DSN Station Clock Synchronization by Maximum Likelihood VLBI

W. J. Hurd
Communications Systems Research Section

The clocks at the DSN ground stations can be accurately synchronized by very long baseline interferometry (VLBI) at lower operational cost than with the existing Moon bounce system. More than an order of magnitude improvement in accuracy can be attained using existing DSN hardware, and ultimate accuracies on the order of 10 nanoseconds are possible. The purpose of the analysis described in this article is to optimize the acquisition and processing of the VLBI data subject to hardware constraints, in order to achieve the best possible time synchronization estimate for a given amount of data, and the most efficient usage of the DSN facilities.

I. Introduction

The clocks at the DSN ground stations can be synchronized by using the station antennas as very long baseline interferometers receiving the random signal from a quasar radio source. Synchronization accuracies more than an order of magnitude better than currently attained by the Moon bounce system are achievable using only existing DSN hardware. A semi-real-time system could be implemented using the TCP 920 computers for data acquisition and the Ground Communications Facility (GCF) to bring the data to a central point for processing. Operational costs should prove to be less than for the existing Moon bounce system.

The fundamental limitations on the time synchronization accuracy achievable by very long baseline interferometry (VLBI) are due primarily to the uncertainties in the station position, the radio point source position, and the propagation delays in the atmosphere. These limitations are currently on the order of 100 nanoseconds (ns), but are expected to be reduced to less than 10 ns within a few years. Hardware restrictions can, of course, place practical limitations on the accuracies which can be achieved with a given implementation. Use of the XDS 920 computer for data acquisition limits both the sampling rate and the number of samples which can be taken phase coherently, and the GCF data rates restrict the amount of data which can be handled in a reasonable length of time to a few million bits. With these restrictions, synchronization accuracies of 300–500 ns are achievable. This is compatible with the current inherent limitations of about 100 ns.

The pertinent analytical problem considered here is the optimization of the processing of the received signals in order to obtain the best possible estimation of the time of

arrival difference with a given amount of data. The amount of data required for a given accuracy of time sync is thus minimized, which is important for the following reasons:

(1) The amount of data which must be stored and buffered at the stations and then transmitted to the computation center is reduced. This is important both because of the storage requirements and because of the low data rates of the GCF.

(2) The number of computations required is proportional to the number of data samples and is thus reduced.

(3) The observation interval is shortened, which reduces the station time required for data acquisition, and reduces the stability requirements of the local oscillators, which must be phase stable over the observation time. The rubidium frequency standards now in the DSN have adequate stability for the system proposed here, but not for existing systems with much lower sampling rates and less efficient processing.

II. Summary of Results

The primary objective of this work is to optimize the filtering, sampling, and processing of the received radio signals for the estimation of time synchronization. The method of analysis also leads to quasi-optimal estimates of signal strength and fringe rate and phase, conditioned on the sample values, although different sampling strategies might be used if these were the parameters of primary interest. This is accomplished by deriving an approximate maximum likelihood estimator (MLE) for the time of arrival difference and the other parameters under appropriate assumptions as to signal statistics and mechanization limitations. Some of the key results of the analysis and the resulting estimator are:

(1) The two orthogonal components of the signal from the radio source should both be utilized. This can be accomplished by dual-channel (phase quadrature) processing at both receivers. Dual-channel processing has an inherent 6-dB advantage over single-channel processing, which appears to be the standard means of processing VLBI signals at the present time. The gain is 6 dB per sample, but only 3 dB per bit of data, since there are twice as many bits per sample. The gain arises from utilization of all of the received signal energy and therefore applies to all VLBI processing, not just to time estimation. For time synchronization, an additional 1.2-dB gain can be realized by properly skewing the sampling times in the two channels.

(2) The MLE results naturally in an optimal method for estimating time synchronization to greater accuracy than the time interval between adjacent samples. This is important because accuracy significantly better than the sampling interval is required. Previous methods for time estimation generally rely on interpolation between the sampling times using the appropriate correlation function.

(3) The MLE procedure properly accounts for the variation in time difference over the observation time. This is important because the change in the time difference can be greater than the desired synchronization accuracy.

A. Estimator SNR

An approximate MLE is derived under the assumptions that the energy emitted by the radio point source is white and gaussian, that the received signals at the two stations are sampled at the same uniform rate, and that pre-filtering assures that all samples at each ground station are independent of one another. Assuming that the local oscillator phases do not change, the approximate MLE procedure is to square law envelope detect for each possible time difference, using weighting coefficients appropriate for each time difference, and to choose the time difference which maximizes the detector output. If insufficient data can be taken while the phase remains constant, the process is repeated and the detector outputs suitably combined. If the envelope detection is performed using N data samples, and this is repeated L times, then the detector output signal-to-noise ratio (SNR) is of the form

$$R = \frac{L}{2}\frac{r}{1+\frac{1}{2r}} \tag{1}$$

where

$$r = KK_Q\rho^2 N \tag{2}$$

In the expression, ρ^2 is the product of the receiver input SNRs and there are NL total sample points at each receiver. The factor K depends on the sampling method and ranges from ⅛ for real samples (one channel at each receiver) to 0.657 for complex (quadrature phase) samples. The factor K_Q depends on the quantization, and ranges from $(2/\pi)^2$ for hard limiting to 1 for no quantization.

When the number of bits of data which can be taken is the limiting factor, then complex sampling is about

4.2 dB better than real sampling. Also, it appears best to hard limit. If the limiting factor is the sampling rate in bits per second, then hard limiting is certainly optimum, since the estimation error is inversely proportional to the sampling rate. Also, it may be better to use real rather than complex samples if the SNR is sufficiently high, but to use complex samples if the SNR is marginal.

B. Time Estimate Error

An approximate relationship for the mean square error in the time estimate is derived in *Section VIII*. The analysis has been confirmed by simulation. Results indicate that the rms error is slightly less than the time between samples divided by $R^{1/2}$.

C. Example System

A system could currently be implemented in the DSN using one 64-m-diameter antenna and one 26-m-diameter antenna, with noise temperatures of 25 K and 40 K, and using an XDS 920 for data acquisition. The computer limits the sampling rate to 500,000 data bits per second, and the total number of bits which can be acquired while the oscillator phases remain constant to about 300,000. Using hard limiting and complex sampling, the detector output SNR is about $1.2LS^2$, where S is the radio source intensity in flux units and the computer memory is filled L times. Of the radio stars which are point sources over the baseline from California to Australia, approximately ten are known with intensities of 2 flux units (fu) or greater, and four with intensities of 3 fu or greater. Time sync to significantly better than the 4-microsecond (μs) sampling interval can be achieved with the strongest four sources using only one batch of data, and with many sources using 5–10 batches of data.

