Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2053064 C 2002/12/10

(11)(21) 2 053 064

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 1991/10/09

(41) Mise à la disp. pub./Open to Public Insp.: 1992/04/16

(45) Date de délivrance/Issue Date: 2002/12/10

(30) Priorité/Priority: 1990/10/15 (07/597,438) US

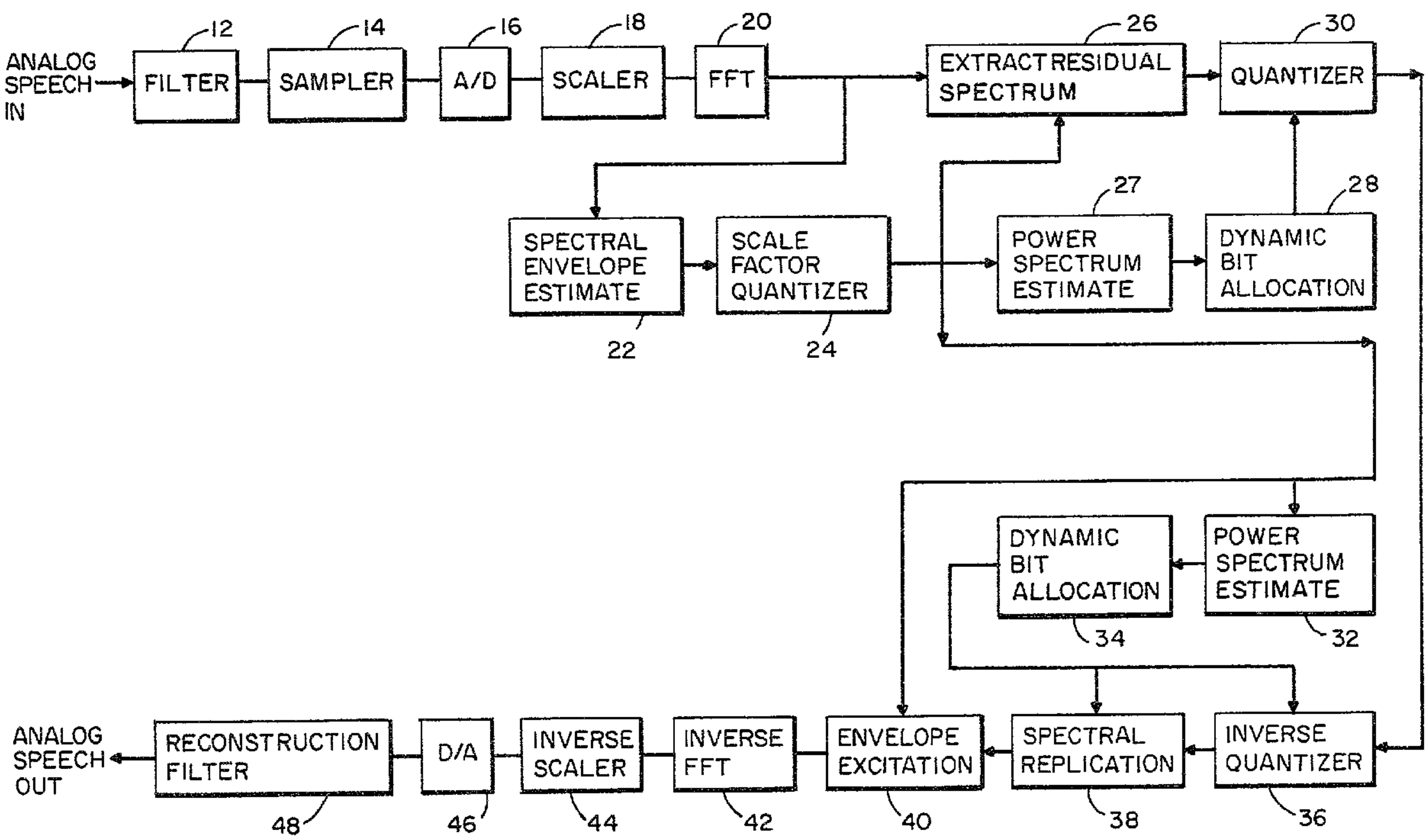
(51) Cl.Int.^{5}/Int.Cl.^{5} G10L 7/06

(72) Inventeur/Inventor:
MAZOR, BARUCH, US

(73) Propriétaire/Owner:
VERIZON LABORATORIES INC., US

(74) Agent: R. WILLIAM WRAY & ASSOCIATES

(54) Titre : METHODE ET APPAREIL DE CODAGE A AFFECTATION DYNAMIQUE DES BITS
(54) Title: DYNAMIC BIT ALLOCATION SUBBAND EXCITED TRANSFORM CODING METHOD AND APPARATUS



(57) Abrégé/Abstract:
In an adaptive subband excited transform speech encoding system, a range of quantizers are available and are dynamically selected for each window of speech. The quantizers designated for individual subbands are determined to minimize mean squared error distortion in the recreated signal while using no more than a predetermined number of quantization bits per window of speech.

2053064

90-3-827

DYNAMIC BIT ALLOCATION SUBBAND EXCITED TRANSFORM CODING METHOD AND APPARATUS

ABSTRACT

In an adaptive subband excited transform speech encoding system, a range of quantizers are available and are dynamically selected for each window of speech. The quantizers designated for individual subbands are determined to minimize mean squared error distortion in the recreated signal while using no more than a predetermined number of quantization bits per window of speech.

DYNAMIC BIT ALLOCATION SUBBAND EXCITED TRANSFORM CODING METHOD AND APPARATUS

Conventional analog telephone systems are being replaced by digital systems. In digital systems, the analog signals are sampled at a rate of greater than or equal to about twice the bandwidth of the analog signals or about eight kilohertz, and the samples are then encoded. In a simple pulse code modulation system (PCM), each sample is quantized as one of a discrete set of prechosen values and encoded as a digital word which is then transmitted over the telephone lines. With eight bit digital words, for example, the analog sample is quantized to 2^8 or 256 levels, each of which is designated by a different eight bit word. Using nonlinear quantization, excellent quality speech can be obtained.

Efforts have been made to reduce the bit rates required to encode the speech and obtain a clear decoded speech signal at the receiving end of the system. The linear predictive coding (LPC) technique is based on the recognition that speech production involves excitation and a filtering process. The excitation is determined by the vocal cord vibration for voiced speech and by turbulence for unvoiced speech, and that actuating signal is then modified by the filtering process of vocal resonance chambers including the mouth and nasal passages. For a particular group of samples, a digital linear filter which simulates the formant effects of the resonance chambers can be defined and the definition can be encoded. A residual signal which approximates the excitation can then be obtained by passing the speech signal through and inverse formant filter, and the residual signal can be encoded. Because sufficient information is contained in the lower-frequency portion of the residual spectrum, it is possible to encode only the low frequency baseband and still obtain reasonably clear speech. At the receiver, a definition of the formant filter and the residual baseband are decoded. The baseband is repeated to complete the spectrum of the residual signal. By applying the decoded filter to the repeated baseband signal, an approximation to the initial speech can be reconstructed.

A major problem of the LPC approach is in defining the formant filter which must be redefined with each window of samples. A problem with such systems is that they do not always provide a satisfactory reconstruction of certain formants such as that resulting from nasal resonance. As a result, the quality of reconstruction from 16,000 bits per second is generally unsatisfactory.

