LOW PASS FILTER DESIGN*

HIROTUGU AKAIKE

(Received Jan. 19, 1968)

1. Introduction and summary

Many of the important problems in statistical analysis of time series are related with some kind of use of low pass filters. The purpose of this paper is to develop a practical and unified approach for the design and evaluation of these low pass filters.

We introduce an index which is concerned with the fidelity of a low pass filter and is closely related with the notion of bandwidth of a filter. The index will be denoted by HWIDTH and is defined in Section 2. Another index CUTOFF which is also closely related with the notion of bandwidth of a low pass filter and is concerned with the concentration of pass band around frequency zero is introduced in Section 2, too.

Interpretations of HWIDTH are discussed in Section 3 and values of HWIDTH and CUTOFF are given in Section 4 for some low pass filters.

By using the two indexes HWIDTH and CUTOFF and another quantity P, to be defined at the end of Section 4, we introduce a two dimensional (CTP, PTH)-space representation of filters, where CTP=CUTOFF/P and PTH=P/HWIDTH. By using this representation we introduce the notion of extremal family to filters. Restricting our attention to the set of extremal families we develop a design principle of low pass filters in Section 5. Numerical computations were done for filters with impulse response function of trigonometric polynomials and the results are given in Section 6. The results show that low pass filters with negative values at both ends of impulse response functions could be good low pass filters, and the validity of this conclusion is exemplified by a numerical example of their application in Figure 18. Also a specification of two simple filters to be denoted by LN1 and LN2 is obtained in this section.

It is almost obvious that the application of spectrum windows in the estimation of spectra is the low pass filtering of Fourier transform of

* This paper was prepared under Grant DA-ARO(D)-31-124-G726 when the author was visiting Stanford University.

sample covariance function in the frequency domain. Thus a review of hitherto proposed spectrum windows is given in Section 7 using the (CTP, PTH)-space representation of low pass filters. It turns out that some of the windows proposed by the present author are situated near the optimum design by the present approach. The implication of the method of fast Fourier transform for the application of low pass filters obtained in this paper for the estimation of spectra is also briefly discussed in this section.

In Section 8 the possibility of extending the present design procedure for the design of data windows is suggested.

2. Definition of HWIDTH and CUTOFF

We consider a linear time invariant filter of which impulse response function $D(t)$ is symmetric around the time $t=0$, i.e., $D(-t)=D(t)$. The frequency response function of this filter is given as $\mathfrak{F}D(f)=\int_{-\infty}^{\infty}\exp(-i2\pi ft)D(t)dt$. We generally assume $D(t)$ to be in $L_1(-\infty,\infty)$ and $L_2(-\infty,\infty)$, but the following definitions are valid if only $D(t)$ is in $L_2(-\infty,\infty)$ and $\mathfrak{F}D(f)$ is its Fourier transform in L_2-sense. Also we shall be able to admit as our $D(t)$ the Dirac's δ-function with corresponding $\mathfrak{F}D(f)=1$. We assume that $D(t)$ is so normalized as to give $\mathfrak{F}D(0)=\int_{-\infty}^{\infty}D(t)dt=1$.

As an index of low frequency characteristic of our filter we adopt HWIDTH which is by definition:

$$\text{HWIDTH}=\frac{2}{\int_{-\infty}^{\infty}\left|\frac{1-\mathfrak{F}D(f)}{f}\right|^2 df}.$$

For an ideal rectangular low pass filter with $\mathfrak{F}D(f)=1$ for $|f|\leqq f_c$ and $=0$ for $|f|>f_c$ we have $\text{HWIDTH}=f_c$. Thus we can see that HWIDTH implies half width of the pass band.

We shall also consider another quantity CUTOFF, which is by definition:

$$\text{CUTOFF}=\left(\frac{3\int_{-\infty}^{\infty}f^2|\mathfrak{F}D(f)|^2df}{\int|\mathfrak{F}D(f)|^2df}\right)^{1/2}.$$

The name CUTOFF does not suggest any direct relation with the notion of cutoff frequency commonly used in the field of engineering, but for the above stated ideal rectangular low pass filter we have $\text{CUTOFF}=f_c$. This definition of CUTOFF or some analogue of it has been fairly commonly adopted as an index of the bandwidth of a filter [1, 6], but as will be seen in the following discussion, CUTOFF itself is not necessarily very much suited for the evaluation of bandwidth of some type of low pass filters.

3. Physical meaning of HWIDTH

As an input to the filter we consider a stationary white noise $n(t)$ with constant spectral density $p_n(f)=1$. To evaluate the fidelity of our low pass filter we take the difference of $n(t)$ and its filtered version, formerly represented by $D*n(t)=\int_{-\infty}^{\infty} n(t-s)D(s)ds$, and measure the power after passing it through an integrator. The resulting quantity would be $\int_{-\infty}^{\infty} |1-\mathfrak{F}D(f)|^2/|2\pi f|^2 df$ and is equal to $(2\pi)^2 \cdot 2/\text{HWIDTH}$. Thus we can see that HWIDTH is closely related with the fidelity of the filtered data to the original data as an input to the integrator.

We can also get the following relation:

$$\int_{-\infty}^{\infty} \left| \frac{1-\mathfrak{F}D(f)}{2\pi f} \right|^2 df = \int_{-\infty}^{\infty} \left[U(t) - \int_{-\infty}^{t} D(s)ds \right]^2 dt ,$$

where

$$U(t)=\begin{cases} 0 & t<0 , \\ 1 & t>0 . \end{cases}$$

For the present case of symmetric $D(t)$ this quantity is equal to $2\int_0^{\infty}\left[\frac{1}{2}-\int_0^t D(s)\,ds\right]^2 dt$. Here it should be noticed that we have normalized $D(t)$ so that $\int_{-\infty}^{\infty} D(t)dt=1$ holds. We can see that the present quantity is directly related with the spread of $D(t)$ in the time domain. The above stated relation can sometimes conveniently be used for the evaluation of HWIDTH.

It should be noticed that in the definition of CUTOFF attention is not paid to the phase characteristics of the filter, while in the definition of HWIDTH it is explicitly taken into account. Also it should be noticed that we get one and the same value of HWIDTH for both of the filters with frequency response functions $\mathfrak{F}D(f)$ and $2-\mathfrak{F}D(f)$. This shows that our present definition of HWIDTH is not sufficient to specify a low pass filter. Rather it is only concerned with the behavior of $\mathfrak{F}D(f)$ in some neighborhood of $f=0$ and its smallness does not necessarily mean that the corresponding filter is a low pass filter. Thus for the evaluation of low pass filters there is a necessity of introducing another quantity such as CUTOFF which is directly concerned with the behavior of $\mathfrak{F}D(f)$ at higher frequencies. By using these two kinds of quantities we can say that if two filters are equivalent with respect to the qualification of low pass characteristic, such as CUTOFF, the one with the larger value of HWIDTH would be considered to be the better one as a low pass filter. This consideration leads us to a design principle which is to be discussed in Section 5.

4. Numerical examples of HWIDTH and CUTOFF

Here we summarize some of the values of HWIDTH and CUTOFF for various types of filters.

$\mathfrak{F}D(f)$	$D(f)$	HWIDTH	CUTOFF
$=1$ for $\lvert f\rvert \leqq f_c$ $=0$ otherwise	$=2f_c\left(\frac{\sin 2\pi f_c t}{2\pi f_c t}\right)$	f_c	f_c
$=1-\frac{\lvert f\rvert}{f_c}$ for $\lvert f\rvert \leqq f_c$ $=0$ otherwise	$=f_c\left(\frac{\sin \pi f_c t}{\pi f_c t}\right)^2$	$\frac{1}{2}f_c$	$\sqrt{\frac{3}{10}}f_c$
$=1-\left(\frac{f}{f_c}\right)^2$ for $\lvert f\rvert \leqq f_c$ $=0$ otherwise	$=\frac{4f_c}{(2\pi f_c t)^2}\left(\frac{\sin 2\pi f_c t}{2\pi f_c t}-\cos 2\pi f_c t\right)$	$\frac{3}{4}f_c$	$\sqrt{\frac{3}{7}}f_c$
$=\frac{\sin 2\pi f L}{2\pi f L}$	$=\frac{1}{2L}$ for $\lvert t\rvert \leqq L$ $=0$ otherwise	$\frac{6}{\pi^2}\frac{1}{2L}$	∞
$=\left(\frac{\sin \pi f L}{\pi f L}\right)^2$	$=\frac{1}{L}\left(1-\frac{\lvert t\rvert}{L}\right)$ for $\lvert t\rvert \leqq L$ $=0$ otherwise	$\frac{10}{\pi^2}\frac{1}{2L}$	$\frac{3}{\pi}\frac{1}{2L}$
$=\frac{1}{1+\left(\frac{2\pi f}{\lambda}\right)^2}$	$=\frac{\lambda}{2}\exp(-\lambda\lvert t\rvert)$ $(\lambda>0)$	$\frac{4}{\pi}\frac{\lambda}{2\pi}$	$\sqrt{3}\frac{\lambda}{2\pi}$

As can be seen from the example of the boxcar filter which has boxcar-like rectangular impulse response function, CUTOFF takes the value ∞ for a $D(t)$ which has some discontinuities. As the boxcar filter is usually considered to give a simple low pass filter, this result shows that CUTOFF is not necessarily a good index of the bandwidth of a low pass filter. Rather it is more sensitive to the higher frequency characteristics of the filter. Thus it can be fairly safe to say that when CUTOFF and HWIDTH show nearly the same values the filter is a low pass filter without significant minor lobes in the higher frequency domain, and if we are going to apply the filter to an input which does not contain extremely significant higher frequency component HWIDTH would be more useful as an index of bandwidth for some wider range of low pass filters than CUTOFF. CUTOFF is concerned with the concentration of the (power) pass band around the frequency zero and HWIDTH is concerned with the fidelity at low frequency. Thus it would be more natural to simultaneously use these two indexes for the specification of a low pass filter.