III. Problem Formulation and Data Sampling

The radio energy emitted by the radio point source is essentially white and gaussian. However, because we can only observe the energy in the bandwidth of our receivers, we can consider the signal to be a narrow-band gaussian process. The signal plus noise at the outputs of the two receivers can be represented as

$$X(t) = [n(t) + s(t)]\cos(\omega_1 t + \phi_1) + [m(t) + r(t)]\sin(\omega_1 t + \phi_1) \quad (3)$$

and

$$Z(t) = [p(t) + s(t - \delta)]\cos(\omega_2 t + \phi_2) + [q(t) + r(t - \delta)]\sin(\omega_2 t + \phi_2) \quad (4)$$

where

t = time

$\delta = \delta(t)$ = difference in time delay

$\omega_1 - \omega_2$ = difference in doppler shift

ϕ_1, ϕ_2 = random phase angles

$s(t), r(t)$ = noise processes representing signal

$n(t), m(t), p(t), q(t)$ = receiver noise

All of the noise processes are assumed independent and band-limited only by the receivers. The difference frequency $\omega_1 - \omega_2$ and phase $\phi_1 - \phi_2$ are assumed to be constant over the observation time; however, the time delay $\delta(t)$ varies due to the rotation of the Earth. We can assume this to be linear and known, $\delta(t) = \delta_0 + \dot{\delta}t$. The difference frequency and phase are essentially constant only because the change in δ is small compared to the reciprocal of the difference frequency.

Suppose now that we observe $X(t)$ beginning at $t = 0$, and $Z(t)$ beginning at $t = \tau$. This time offset τ is not precisely known, because the clocks at the two stations are not precisely synchronized. We desire to form an estimate $\hat{\tau}$ of τ from the received signals, and to use this estimate to synchronize the clocks.

In order to extract the maximum information from the received signals, both the sine and cosine components of the random processes must be processed. The received signals are thus demodulated to baseband in two channels, using quadrature phase reference signals derived from rubidium frequency standards which we require to be frequency and phase stable over the observation interval. The signals are then filtered and sampled, with the filtering assuring that all samples in each channel are independent of one another. The demodulated and filtered signals, with $*$ denoting convolution, are

$$x(t) = [X(t)\cos(\omega_3 t + \phi_3)] * h_x(t) \quad (5)$$

$$y(t) = [X(t)\sin(\omega_3 t + \phi_3)] * h_y(t) \quad (6)$$

at the X receiver, and

$$z(t) = [Z(t)\cos(\omega_4 t + \phi_4)] * h_z(t) \quad (7)$$

$$w(t) = [Z(t)\sin(\omega_4 t + \phi_4)] * h_w(t) \quad (8)$$

at the Z receiver. We have represented the filtering by convolutions with h_x, h_y, h_z, and h_w, the filter weighting functions.

Since the frequency and phase reference for a narrow-band process can be chosen arbitrarily, we can choose the frequency and phase reference of either X or Z arbitrarily. For convenience, we chose $\omega_1 = \omega_3$ and $\phi_1 = \phi_3$, and we define $\omega = \omega_2 - \omega_4$ and $\phi = \phi_2 - \phi_4$. The difference frequency ω, also called the stopped fringe rate, is determined by the relative doppler between X and Z, as reflected by ω_2, and by the reference ω_4. The difference or fringe phase ϕ is random, and uniformly distributed. With this simplification, the observed processes are

$$x(t) = [n(t) + s(t)] * h_x(t) \tag{9}$$

$$y(t) = [m(t) + r(t)] * h_y(t) \tag{10}$$

$$z(t) = \{[p(t) + s(t-\delta)]\cos(\omega t + \phi) + [q(t) + r(t-\delta)]\sin(\omega t + \phi)\} * h_z(t) \tag{11}$$

$$w(t) = \{[q(t) + r(t-\delta)]\cos(\omega t + \phi) - [p(t) + s(t-\delta)]\sin(\omega t + \phi)\} * h_w(t) \tag{12}$$

The four observables are now sampled, all at a uniform and identical rate, with a sampling interval Δ. Independence of the samples in each channel is assured by having the weighting functions be zero outside of the interval $(0, \Delta)$, and by the whiteness of the noise processes. A remaining parameter which can be varied is the relative times of the samples in the sine and cosine channels, so we leave this arbitrary. As references, we assume that the sampling of $x(t)$ begins at $t = 0$, and the sampling of $z(t)$ begins at $t = \tau$, i.e., at the delay we wish to estimate. The samples of y and w occur Δ_1 and Δ_2 after the samples of x and z. Thus, the samples are

$$X_j = x(j\Delta)$$

$$Y_j = y(j\Delta + \Delta_1)$$

$$Z_j = z(j\Delta + \tau)$$

$$W_j = w(j\Delta + \tau + \Delta_2)$$

At this point we make the further assumption that ω is a very low frequency compared to the sampling rate, so that the factors $\cos(\omega t + \phi)$ are constant over Δ and can be brought outside of the convolution integrals. This assumption is reasonable, since ω can be chosen by the experimenter.

We now normalize the observables to unit variance, and define the input SNR to be

$$\rho^2 = \frac{S_x S_z}{(N_x + S_x)(N_z + S_z)} \tag{13}$$

where S_x, S_y, N_x, and N_y are the input signal and noise spectral densities. In practice, ρ will not exceed 10^{-2}. The observable covariances can then be expressed as

$$E\{X_i Z_j\} = A_{ij} = \rho a_{ij}\cos(j\Delta\omega + \phi) \tag{14}$$

$$E\{X_i W_j\} = B_{ij} = -\rho b_{ij}\sin(j\Delta\omega + \phi) \tag{15}$$

$$E\{Y_i Z_j\} = C_{ij} = \rho c_{ij}\sin(j\Delta\omega + \phi) \tag{16}$$

$$E\{Y_i W_j\} = D_{ij} = \rho d_{ij}\cos(j\Delta\omega + \phi) \tag{17}$$

The a_{ij}, b_{ij}, c_{ij}, d_{ij} reflect the dependence on $\tau - \delta(t)$, and are constant for fixed $i - j$ when $\tau - \delta$ is constant. In any case, they vary slowly in $i - j$. Also, the sinusoidal variation in the covariances is slow in j, because $\omega\Delta << 1$. Thus, for each $i - j$ there is a range of j for which the covariances are essentially constant.