Another speech coding scheme which exploits the concepts of excitation-filter separation and excitation baseband transmission is described by Zibman in U.S. patent number 4,914,701. In that approach, speech is encoded by first performing a Fourier transform of a window of speech. The Fourier transform coefficients are normalized by first defining a piecewise-constant approximation of the spectral envelope and then scaling the frequency coefficients relative to the approximation. The normalization is accomplished first for each formant region and then repeated for smaller subbands. Quantization and transmission of the spectral envelope approximations amount to transmission of a filter definition. Quantization and transmission of the scaled frequency coefficients associated with either the lower or upper half of the spectrum amounts to transmission of a "baseband" excitation signal. At the receiver, the full spectrum of the excitation signal is obtained by adding the transmitted baseband to a frequency translated version of itself. Frequency translation is performed easily by duplicating the scaled Fourier coefficients of the baseband into the corresponding higher or lower frequency positions. A signal can then be fully recreated by inverse scaling with the transmitted piecewise-constant approximations. This coding approach can be very simply implemented and provides good quality speech at 16 kilobits per second. However, it performs poorly with non-speech voice-band data transmission.

A modification of the Zibman coding technique is presented by Mazor, et al. in U.S. Patent 4,790,016. In that approach, the transform spectrum is divided into a plurality of subbands of coefficients. The approximate envelope is defined for each subband and each envelope definition is encoded for transmission. As in the Zibman approach, each spectrum coefficient is scaled relative to the defined envelope of the respective subband. In the Mazor, et al. improvement, the number of bits to which each coefficient is encoded is determined by the defined envelope of its subband. Specifically, the four subbands having the largest initial peak energy, and thus the largest envelope definition, are quantized to seven bits for each coefficient. The four subbands having the next smaller envelope definitions are quantized to six bits per coefficient, and the four next smaller subbands are quantized to four bits per coefficient. The coefficients of the remaining subbands are not transmitted; that is, they were quantized to zero bits per coefficient. At the receiver, the transmitted subbands are replicated to define coefficients of frequencies which are not transmitted.

The prior Mazor, et al. system resulted in much less than optimal bit allocation with certain types of signals. For example, a narrow band signal such as a tone might be best treated by allocating all available bits to only one or two subbands. In the prior Mazor, et al. system, bits were allocated to 12 subbands regardless of the signal, so the bits allocated to a number of the subbands would be wasted with the narrow band signal. On the other hand, with wide band signals like white noise all subbands have about the same signal level, and it would therefore be better to evenly distribute all available bit across the full spectrum. In the prior Mazor, et al. system the bits would always be allocated in the same way to 12 subbands.

The present invention is an improvement to the Mazor, et al. approach.

According to one aspect of the invention, there is provided a speech coding system comprising: transform means for performing a discrete transform of a window of speech to generate a discrete transform spectrum of coefficients; envelope defining and encoding means for defining an approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and for encoding the defined envelope of each subband of coefficients; means for scaling each spectrum coefficient relative to the defined envelope of the respective subband of coefficients; a plurality of quantizers of different bit lengths for encoding scaled spectrum coefficients within each subband; and means for determining a quantizer, if any, to be used to encode the coefficients of each subband within each window of speech wherein the selection of available quantizers used in successive windows is variable.

According to another aspect of the invention, there is provided a method of coding speech comprising: performing a discrete transform of a window of speech to generate a discrete spectrum of coefficients; defining an approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and digitally encoding the defined envelope of each subband of coefficients; and with at least one of a plurality of quantizers of different bit lengths, encoding each of scaled coefficients within subbands into a number of bits, the quantizer used for each subband being determined for each window of speech to minimize distortion.

More particularly, we provide a method wherein the quantizers are determined by computing the allocation of bits as a number of bits greater than or equal to zero which is derived from $\log_2 (P_i/D)$ for each subband where P_i is a power density estimate for the subband and D is a distortion error estimate for the window of speech. Also, more particularly, we provide a method wherein P_i is estimated for each subband from the defined envelope of the respective subband of coefficients.

Additionally, we provide a method wherein the allocation of bits is tuned toward the number of available bits by adding bits to higher energy subbands and subtracting bits from lower energy subbands.

Some embodiments of the invention will now be described by way of example with reference to the accompanying drawings in which:

FIG. 1 is a block diagram of a speech encoder and corresponding decoder of a coding system embodying the present invention.

FIG. 2 is an example of a magnitude spectrum of the Fourier transform of a window of speech illustrating principles of the present invention.

FIG. 3 is an example spectrum normalized from that of FIG. 2 based on principles of the present invention.

FIG. 4 schematically illustrates the iterative process for determining the distortion and the bit allocation.

FIG. 5 is a flow chart of the process for determining bit allocation for each window of speech.

In the accompanying drawings like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

A block diagram of the coding system is shown in FIG. 1. Prior to compression, the analog speech signal is low pass filtered in filter 12 at 3.4 kilohertz, sampled in sampler 14 at a rate of 8 kilohertz, and digitized using a 12 bit linear analog to digital converter 16. It will be recognized that the input to the encoder may already be in digital form and may require conversion to the code which can be accepted by the encoder. The digitized speech signal, in frames of N samples, is first scaled up in a scaler 18 to maximize the numerical resolution in the ensuing processing steps. The scaled input samples are then Fourier transformed in a fast Fourier transform device 20 to obtain a corresponding discrete spectrum represented by (N/2)+1 complex frequency coefficients.

In a specific implementation, the input frame size equals 180 samples and corresponds to a frame every 22.5 milliseconds. However, the discrete Fourier transform is performed on 192 samples, including 12 samples overlapped with the previous frame, preceded by trapezoidal windowing with a 12 point slope at each end. The resulting output of the FFT includes 97 complex frequency coefficients spaced 41.667 Hertz apart.

An example magnitude spectrum of a Fourier transform output from FFT 20 is illustrated in FIG. 2. Although illustrated as a continuous function, it is recognized that the transform circuit 20 actually provides only 97 incremental complex outputs.

Following the basic approach of Mazor, et al. presented in U.S. patent number 4,790,016, the magnitude spectrum of the Fourier transform output is equalized and encoded. To that end, the spectrum is partitioned into L contiguous subbands and a spectral envelope estimate is based on a piecewise-constant approximation of those subbands at 22. In a specific implementation the spectrum is divided into twenty subbands, each including four complex coefficients. Frequencies above 3291.67 Hertz are not encoded and are set to zero at the receiver. To equalize the spectrum, the spectral envelope of each subband is assumed constant and is defined by the peak magnitude in each subband as illustrated by the horizontal lines in FIG. 2. Each magnitude, or more correctly the inverse thereof, can be treated as a scale factor for its respective subband. Each scale factor is quantized in a quantizer 24 to q bits. For example, q may equal 4.

By then multiplying at 26 the magnitude of each coefficient of the spectrum by the scale factor associated with that coefficient, the flattened residual spectrum of FIG. 3 is obtained. This flattening of the spectrum is equivalent to inverse filtering the signal based on the piecewise-constant estimate of the spectral envelope.

Only selected subbands of the flattened spectrum of Figure 3 are quantized and transmitted, and the selection is based on the spectral envelope of the subbands. In accordance with the prior Mazor, et al. approach, the number of transmitted subbands and the bit allocation to the transmitted subbands are both predetermined (off-line) and kept fixed. Specification of the subbands which are transmitted using the predetermined bit allocation is adapted to each window of speech.

