifying or removing alarm data from the industrial process data and the sensor data.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

21. A method of processing at least one database obtained from at least one of an industrial process and at least a first and second sensor for determining fault conditions of data characteristic of a process, comprising the steps of:

collecting data from the at least first and second sensor to accumulate comparative data for at least one physical variable of the industrial process to provide a first data signal from said first sensor and a second data signal from said second sensor, each said data signal being characteristic of the one physical variable;

obtaining a difference function of the data characteristic of the arithmetic difference pair wise between said first data signal and said second data signal at each of a plurality of different times of sensing the one physical variable;

obtaining a frequency domain transformation of said first difference function to procure Fourier coefficients corresponding to Fourier frequencies;

generating a composite function over a time domain using the Fourier coefficients;

obtaining a residual function over time by determining the arithmetic difference between the difference function and the composite function, the residual function having reduced serially correlated noise;

operating on the residual function using computer means for performing a statistical analysis technique to determine whether alarm condition data is present in the at least one database from the industrial process and the first and second sensor, the residual function including white noise characteristics of an uncorrelated function of reduced skewness relative to the difference function and input to the statistical analysis technique; and

producing said at least one database having data with identified alarm conditions associated therewith, allowing separation of said data with identified alarm conditions.

22. The method as defined in claim 21 wherein said first and second sensors measure wear out conditions in machinery.

23. A system for accumulating a database with validated signals from data obtained from at least one of an industrial process and a sensor for determining a fault condition therein, comprising:

first means for providing a first set of data characteristic of at least one physical variable of the industrial process;

second means for providing a second set of data related to the one physical variable of the industrial process;

third means for determining a difference function of data characteristic of the arithmetic difference pair wise between said first set of data and said second set of

United States Patent	5,410,492
Gross , et al.	April 25, 1995

data at each of a plurality of different times of the one physical variable being sensed;

means for generating composite function data over a time domain;

fourth means for obtaining a residual function of data over time by means for determining the arithmetic difference between the difference function data and the composite function data, the residual function data having reduced serially correlated noise; and

fifth means for operating on the residual function data including a computer means for performing a statistical analysis technique and for determining whether degraded data is present in at least one of the data from the industrial process and from the sensor and with said third means, said fourth means, and said fifth means cooperatively providing the residual function data with white noise characteristics of an uncorrelated signal of reduced skewness relative to the difference function as an input to the statistical analysis technique to provide the data base with validated signals.