IV. Derivation of Approximate Maximum Likelihood Estimator

The general procedure of maximum likelihood estimation is to maximize the *a posteriori* probability density function (pdf) of the observables, conditioned on the unknown parameters. The values of the parameters which maximize the pdf for the given set of observables are chosen as the maximum likelihood (ML) estimates. The parameters to be estimated here are ρ, τ, ϕ, and ω. In this section, we derive approximate maximizations of the pdf with respect to ρ and ϕ. The resulting function must then be maximized numerically with respect to τ and ω in order to obtain estimates of all the parameters.

The first step in our problem is to find the joint pdf of the observables X_i, Y_i, Z_i, and W_i, conditioned on the unknown parameters ρ, ϕ, τ, and ω. This pdf depends only on the conditional covariance matrix, since the observables are jointly gaussian and zero mean. Suppose we define a row vector U having as its components all of the observables:

$$U = (X_1, X_2, \cdots, X_N, Y_1, Y_2, \cdots, Y_N, Z_1, Z_2, \cdots, Z_N, W_1, W_2, \cdots, W_N) \tag{18}$$

where N is the number of samples of each variable.

Then the covariance matrix of U is

$$\Lambda = \begin{pmatrix} I & 0 & A & B \\ 0 & I & C & D \\ A^t & C^t & I & 0 \\ B^t & D^t & 0 & I \end{pmatrix} \tag{19}$$

where A, B, C, and D are the covariance matrices with elements A_{ij}, B_{ij}, etc., given by Eqs. (14) through (17), and the conditional pdf of the observables is

$$P(U|\rho, \phi, \tau, \omega) = \frac{c}{|\Lambda|^{1/2}} \exp\left\{-\frac{1}{2} U\Lambda^{-1}U^t\right\} \quad (20)$$

The covariance matrix Λ depends on the parameters ρ, ϕ, τ, and ω, and c is a constant.

The major problem at this point is to invert the covariance matrix. We can do this only in series form, and it is the truncation of this series in the maximization procedure which causes our estimator to be only approximately maximum likelihood.

To proceed we define a matrix P such that

$$\Lambda = I + P \quad (21)$$

The matrix P has at most four nonzero elements in each row and column, because A, B, C, and D have at most two nonzero elements in each row and column. Furthermore, the nonzero elements of P are proportional to ρ and do not exceed ρ in absolute value. Thus we can expand Λ^{-1} in a power series, and bound the terms:

$$\Lambda^{-1} = I - P + P^2 - P^3 + \cdots \quad (22)$$

Since the two principal quadrants of P are zero, the principal diagonal elements of P^n are zero for odd n. The other elements are bounded by

$$\begin{aligned}\max_{i,j} \left|(P^n)_{ij}\right| &\leq 4\rho \max_{i,j} \left|(P^{n-1})_{i,j}\right| \\ &\leq 4^{n-1}\rho^n \end{aligned} \quad (23)$$

where $(P^n)_{ij}$ denotes the ij element of P^n.

Closer bounds can be obtained utilizing properties of the cross covariances for particular cases.

The conditional pdf can now be written as

$$P(U|\rho, \phi, \tau, \omega) = c \exp\left\{-\frac{1}{2} U(I+P)^{-1}U^t - \frac{1}{2}\log\det(I+P)\right\} \quad (24)$$

Using a well-known matrix identity,

$$\begin{aligned}\log\det(I+P) &\equiv \mathrm{Tr}\log(I+P) \\ &= \mathrm{Tr}\left(P - \frac{P^2}{2} + \frac{P^3}{3} - \frac{P^4}{4} + \cdots\right) \end{aligned} \quad (25)$$

The odd power terms can be deleted, since the principal diagonal of P^n is zero for odd n. Thus

$$\log\det(I+P) = -\mathrm{Tr}\left(\frac{P^2}{2} + \frac{P^4}{4} + \cdots\right) \quad (26)$$

We now define a likelihood function $L_1(U|\rho, \phi, \tau, \omega)$ as the exponent of the conditional pdf, and maximization of L_1 is equivalent to maximization of the pdf.

$$\begin{aligned}L_1(U|\rho, \phi, \tau, \omega) &= -\frac{1}{2} U(I - P + P^2 - \cdots) U^t \\ &\quad + \frac{1}{2}\mathrm{Tr}\left(\frac{P^2}{2} + \frac{P^4}{4} + \cdots\right) \end{aligned} \quad (27)$$

It is not feasible to maximize L_1 analytically with respect to any of the parameters without neglecting terms in P of higher order than P^2. With this approximation, we can maximize with respect to ρ and ϕ. Since normally τ and ω are the parameters of primary interest, the approximate solutions for ρ and ϕ usually suffice, but greater accuracy can be obtained numerically if required.

To proceed, we define a new matrix Q by

$$Q = \frac{1}{\rho}P = \frac{1}{\rho}\begin{pmatrix} \bigcirc & \begin{matrix} A & B \\ C & D \end{matrix} \\ \begin{matrix} A^t & C^t \\ B^t & D^t \end{matrix} & \bigcirc \end{pmatrix} \quad (28)$$

Next we drop the UIU^t term in L_1, which is independent of the parameters, to obtain

$$L_2(U|\rho, \phi, \tau, \omega) \approx \frac{1}{2} U(\rho Q - \rho^2 Q^2) U^t + \frac{1}{4}\rho^2 \mathrm{Tr}(Q^2) \quad (29)$$

By differentiating with respect to ρ, we see that L_2 is maximized for the conditional estimate of ρ

$$\hat{\rho} = \frac{UQU^t}{2UQ^2U^t - \mathrm{Tr}(Q^2)} \quad (30)$$

The denominator of this expression can be approximated by its mean, which is $\mathrm{Tr}(Q^2)$, so

$$\hat{\rho} \approx \frac{UQU^t}{\mathrm{Tr}(Q^2)} \quad (31)$$

The variance of the denominator of Eq. (30) is also on the order of $\mathrm{Tr}(Q^2)$. Therefore, since $\mathrm{Tr}(Q^2) \approx 4N$, the

approximation is good when N is large, say 10^4 or greater, which will always be true in VLBI problems.

