Threshold Analysis for VLBI Delay and Doppler

J. W. Layland
Communications Systems Research Section

This article considers the problem of estimating the delay difference and doppler frequency difference for the arrival of a white noise signal from a distant Quasar at two widely separated receiving stations on Earth. The nonlinear Barankin bound is reviewed as it applies to the very long baseline interferometry (VLBI) problem, and used to evaluate the signal-to-noise ratio threshold for VLBI Estimates. We conclude by comparing the performance of several sampling strategies for VLBI data.

I. Introduction

This article considers the problem of estimating the delay difference and doppler frequency difference for the arrival of a white noise signal from a distant Quasar at two widely separated receiving stations on Earth. At both receiving stations, the signal is corrupted by additive white Gaussian noise, and further, by filtering, and perhaps other processing. This additive noise is statistically independent of the signal, and the noises at the two stations are independent of each other. Our parameter estimator is that parameter transformation which maximizes the cross correlation between the corrupted signals received at the two stations. If the signal-to-noise ratio (SNR) in this estimator is sufficiently high, the maximizing parameters are well-defined, and the calculation of estimator variance is quasi-linear. As the estimator SNR decreases, the problem becomes distinctly nonlinear, and bounds or approximations are required. The SNR at which the nonlinear effects become important is called the threshold of the system.

The bounds derived by Barankin (Ref. 1) for estimation error under a wide variety of conditions apply here. These bounds include the quasi-linear Cramer-Rao bound as a special case. Swerling (Ref. 2) has applied the Barankin bounds to the variance of the estimate of time delay and doppler shift for a radar signal, and McAulay and Seidman (Ref. 3) have applied it to a threshold analysis for pulse-position modulation. The VLBI problem can be transformed in a systematic way to the problem of estimating the parameters of a "known" wideband Gaussian signal in wideband Gaussian noise, so that much of this above-cited work can be directly applied.

II. The Barankin Bound

For the approximate application of the Barankin bound to the VLBI problem, it is convenient to follow the notation and arguments of Swerling (Ref. 2). A more rigorous and exact formulation could be built from the formulations used by Barankin (Ref. 1) or Kiefer (Ref. 4).

Let $\{F(t,\xi)\}$ be a family of real-valued functions of time t and defining parameter vector ξ. We wish to estimate some of the components of ξ by observing functions $Z(t)$ over $T_1 \leq t \leq T_2$ where

$$Z(t) = F(t,\xi) + x(t) \tag{1}$$

and where $x(t)$ is a real-valued Gaussian random process with zero-mean and covariance function $\psi_x(s,t)$. Given appropriate conditions, $x(t)$ can be expanded in the Karhunen-Loueve Series

$$x(t) = \sum_{i=1}^{\infty} \frac{x_i}{\sqrt{\chi_i}} \phi_i(t) \tag{2}$$

where the x_i are independent, zero-mean, unit-variance Gaussian random variables, and χ_i, $\phi_i(t)$ are the eigenvalues and orthonormal eigenfunctions of the integral equations

$$\phi(s) = \chi \int_{T_1}^{T_2} \psi_x(s,t)\, \phi(t)\, dt \tag{3}$$

We further suppose that for all ξ, we can express

$$F(t,\xi) = \sum_{i=1}^{\infty} \beta_i(\xi) \cdot \phi_i(t) \tag{4}$$

over the observation interval, where

$$\beta_i(\xi) = \int_{T_1}^{T_2} F(t,\xi)\, \phi_i(t)\, dt \tag{5}$$

Barankin's result for the greatest lower bound to the variance of any estimator for ξ is

$$\sigma^2 \geq \frac{\left[\sum_{K=1}^{K} (\xi'_k - \xi_0) \cdot a_k\right]^2}{\sum_{k=1}^{K} \sum_{l=1}^{K} G(\xi'_k, \xi''_l \,|\, \xi_0) \cdot a_k a_l} \tag{6}$$

where ξ_0 is the true parameter value, K is any finite integer, the ξ'_k are alternate parameter values, and the a_k are arbitrary real numbers. In the above notation, the function $G(\cdot,\cdot|\cdot)$ is defined by

$$G(\xi',\xi''|\xi_0) = \exp\left\{\sum_{i=1}^{\infty} \chi_i \left[\beta_i(\xi') - \beta_i(\xi_0)\right]\left[\beta_i(\xi'') - \beta_i(\xi_0)\right]\right\} \tag{7}$$

This bound may be made arbitrarily tight by appropriate choices of K and $\{a_k, \xi'_k\}$.