In accordance with the present invention, the number of subbands to be transmitted and the bit allocation to the subbands are recalculated (on-line) for each coding frame (speech window) to match the signal short-term spectral power distribution. The procedure used to calculate the per-frame allocation is the dynamic bit-allocation 28. Given the frame spectral envelope estimate and the total number of available quantization bits, the algorithm attempts to minimize the residual coefficients average quantization error (in the mean squared-error, mse, sense) for that frame. This process automatically adjusts the bit allocation to match the spectral characteristics of the signal and the specified compression rate.

Given a total number of available quantization bits B_T and an available set of quantizers for quantizing coefficients, the system searches for a bit allocation $B = (b_1, b_2, \ldots, b_i, \ldots, b_L)$ where b_i is the number of bits per coefficient in the i-th subband. The available quantizers quantize to a number of bits from the range [0, minb, minb+1, minb+2,...., maxb]. The bit allocation is such that (1) b_i is a nonnegative integer in the range of available quantizers, (2) $b_1 + b_2 + \ldots + b_L \leq B_T$, and (3) the average quantization error is minimized.

The search for optimal allocation is based on the unconstrained solution $r_i = \max\,[0, \operatorname{Log}_2\,(P_i/D)] = \max[0, LP_i - LD)]$, where D is the expected mse distortion for the window of speech, r_i is a nonnegative real number, and LP_i and LD are respectively the log-power and log-mse terms. The solutions r_i are rounded to the available quantization levels to provide b_i. The power of spectrum density $P = (P_1, P_2, \ldots, P_i, \ldots, P_L)$ can be derived from the spectral envelope coefficients. In accordance with the present approach, the expected distortion D is estimated and the bit allocation vector B is calculated from the unconstrained solution of r_i. In an iterative process, the distortion D is then re-estimated and the bit allocation vector B is recomputed to obtain the bit allocation for which the sum $B_S = b_1 + b_2 + \ldots + b_L$ is as close as possible to B_T. The bit allocation is then fine-tuned to make full use, or nearly full use, of the available bits B_T.

Although the power density P_i could be computed directly from the four coefficients of each subband, it is in this case only estimated or derived at 27 from the i-th subband spectral envelope. From the spectral envelope estimate obtained at 22, the power density is obtained from a precalculated conversion table. As a result, no additional side information is needed to transmit the power spectrum density.

The iterative process 28 of determining the expected mse distortion and thus computing the bit allocation vector B from the expected distortion and power spectrum density P is illustrated in FIG. 4. The system searches for the best value LD using a bisecting search. Assume that the expected mse distortion which allows for full use of the available quantization bits is LD_{exp}. The system begins with an estimated value LD midway between the maximum and minimum possible values for LD. LD_{max} and LD_{min} are equal to the natural log of the maximum and minimum power respectively of the power spectrum P as specified in the look up table or by the nature of a specific signal and bit rate. (P_{min} should be greater than zero, because log 0 is undefined.) From the first estimate of LD, $LD^{\circ} = (LD_{min} + LD_{max})/2$, the bit allocation vector B is computed from r_i, where b_i is the value r_i rounded to available quantizer levels. The total number of bits thus allocated is compared to the number of available bits. If the solution indicates that more bits may be used, indicating that a lower level of distortion may be obtained, the value LD is decreased. On the other hand, if the solution indicates that more than the available bits B_T have been allocated, a greater distortion must be accepted to reduce the number of allocated bits.

In the example of FIG. 4, assume that the best solution is LD_{exp}. The first solution of B from LD° would indicate that too many bits were allocated. Therefore, the distortion would be increased from $(LD_{max} + LD_{min})/2$ to a level LD^1 toward LD_{max}. Specifically, LD^1 is the point bisecting the interval $[LD^{\circ}, LD_{max}]$ computed as $LD^1 = LD^{\circ} + (LD_{max} - LD_{min})/4$. The bit allocation would be recomputed, and it would be determined from that new solution in this example that additional bits would be available and LD should be decreased. A new distortion level LD^2 bisecting the prior two distortion levels would then be used to solve again for the bit allocation B.

This iterative process would continue until the number of allocated bits was equal to, or very near to, the available bit allocation B_T. Because the solution must be of integer values, the calculations may result in an oscillation about the level LD_{exp}. As discussed below, the system avoids such oscillation and fine tunes the bit allocation in a heuristic fashion.

FIG. 5 illustrates the dynamic bit allocation procedure in greater detail. As noted above, power density for each subband is estimated by table lookup from the spectral envelope parameters determined for each subband. For purposes of determining the bit allocation, the subbands are then ordered in decreasing power density at 50. The system is initialized at 52. The initial distortion factor $\log_2 D^\circ = LD^\circ$ is computed as $(LD_{min} + LD_{max})/2$. An initial distortion increment value d_0 is set equal to $(LD_{max} - LD_{min})/2$. Two counter j and k for limiting the duration of searches are set at 0.

In a first loop 54 for performing the search illustrated in Fig. 4, j is set equal to j+1 and d_j is set equal to $d_{j-1}/2$ at 56. To limit the search once the value LD is very near to the expected value LD_{exp}, j is compared to a value j_{max} at 58. If the system has reached the maximum value of j, it goes to a heuristic fine-tuning step at II. Otherwise, the unconstrained solution of r is computed at 60. That solution is then rounded at 62 to b_i.

In the preferred available range of quantizer values, quantizers from minb equal to 2 bits per coefficient through maxb equal to 9 bits per coefficient are available. A zero bit quantizer is always included in the set of quantizers. Although it is preferred that quantizers are available at all integer values between minb and maxb, such is not a requirement. All solutions r_i are rounded to the nearest available quantizer value. Specifically, in the present implementation if r_i is computed to be .8, that value is rounded to b = 0 since a 1 bit quantizer is not available. Similarly, if r_i is computed to be 1.1, it is rounded to $b_i = 2$, the nearest available quantizer level.

At 64, the sum of the bit allocations B_S is computed, as are the difference between the total available bits B_T and B_S and the absolute value of that difference. If it is determined at 68 that B_T is equal to B_S, the system has the solution and stops. If the value is greater than zero, indicating that additional bits are available and the distortion level can be decreased, LD is set equal to LD - d_j at 70 and the system returns to the beginning of loop 54. On the other hand, if the value $B_T - B_S$ is less than zero, indicating that too many bits were allocated and that a greater distortion must be accepted, LD is set equal to LD + d_j and the system goes to the beginning of the loop 54.

Once the system has processed through the loop 54 to the point where j is greater than jmax at 58, a heuristic routine II is begun. From the prior loops at 54, the LD associated with the smallest value of the absolute value of $B_T - B_S$ is selected at 74 as the nearest solution. The solution of the B is then calculated again at 76 and rounded at 78. B_S and $B_T - B_S$ are calculated at 80. Counters j and i are set to zero and L + 1 respectively at 81. IF $B_T - B_S$ is greater than zero, indicating that additional bits are available, one bit is added to the prerounded value r_j beginning with the highest energy subband where j equals 1. To that end, counter j is incremented at 84. One bit is added to r_j at 85. The new solution is then rounded at 86 and B_S and $B_T - B_S$ are computed at 88. If $B_T - B_S$ is then equal to zero the search stops. If it is still greater than zero the system loops back to 84 to increment the next value r_j.