United States Patent	5,410,492
Gross , et al.	April 25, 1995

Fig. 1

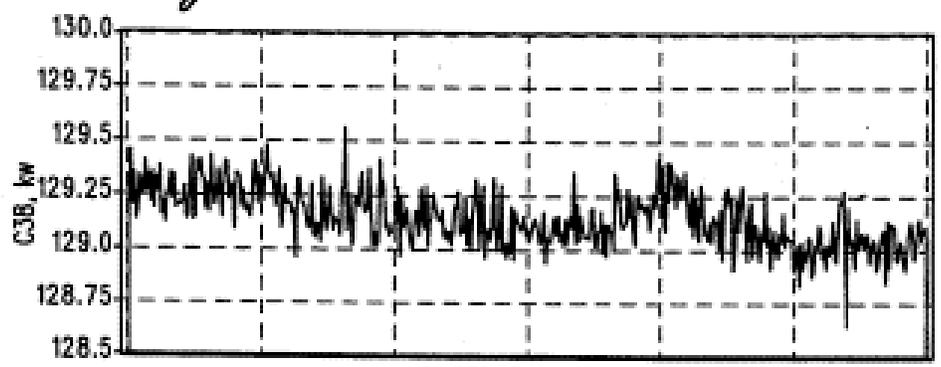


Fig. 2

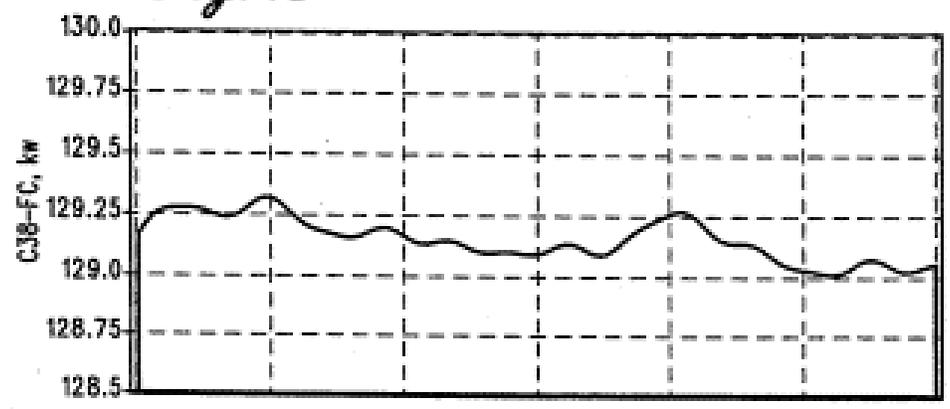