A new likelihood function is now obtained by substituting the value of $\hat{\rho}$ into Eq. (29), and again approximating UQ^2U^t by $\mathrm{Tr}\,(Q^2)$:

$$L_3(U|\rho, \phi, \tau, \omega) \approx \frac{[UQU^t]^2}{\mathrm{Tr}\,(Q^2)} \tag{32}$$

Since the elements of Q vary slowly except for the sinusoidal variation, $\mathrm{Tr}\,(Q^2)$ is essentially independent of Q and ω so long as $N\Delta\omega >> \pi$. This can be assured by controlling ω by selecting the local oscillator frequencies. Neglecting any slight variation of $\mathrm{Tr}\,(Q^2)$, L_2 can be maximized over ϕ. To do this, Q is expressed

$$Q = R\cos\phi + S\sin\phi \tag{33}$$

where R and S do not depend on ϕ and are given by

$$R = \begin{pmatrix} O & R_0 \\ R_0^t & O \end{pmatrix} \tag{34}$$

$$S = \begin{pmatrix} O & S_0 \\ S_0^t & O \end{pmatrix} \tag{35}$$

where

$$R_0 = \begin{pmatrix} (a_{ij}\cos j\Delta\omega) & -(b_{ij}\sin j\Delta\omega) \\ (c_{ij}\sin j\Delta\omega) & (d_{ij}\cos j\Delta\omega) \end{pmatrix} \tag{36}$$

$$S_0 = \begin{pmatrix} -(a_{ij}\sin j\Delta\omega) & -(b_{ij}\cos j\Delta\omega) \\ (c_{ij}\cos j\Delta\omega) & -(d_{ij}\sin j\Delta\omega) \end{pmatrix} \tag{37}$$

The derivative of the likelihood ratio with respect to ϕ is then

$$\frac{d}{d\phi}L_3 = \frac{2(UQU^t)\,U(S\cos\phi - R\sin\phi)\,U^t}{\mathrm{Tr}\,(Q^2)} \tag{38}$$

and the value of ϕ which maximizes L_3 is

$$\hat{\phi} = \mathrm{Arctan}\frac{USU^t}{URU^t} \tag{39}$$

The new likelihood ratio is the maximum of L_3,

$$L(\hat{\rho}, \hat{\phi}, \tau, \omega) = \frac{(URU^t)^2 + (USU^t)^2}{\mathrm{Tr}\,(Q^2)} \tag{40}$$

This is as far as we can proceed analytically. To find the final approximate ML estimates of all the parameters, L is maximized numerically over τ and ω. When only $\hat{\tau}$ is required, ω is usually known *a priori*, so that the numerical maximization is only over one parameter, τ.

V. Probability Distribution and SNR of the Estimator

The estimation procedure is to compute $L(\hat{\rho}, \hat{\phi}, \omega_j, \tau_i)$ for all possible values of ω_j and τ_i in the regions of uncertainty of ω and τ, and to choose as $\hat{\omega}$ and $\hat{\tau}$ the values of ω_j and τ_i which maximize L. In order that the estimation error be small, it is important that the maximum of L occurs near ω and τ. For example, suppose that ω is known, but the uncertainty in τ is over a range of $M\Delta$, and we compute L for $\tau_i = \tau + i\Delta + \epsilon$, $i = -M/2, -M/2+1, \cdots, M/2$, where ϵ is small. L is computed for M independent incorrect values of τ_i, and for one value, $\tau + \epsilon$, near the correct value τ. If any of the M values of L for incorrect τ_i exceeds the value for $\tau + \epsilon$, there will be a large error in the estimate. It is important that the probability of this occurring be small, and to estimate this probability we must know the distribution of the estimator.

In the expression for L, Eq. (40), only the components of the random vector U depend on the actual parameters. The matrices R and S and the denominator term $\mathrm{Tr}\,(Q^2)$ depend only on the filter weighting functions, the sampling times, and the assumed parameter values ω_j and τ_i. The terms URU^t and USU^t are weighted sums of large numbers of random variables. They are therefore approximately gaussian, by the Central Limit Theorem, and, as we show later, their variances are approximately the same. Thus L is approximately the sum of squares of two independent gaussian variables with the same variance, and is therefore approximately Chi-squared distributed with one degree of freedom. Letting v denote $L(\hat{\rho}, \hat{\phi}, \omega_j, \tau_i)$, and $m_r, m_s, \sigma_r^2, \sigma_s^2$ denote the means and variances of URU^t and USU^t, and assuming $\sigma = \sigma_r = \sigma_s$, the density of the estimator conditioned on the actual values ω and τ is approximately

$$p(v|\tau, \omega) \approx \frac{1}{2\Psi}\exp\left\{-\frac{v + a^2}{2\Psi}\right\} I_0\left(\frac{a\sqrt{v}}{\Psi}\right) \tag{41}$$

where

$$\Psi = \frac{\sigma^2}{\mathrm{Tr}\,(Q^2)} \tag{42}$$

$$a^2 = \frac{m_r^2 + m_s^2}{\mathrm{Tr}\,(Q^2)} \tag{43}$$

In evaluating the required statistics, we observe that the coefficients a_{ij}, b_{ij}, c_{ij}, and d_{ij} depend on the filter weighting functions and sampling times, as well as on the assumed values of the parameters, ω_j and τ_i. For convenience, we suppress this dependence, and also the dependence on ω_j, since estimation of τ is our primary interest. We explicitly carry the dependence on τ and τ_i, the actual and assumed values of the time difference.

The normalizing factor $\mathrm{Tr}\,(Q^2)$ is equal to the sum of the squares of the elements of Q. Each weighting coefficient appears twice in Q, so

$$\mathrm{Tr}\,(Q^2) = 2 \sum_{m,n=1}^{N} [a_{mn}^2(\tau_i) + d_{mn}^2(\tau_i)] \cos^2(n\Delta\omega + \phi) + [b_{mn}^2(\tau_i) + c_{mn}^2(\tau_i)] \sin^2(n\Delta\omega + \phi) \quad (44)$$

Since the coefficients vary slowly in n for fixed τ_i and fixed $m - n$, the trigonometric terms average out approximately to ½, and

$$\mathrm{Tr}\,(Q^2) \approx \sum_{m,n=1}^{N} a_{mn}^2(\tau_i) + b_{mn}^2(\tau_i) + c_{mn}^2(\tau_i) + d_{mn}^2(\tau_i) \quad (45)$$

To evaluate the means and variances, we write the variables in terms of the original observables. First,

$$URU^t = 2 \sum_{m,n=1}^{N} [X_m Z_n a_{mn}(\tau_i) + Y_m W_n d_{mn}(\tau_i)] \cos n\Delta\omega + [-X_m W_n b_{mn}(\tau_i) + Y_m Z_n c_{mn}(\tau_i)] \sin n\Delta\omega \quad (46)$$