We have also another natural measure for the specification of a filter. This is the power P of the output of the filter when it is excited by a white noise with unit spectral density and is by definition:

$$P=\int_{-\infty}^{\infty} |\mathfrak{F}D(f)|^2 df .$$

Obviously this quantity is the area covered by $|\mathfrak{F}D(f)|^2$ and thus also related with the notion of bandwidth. When a low pass filter is used with complex demodulation technique [8] to provide an estimate of a power spectral density, this quantity P is generally directly concerned with the sampling variability of the estimate. In the next section we shall use these three quantities P, CUTOFF and HWIDTH to develop a design principle of low pass filters.

5. A design principle of low pass filters

For the purpose of normalization we shall consider a family of filters of which impulse response functions $D_L(t)$ are specified by a positive scale parameter L and satisfy the relation

$$D_L(t)=\frac{1}{L}D_1\left(\frac{t}{L}\right) .$$

We are assuming that $\int_{-\infty}^{\infty} D_1(t)dt=1$ holds.

Here if we denote by $P(D_L)$, CUTOFF (D_L) and HWIDTH (D_L) the values of P, CUTOFF and HWIDTH of a filter with impulse response function $D_L(t)$, respectively, then from the definition we get the relation:

$$P(D_L)=\frac{1}{L}P(D_1) ,$$

$$\text{CUTOFF }(D_L)=\frac{1}{L}\text{CUTOFF }(D_1)\,,$$

$$\text{HWIDTH }(D_L)=\frac{1}{L}\text{HWIDTH }(D_1)\,.$$

Now if we ask that $P(D_L)=P_0\ (>0)$, then we have $\frac{1}{L}=\frac{P_0}{P(D_1)}$ and thus CUTOFF $(D_L)=P_0\frac{\text{CUTOFF }(D_1)}{P(D_1)}$ and HWIDTH $(D_L)=P_0\frac{\text{HWIDTH }(D_1)}{P(D_1)}$. The quantities $\frac{\text{CUTOFF }(D_L)}{P(D_L)}$ and $\frac{\text{HWIDTH }(D_L)}{P(D_L)}$ are independent of the scale factor L and can be considered to give a specification of the family to which $D_L(t)$ belongs. For the sake of simplicity of typography and analytical treatment we define CTP(D) and PTH(D) for a filter with impulse response function $D(t)$ or for the family to which it belongs, which will hereafter simply be denoted by D, by

$$\text{CTP}(D)=\frac{\text{CUTOFF }(D)}{P(D)}\,,$$

$$\text{PTH}(D)=\frac{P(D)}{\text{HWIDTH }(D)}\,.$$

By using these two quantities, we can make each family of filters correspond to a point (CTP(D), PTH(D)) in a two-dimensional space, which we shall call (CTP, PTH)-space.

Now for the purpose of evaluation, we are going to introduce a partial order $<$ into the space of families of filters, which is by definition:

$$D^{(2)}<D^{(1)}\Longleftrightarrow\begin{cases} & \text{CTP}(D^{(2)})\leqq\text{CTP}(D^{(1)}) \\ \text{and} & \\ & \text{PTH}(D^{(2)})\leqq\text{PTH}(D^{(1)}) \\ \text{where at least one equality holds} & \end{cases}$$

and we shall call $D^{(2)}$ to be sharper than $D^{(1)}$. Let us consider the simple case where two families of filters $D^{(1)}$ and $D^{(2)}$ satisfy the relations $\text{PTH}(D^{(1)})=\text{PTH}(D^{(2)})$ and $\text{CTP}(D^{(2)})<\text{CTP}(D^{(1)})$. In this case it is easy to see that if two filters, one from $D^{(1)}$ and another from $D^{(2)}$, are asked to have one and the same value of HWIDTH, then the one from $D^{(2)}$ will have a smaller value of CUTOFF. In this case, from the definition of CUTOFF, it will be reasonable to expect that the one from $D^{(2)}$ will show a better concentration around the frequency zero and, in this sense, will be the better one. The same kind of argument can be extended to the general case of $D^{(2)}<D^{(1)}$ and we adopt the standpoint that the

sharper filter is the better one as a low pass filter.

Now that we have got a method of representation of a family of filters as a point in (CTP, PTH)-space and got an ordering or evaluation of families by it, we are in a position to further the investigation of low pass filters systematically.

Within a set of families of filters we call a D to be extremal if there is not any other family which is sharper than D. If the set of families of filters is forming a convex set in (CTP, PTH)-space, an extremal D will be characterized as the one which gives the minimum of

$$R(U)=U(\mathrm{CTP}(D))^2+(1-U)(\mathrm{PTH}(D))^2$$

for some U satisfying $0 \leqq U \leqq 1$. It seems quite natural to first focus our attention to the set of all extremal families to find good low pass filters. To further concentrate our attention, let us consider the situation when an extremal D is given for practical application. The most probable procedure will be first to select a value of L so as to make HWIDTH $(D_L)=H_0$, where H_0 is determined from the requirement of the width of pass band. In this case it holds that CUTOFF $(D_L)=H_0\times$ CTH(D), where

$$\begin{aligned}\mathrm{CTH}(D)&=\frac{\mathrm{CUTOFF}\ (D)}{\mathrm{HWIDTH}\ (D)}\\&=\mathrm{CTP}(D)\times\mathrm{PTH}(D)\,.\end{aligned}$$

Following the same reasoning as we have done in evaluating a sharper family as a family of better low pass filters, we can expect that within the set of extremal families the ones with smaller values of CTH(D) will give better low pass filters.

Summarizing this, we get a rather simple design principle of low pass filters: *to restrict our attention to the families which are nearly extremal and with small values of* CTH(D).

It would be needless to say that various further considerations would be necessary to make this principle a practical one, but, as will be seen in the following sections, this simple approach gives a fairly good insight into the nature of good low pass filters.

6. Design of trigonometric filters

In this section we consider a set of families of filters for which $D_L(t)$ is given as follows:

$$D_L(t)=\begin{cases}\dfrac{1}{2L}\left(1+2\sum\limits_1^k a_n\cos\left(2\pi\dfrac{n}{2L}t\right)\right) & \text{for } |t|\leqq L\,,\\ 0 & \text{otherwise.}\end{cases}$$

Throughout the following discussion we shall assume $D_L(\pm L)=\frac{1}{2L}\left(1+2\sum_1^k(-1)^n a_n\right)=0$ to avoid the case where CUTOFF $(D_L)=+\infty$. We have

$$P(D_L)=\frac{1}{2L}\left(1+2\sum_{n=1}^K a_n^2\right)$$

$$\text{CUTOFF }(D_L)=\frac{1}{2L}\left(\frac{6\sum_{n=1}^K n^2a_n^2}{1+2\sum_1^K a_n^2}\right)^{1/2}$$

$$\text{HWIDTH }(D_L)=\frac{1}{2L}\left\{\sum_{n=1}^K\frac{(a_n-1)^2}{n^2}+\frac{\pi^2}{6}-\sum_{n=1}^K\frac{1}{n^2}\right\}^{-1}.$$

From these relations we get

$$\text{CTP}(D_L)=\left(6\sum_{n=1}^K n^2a_n^2\right)^{1/2}\left(1+2\sum_1^K a_n^2\right)^{-3/2}$$

$$\text{PTH}(D_L)=\left\{\sum_{n=1}^K\frac{(a_n-1)^2}{n^2}+\frac{\pi^2}{6}-\sum_{n=1}^K\frac{1}{n^2}\right\}\left(1+2\sum_{n=1}^K a_n^2\right).$$

Thus, assuming the convexity of the set of all families, the search for extremal families is reduced to the minimization problem of

$$R(U)=U(\text{CTP}(D_L))^2+(1-U)(\text{PTH}(D_L))^2$$

in a K-dimensional space of $(a_1, a_2, \cdots, a_K)$ for the values of U lying in $[0, 1]$. We shall also give our attention to a subset of the present set of families, members of which are assumed to satisfy the condition

$$\int_{-L}^{L} t^2D_L(t)dt=0\,,$$

or equivalently,

$$\left\{\frac{1}{12}+\frac{1}{\pi^2}\sum_{n=1}^K a_n\frac{(-1)^n}{n^2}\right\}=0\,.$$

In the following we shall symbolically designate a quantity related with this subset by putting $N=2$ and that of the former set by $N=1$. Also we shall designate by $N=0$ the case of the boxcar

$$D_L(t)=\begin{cases}\dfrac{1}{2L} & \text{for } |t|\leqq L\,,\\ 0 & \text{otherwise}\,.\end{cases}$$

From the definition, the members of the subset $N=2$ are considered to be distortion free to a locally quadratic variation in time of the input. In the following analysis we can see the effect of this restriction in the design of low pass filters.

We shall also give attention to the change of the result due to the change of the value of K, the degree of trigonometric polynomial. The practical use of a result which is heavily dependent on the higher values of K would be quite dubious. Thus the family of our good filter should be located in a region which is not very much sensitive to the increase of K beyond a certain limit.

Numerical results. First of all it should be mentioned that the numerical results to be used in this section were obtained mainly by using various methods of successive approximations. The results are generally expected to be reliable up to three digits, but there may still remain some quite significant errors. Even if there may be some defects in the following reasoning due to these possible errors, the concrete shapes of the filters are, if once obtained, free from this consideration of numerical errors.