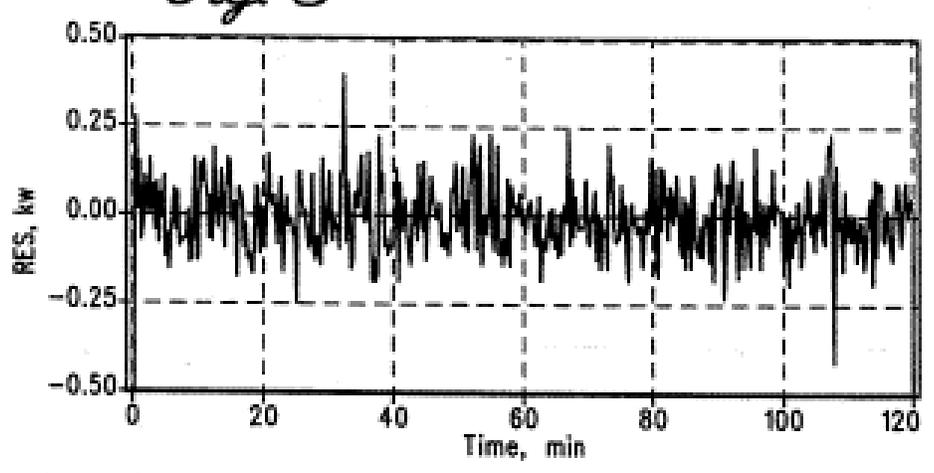
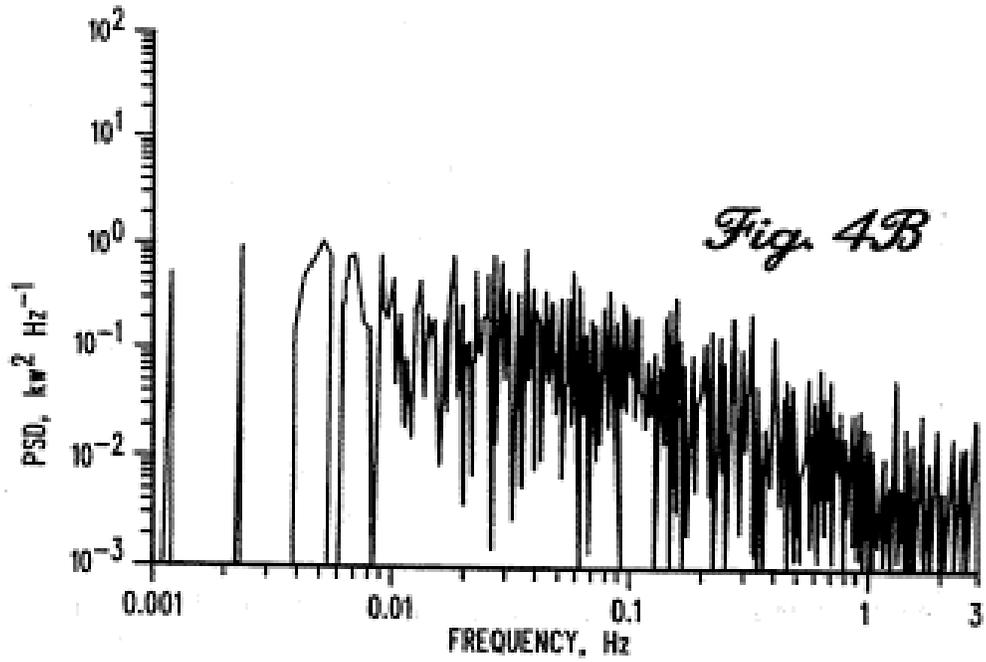
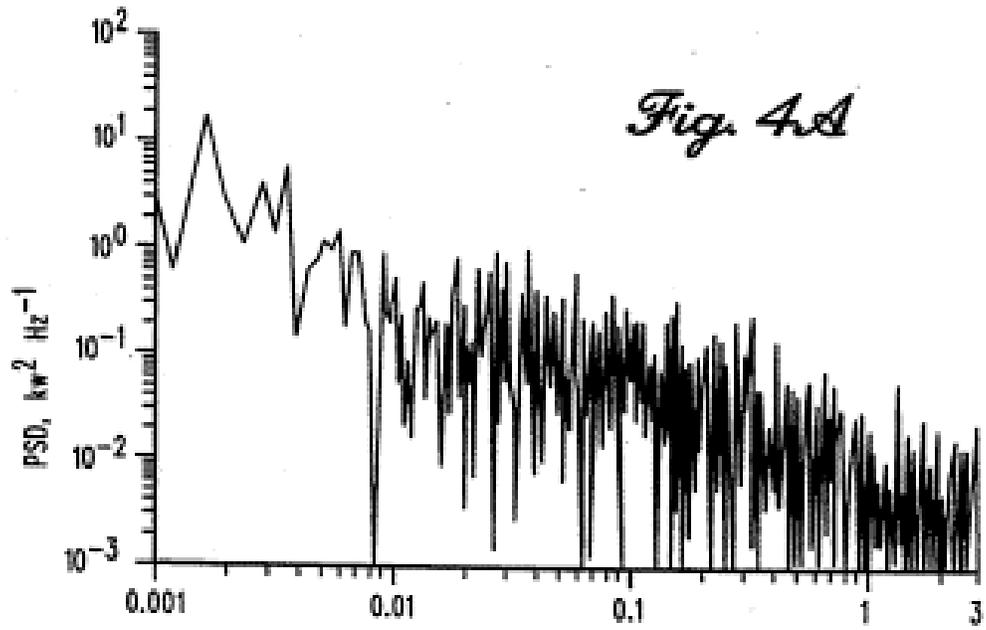