It is possible for the addition of 1 bit at 85 to increase the number of rounded bits by 2. For example, if the prerounded r_j had previously been 0.6, resulting in zero bits being allocated for that subband, the addition of one bit results in an unrounded value of 1.6 which results in two bits for that subband. If the addition of the bit results in $B_T - B_S$ being less than zero, the value k is incremented at 92 and compared to k_{max} at 94. If k has not reached the maximum value the system moves to a decrementing loop beginning at 96. to avoid oscillation between incrementing and decrementing, once the value k_{max} is reached, the system goes to a final step III.

If $B_T - B_S$ is found to be less than zero at either 82 or 90, bits must be decremented. At 96 the counter i is decremented. Decrementing of bits at 98 is then prerounded r_i beginning with i = L, the lowest energy subband. The solution is rounded at 99, and B_S and $B_T - B_S$ are recomputed at 100. So long as $B_T - B_S$ is found to be less than zero at either 82 or 102, the next lowest energy value r_i is decremented at 98. If $B_T - B_S$ is found to be equal to zero at 102, the allocation is stopped. If it is found to be less than zero, k is incremented at 104 and compared to k_{max} at 106. Again, once k_{max} is reached the system goes to routine III. Otherwise the system goes to the incrementing step 84.

If the system fails to provide a B_S equal to B_T in step II, routine III is followed. A counter n is set at L + 1 at 108 and decremented at 110. If it is determined at 112 that B_T is greater than B_S the system stops. One or more available bits may not be used. If the number of allocated bits is greater than B_T, the system sets the lower energy subband bits b_n to zero at 114 beginning with n equal to L until $B_T - B_S \geq 0$.

As an alternative to incrementing and decrementing by power level, the system could increment or decrement subbands by order of frequency. Thus, for example, the additional bits in routine II could be added to the low frequency subbands and subtracted from the high frequency subbands.

From the computed bit allocation B, appropriate quantizers are applied to coefficients of selected subbands of the residual spectrum at 30. In a preferred implementation, the 2D quantizers range from minb = 2 bits per coefficient to maxb = 9 bits per coefficient, and each of these quantizers is designed using an approach presented by Linde, et al., "An Algorithm for Vector Quantizer Design," IEEE trans on Commun. Vol. COM 28, pp. 84-95, January 1980.

At the receiver, the spectral envelope estimate which is transmitted in frequency sequence is used to recalculate at 32 the signal power spectrum density P = $(P_1, P_2, \ldots, P_L)$. The power spectrum density is used to determine at 34 the corresponding bit allocation which, in turn, controls the inverse quantization process 36 as well as the spectral replication process 38. As in the prior Mazor, et al. approach, the coefficients of the subbands which are not transmitted are approximated by replication at 38 of transmitted subbands. Once the equalized spectrum of Figure 3 is recreated by replication of subbands, a reproduction of the spectrum of Figure 2 can be generated at 40 by applying the transmitted scale factors to the equalized spectrum. From that reproduction of the original Fourier transform, the speech can be obtained through an inverse FFT 42, an inverse scaler 46, a digital to analog converter and a reconstruction filter 48.

The important benefits associated with the dynamic bit allocation coding technique are (1) the coder is automatically optimized to match the specified coding rate, and (2) the coder's bit allocation is dynamically adjusted to track the signal short term spectrum. With these added capabilities the coder handles more effectively than the prior coder non-stationary signals like speech, non-speech signals like voiceband data, and variable rate coding.

The coding technique provides for excellent speech coding and reproduction at rates as low as 9.6kb/s. In terms of signal to noise ratio (SNR), the algorithm with the above specific implementation, outperforms the prior Mazor, et al. algorithm by over 2 dB at 16 kb/s and by more than 1 dB at 9.6 kb/s. The improvement in performance is even more dramatic when voiceband data and DTMF (dual tone multifrequency) signals are considered. This coder handles very well up to 2400-b/s data at 16-kb/s compression rate and up to 1200-b/s data at 12-kb/s compression rate. On the other hand, the prior coder can only handle up to 1200-b/s data at 16-kb/s compression rate. In terms of SNR, this coder outperforms the prior coder by over 5 dB at 16-kb/s and 3dB at 12 kb/s for voiceband data, and by over 6 dB at 16 kb/s as well as at 12 kb/s for DTMF signals.

While this invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

For example, transforms other than the Fourier Transform, such as the cosine transform, may be used. Also, the specific choice of available quantizers can vary. The encoding and decoding may be performed by dedicated hardware or in a software controlled system.

THE EMBODIMENTS OF THE INVENTION FOR WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A speech coding system comprising:
transform means for performing a discrete transform of a window of speech to generate a discrete transform spectrum of coefficients;
envelope defining and encoding means for defining an approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and for encoding the defined envelope of each subband of coefficients;
means for scaling each spectrum coefficient relative to the defined envelope of the respective subband of coefficients;
a plurality of quantizers of different bit lengths for encoding scaled spectrum coefficients within each subband; and
means for determining a quantizer to be used to encode the coefficients of each subband within each window of speech wherein the selection of available quantizers used in successive windows is variable.

2. A speech coding system as claimed in Claim 1 wherein the means for determining allocates bits to minimize quantization error for each window of speech.

3. A speech coding system as claimed in Claim 1 wherein the means for determining iteratively estimates distortion and computes for each distortion estimate a bit allocation across the spectrum to allocate a predetermined number of bits.

4. As speech coding system as claimed in Claim 3 wherein the means for determining computes the allocation of bits as number of bits greater than or equal to zero which is derived from $\log_2(P_i/D)$ for each subband where P_i is a power density estimate for the subband and D is a distortion error estimate for the window of speech.

5. A speech coding system as claimed in Claim 4 wherein P_i is estimated for each subband from the defined envelope of the respective subband of coefficients.

6. A speech coding system as claimed in Claim 4 wherein the means for determining tunes the computed bit allocation toward the number of available bits by adding bits to higher energy subbands or subtracting bits from lower energy subbands.

7. A speech coding system as claimed in Claim 6 wherein P_i is estimated for each subband from the defined envelope of the respective subband of coefficients.

8. A speech coding system as claimed in Claim 1 wherein the transform means performs a discrete Fourier transform.

9. A method of coding speech comprising:

performing a discrete transform of a window of speech to generate a discrete spectrum of coefficients;

defining an approximate envelope of the discrete spectrum in each of a plurality of subbands of coefficients and digitally encoding the defined envelope of each subband of coefficients; and

with at least one of a plurality of quantizers of different bit lengths, encoding each of scaled coefficients within subbands into a number of bits, the quantizer used for each subband being determined for each window of speech to minimize distortion, wherein the selection of available quantizers used in successive windows is variable.

10. A method as claimed in Claim 9 wherein the quantizers are determined to minimize quantization error for each window of speech.

11. A method as claimed in Claim 9 wherein the quantizers are determined by iteratively estimating distortion and computing for each distortion estimate a bit allocation across the spectrum to allocate a predetermined number of bits.

12. A method as claimed in Claim 11 wherein the quantizers are determined by computing the allocation of bits as a number of bits greater than or equal to zero which is derived from $\log_2 (P_i/D)$ for each subband where P_i is a power density estimate for the subband and D is a distortion error estimate for the window of speech.