Also it should be mentioned that in observing the figures illustrating $\mathfrak{F}D(f)$ one should take care of the possible changes of scale in various parts of the figures.

Necessary information for the understanding of the subsequent figures is summarized in Table 1. Here is included a quantity $RA=1/(2L\cdot \text{HWIDTH})$ for the purpose of evaluation of the spread of $D(t)$ over the time axis. (CTP(D), PTH(D)) which corresponds to a minimum of $R(U)$ is plotted in Figure 1, where U is successively changed from 0 to 0.99. We shall denote the filters obtained in this way by $\min R(U)$. The result suggests that $\min R(U)$ gives an extremal family and that the curves in the figure represent the graphs of the sets of extremal families. The figure also suggests that the results corresponding to smaller values of U are sensitive to the change of K and thus will not give practically useful filters. To get a feeling of the effect of changing the value of U, $D(t)$ and $\mathfrak{F}D(f)$ of $\min R(U)$ with $U=0$ and 0.99 are illustrated in Figures 2, 3, 4 and 5 for the cases of $N=1$ and $N=2$. The shapes of filters corresponding to the minima of CTH for ($N=1$, $K=4$), ($N=1$, $K=10$) and ($N=2$, $K=4$) are obtained by successive approximation using some of $\min R(U)$ as initial values and are illustrated in Figures 6, 7 and 8, respectively. This type of filter may hereafter be denoted as min CTH. By comparing Figures 6 and 7 we can see that the increase of K is mainly contributing to the improvement of phase characteristic of the filter at higher frequencies. The shapes of these filters are mutually quite similar and thus the value $\int_{-L}^{L} t^2 D_L(t)dt$ is very small even for those with $N=1$. It is remarkable that this min CTH introduced a $D(t)$ with negative values at both ends of $D(t)$ even without the restriction of $\int_{-L}^{L} t^2 D_L(t)dt=0$ or the distortion free requirement for the locally quadratic variation in time of the input.

In practical applications the cost of computation should also be taken into account. In the present case, $P(D_L)=\frac{1}{L}P(D_1)$ holds and we can use $P(D_1)=\frac{1}{2}\left(1+2\sum_{n=1}^{K} a_m^2\right)$ as an index of the efficiency of computation. The smaller the value of $P(D_1)$ we can expect that $D_L(t)$ is better spread all over $[-L, L]$. We have tried to take into account this factor and computed the values of $(a_1, a_2, \cdots, a_K)$ which give the minimum of $R(U)+W\times P(D_1)$ for a decreasing sequence of W for some selected values of U. Starting at $W=12.8$ we have successively reduced the value of W by a factor 1/2 and the computation was done by using a modified optimum gradient method [4] with initial values equal to the final values of the preceding step. The result is shown in Figure 9. It is quite interesting to note that in all of the cases the locus of solutions behaves as if it is strongly attracted toward the one corresponding to $U=1$ when W departs from zero. In this figure the sequences corresponding to the minimum of $R(1)+W\times P(D_1)$ are converging to the points corresponding to ($a_1=0.5$, $K=1$, $N=1$) and ($a_1=0.92995$, $a_2=0.42995$, $K=2$, $N=2$) for $N=1$ and $N=2$, respectively, as W tends to zero. Though this result is obtained only numerically, it seems that this is giving a specification of the two filters given by these two sets of coefficients. These two are also characterized as the lowest degree cases of trigonometric filters satisfying the conditions $N=1$ and $N=2$, respectively. We shall hereafter denote these two filters by LN1 and LN2. It also should be noticed that in the present procedure we could not approach a solution which is lying near the min CTH solution for the case of $N=1$. This suggests that so long as we try to keep $P(D_1)$ small it is difficult to imagine a filter nearly equal to min CTH ($N=1$) as a good low pass filter. And the fact that LN2 is situated quite near to this min CTH suggests that the requirement of the distortion free characteristic to the locally quadratic input could play a fundamental role in introducing a change in our notion of a good low pass filter.

In Figures 10 to 17 are given the shapes of various filters. For the purpose of comparison the boxcar, a filter with impulse response function of the shape of a boxcar, is illustrated in Figure 10. Figures 11 and 15 illustrate the filters which give the minimum values of CUTOFF (D_1). Figure 13 illustrates the case corresponding to the minimum of $R(U)$ with $U=0.89$ and is almost equal to min CTH of Figure 6. Figure 14 shows that by slightly modifying the value of U from 0.89 to $U=0.84$ we can keep the maximum of $|\mathfrak{F}D(f)|$ less than 1.01 without introducing significant changes of $|\mathfrak{F}D(f)|$ at the higher frequencies. Figures 12 and 16 give the shapes of LN1 and LN2. Figure 17 gives the shape of the filter giving the minimum of $R(U)$ for $U=0.86$. This is almost identical to min CTH of Figure 8. By comparing Figures 16 and 17 we can see that LN2 will be welcome in most ordinary applications.

The figures are all normalized by HWIDTH in frequency, and they strongly suggest that this normalization is quite useful. For almost all the good low pass filters in these figures $|\mathfrak{F}D(f)|$ remains less than 0.01 in the range $|f|>2$ HWIDTH and greater than 0.30 in the range $|f|<$ HWIDTH. Thus if only the length of the sampling interval Δt satisfies the relation $\frac{1}{2\Delta t}>2$ HWIDTH (or preferably, 3 HWIDTH), $\{D_L(n\Delta t); n=0, \pm1, \cdots, \pm M\}$ with $2L=(2M+1)\Delta t$ will be used as a good digital low pass filter for ordinary applications when $D_L(t)$ is a good low pass filter in the present sense. This shows that the minimizing CTH principle has a practical basis of reasoning, as it is based on the notion of HWIDTH which is proved to have a practical meaning.

The fact that minimum CTH design principle has introduced a $D(t)$ with negative values might yet make us feel somewhat dubious about the effectiveness of the principle. Although the foregoing discussions and figures have already provided definite answers to possible questions, the result illustrated in Figure 18 might be of some use to help our understanding. The original signal $\sin\left(\frac{2\pi}{60}n\right)$ was contaminated with a Gaussian white noise with standard deviation equal to 0.1 and fed into the various filters. A portion of the outputs is illustrated in Figure 18. With one exception of the case of boxcar with $2L=25$ all the values of $2L$ are chosen so as to make HWIDTH (D_L) approximately equal to one and the same constant. The result shows clearly that there is enough possibility of this kind of situation occurring in practical applications. The only practical limitation of LN2 and others nearly equal to min CTH is the requirement of the length of the span of $D_L(t)$. Thus we can say that in case $D_L(t)$ is used as a filter of an infinitely long series these filters will work quite efficiently.

If we take into account the fact that the Fourier transform of a sampled sequence of finite length is a periodic function in frequency domain, our present observation strongly suggests that the applications of these filters which are nearly equal to min CTH for the smoothing of a periodogram will be quite successful for the estimation of spectra if it is combined with the use of the method of fast Fourier transform.

7. Review of spectrum windows

In the estimation of spectra it has been a common practice to multiply a sample (auto or cross) covariance function $C(t)$ with a function $G(t)$, which is called a lag window [1], [5] and to compute the Fourier transform of $G(t)C(t)$ to get an estimate of the spectrum. Though the situation may be changing due to the introduction of the method of fast Fourier transform [8] it would be of some interest to review the performance characteristics of various lag windows hitherto proposed for practical use from our present point of view of the design of low pass filters. Obviously the lag window $G(t)$ plays the role of $\mathfrak{F}D(f)$ in our present discussion and the spectrum window $\mathfrak{F}G(f)$ (the Fourier transform of $G(t)$) plays the role of $D(t)$ with the corresponding change of roles between t and f.

In Figure 19 some of the well-known windows, including the three windows proposed by the present author, are represented as points in (CTP, PTH)-space. Definitions of these windows are given in Table 2. It is interesting to note that the primitive lag window is situated distantly from the set of min $R(U)$ filters. This means that the ideal rectangular low pass filter is not highly evaluated from our present point of view. Also Bartlett's window is distant from min $R(U)$ filters and we can see a fairly significant improvement of performance in other windows especially designed for the estimation of spectra.

Among the windows W_1, W_2 and W_3, which are proposed by the present author [1], [2], W_2 and W_3, which were designed so as to be bias free for locally quadratic variation of the spectrum, are fairly closely located to the points which correspond to min CTH filters. These windows were once criticized by E. Parzen [7] of having negative side lobes; but so long as the present procedure of multiplication by lag window can be considered as a low pass filtering in frequency domain, our present result suggests that these widows still have their reason of existence, especially when we recall the fact that in this case to widen the range of spread of the low pass filter or the spectrum window does not mean any practical limitation of the window, rather it is profitably realized by simply decreasing the maximum number of lags in the computation of sample covariance functions. This observation is in good agreement with the experience of applying these windows to practical problems, and their efficient use for the estimation of power spectral densities is discussed in a paper by the present author [3]. The content of the paper also suggests a new direction of the application of results of the present paper for the detection of bias introduced by the application of low pass filters.

It should be mentioned here that by the introduction of the method of fast Fourier transform it has become practicable to compute numerical Fourier transform of an observed time series [8], and we can formulate the problem of spectrum estimation directly as a low pass filtering problem in the frequency domain. Thus the filters designed in this paper are readily applicable as spectrum windows for this purpose. The negative side lobes of some of the filters may cause some trouble, but this will be restricted to the frequency range where the spectrum shows very low values. To avoid the situation some kind of prewhitening or taking the logarithm of unsmoothed spectrum will be useful. In this situation it may also be useful to apply a properly selected data window (c.f. [8] and the next section of this paper). By taking logarithms, we shall generally have to expect larger sampling variations of our estimate, but will be able to reduce the amount of bias in the low power frequency range. Also the negative side lobes of spectrum windows does not cause a direct annoyance in this case. In this approach we could use the boxcar or Daniell window. It has $\mathrm{PTH}=\frac{\pi^2}{6}$ ($\doteqdot 1.645$) and $\mathrm{CTP}=\infty$, and is not highly evaluated from our present point of view.