Fig. 3



United States Patent	5,410,492
Gross , et al.	April 25, 1995



United States Patent	5,410,492
Gross, et al.	April 25, 1995

Fig. 5A

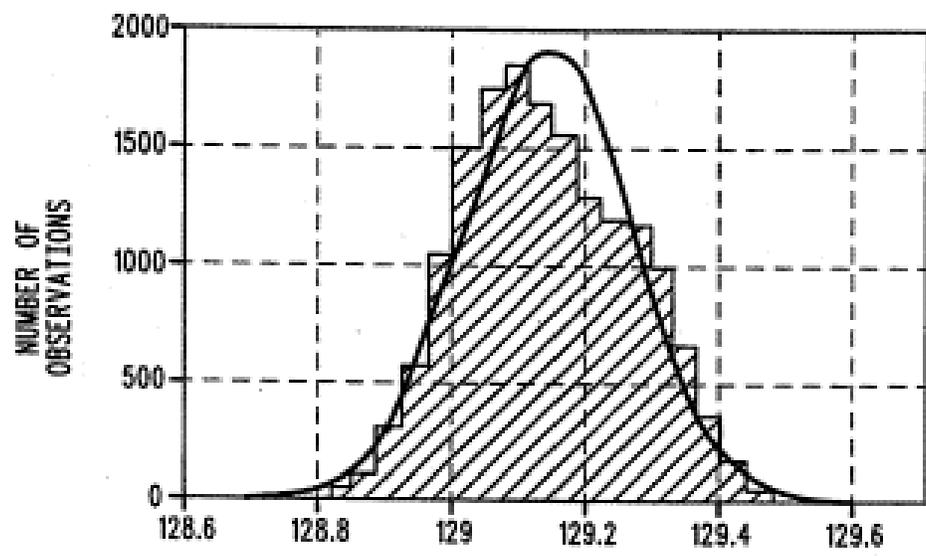