Using the covariances given by Eqs. (14) through (17), the mean of URU^t is

$$\begin{aligned} m_r(\tau_i, \tau) &= 2\rho \sum_{m,n=1}^{N} [a_{mn}(\tau)\, a_{mn}(\tau_i) + d_{mn}(\tau)\, d_{mn}(\tau_i)] \\ &\quad \times \cos(n\Delta\omega) \cos(n\Delta\omega + \phi) \\ &\quad + [b_{mn}(\tau)\, b_{mn}(\tau_i) + c_{mn}(\tau)\, c_{mn}(\tau_i)] \\ &\quad \times \sin(n\Delta\omega) \sin(n\Delta\omega + \phi) \quad (47) \\ &\approx \rho \cos\phi\, F(\tau_i, \tau) \quad (48) \end{aligned}$$

where

$$F(\tau_i, \tau) = \sum_{m,n} a_{mn}(\tau)\, a_{mn}(\tau_i) + b_{mn}(\tau)\, b_{mn}(\tau_i) + c_{mn}(\tau)\, c_{mn}(\tau_i) + d_{mn}(\tau)\, d_{mn}(\tau_i) \quad (49)$$

A similar result holds for m_s with cos ϕ replaced by sin ϕ, and

$$m_r^2 + m_s^2 \approx \rho^2 F^2(\tau_i, \tau) \quad (50)$$

Since

$$\mathrm{Tr}\,(Q^2) = F(\tau_i, \tau_i) \quad (51)$$

we observe that when the assumed value τ_i is equal to the correct value, τ,

$$m_r^2 + m_s^2 = \rho^2\, \mathrm{Tr}^2(Q^2) \quad (52)$$

and is zero when $\rho = 0$ or when τ_i is so far from τ that $a_{mn}(\tau)\, a_{mn}(\tau_i) = 0$, etc. The product random variables $X_m Z_n$, $Y_m W_n$, $X_m W_n$ and $Y_m Z_n$ are all independent, and their variances are essentially unity since $\rho^2 << 1$. Thus the variances of weighted sums of these variables are just the sums of the squares of the weighting coefficients, so

$$\begin{aligned} \sigma_r^2 &= \sigma_s^2 = \sigma^2 \\ &\approx 4 \cdot \frac{1}{2} F(\tau_i, \tau_i) \\ &\approx 2\, \mathrm{Tr}\,(Q^2) \quad (53) \end{aligned}$$

The factor of four arises because each term in the summation occurs twice, and the factor of one-half arises from the trigonometric factors.

Finally,

$$\Psi = \frac{\sigma^2}{\mathrm{Tr}\,(Q^2)} \quad (54)$$

and

$$a^2 = \begin{cases} 0, & \text{for } |\tau_i - \tau| >> \Delta \\ \rho^2\, \mathrm{Tr}^2(Q^2), & \text{for } \tau_i = \tau \end{cases} \quad (55)$$

A. Estimator SNR

A convenient figure of merit is the signal-to-noise ratio of the estimator function, which we define to be the ratio of the squared difference in the means of L (or v) for $\tau_i = \tau$ and for $|\tau_i - \tau| >> \Delta$, to the variance. The mean and variance are

$$E\{L\} = 4 + a^2 \quad (56)$$

and

$$\mathrm{Var}\,\{L\} = 16 + 8a^2 \quad (57)$$

so the SNR is

$$R = \frac{r^2}{1+2r} = \frac{1}{2}\,\frac{r}{1+\dfrac{1}{2r}} \tag{58}$$

where

$$r = \frac{\rho^2}{4}\,\mathrm{Tr}\,(Q^2) = \frac{a^2}{4} \tag{59}$$

Clearly the distribution of L depends only on the SNR.

B. SNR Required for Small Estimator Errors

Suppose instead of calculating $L(\tau_i)$ for all possible τ_i and choosing the maximum, $L(\tau_i)$ is calculated for various values until a threshold, say T, is exceeded. Then the error in τ will be small if $L(\tau+\epsilon)$ exceeds T for $|\epsilon| < \Delta$, but no other $L(\tau_i)$ calculated exceeds T. The probability that this occurs can be calculated from the distributions of L for the correct and incorrect values of τ_i. Figure 1 shows the probability that $L(\tau)$ fails to exceed various thresholds, T, as a function of the SNR. For example, if the SNR is 12, a threshold of 32 will be exceeded by $L(\tau)$ with probability about $1 - 10^{-3}$. For $|\tau_i - \tau| >> \Delta$ or for $\rho = 0$, the density of L is exponential with mean 4, so the probability that T is exceeded is

$$P(L > T \mid \rho = 0) = \exp\{-T/4\} \tag{60}$$

For $T = 32$, this probability is 0.33×10^{-3}, and it is fairly likely that T will be exceeded on noise alone if L is calculated for too many incorrect values of τ_i before it is calculated for a τ_i close to τ. Clearly there is a tradeoff between SNR, threshold value, and the initial uncertainty in τ. A SNR of 10 to 20 should prove adequate when the initial uncertainty is not too large.

VI. Effect of Quantization or Limiting of Samples

We have shown that maximization of L results in approximately ML estimation when the original data samples X_j, Y_j, Z_j and W_j are gaussian. This is no longer true if the samples are quantized, but the same estimator function can certainly still be used. The only effect is to decrease the estimator SNR and increase the error by decreasing the correlation coefficients. For hard limiting, the effect is to replace ρ by $(2/\pi)\rho$ throughout, which causes a loss of approximately $(2/\pi)^2$ or 4 dB in R. The reason that ρ is replaced by $(2/\pi)\rho$ is that the means of the cross products of the data samples are so reduced. This can be shown by considering two normal variables, say x and y, with correlation ρ, and showing that for $\rho << 1$, the expected value of sgn (xy) is approximately $(2/\pi)\rho$.

VII. Selection of Sampling Times and Filter Weighting Functions

The problem of choosing an optimum set of sampling times and filter weighting functions is considerably more complex than that of deriving the approximate ML estimator as a function of these variables. In particular, what we really wish to optimize is some cost function of the error in the estimate of τ—but we have already admitted failure in this respect by limiting our choice to the MLE. However, given our choice of estimator, the parameter R is clearly a significant figure of merit, as it is closely related to the probability that the estimation error is no worse than one sampling interval.