During the proofreading of the present paper, Professor Parzen suggested the introduction of a new lag window $G(t)$ which is by definition:

$$G(t)=(1+6t^2)G_P(t)\,,$$

where $G_P(t)$ is the original PARZEN window given in Table 2. The present window is the lowest degree modification of the original window into the one which has vanishing second order derivative at $t=0$ and thus will be bias free for locally quadratic variations of spectra.

It turned out that this window, which we shall denote by PARZEN 2, is situated very close to min CTH ($N=1$) in (CTP, PTH)-space of Figure 19. This gives another evidence of the fact that the requirement of distortion free characteristic for a locally quadratic input is playing a definite role in shifting our attention from those windows around HANNING and HAMMING to those around min CTH in Figure 19. It is quite possible that PARZEN 2 will find interesting applications in the estimation of spectra, especially in the line of approach described in [3], and thus will give another proof of the usefulness of our (CTP, PTH)-space representation of filters.

8. Application to data window design

It has been suggested by J. W. Tukey [8] that to avoid an annoying effect due to the sudden truncation of observation of a time series it is desirable to attenuate both ends of a set of data before taking its

Fourier transform. This is necessary when there exists a significantly periodic component in the series. The procedure is realized by multiplying the time series $X(t)$ by a function $G(t)$, which is called a data window, and taking its Fourier transform.

The present author proposed ([8], p. 42) the use of $G(t)$ which can be decomposed into the form,

$$G(t)=\int_{-\infty}^{\infty} V(t-s)U(s)ds$$

where $U(t)=1$ for $|t|\leqq M+\frac{1}{2}L$ and $=0$ otherwise, and $V(-t)=V(t), V(t)$ $=0$ for $|t|>\frac{L}{2}$ and $\int_{-\infty}^{\infty} V(t)dt=1$. It is assumed that M and L are positive and $M\geqq 2L$. It turns out that for a data window of this type to be a good one the Fourier transform $\mathfrak{F}V(f)$ of $V(t)$ should be fairly equal to 1 for some given range of f around $f=0$ and then rapidly go to zero outside the range. This is quite similar to the requirement that $V(t)$ should be the impulse response function of a good low pass filter in our sense. In the present case it is further required that L should be kept as small as possible, and consequently, $V(t)$ should be spread over $[-L, L]$ as well as possible. This means that in the desigh of this $V(t)$, the computational efficiency in the sense of the former section should be taken into account. Thus it seems quite possible that the filters LN1 and LN2 obtained in that section may provide good candidates for this use.

We shall here content ourselves by only suggesting the possibility of this kind of use of our present design procedure and leave the detailed discussion for a future occasion.

Acknowledgement

The author withes to express his hearty thanks to Professor E. Parzen for his encouragement and helpful discussion of the present study.

Extensive use of on-line facilities of ACME (Advanced Computer for Medical Research) of Stanford Medical School was made, without which the present work would not have been finished.

Stimula to this study were rooted in the discussions at the Committee on Random Vibration Analysis of Japan Mechanical Engineers Society and in the seminar on non-stationary time series at Princeton University. The author's thanks are due to Professor H. Sato of the Institute of Industrial Science, University of Tokyo, and to Professor J. W. Tukey of Princeton University for directing his attention to this problem.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] H. Akaike, "On the design of lag windows for the estimation of spectra," *Ann. Inst. Statist. Math.*, 14 (1962), 1-21.

[2] H. Akaike, "Statistical measurement of frequency response functions," *Ann. Inst. Statist. Math.*, Supplement III (1964), 5-17.

[3] H. Akaike, "Note on the use of an index of bias in the estimation of power spectra," *Ann. Inst. Statist. Math.*, 20 (1968), 55-69.

[4] H. Akaike, "On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method," *Ann. Inst. Statist. Math.*, 11 (1959), 1-16.

[5] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communications engineering," *Bell System Technical Journal*, 37 (1958), 185-282, 485-569, (also published separately by Dover (1959)).

[6] E. Parzen, "Mathematical considerations in the estimation of spectra, *Technometrics*, 3 (1961), 167-190.

[7] E. Parzen, "On empirical multiple time series analysis," *Proceedings of the Fifth Berkeley Symposium*, 1 (1967), 305-340.

[8] J. W. Tukey, "An introduction to the calculations of numerical spectrum analysis," Paper in *Spectral Analysis of Time Series* (edited by Bernard Harris), New York: John Wiley and Sons, Inc., (1967), 25-46.

Table 1. Key to the figures

$$D(t)=\frac{1}{2L}\left(1+2\sum_{n=1}^{K} a_n \cos\left(2\pi\frac{n}{2L}t\right)\right) \qquad \text{for } |t|\leqq L$$
$$=0 \qquad \text{for } |t|>L$$

N:

$$N=1 \longleftrightarrow D(L)=0$$
$$N=2 \longleftrightarrow D(L)=0 \quad \text{and} \quad \int_{-L}^{L} t^2 D(t)dt=0$$
$$N=0 \longleftrightarrow D(L)\neq 0$$

$$\mathfrak{F}D(f)=\int_{-\infty}^{\infty} \exp(-i2\pi f t)D(t)dt$$

$$\text{HWIDTH}=\frac{2}{\int_{-\infty}^{\infty}\left|\frac{1-\mathfrak{F}D(f)}{f}\right|^2 df}$$

$$P=\int_{-\infty}^{\infty}|\mathfrak{F}D(f)|^2 df$$

$$\text{CUTOFF}=\sqrt{\frac{3\int_{-\infty}^{\infty}|f\mathfrak{F}D(f)|^2 df}{P}}$$

CTP=CUTOFF/P

PTH=P/HWIDTH

CTH=CUTOFF/HWIDTH

RA=(1/2L)(1/HWIDTH)

$R(U)=U(\text{CTP})^2+(1-U)(\text{PTH})^2$

Table 2. Specifications of lag windows

PRIMITIVE	$G(t)=1$		for $\|t\|\leqq 1$
	$=0$		otherwise
BARTLETT	$G(t)=1-\|t\|$		for $\|t\|\leqq 1$
	$=0$		otherwise
PARZEN	$G(t)=1-6t^2+6\|t\|^3$		for $\|t\|\leqq 1/2$
	$=2(1-\|t\|)^3$		for $1/2\leqq\|t\|\leqq 1$
	$=0$		otherwise
GAUSS	$G(t)=\exp(-t^2)$		
TRIGONOMETRIC	$G(t)=a_0+2\sum_{n=1}^{k} a_n\cos(n\pi t)$		for $\|t\|\leqq 1$
	$=0$		otherwise
HANNING	$a_0=0.5$,	$a_1=0.25$,	$K=1$
HAMMING	$a_0=0.54$,	$a_1=0.23$,	$K=1$
W_1	$a_0=0.5132$,	$a_1=0.2434$,	$K=1$
W_2	$a_0=0.6398$,	$a_1=0.2401$,	
	$a_2=-0.0600$		$K=2$
W_3	$a_0=0.7029$,	$a_1=0.2228$,	
	$a_2=-0.0891$,	$a_3=0.0149$,	$K=3$