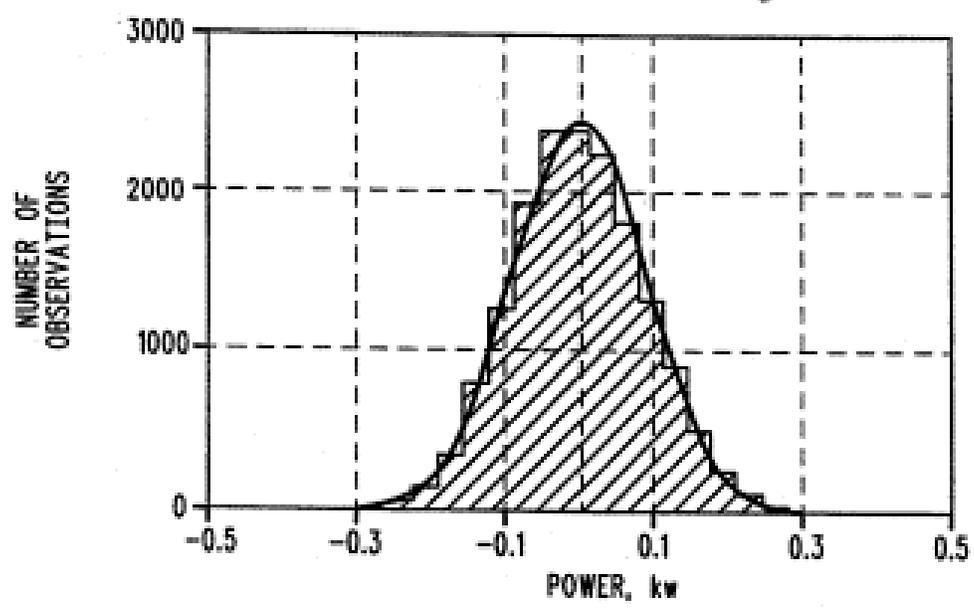
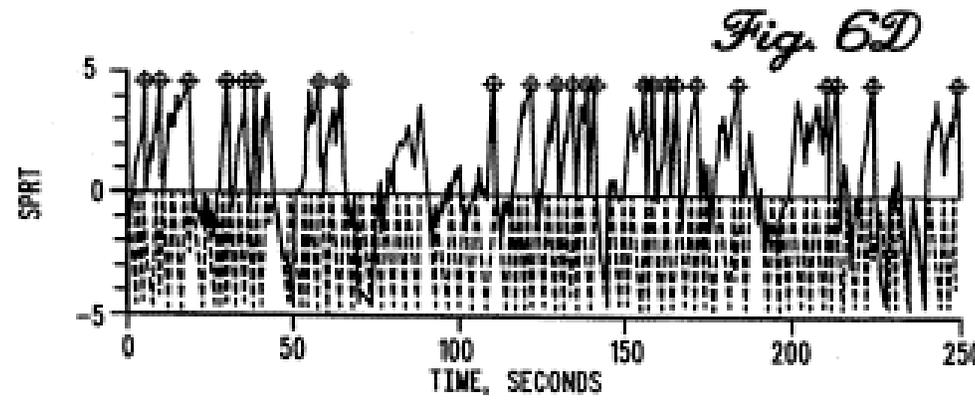
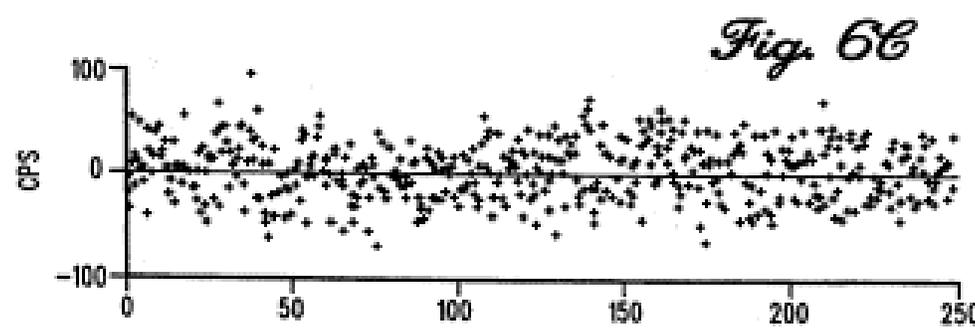
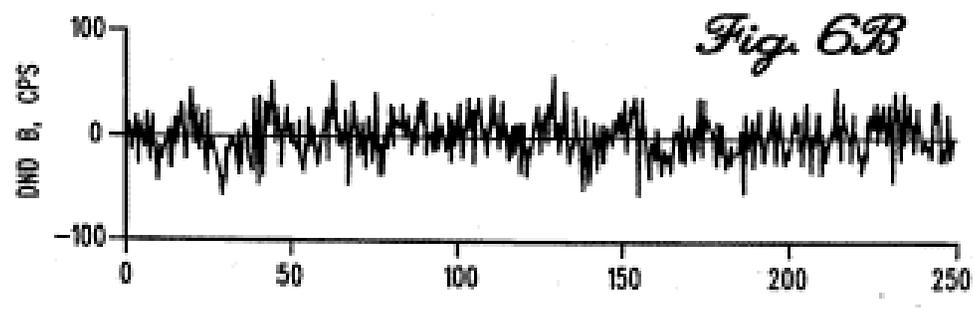
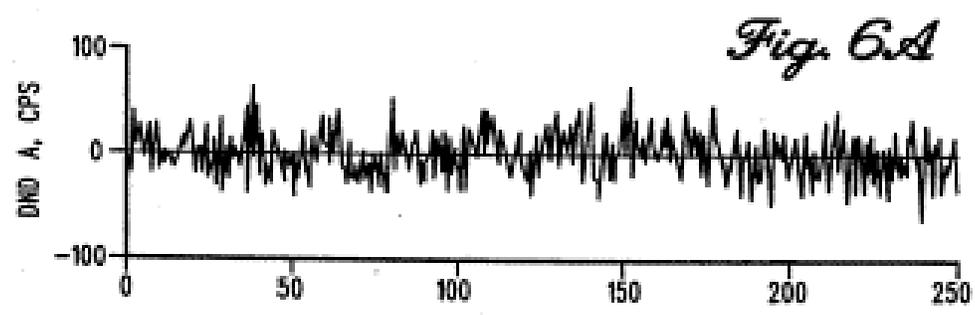