13. A method as claimed in Claim 12 wherein P_i is estimated for each subband from the defined envelope of the respective subband of coefficients.

14. A method as claimed in claim 12 wherein the allocation of bits is tuned toward the number of available bits by adding bits to higher energy subbands and subtracting bits from lower energy subbands.

15. A method as claimed in Claim 12 wherein P_i is estimated for each subband from the defined envelope of the respective subband of coefficients.

16. A method as claimed in Claim 9 wherein the transform is a discrete Fourier transform.

R. WILLIAM WRAY & ASSOCIATES
BOX 2760 - STATION D
OTTAWA, CANADA K1P 5W8
PATENT AGENT FOR THE APPLICANT

2053064

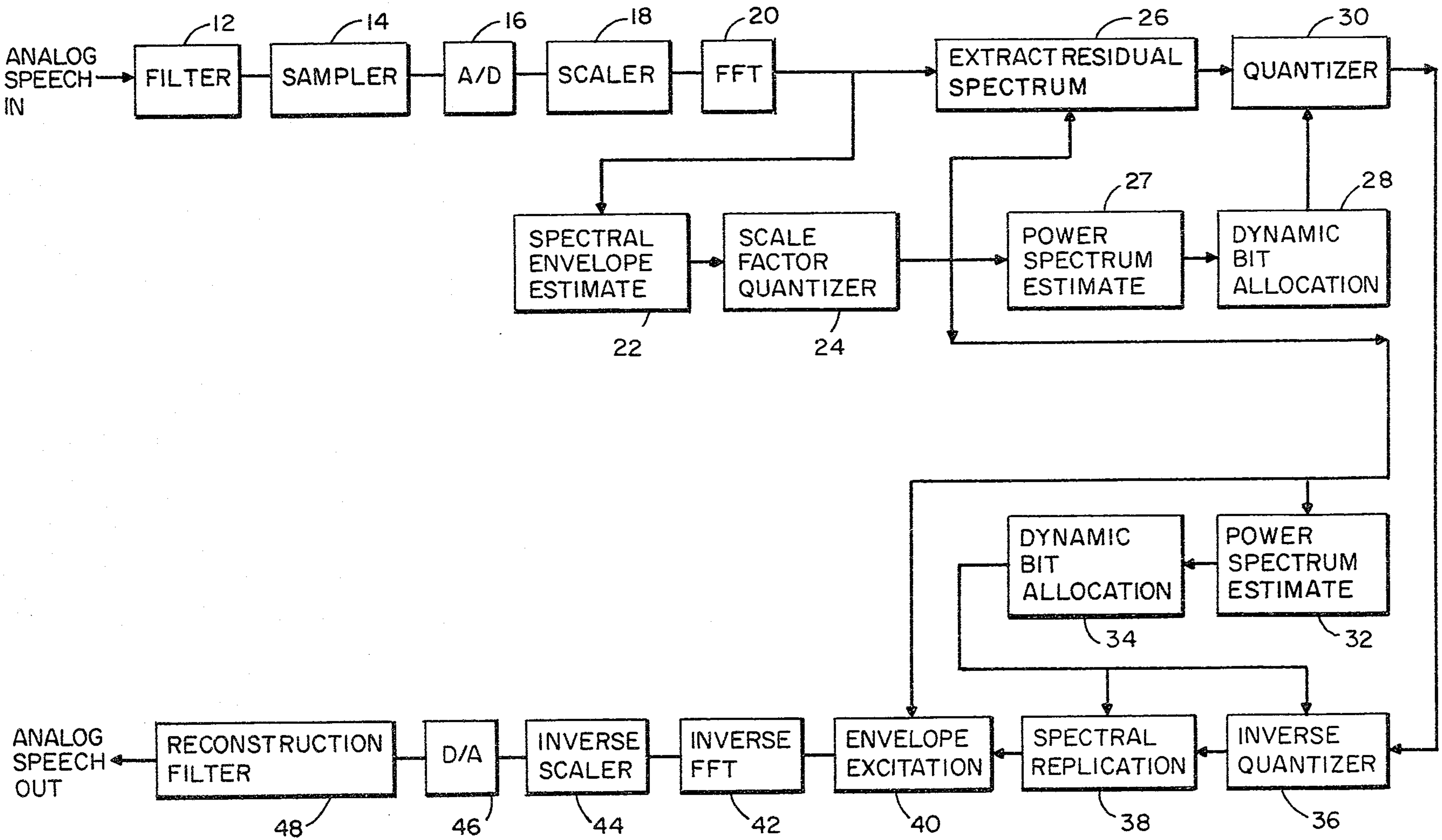


FIG. I

R. William Wray & Associates

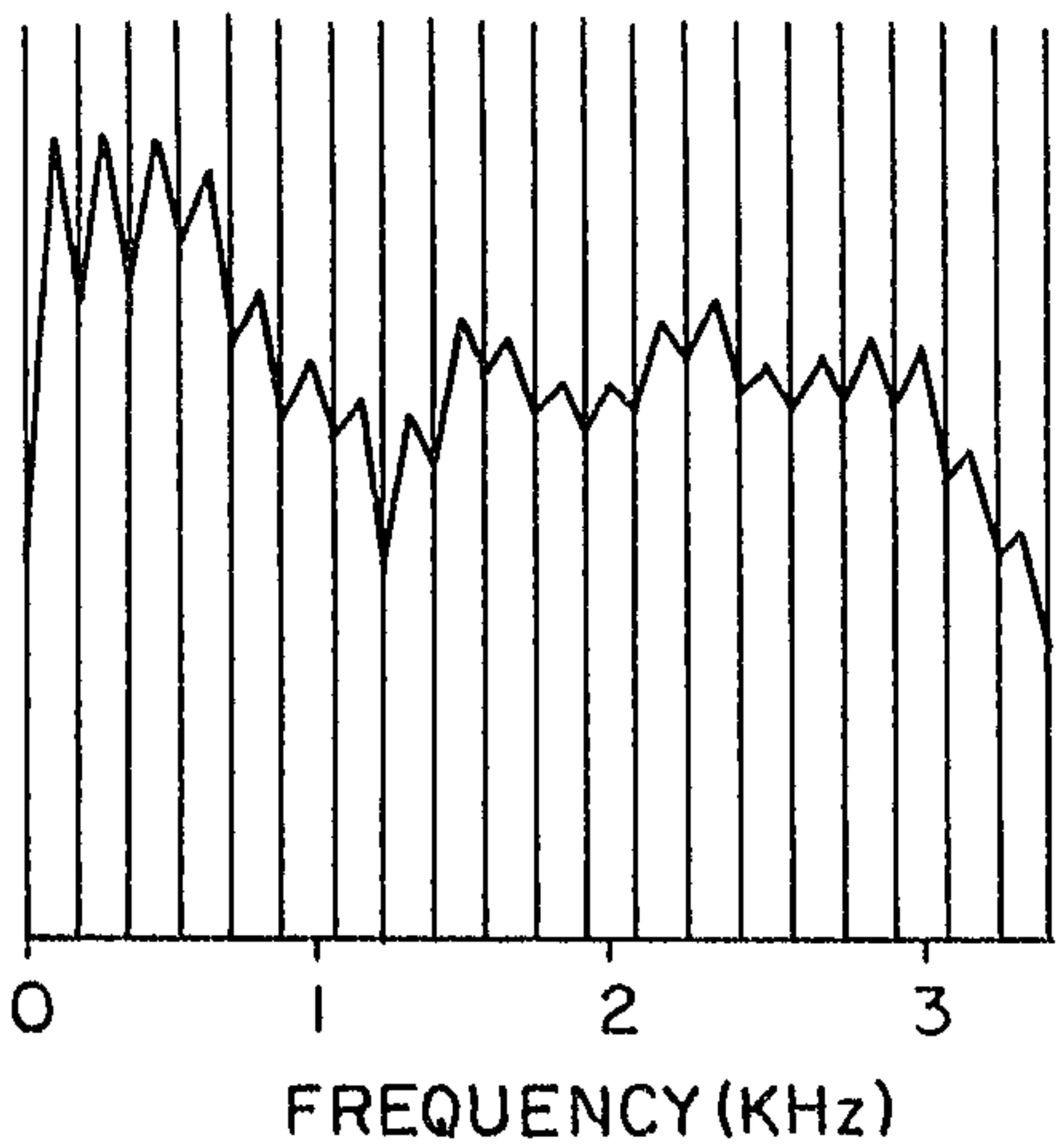


FIG. 2

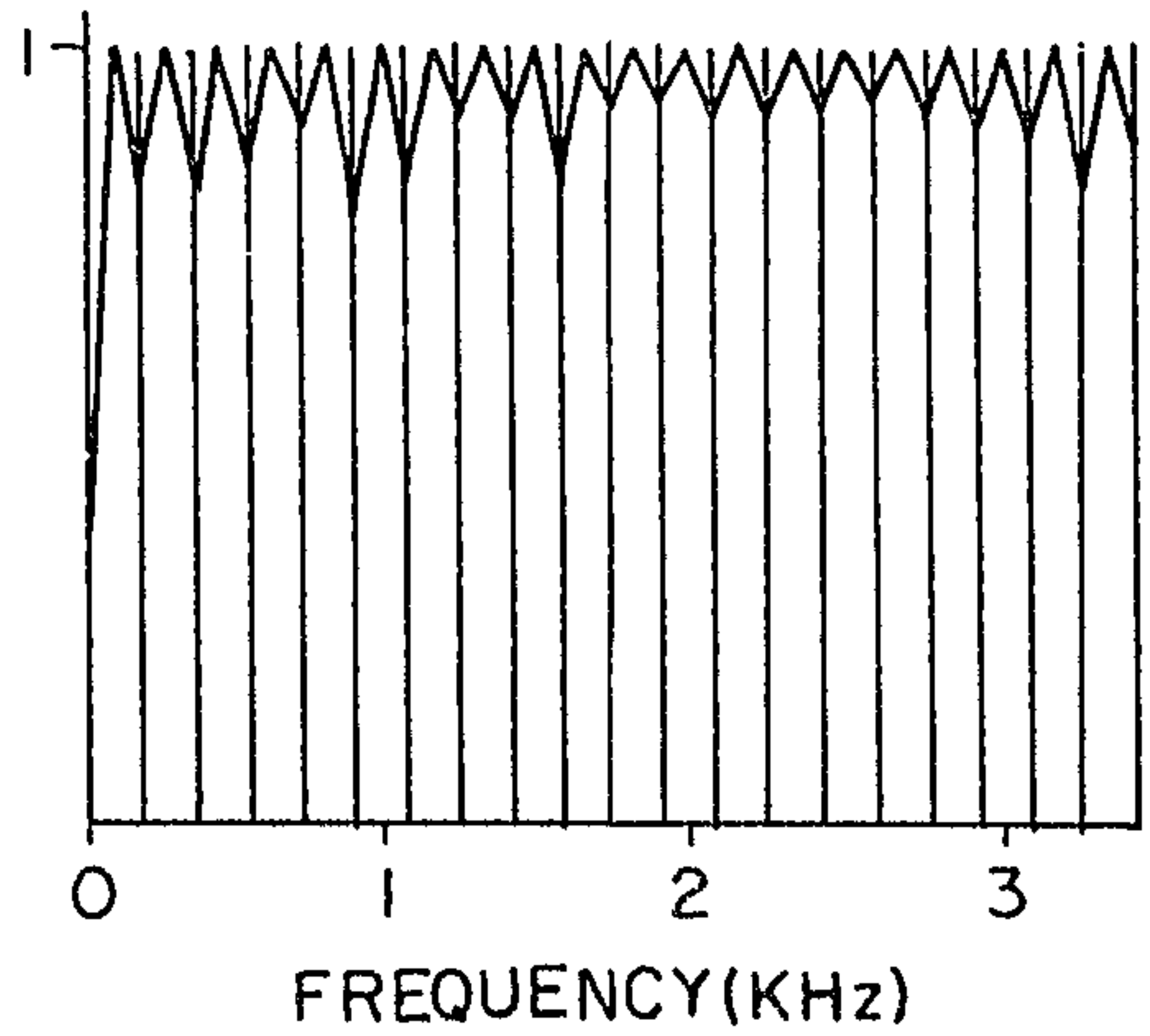


FIG. 3

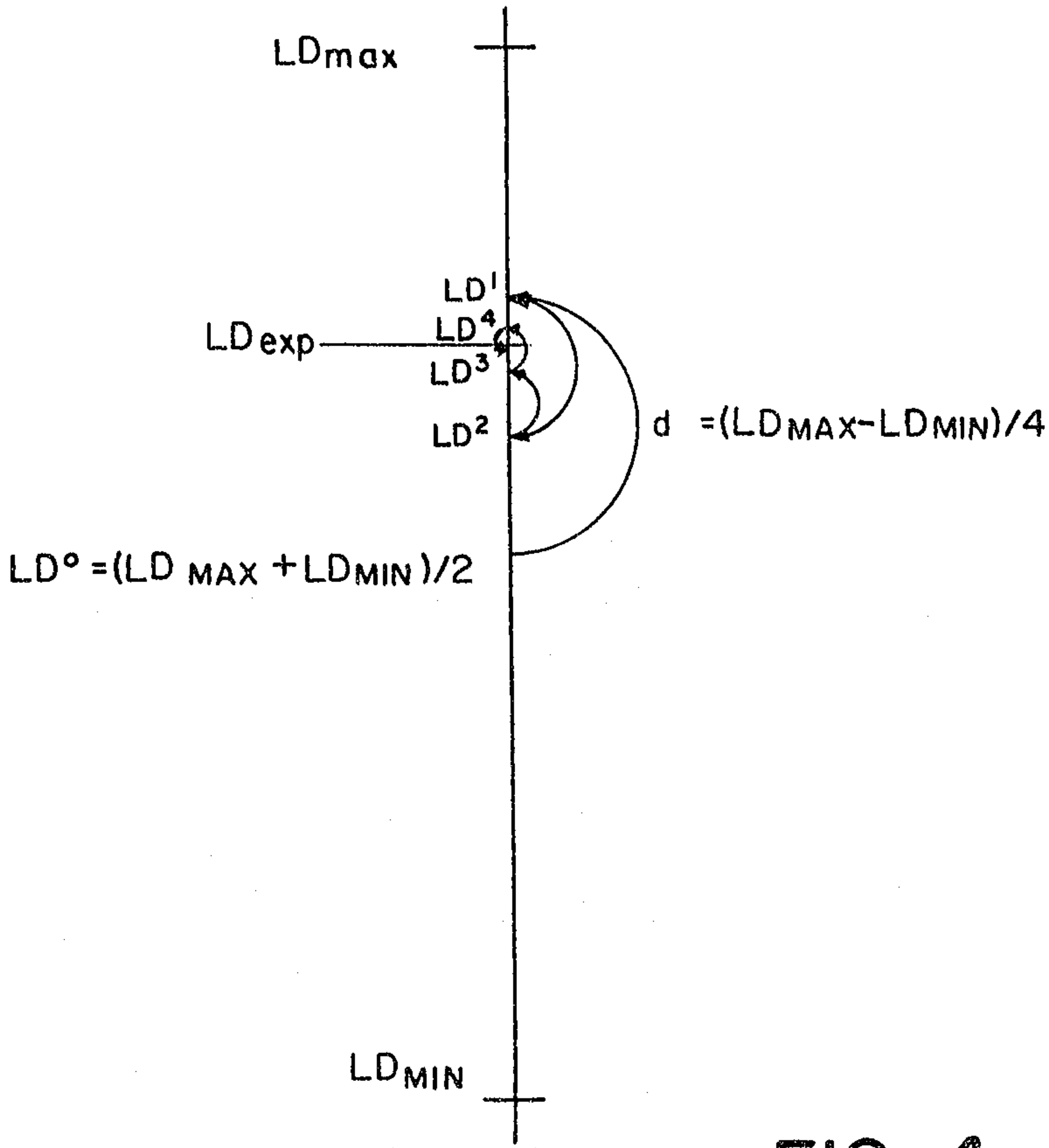