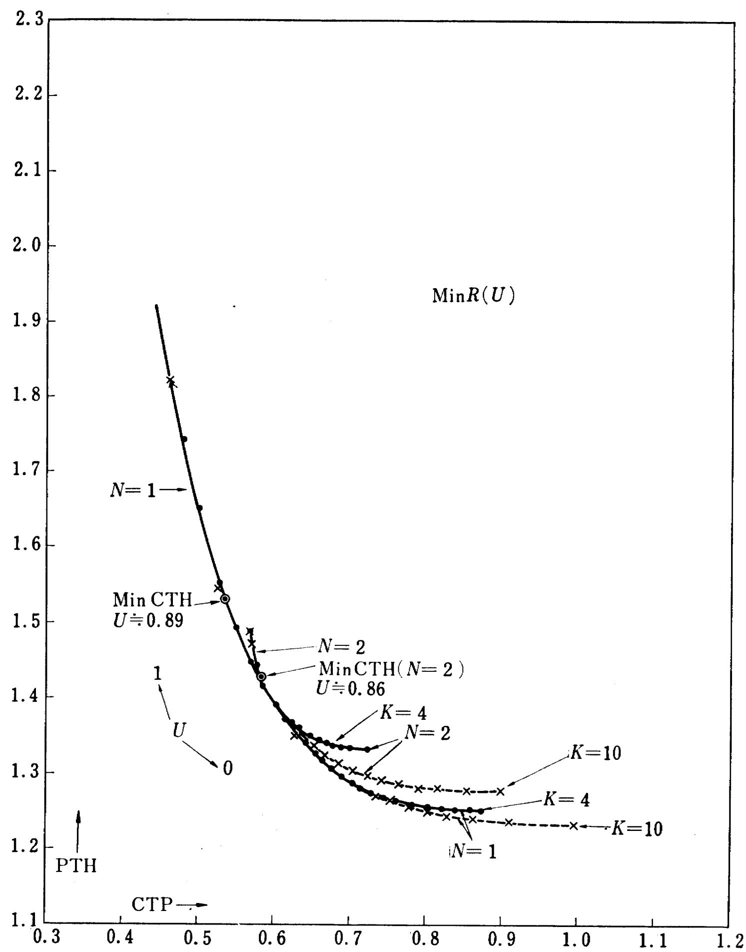


Fig. 1

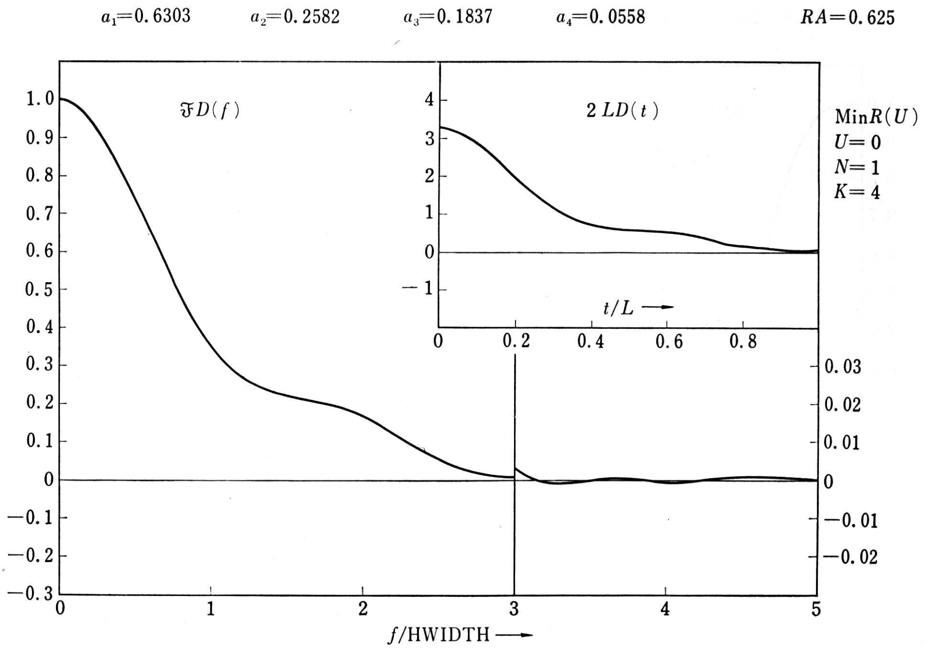

a_1=0.6303
a_2=0.2582
a_3=0.1837
a_4=0.0558
RA=0.625
𝔉D(f)
2 LD(t)
Min R(U)
U= 0
N= 1
K= 4
t/L
f/HWIDTH

Fig. 2

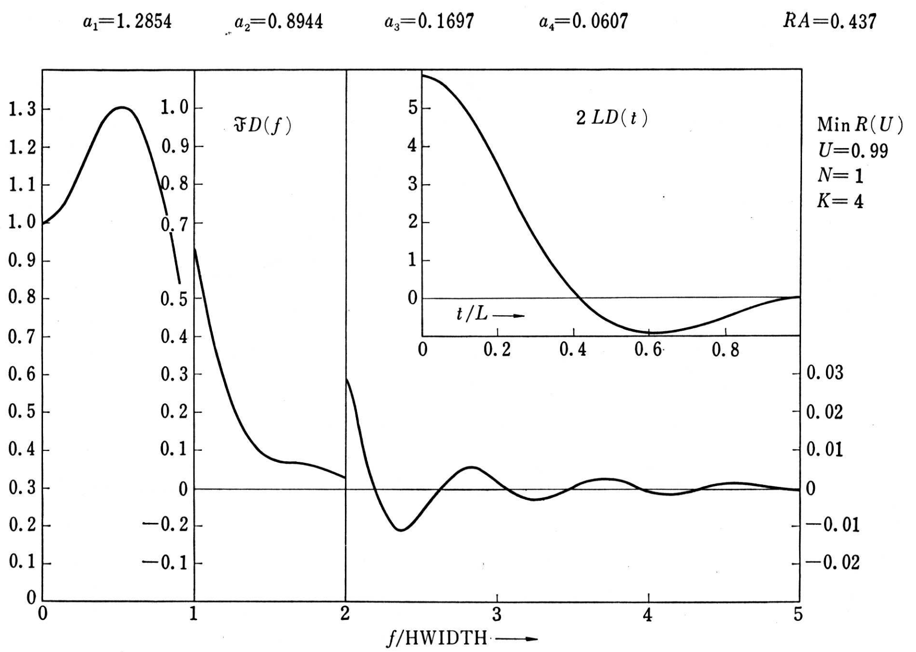

a_1=1.2854
a_2=0.8944
a_3=0.1697
a_4=0.0607
RA=0.437
𝔉D(f)
2 LD(t)
Min R(U)
U=0.99
N= 1
K= 4
t/L
f/HWIDTH