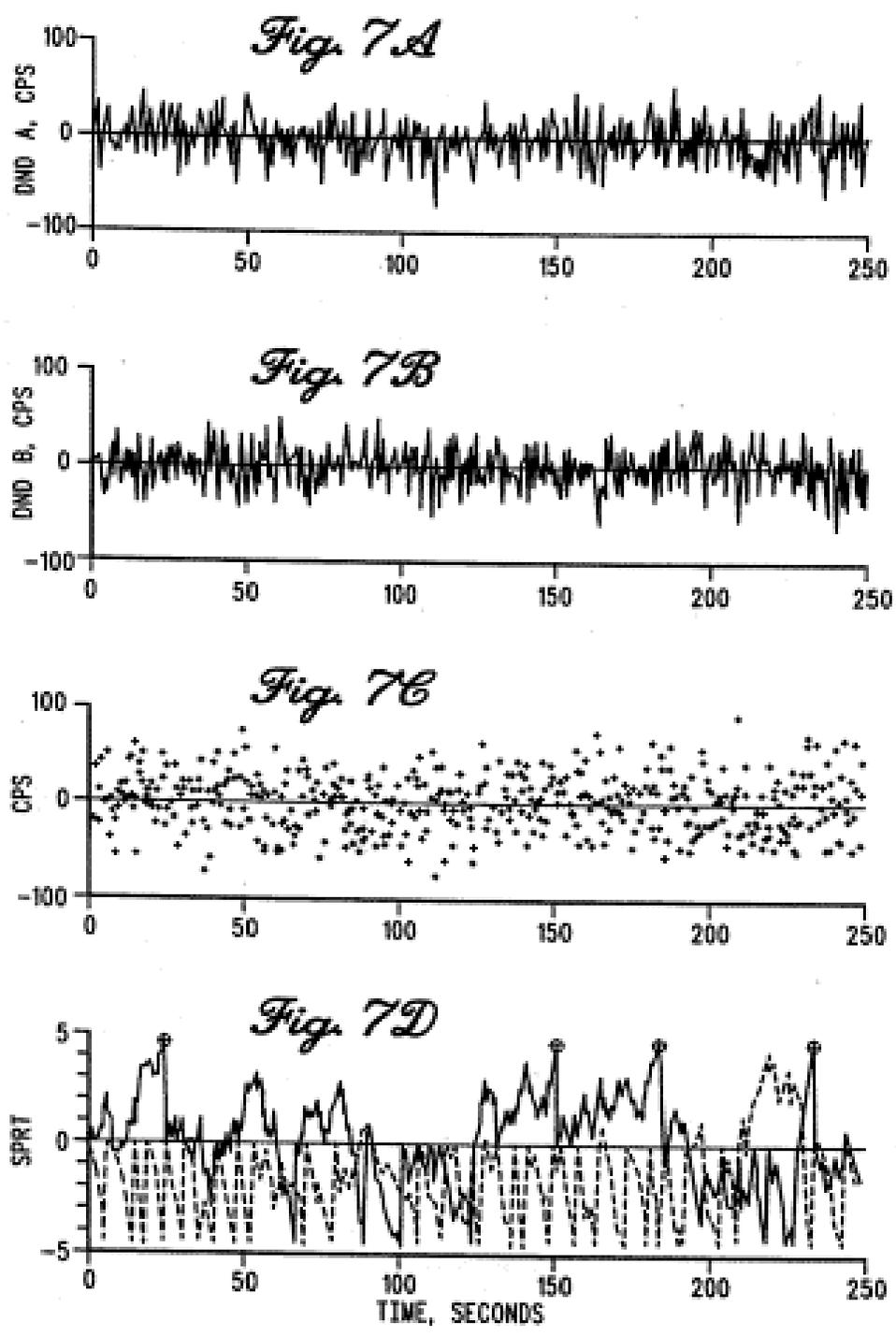
Fig. 5B



United States Patent	5,410,492
Gross, et al.	April 25, 1995

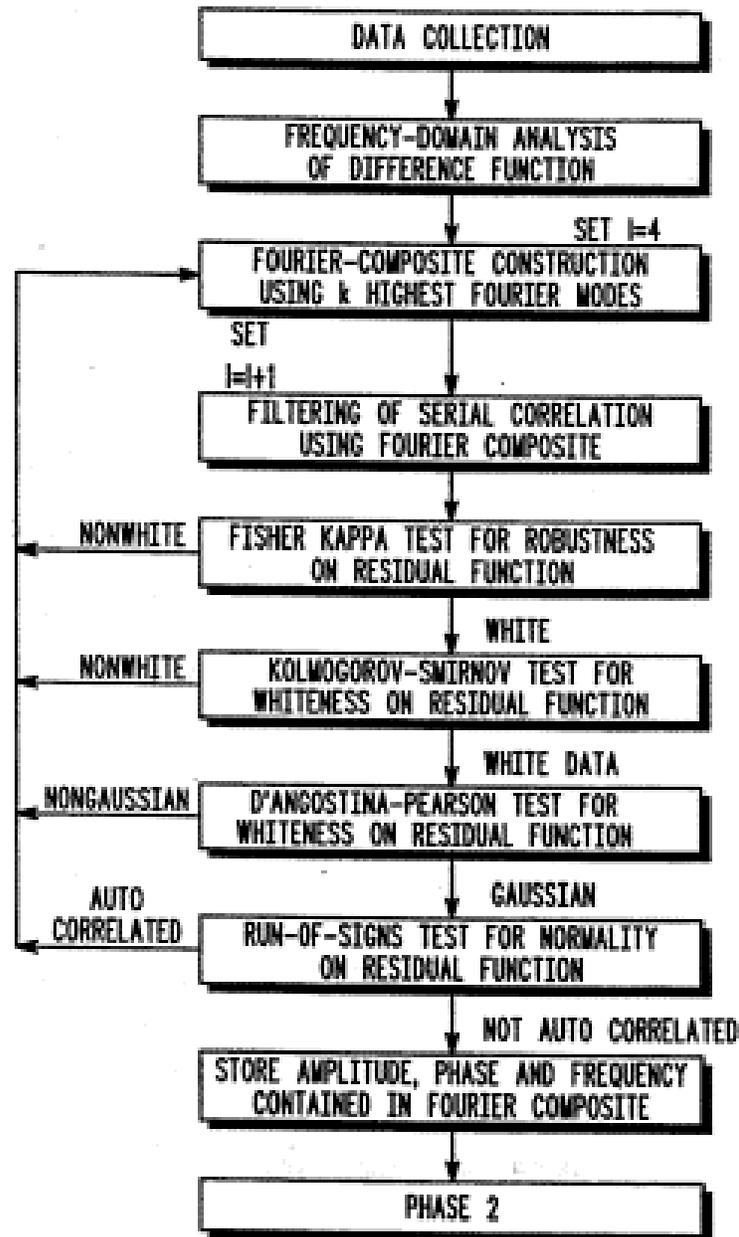


United States Patent	5,410,492
Gross, et al.	April 25, 1995



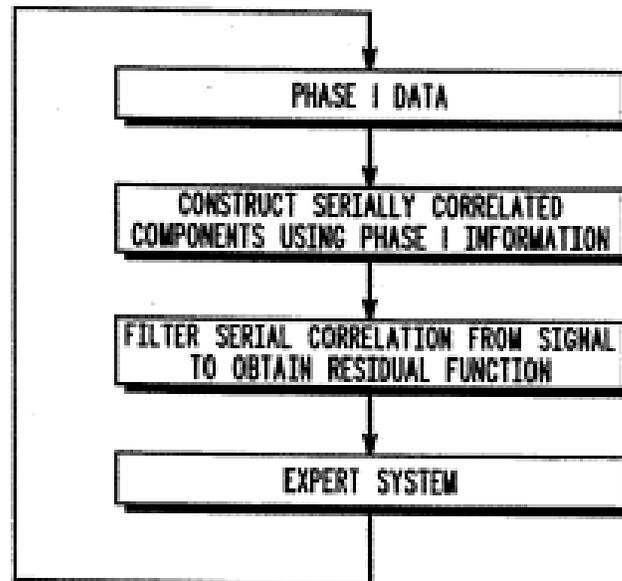
United States Patent	5,410,492