The Cramer-Rao bound can be derived from Eq. (6) by taking a limit as two of the ξ'_k smoothly approach each other at ξ_0, and their corresponding a_k grow equally and oppositely to infinity. The resulting zeroth-index derivative term can be included in the bound (6) as (Ref. 3)

$$\sigma^2 \geq \frac{\left[a_0 + \sum_{k=1}^{\infty} a_k(\xi'_k - \xi_0)\right]^2}{\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_k a_l\, d_{k,l}} \tag{8}$$

where

$$d_{00} = \frac{\partial}{\partial \xi'} \frac{\partial}{\partial \xi''} G(\xi',\xi''|\xi_0)\big|_{\xi'=\xi''=\xi_0}$$

$$d_{0k} = d_{k0} = \frac{\partial}{\partial \xi'} G(\xi', \xi''_k \,|\, \xi_0)\big|_{\xi'=\xi_0, k\epsilon[1,K]}$$

$$d_{k,l} = G(\xi'_k, \xi''|\xi_0)\big|_{k,l\epsilon[1,K]} \tag{9}$$

Equation 8 is the Cramer-Rao Bound if $a_i = 0$, for all $i > 0$. Continuing to follow Ref. 3, let A denote the column vector of the a_i, and N denote the column vector of the n_i defined by $n_0 = 1$, $n_i = \xi'_i - \xi_0$ for $i\epsilon[1,K]$. If D denotes the matrix of $\{d_{kl} | k, l\epsilon 0, K\}$, then the bound may be rewritten as

$$\sigma^2 \geq A^t N N^t A / A^t D A \tag{10}$$

The right-hand-side of Eq. (10) is maximized by letting $A = \lambda D^{-1} N$, for any λ, so that the greatest lower bound on the variance is

$$\sigma^2 \geq N^t D^{-1} N \tag{11}$$

We will apply this form of the Barankin Bound to the VLBI estimation problem.

III. Very Long Baseline Interferometry Detection and Parameter Estimation

Figure 1 shows an overview of the features of the VLBI problem. The white noise signal from the radio star is received at the two stations with some relative delay and doppler offset which can be related to Earth's rotational position and velocity. Local ambient noise, usually much stronger than the radio-star signal, is added at each receiver, and the resultant is filtered, translated by a local oscillator to a manageable frequency and bandwidth, and then sampled, hard-clipped, and recorded. The recorded signals are then brought together at leisure, modified by an a priori estimate of the relative delay and doppler, and cross-correlated. Refined estimates of the delay and doppler are extracted from this cross-correlation.

The input to the cross-correlator from station 1 may be denoted as $x_{j+k} = \text{Sign}\{S(j+k,\hat{f}) + n_1(j+k,\hat{f})\}$, and the input from station 2 may be denoted $y_j = \text{Sign}\{S^*(j+d,f_0) + n_2(j,f_0)\}$. Here $S^*(\cdot,\cdot)$ is the naturally delayed and doppler-shifted version of $S(\cdot,\cdot)$, and k and $\hat{f}$ are our a priori estimates of this delay and doppler, respectively. Because of their origins as white noise, and subsequent filtering, $S(\cdot,\cdot)$ and $n_1(\cdot,\cdot)$ have identical autocorrelation functions, as do $S^*(\cdot,\cdot)$, and $n_2(\cdot,\cdot)$.The cross-correlation

$$Z(k,f) = \sum_j x_{j+k} \cdot y_j \tag{12}$$

is evaluated as a function of k and f near its maximum. The final estimate of the actual delay and doppler is made much finer than the integral steps in k.