Fig. 3

$a_1=0.9267$ $a_2=0.5024$ $a_3=0.3424$ $a_4=0.2668$ $RA=0.370$

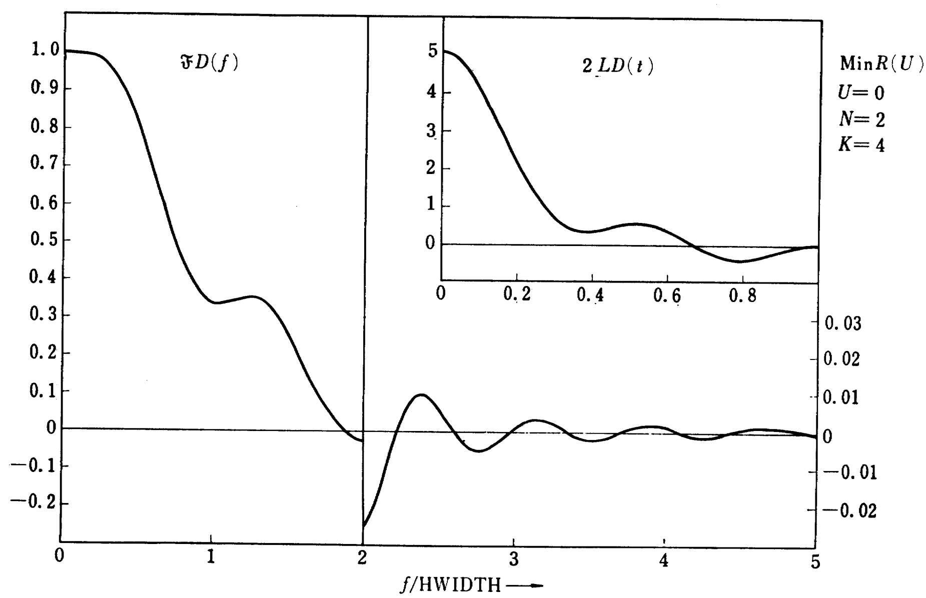


Fig. 4

$a_1=0.9333$ $a_2=0.4493$ $a_3=0.0102$ $a_4=-0.0059$ $RA=0.474$

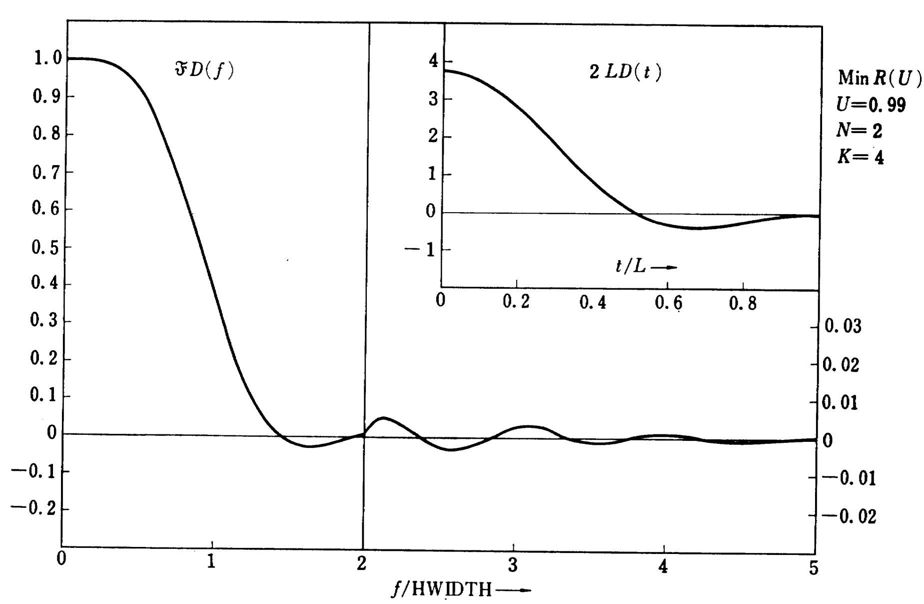


Fig. 5

$a_1=1.0225$ $a_2=0.6146$ $a_3=0.1613$ $a_4=0.0691$ $RA=0.391$

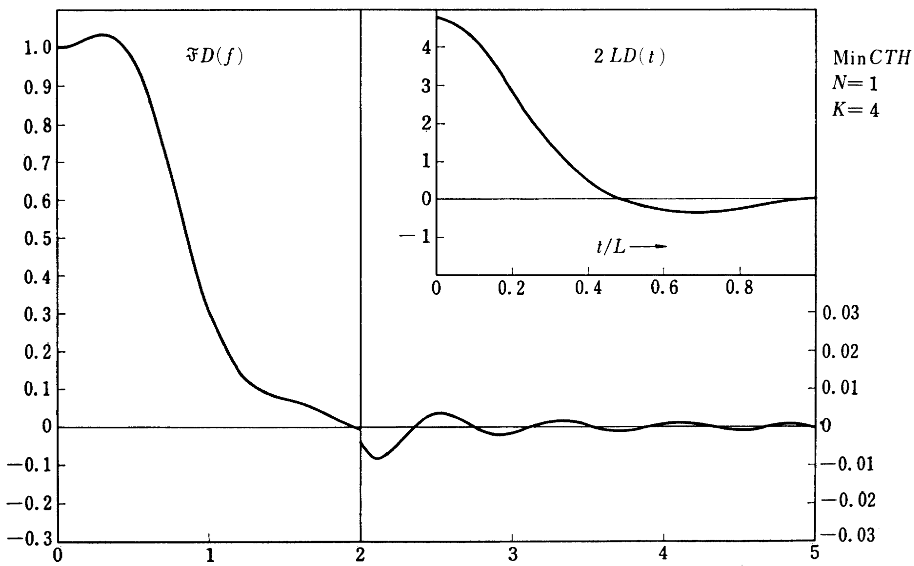


Fig. 6

$a_1=1.0303$ $a_2=0.6600$ $a_3=0.1990$ $a_4=0.0768$ $a_5=0.0255$ $RA=0.372$

$a_6=0.0166$ $a_7=0.0056$ $a_8=0.0058$ $a_9=0.0016$ $a_{10}=0.0027$

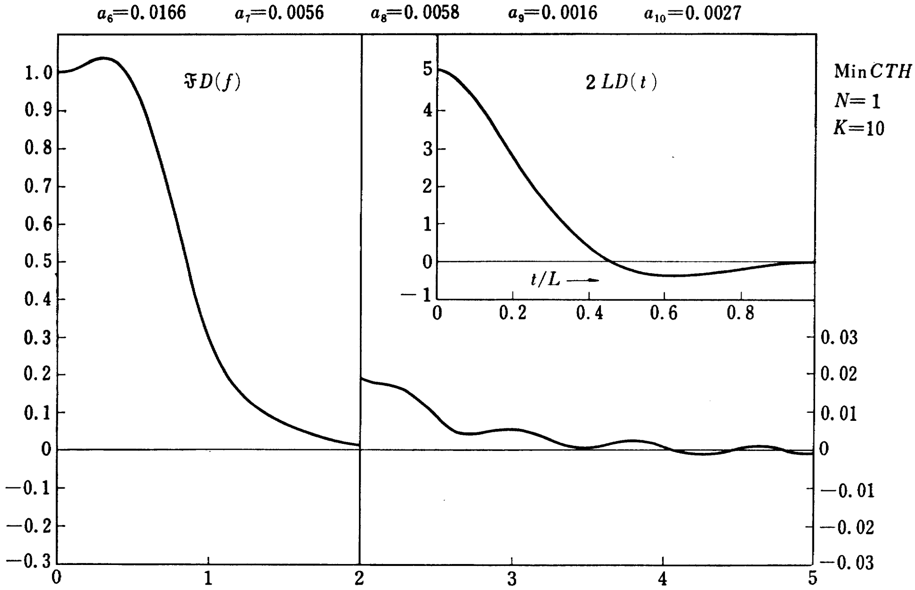


Fig. 7

$a_1=0.9392$ $a_2=0.5039$ $a_3=0.1067$ $a_4=0.0421$ $RA=0.433$

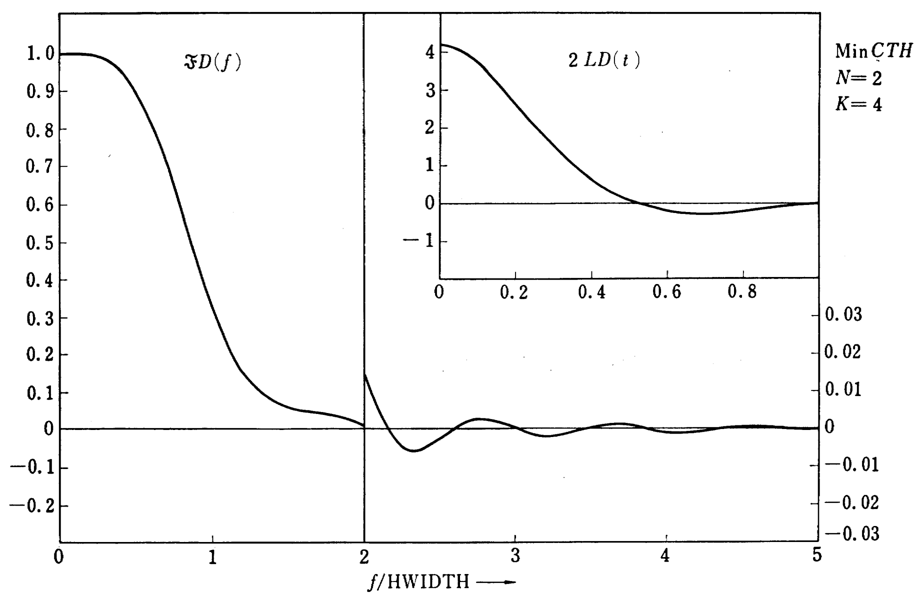
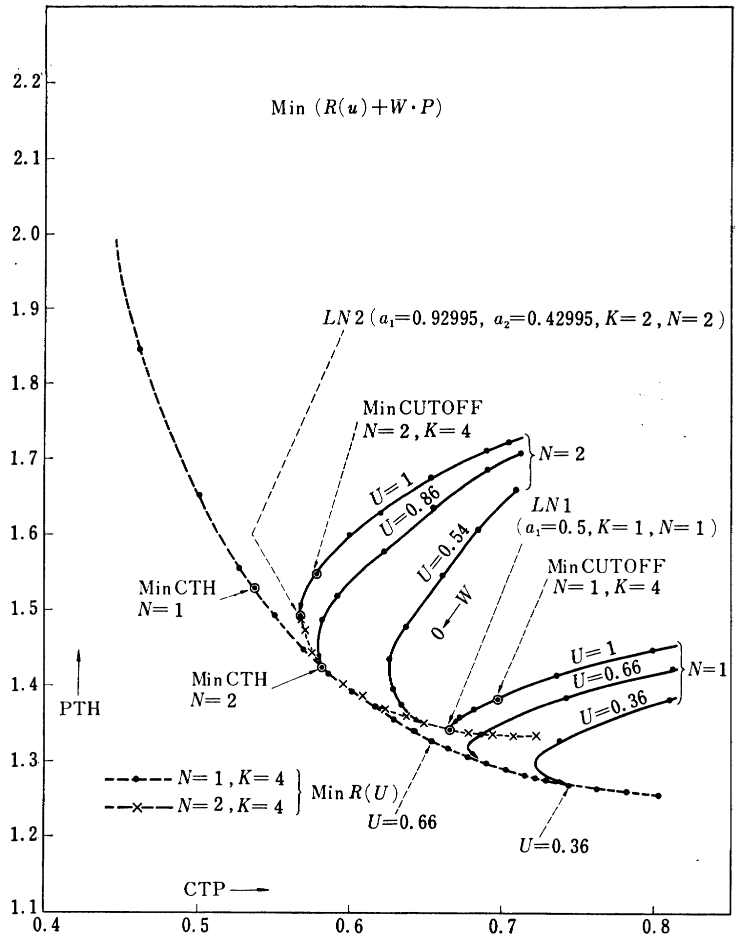
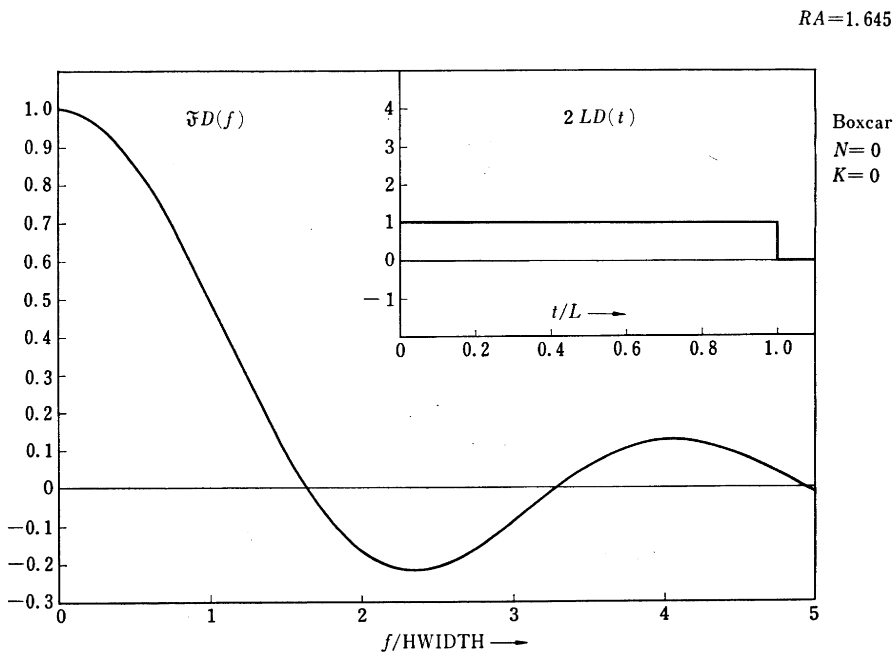