Because we cannot truly optimize, we will make some reasonable assumptions which are compatible with practical implementation, and show that we can do quite well. In particular, R depends on the actual value of $\tau - \delta$, so we must be concerned with the minimum of R over $\tau - \delta$. We will obtain a minimum which is almost as large as possible.

The estimator SNR increases with the input SNR and $\mathrm{Tr}\,(Q^2)$, which in turn depends on the number of samples, on $\tau - \delta(t)$, and on the sampling and filtering. Since $\mathrm{Tr}\,(Q^2)$ is roughly proportional to N, we define K such that

$$K = \frac{\mathrm{Tr}\,(Q^2)}{4N} = \frac{1}{4N}\sum_m \sum_n a_{mn}^2 + b_{mn}^2 + c_{mn}^2 + d_{mn}^2 \tag{61}$$

The functions a, b, c, and d depend on $\tau - \delta(t)$. However, since they are constant for fixed $m - n$ when $\tau - \delta$ is constant, the minimum over τ of K is clearly smallest when δ is constant. Then, since the coefficients are nonzero only for $|m - n| << N$

$$K(\tau, \delta) \approx \frac{1}{4}\sum_{m-n} a_{mn}^2 + b_{mn}^2 + c_{mn}^2 + d_{mn}^2 \tag{62}$$

We must now consider the effect of the sampling and filtering on a, b, c, and d.

From here on, we restrict ourselves to using identical filters in the four channels:

$$h(t) = h_x(t) = h_y(t) = h_z(t) = h_w(t) \tag{63}$$

The justification for this restriction is that this is the only way that any cross correlation can be as high as ρ^2, and it

assures that this occurs whenever the signal components in the two receivers are in the same phase relationship to the sampling times.

The functions a, b, c, and d then become identical except for the effect of the sample point skewing, reflected by Δ_1 and Δ_2. Furthermore, the shape of the function is the same as the correlation function at the output of the filter; i.e.,

$$r(\beta) = \int_{-\infty}^{\infty} h(\alpha)\, h(\beta - \alpha)\, d\alpha \tag{64}$$

and, letting $k = m - n$,

$$\left.\begin{aligned} a_{mn}(\tau - \delta) &= r(\tau - \delta - k\Delta) \\ b_{mn}(\tau - \delta) &= r(\tau - \delta + \Delta_2 - k\Delta) \\ c_{mn}(\tau - \delta) &= r(\tau - \delta - \Delta_1 - k\Delta) \\ d_{mn}(\tau - \delta) &= r(\tau - \delta + \Delta_2 - \Delta_1 - k\Delta) \end{aligned}\right\} \tag{65}$$

Because $h(\alpha)$ is nonzero only for $0 \leq \alpha < \Delta$, $r(\tau)$ is nonzero only for $|\tau| < 2\Delta$. This means that a_{mn}, b_{mn}, c_{mn}, and d_{mn} are nonzero for at most two values of $k = m - n$. We now define

$$A(\beta) = \sum_k r^2(\beta - k\Delta) \tag{66}$$

Because of the symmetry of $r(\beta)$, this function depends only on the minimum distance of β from an integer multiple of Δ.

For a given filter, K depends only on $\tau - \delta, \Delta_1$, and Δ_2 as

$$\begin{aligned} K(\tau - \delta, \Delta_1, \Delta_2) = \frac{1}{4}[&A(\tau - \delta) + A(\tau - \delta + \Delta_2) \\ &+ A(\tau - \delta - \Delta_1) \\ &+ A(\tau - \delta + \Delta_2 - \Delta_1)] \end{aligned} \tag{67}$$

We are interested in maximizing the minimum value of K with respect to $\tau - \delta$ by selection Δ_1 and Δ_2. If the samples are not skewed, i.e., $\Delta_1 = \Delta_2 = 0$, the minimum of K is the minimum of A, which is the worst we can do. The best we can do is the average of A, and the closest we can come to this average is to pick different nonzero values of Δ_1 and Δ_2.

Garsia, Rodemich, and Rumsey (Ref. 1) have shown that for all normalized correlation functions $c(t)$ which are zero outside of the interval $(-1, 1)$,

$$\int_{-1}^{1} c^2(t)\, dt \leq 0.68698\ldots \tag{68}$$

This bound is exactly the upper bound on the average of $A(\beta)$.

We can now show that the moving window filter,

$$\begin{aligned} h(t) &= \frac{1}{\Delta}, \qquad \text{for } 0 \leq t \leq \Delta \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{69}$$

is almost optimum. For this filter, the average of A is ⅔—very close to the upper bound. We achieve a minimum K of 21/32 by choosing $\Delta_1 = \Delta/2$ and $\Delta_2 = \Delta/4$, and a minimum of ⅝ for $\Delta_1 = \Delta_2 = \Delta/2$, or $\Delta_1 = \Delta/2$, $\Delta_2 = 0$. The loss relative to the upper bound is only 0.2 dB in the first case, and 0.4 dB in the second case.

We define the best minimum value of K to be K_c:

$$K_c = \frac{21}{32} \approx 0.657 \tag{70}$$

If no skewing is used, the minimum of K is ½; skewing improves the SNR by 1.0 to 1.2 dB.

A. Real Versus Complex Sampling

By real rather than complex sampling, we mean using only the sine or cosine components at the receivers. For example, we use only the X_j and Z_j, rather than X_j, Y_j, Z_j, and W_j. This may be desirable when memory size or access time is the limiting factor in implementation.

Using only the X_j and Z_j is equivalent to letting the weighting functions in the Y and W channels be zero. Then b_{mn}, c_{mn}, and d_{mn} are all zero, and

$$K = \frac{1}{4N} \sum_{m,n} a_{mn}^2 \tag{71}$$

The worst case is again when δ is constant, and now sample time skewing cannot be used to average K over $\tau - \delta$. The minimum value of K over $\tau - \delta$, which we denote by K_r, is

$$K_r = \frac{1}{8} \tag{72}$$

Since R is proportional to K, the loss for real sampling compared to complex sampling is

$$\frac{K_r}{K_c} = \frac{0.125}{0.657} = 0.19 \tag{73}$$

or approximately 7.2 dB.

If the fundamental limitation is memory size, then the loss is 3 dB less than this, or 4.2 dB, since twice as many real as complex samples can be taken.

If the sampling rate is limited by the memory speed rather than by the receiver bandwidth, we must realize that the sampling rate can be twice as high with real sampling rather than complex. Real sampling has a distinct advantage here, because for a fixed estimator SNR, the error in time estimation is inversely proportional to sampling rate. Real sampling might be better in spite of the 4.2-dB loss in SNR, provided that the SNR is high enough so that τ is almost certainly resolved to better than one sampling interval.