Let us denote by T_{n1} and T_{n2} the noise temperatures of the environment at stations 1 and 2, respectively, and by T_{s1} and T_{s2} the increment to these noise temperatures caused by the radio-star "signal" at the two stations. T_{s1} and T_{s2} may differ as a result of differing antenna collecting areas, efficiencies, etc. Denote as r_j and m_j the signal and noise terms respectively in x_j, and as r'_j and m'_j the signal and noise terms in y_j. By definition of the hard limiting, the average values of $(r_j + m_j)^2 = (x_j)^2 = 1$ and $(r'_j + m'_j)^2 = (y_j)^2 = 1$. Furthermore, $E_r\{(E_m\{r_j + m_j\})^2\} = (2/\pi)(T_{s1}/T_{n1})$, and $E_r\{(E_m\{r'_j + m'_j\})^2\} = (2/\pi)(T_{s2}/T_{n2})$, where E_α denotes expected value over the distribution of the random variable α, and where we assume $T_{si} \ll T_{ni}$.

Because the number of samples in the summation (12) is large, $Z(\cdot,\cdot)$ will be approximately Gaussian (by the central limit theorem), and its distribution specified by its first two moments. The first moment is

$$\begin{aligned}\overline{Z(k,f)} &= \sum_j \overline{x_{j+k} \cdot y_j} \\ &= \sum_j \overline{r_{j+k} \cdot r'_j} \\ &= N \cdot \frac{2}{\pi}\sqrt{\frac{T_{s1}}{T_{n1}} \cdot \frac{T_{s2}}{T_{n2}}} \cdot \phi(k-d, \hat{f} - f_0)\end{aligned} \tag{13}$$

where N is the number of samples, and $\phi(\cdot,\cdot)$ is the normalized autocorrelation (ambiguity) function for the filtered signal process.

The second moment is

$$\overline{Z(k,f) \cdot Z(\ell,g)} = \sum_j \sum_i \overline{[r_{j+k} + m_{j+k}][r'_j + m'_j][r_{i+\ell} + m_{i+\ell}][r'_i + m'_i]\}} \tag{14}$$

If we expand the right hand side of Eq. (14), and drop those terms with zero expected values,

$$\begin{aligned}\overline{Z\{k,f) \cdot Z(\ell,g)} = \sum_j \sum_i \{&\overline{r_{j+k} \cdot r_{i+\ell} \cdot r'_j \cdot r'_i} \\ &+ \overline{m_{j+k} \cdot m_{i+\ell}} \cdot \overline{r'_j \cdot r'_i} \\ &+ \overline{r_{j+k} \cdot r_{i+\ell}} \cdot \overline{m'_j \cdot m'_i} \\ &+ \overline{m_{j+k} \cdot m_{i+\ell}} \cdot \overline{m'_j \cdot m'_i}\}\end{aligned} \tag{15}$$

In order to get the most information possible on a per-sample basis, the sampling is performed at a rate such that $\overline{x_j x_k} \approx 0$ and $\overline{y_j y_k} \approx 0$ for $j \neq k$. Hence at the sample points

$$\begin{aligned}\overline{m_{j+k} \cdot m_{i+\ell}} &= \beta \cdot \overline{r_{j+k} \cdot r_{i+\ell}} \\ \overline{m'_j \cdot m'_i} &= \beta' \cdot \overline{r'_j \cdot r'_i}\end{aligned} \tag{16}$$

For some constants β, β' and any particular i, j, k, ℓ. We will assume that Eq. (16) holds at interpolated values also. With this assumption, Eq. (15) may be rewritten in a statistically-equivalent but simpler form:

$$\overline{Z(k,f)\,Z(\ell,g)} = \sum_j \sum_i \left\{\overline{r_{j+k} \cdot r_{i+\ell} \cdot r'_j \cdot r'_i} + \overline{m_{j+k} \cdot m_{i+\ell}} \cdot \overline{r'_j \cdot r'_i} \cdot \left(1 + \frac{\beta'}{\beta} + \beta'\right)\right\} \tag{17}$$