Fig. 8



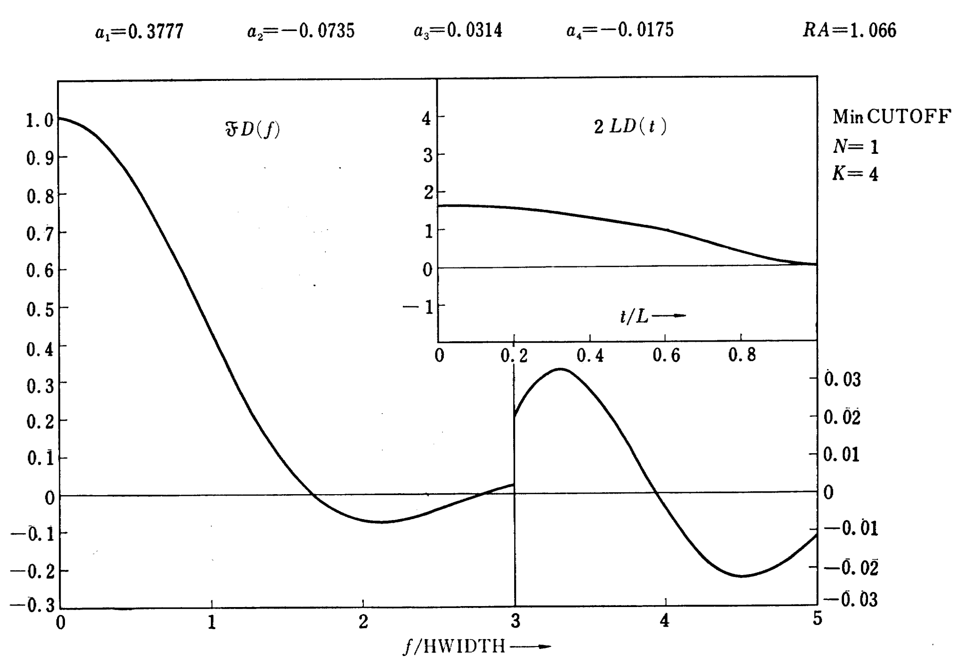
Min (R(u)+W·P)
LN 2 (a1=0.92995, a2=0.42995, K=2, N=2)
Min CUTOFF
N=2, K=4
N=2
U=1
U=0.86
U=0.54
LN 1
(a1=0.5, K=1, N=1)
Min CUTOFF
N=1, K=4
Min CTH
N=1
0←W
U=1
U=0.66
N=1
U=0.36
Min CTH
N=2
PTH
N=1, K=4
N=2, K=4
Min R(U)
U=0.66
U=0.36
CTP
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
0.4
0.5
0.6
0.7
0.8

Fig. 9



RA=1.645
𝔉D(f)
2 LD(t)
t/L
Boxcar
N=0
K=0
f/HWIDTH

Fig. 10



a1=0.3777
a2=−0.0735
a3=0.0314
a4=−0.0175
RA=1.066
𝔉D(f)
2 LD(t)
t/L
Min CUTOFF
N=1
K=4
f/HWIDTH