B. Final Estimator SNR

Considering the effects of quantization, filtering, and sampling, the minimum SNR which can be guaranteed is

$$R_{\min} = \frac{1}{2}\frac{r}{1+\frac{1}{2r}} \tag{74}$$

where

$$r = K_Q K \rho^2 N \tag{75}$$

The factor K_Q is due to quantization, and varies from $(2/\pi)^2$ for hard limiting to 1 for no quantization. The factor K depends on the sampling method. If real rather than complex sampling is used, then

$$K = K_r = \frac{1}{8} \tag{76}$$

For complex sampling with proper skewing,

$$K = K_c = \frac{21}{32} = 0.657 \tag{77}$$

VIII. Mean Square Error of Time Estimate

In this section we derive an approximation to the mean square error in estimation of τ using a method similar to that used by Helstrom (Ref. 2) for ML estimation of the time of arrival of radar signals. Since the method is valid only when the estimator SNR is sufficiently high, a simulation of the received signals and the processing is conducted to test the validity of the approximations at fairly low SNRs. The simulation confirms the analysis, with the mean square errors being slightly greater than calculated, as would be expected considering the approximations used.

A. Calculation of Mean Square (MS) Error

The first step is to write L in terms of signal and noise components. We explicitly carry the dependence on τ and on the assumed value τ_i only when deleting this dependence might be confusing. Let

$$URU^t = r(\tau_i, \tau) + m_r(\tau_i) \tag{78}$$

$$USU^t = s(\tau_i, \tau) + m_s(\tau_i) \tag{79}$$

where, as before, m_r and m_s are the means of URU^t and USU^t, and r and s are the noise portions. Any slight dependence of r and s on τ is neglected. Then

$$L(\tau_i, \tau) = \frac{m_r^2 + m_s^2}{\mathrm{Tr}(Q^2)} + \frac{2(rm_r + sm_s)}{\mathrm{Tr}(Q^2)} + \frac{r^2 + s^2}{\mathrm{Tr}(Q^2)} \tag{80}$$

When R is sufficiently high, the quadratic noise terms become insignificant, and L can be approximated by

$$L(\tau_i) \approx L_0(\tau_i, \tau) + M(\tau_i, \tau) \tag{81}$$

where

$$L_0(\tau_i, \tau) = \frac{m_r^2 + m_s^2}{\mathrm{Tr}(Q^2)} = \rho^2 \frac{[F(\tau_i, \tau)]^2}{\mathrm{Tr}(Q^2)} \tag{82}$$

and

$$M(\tau_i, \tau) = \frac{2(rm_r + sm_s)}{\mathrm{Tr}(Q^2)} \tag{83}$$

Since the estimate $\hat{\tau}$ of τ is the value of τ_i which maximizes L, the derivative of L with respect to τ_i is zero at $\tau_i = \hat{\tau}$; i.e.,

$$0 = L'(\tau_i, \tau)\Big|_{\tau_i = \hat{\tau}} \approx L_0'(\tau_i, \tau) + M'(\tau_i, \tau)\Big|_{\tau_i = \hat{\tau}} \tag{84}$$

where primes denote differentiation with respect to τ_i. We now expand the L_0' term in a Taylor series about $\tau_i = \tau$, and retain only the first order term:

$$L_0'(\tau_i, \tau) + (\hat{\tau} - \tau)L_0''(\tau_i, \tau) + M'(\tau_i, \tau)\Big|_{\tau_i = \tau} \approx 0 \tag{85}$$

Since the signal only term L_0 is maximized at $\tau_i = \tau$, its derivative there is zero, and

$$\tau - \hat{\tau} \approx \frac{M'(\tau_i, \tau)\Big|_{\tau_i = \tau}}{L''(\tau_i, \tau)\Big|_{\tau_i = \tau}} = \frac{M'(\tau, \tau)}{L_0''(\tau, \tau)} \tag{86}$$

Finally, since L_0 does not depend on noise, the mean square error in the estimate of τ is

$$E\{(\tau - \hat{\tau})^2\} = \frac{E\{[M'(\tau, \tau)]^2\}}{[L_0''(\tau, \tau)]^2} \tag{87}$$

The terms $E\{M'^2\}$ and L_0'' can be evaluated for specific coefficient functions, with these functions being determined from Eqs. (64) and (69). The coefficients for the optimum skewing, $\Delta_1 = \Delta/2$ and $\Delta_2 = \Delta/4$, and normalizing to $\delta = 0$ and $\Delta = 1$, are shown in Fig. 2.

For these parameters

$$L_0''\Big|_{\tau_i = \tau} = \frac{\rho^2}{4}\left[\frac{2[F'(\tau, \tau)]^2 - 16N \operatorname{Tr}(Q^2)}{\operatorname{Tr}(Q^2)}\right] \tag{88}$$

and

$$\begin{aligned} E\{[M'(\tau, \tau)]^2\} &= \frac{\rho^2}{4 \operatorname{Tr}(Q^2)}(16N \operatorname{Tr}(Q^2) - 2[F'(\tau, \tau)]^2) \\ &= \frac{\rho^2}{4 \operatorname{Tr}(Q^2)}(42N^2) \end{aligned} \tag{89}$$

Finally, the mean square error is

$$\begin{aligned} \epsilon^2(\tau) = E\{(\hat{\tau} - \tau)^2\} &= \frac{4}{\rho^2}\frac{\operatorname{Tr}(Q^2)}{42N^2} \\ &= \frac{4}{\rho^2 N}\frac{8\tau^2 - 2\tau + \frac{11}{4}}{42} \end{aligned} \tag{90}$$

(τ taken modulo ¼)

Averaging over τ, the mean square (MS) error is

$$\epsilon_{\text{ave}}^2 = \frac{16}{63\rho^2 N} \qquad \text{(no limiting)} \tag{91}$$

This MS error is of course increased by a factor of $(\pi/2)^2$ when hard limiting is used, to

$$\epsilon_{\text{ave}}^2 = \frac{4\pi^2}{63\rho^2 N} = \frac{0.626}{\rho^2 N} \qquad \text{(hard limiting)} \tag{92}$$

B. Comparison of *R* and MS Error

It is interesting that for high SNRs, both $R(\tau)$ and $\epsilon^2(\tau)$ vary with τ in direct proportion to $\operatorname{Tr}(Q^2)$. This is not what one would intuitively expect, since one would expect the MS error to be inversely proportional to R. The explanation for this is that for high SNRs, the variation of ϵ^2 with τ depends on the second derivative of the mean of the estimator function, which introduces the particular dependence on $\operatorname{Tr}(Q^2)$. In any case, the effect is very minor, since $\operatorname{Tr}(Q^2)$ varies only plus 3 percent and minus 1.5 percent about its average value.