Equations (13) and (17) are representative of the detection of a known Gaussian signal, $\{r_i\}$ embedded in Gaussian noise. The constants β, and β' are

$$\begin{aligned}\beta &= 1\bigg/\left(\frac{2}{\pi}\frac{T_{S1}}{T_{N1}}\right) - 1 \\ \beta' &= 1\bigg/\left(\frac{2}{\pi}\frac{T_{S2}}{T_{N2}}\right) - 1\end{aligned} \tag{18}$$

On a per-sample basis, the noise-to-signal ratio for this detection problem is

$$NSR_j = \beta + \beta' + \beta \cdot \beta'$$
$$= \frac{1}{\left(\frac{2}{\pi}\right)^2 \frac{T_{S1}}{T_{N1}} \cdot \frac{T_{S2}}{T_{N2}}} - 1 \qquad (19)$$

It is convenient to consider the detection problem as if the interfering noise samples have unit variance. If it is so normalized, each of the "signal" samples has energy given by $E_s = 1/NSR_j$, or

$$E_s = \frac{\left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}}{T_{N1} \cdot T_{N2} - \left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}} \qquad (20)$$

With the VLBI parameter estimation problem thus transformed to one of parameter estimation for a known signal in noise, the formulation of the Barankin bound (Eq. 8) is directly applicable. Since the sequence of samples of the signal and noise processes at both stations are virtually independent from sample-to-sample, we take as the orthonormal basis for the signal set, the sequence of sample functions themselves. We normalize so that the interfering noise samples have unit variance, and hence so $\chi_j = 1$, for all sample indices j. The signal projections $\beta_j(\cdot)$ are a scaled version of the r_j. The number of samples, N from which the estimates of delay and doppler are made, is large, so that central limit theorem applies, and

$$\sum_j \chi_j \beta_j(\xi) \cdot \beta_j(\xi') \approx N \cdot E\{\beta_j(\xi) \cdot \beta(\xi')\} \qquad (21)$$

Or using the normalized autocorrelation function $\phi(\cdot,\cdot)$:

$$\sum_{j=1}^{N} \chi_j \beta_j(\xi) \cdot \beta_j(\xi') \simeq N \frac{\left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}}{T_{N1} T_{N2} - \left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}} \phi(\xi - \xi') \qquad (22)$$

Note that here ξ denotes a vector parameter which can include delay or doppler offsets, or both.

The $G(\cdot,\cdot|\cdot)$ function for VLBI (Eq. 7) is defined by

$$G(\xi', \xi'' | \xi^0) = \exp\{R \cdot [\phi(\xi' - \xi'') - \phi(\xi' - \xi^0) - \phi(\xi'' - \xi^0) + \phi(\xi^0 - \xi^0)]\} \qquad (23)$$

where

$$R = N \cdot \frac{\left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}}{T_{N1} \cdot T_{N2} - \left(\frac{2}{\pi}\right)^2 T_{S1} \cdot T_{S2}}$$

The D matrix of Eqs. (9) to (11) is defined from Eq. (23) as

$$D = \{d_{kl} | k, l = 0, \cdots, K\}$$
$$d_{k,l} = \exp\{R[\phi(0) - \phi(\xi'_k - \xi_0) - \phi(\xi'' - \xi_0) + \phi(\xi'_k - \xi'')]$$

for $k, l \epsilon [1, K]$ indices of alternate ξ.

$$d_{k,0} = d_{0,k} = R \cdot \frac{\partial}{\partial \xi'} [\phi(\xi' - \xi''_k) - \phi(\xi' - \xi_0)|_{\xi' = \xi_0}$$
for $k \epsilon [1, K]$

$$d_{00} = R \cdot \frac{\partial}{\partial \xi'} \frac{\partial}{\partial \xi''} \phi(\xi' - \xi'')|_{\xi' = \xi'' = \xi_0} \qquad (24)$$