FIG. 4

R. William Wray & Associates

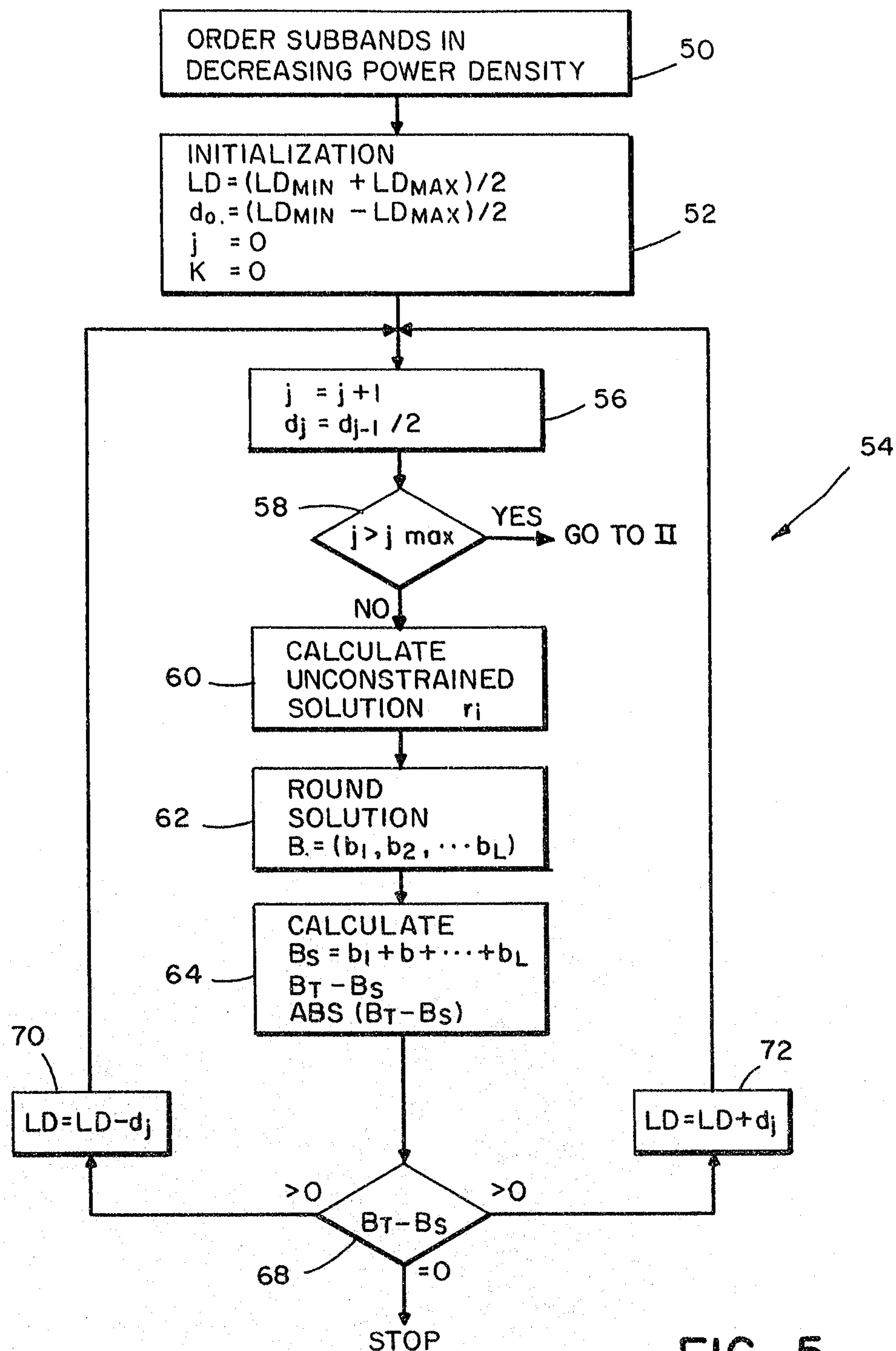

ORDER SUBBANDS IN DECREASING POWER DENSITY
50
INITIALIZATION
LD = (LDMIN + LDMAX)/2
d0 = (LDMIN - LDMAX)/2
j = 0
K = 0
52
j = j+1
dj = dj-1 /2
56
54
58
j > j max
YES
GO TO II
NO
CALCULATE UNCONSTRAINED SOLUTION ri
60
ROUND SOLUTION
B = (b1, b2, ··· bL)
62
CALCULATE
BS = b1 + b + ··· + bL
BT - BS
ABS (BT - BS)
64
70
LD = LD - dj
72
LD = LD + dj
>0
>0
BT - BS
=0
68
STOP