Gross , et al.	April 25, 1995

Fig. 8A



United States Patent	5,410,492
Gross , et al.	April 25, 1995

Fig. 8B



United States Patent	5,410,492
Gross , et al.	April 25, 1995

Fig. 9

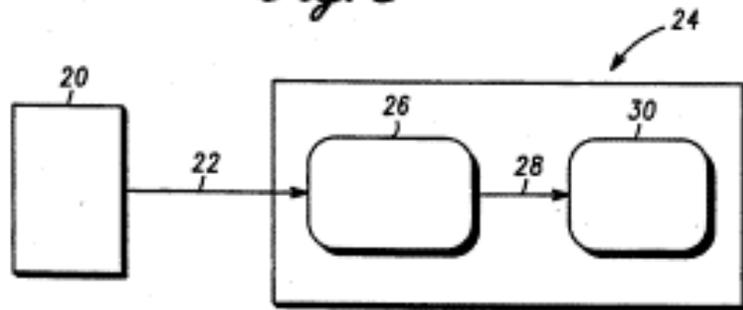


Fig. 10

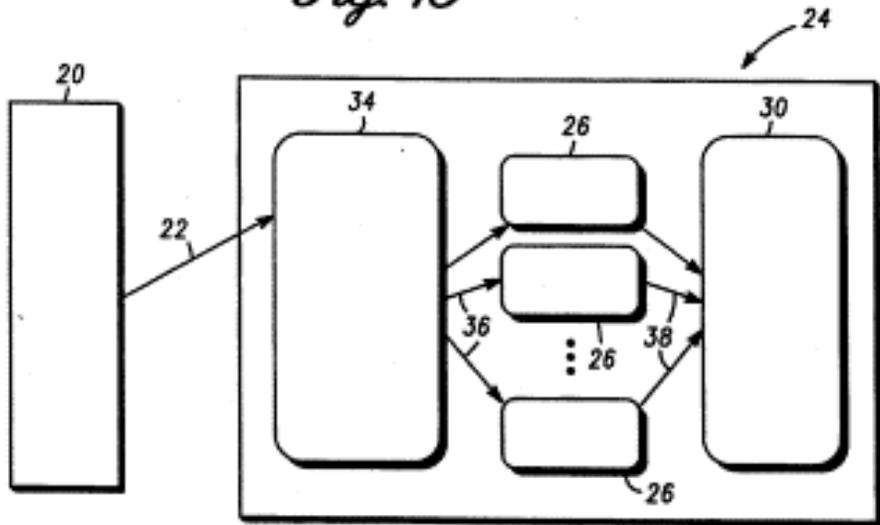


Fig. 11