Furthermore, we note that ξ_0 is presumed to be the global maximum of $\phi(\cdot)$ so that the second term of d_{0k} is trivially equal to zero. If we restrict the set of alternate ξ'_k to local maxima of $\phi(\cdot)$, the first term of d_{0k}, and hence d_{0k} itself, is zero for all k. With this restriction on D, the Barankin bound for σ^2 becomes

$$\sigma^2 \geq \frac{1}{d_{0,0}} + N_1^t D_1^{-1} N_1 \qquad (25)$$

where the subscripts 1 in the second term of Eq. (25) are meant to imply that the zero-index row and column have been dropped. The first term of Eq. (25) is the Cramer-Rao Bound; the second is the nonlinear threshold term.

IV. A Very Long Baseline Interferometry Example

The wideband quasar noise signal is processed at the two receiving stations into a form which is suitable for recording. This processing corresponds to constraints on available equipment bandwidths, recording rates, data storage capacities, etc. The basic constraint is the bandwidth of the low-noise Maser amplifier, which we will hypothesize to be 40 MHz centered at 2 or 8 GHz for present systems. This is still too wide for any but the fastest recorders, and is further reduced by filtering several narrow channels from the passband which span the full bandwidth available (bandwidth synthesis); or by sampling the full bandwidth into a buffer memory, and then recording at a slower rate once that buffer is filled (burst mode). Both sampling modes have side-lobes in the delay-estimator which arise at roughly the reciprocal of the spanned bandwidth (10^{-8} sec). There are also side-lobes in the frequency estimator at roughly the reciprocal of the recording or sampling interval (10^{-2} Hz). Burst sampling will generate additional sidelobes spaced a few Hz away from the true frequency, but these should be able to be removed by a priori knowledge, as should some of the more widely spaced side lobes in time-delay.

Bandwidth synthesis systems which have been proposed utilize two channels at the extremes of the available bandwidth, plus a third channel placed somewhere between to reduce the sidelobe amplitudes. The exact sidelobe structure is sensitive both to the position of this third channel, and to the bandwidth of the channels. For an overall spanned bandwidth of 40 MHz, a reasonably tolerable side-lobe structure is achieved if the third channel is located 5 MHz from the passband center. While this position is perhaps not optimum, it offers near optimum side-lobe suppression which is far better than that provided at many alternative positions, such as at, e.g., 6 MHz. Figure 2 shows the effect of varying the individual channel widths for a 40-MHz spanned bandwidth with the third channel located at 5 MHz. The function plotted is the aggregate autocorrelation function for the three recorded channels, which results with optimum use of all available information. The recording bandwidths are 6 MHz, 4 MHz and 2 MHz for Figs. 2a, 2b, and 2c, respectively. The corresponding Barankin lower bound to delay estimate error is shown in Fig. 3, lines B, C, and D, respectively. As could be anticipated from the sidelobe structure of Fig. 2, the threshold performance degrades seriously as the individual channel bandwidth is narrowed, and would continue to degrade if the recording bandwidth were reduced further. At Estimator signal-to-noise ratios well above threshold, the delay estimate error depends only minimally upon the individual channel bandwidths.

Figure 3 also shows the Barankin lower bound to the delay estimate error achieved by a full bandwidth 40-MHz Burst Mode System (line A). The autocorrelation function is the familiar $\sin(x)/x$. Threshold performance of the burst mode sampling is better by about 2 to 3 dB than that for bandwidth synthesis with 6-MHz channels, and better by about 10 dB than that for bandwidth synthesis with 2-MHz channels. At high SNRs, however, the delay estimate error is 2 dB worse for burst mode than for bandwidth synthesis.

The doppler estimator error has not been calculated per se, but its shape can be inferred as follows: The recording of data is approximately uniform in time over some fixed interval. Accordingly the transform of this interval is the $\sin(x)/x$ function and the doppler estimate error is identical to line A of Fig. 3 with the abscissa relabeled as appropriate: for example, if the recorded sample occupied 40 sec, then $\sigma_\tau = 10^{-9}$ sec on Fig. 3, line A is comparable to $\sigma_f = 10^{-3}$ Hz.