FIG. 5
SHEET 1

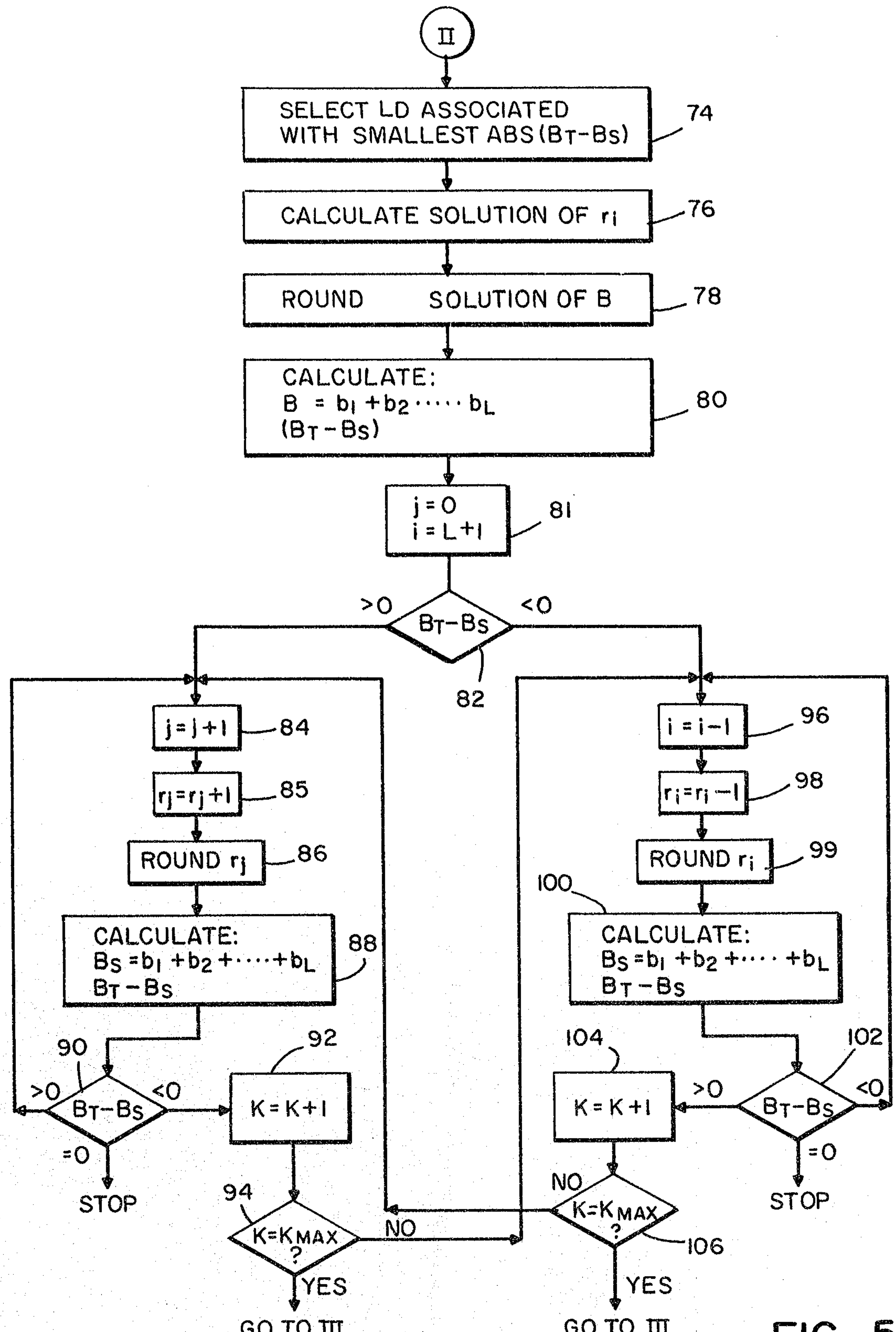


FIG. 5

SHEET 2

R. William Wray & Associates

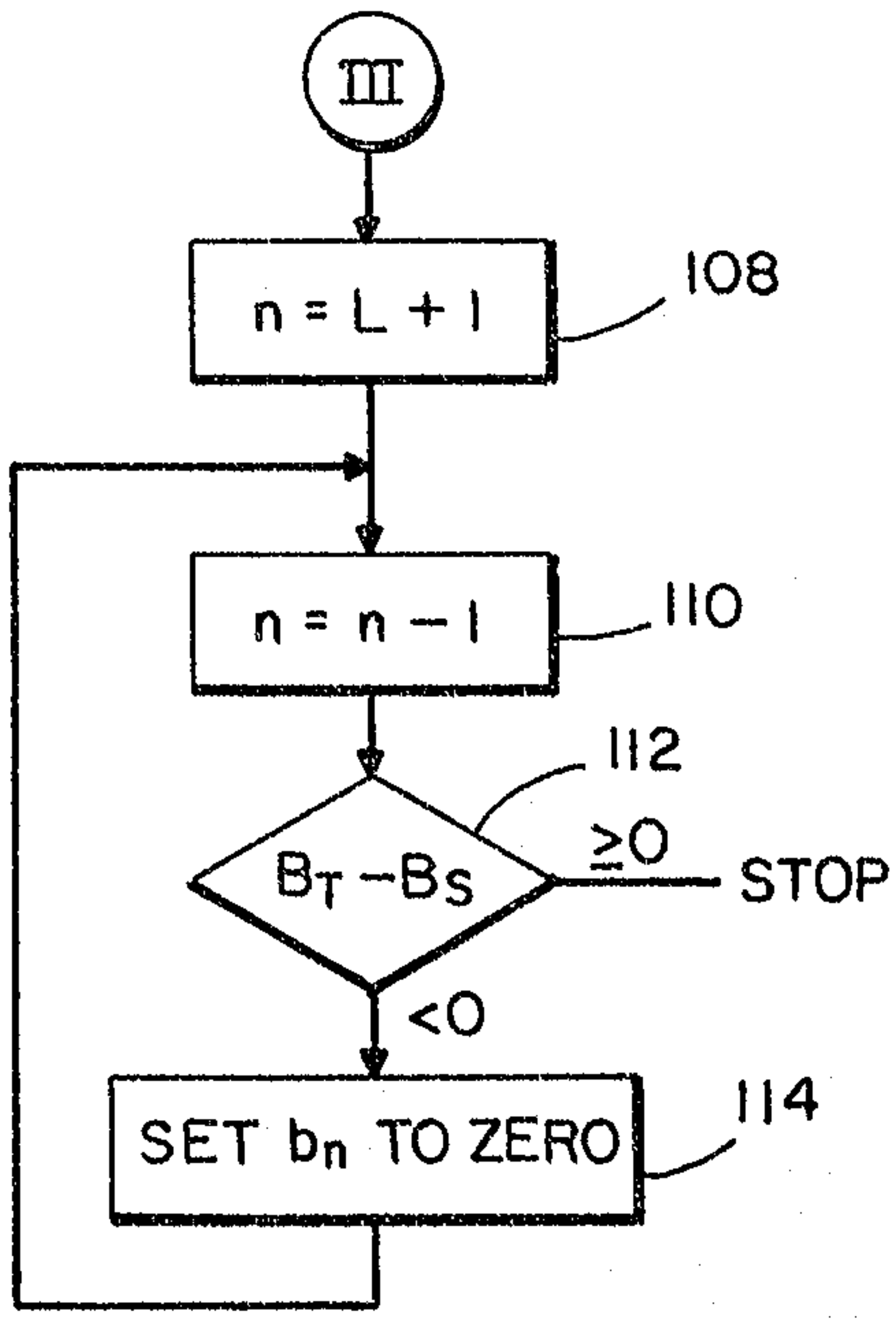


FIG. 5
SHEET 3

ANALOG SPEECH IN
12
FILTER
14
SAMPLER
16
A/D
18
SCALER
20
FFT
26
EXTRACT RESIDUAL SPECTRUM
30
QUANTIZER
SPECTRAL ENVELOPE ESTIMATE
22
SCALE FACTOR QUANTIZER
24
27
POWER SPECTRUM ESTIMATE
28
DYNAMIC BIT ALLOCATION
DYNAMIC BIT ALLOCATION
34
POWER SPECTRUM ESTIMATE
32
ANALOG SPEECH OUT
RECONSTRUCTION FILTER
48
D/A
46
INVERSE SCALER
44
INVERSE FFT
42
ENVELOPE EXCITATION
40
SPECTRAL REPLICATION
38
INVERSE QUANTIZER
36