Fig. 11

a_1=0.5 RA=0.895

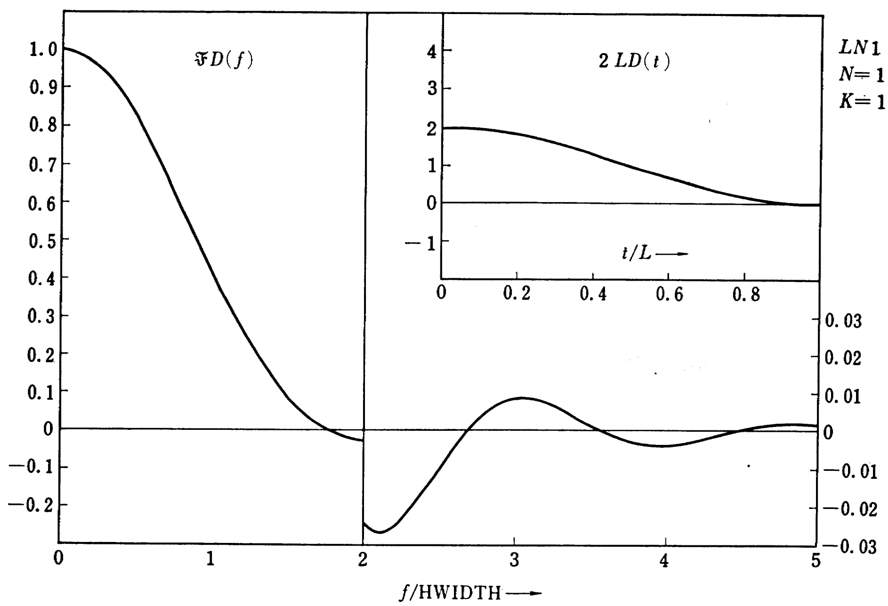


Fig. 12

a_1=1.0220 a_2=0.6142 a_3=0.1609 a_4=0.0688 RA=0.391

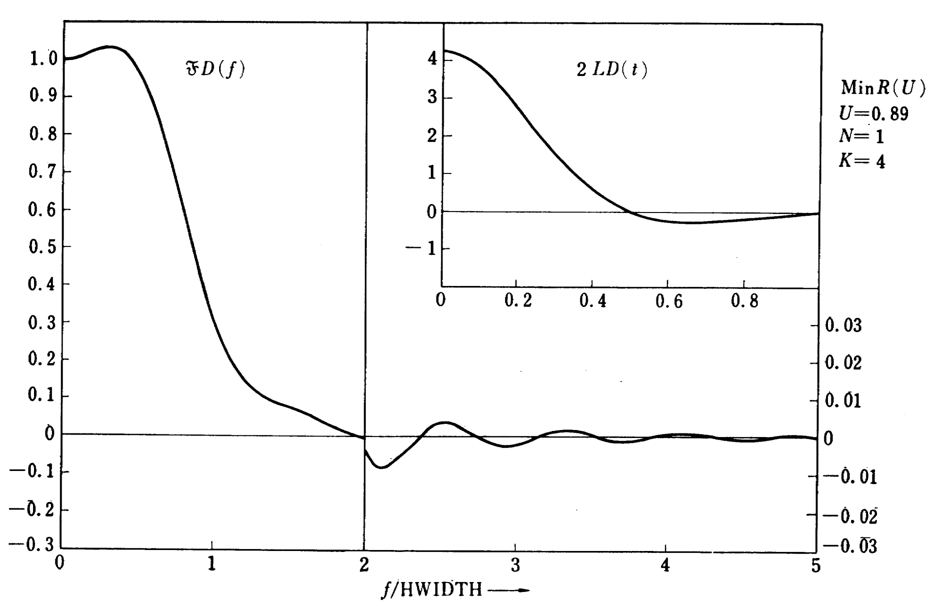


Fig. 13

a_1=0.9728 a_2=0.5553 a_3=0.1496 a_4=0.0671 RA=0.406

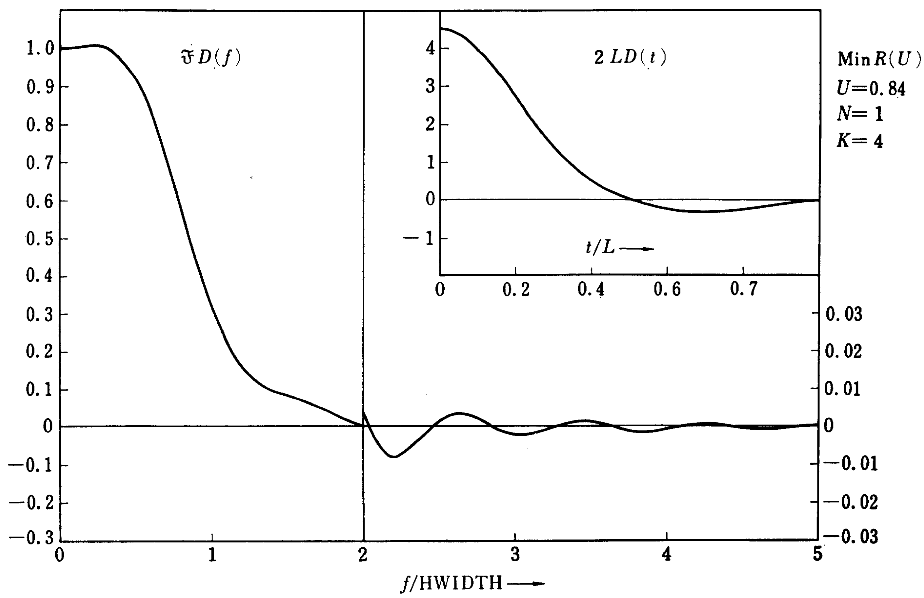


Fig. 14

a_1=0.8999 a_2=0.2556 a_3=−0.0919 a_4=0.0521 RA=0.559

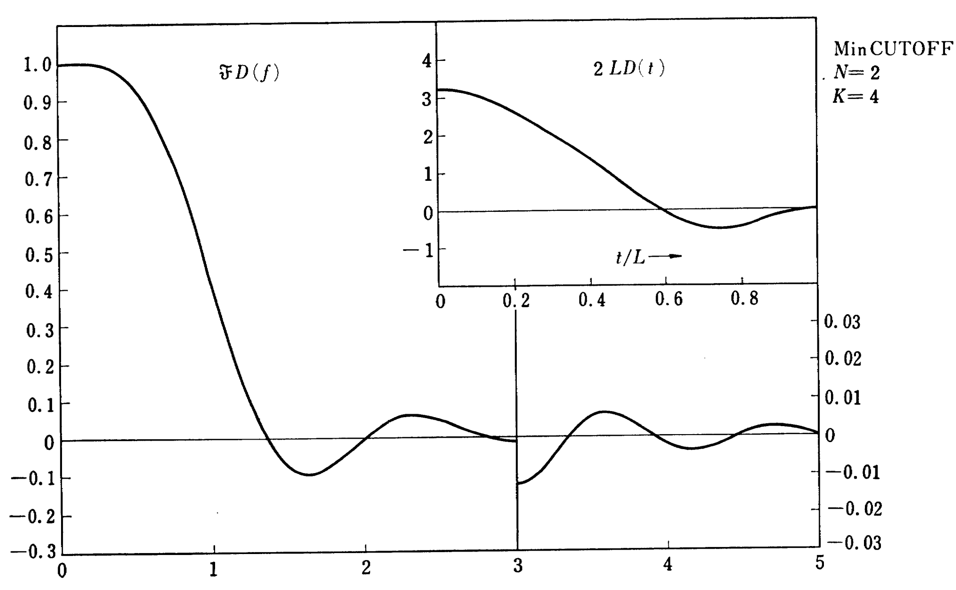


Fig. 15

$a_1=0.92995$ $a_2=0.42995$ $RA=0.481$

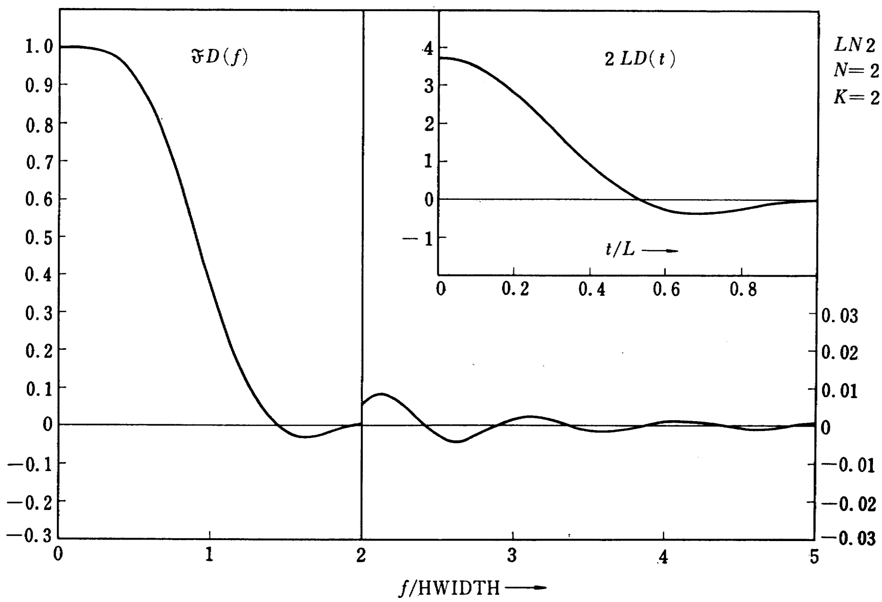


Fig. 16

$a_1=0.9425$ $a_2=0.5241$ $a_3=0.1212$ $a_4=0.0396$ $RA=0.425$

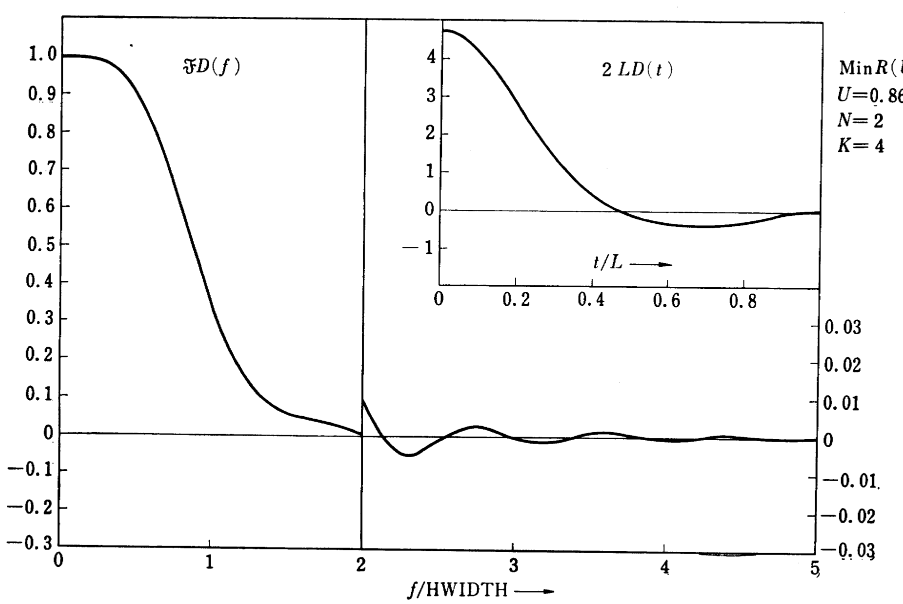


Fig. 17

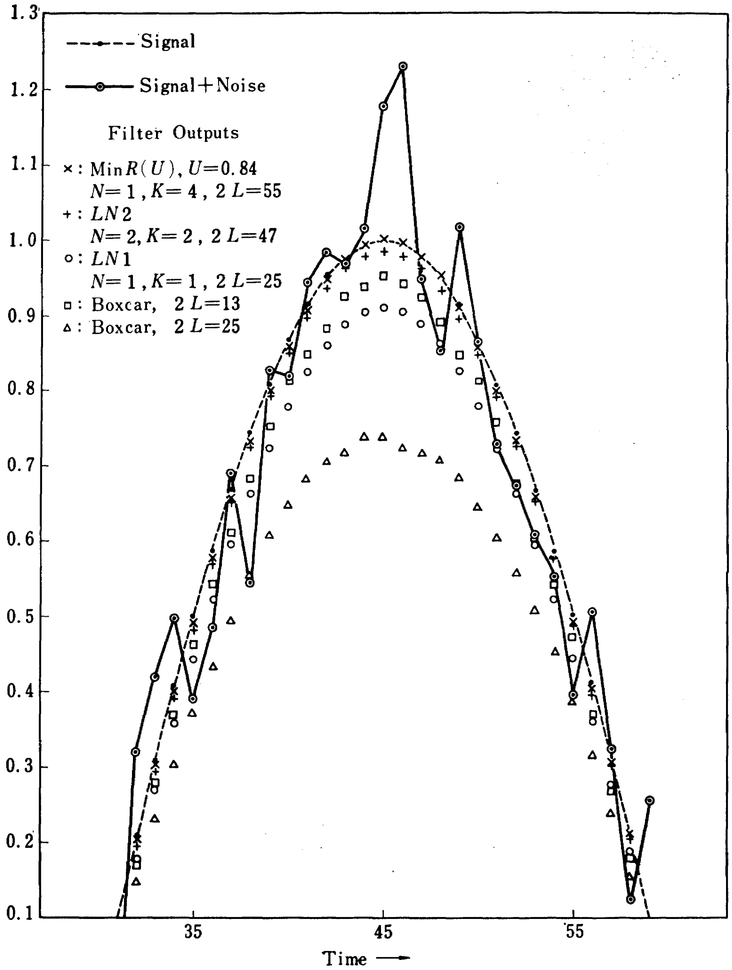

Signal
Signal+Noise
Filter Outputs
×: Min R(U), U=0.84
N=1, K=4, 2L=55
+: LN2
N=2, K=2, 2L=47
○: LN1
N=1, K=1, 2L=25
□: Boxcar, 2L=13
△: Boxcar, 2L=25
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
35
45
55
Time

Fig. 18

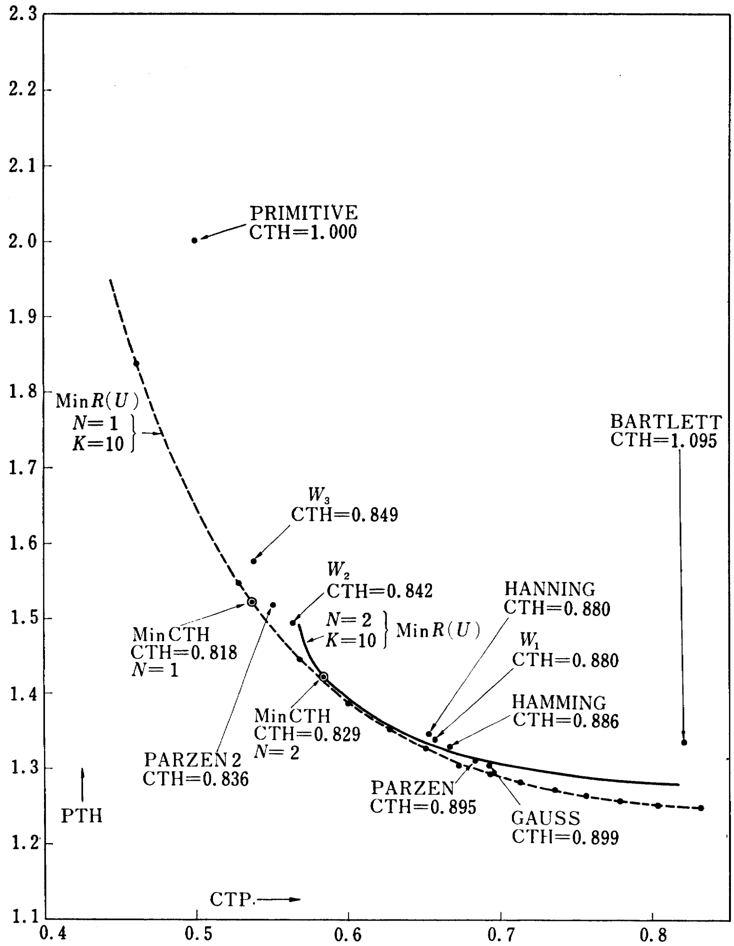

2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
0.4
0.5
0.6
0.7
0.8
PRIMITIVE
CTH=1.000
Min R(U)
N=1
K=10
BARTLETT
CTH=1.095
W_3
CTH=0.849
W_2
CTH=0.842
N=2
K=10
Min R(U)
HANNING
CTH=0.880
W_1
CTH=0.880
HAMMING
CTH=0.886
Min CTH
CTH=0.818
N=1
Min CTH
CTH=0.829
N=2
PARZEN 2
CTH=0.836
PARZEN
CTH=0.895
GAUSS
CTH=0.899
PTH
CTP

Fig. 19