V. System Implications

One can draw a variety of conclusions from Figure 3, depending upon where and how he looks. Above threshold, for example, a three-fold improvement in delay estimate error requires roughly a 10-dB increase in estimator SNR, or a hundred-fold increase in the number of samples processed. This is a potentially expensive way to gain accuracy. This improvement in accuracy could perhaps be better achieved by other means, such as increasing the spanned bandwidth, if possible, and operating the system at some nominal margin above its threshold. We would like furthermore to be able to operate with a performance curve like Fig. 3A or Fig. 3B with a threshold as low as possible to minimize the amount of data needed to ensure above-threshold operation. Line 3B corresponds to a recording channel bandwidth of 15 percent of the spanned bandwidth, and for spanned bandwidths which are significantly larger than the 40 MHz considered here, it suggests that the use of a combined burst/bandwidth synthesis recording is desirable. For example, it does not appear reasonable to burst-sample a 200-MHz spanned bandwidth with present-day equipment, but 15 percent bandwidth channels, each 30 MHz wide, could be easily burst-sampled and recorded. Such operation would lead to a performance curve of the type of line B, Fig. 3, but with the delay-estimate error in the above-threshold region reduced five-fold by the increased spanned bandwidth.

References

1. Barankin, E. W., "Locally Best Unbiased Estimates," *Ann. Math. Stat.*, Vol. 20, pp. 477–501, 1949.

2. Swerling, P., "Parameter Estimation for Waveforms in Additive Gaussian Noise," *J. SIAM*, Vol. 7, pp. 154–166, 1959.

3. McAulay, R. J. and Seidman, L. P., "A Useful Form of the Barankin Lower Bound and Its Application to PPM Threshold Analysis," *IEEE Transactions on Information Theory*, Vol. IT-15, No. 2, pp. 273–279, March 1969.

4. Kiefer, J., "On Minimum Variance Estimators," *Ann. Math Stat.*, Vol. 23, pp. 627–629, 1952.

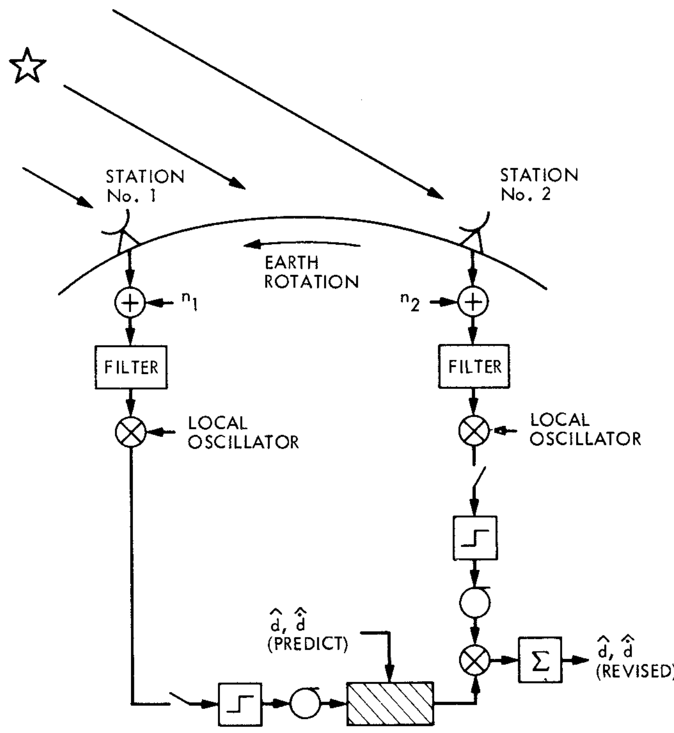

Fig. 1. Overview of VLBI receiving/processing functions

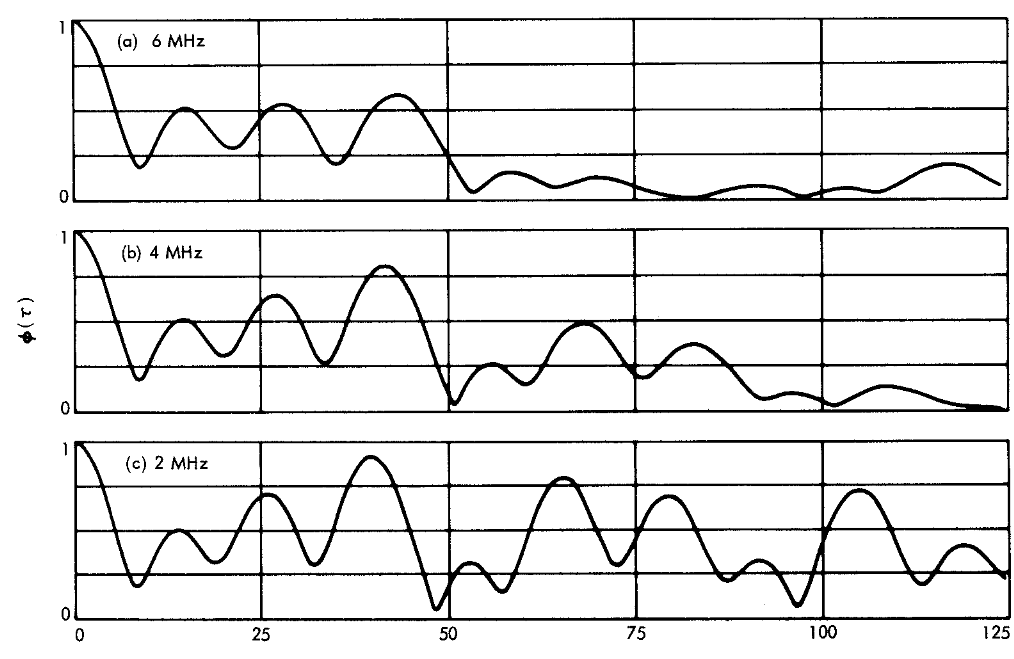

Fig. 2. Aggregate autocorrelation functions for 3-channel bandwidth synthesis with: (a) 6 MHz channels, (b) 4 MHz channels, (c) 2 MHz channels

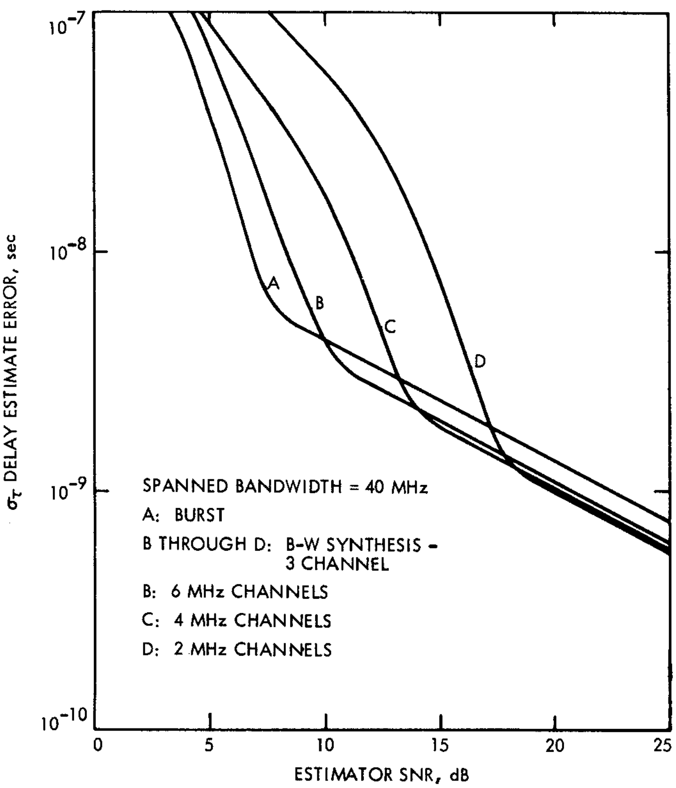

Fig. 3. Delay estimate error vs. estimator SNR in dB