C. Simulation Results

A simulation was conducted for two different SNRs and two values of τ, with the resulting rms errors compared to the calculated values in Table 1. The rms errors are the average of 160 cases for each set of parameters, using $N = 19840$.

The rms errors obtained in the simulation were typically about ten percent higher than the calculated values. This is because the calculated values are not accurate at the relatively low SNRs used, primarily because typical errors in $\hat{\tau}$ are large enough that truncation of the power series for L_0 causes significant error. There is also an error in the MS error of approximately $1/(4R)$ percent due to neglecting the quadratic noise terms. In Table 1, the observed errors are higher for $\tau = 0.125$ than for $\tau = 0$, whereas the calculated errors are higher for $\tau = 0$. We again feel that the calculated values are in error at the SNRs in the simulation, because of the series approximation. Unfortunately, it was not practical to conduct further simulations at higher SNRs, because of computer time limitations. Higher SNRs require more computer time, as N must be increased rather than ρ, for increasing ρ would render the assumed degradation of $(\pi/2)^2$ due to hard limiting to be in error.

IX. Performance of Proposed System

The system proposed for DSN time synchronization is performance limited by the antenna sizes and receiver noise temperatures and by the restriction of using an XDS 920 computer at each site for data buffering. The 920 computers limit the sampling rate to 500 kbps, and the total number of bits of data to 300,000. The antennas assumed are one of 64 m (210 ft), with a noise temperature of 25 K, and one of 26 m (85 ft), with a noise temperature of 40 K.

A radio star with a flux density of 1 fu causes an increase in system temperature of 0.6 K and 0.1 K for 64 m

and 26 m antennas, respectively. Therefore, letting S be the source intensity, the input SNR is approximately

$$\rho^2 = \frac{0.06\,S^2}{40 \times 25} = 6 \times 10^{-5}\,S^2 \tag{93}$$

and, for complex sampling,

$$\rho^2 N \approx 9\,S^2 \tag{94}$$

For a system using complex sampling and hard limiting,

$$\begin{aligned} r &= KK_Q\rho^2 N \\ &= (0.657)\left(\frac{2}{\pi}\right)^2 \rho^2 N \\ &= 2.4\,S^2 \end{aligned} \tag{95}$$

The minimum output SNR is

$$\begin{aligned} R_{\min} &= \frac{1}{2}\frac{r}{1+\frac{1}{2r}} \\ &\approx \frac{r}{2} = 1.2\,S^2 \end{aligned} \tag{96}$$

If real samples are used, N doubles but K drops to ⅛, and

$$r \approx 0.91\,S^2 \tag{97}$$

$$R \approx 0.45\,S^2 \tag{98}$$

A. Post Detection Integration

The final SNR can be improved by filling the memory in the computer at each station several times, writing out onto magnetic tape between fills. Since it would take several seconds to dump the core onto tape, the local oscillator phase may change between batches of data. Nevertheless, the SNR can be increased in direct proportion to the number of batches of data. If L batches of data are used, the final SNR is

$$\frac{L}{2}\frac{r}{1+\frac{1}{2r}} \tag{99}$$

where r is as before. With ten batches of data and complex sampling, a SNR of 10 could be obtained from a 1-fu source, and an SNR of over 40 from a 2-fu source.

References

1. Garsia, A., Rodemich, E., and Rumsey, H., "On Some Extremal Positive Definite Functions," *J. Mathematics Mechanics*, Vol. 18, No. 9, pp. 805–834, Mar. 1969.
2. Helstrom, C. W., *Statistical Theory of Signal Detection*, Chapter VIII. Pergamon Press, New York, 1960.

Table 1. Simulation results

ρ^2K	τ	$R(\tau)$	RMS estimation error	
			Simulation	Calculated
71.42	0	9.53	0.104	0.0950
71.42	0.125	9.10	0.114	0.0927
71.42	Average	9.25	0.109	0.0935
285.7	0	38.8	0.0512	0.0475
285.7	0.125	37.1	0.0555	0.0464
285.7	Average	37.7	0.0534	0.0418

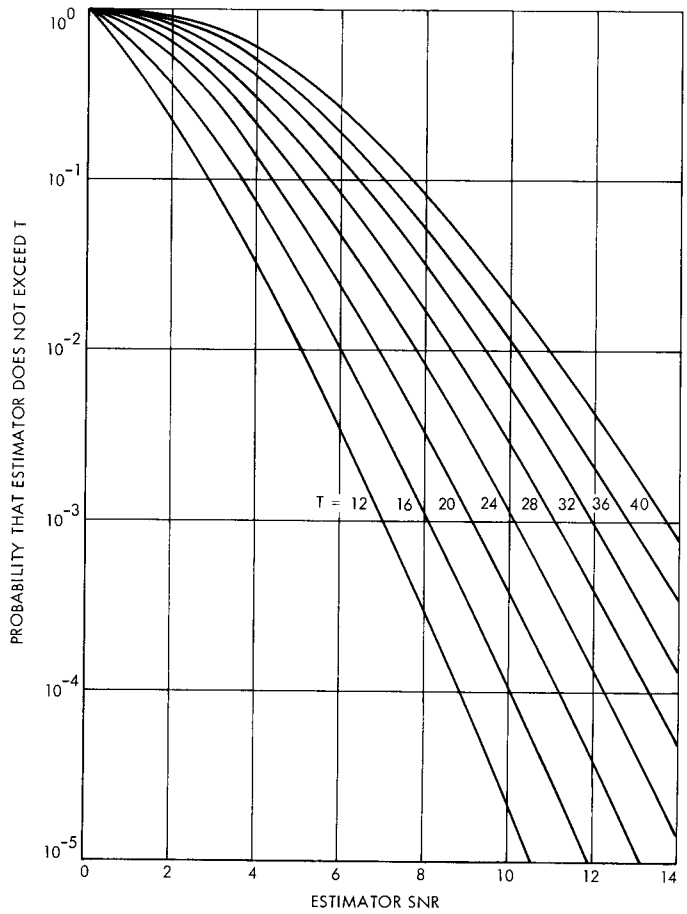

Fig. 1. Probability that estimator does not exceed threshold for correct value of τ_i

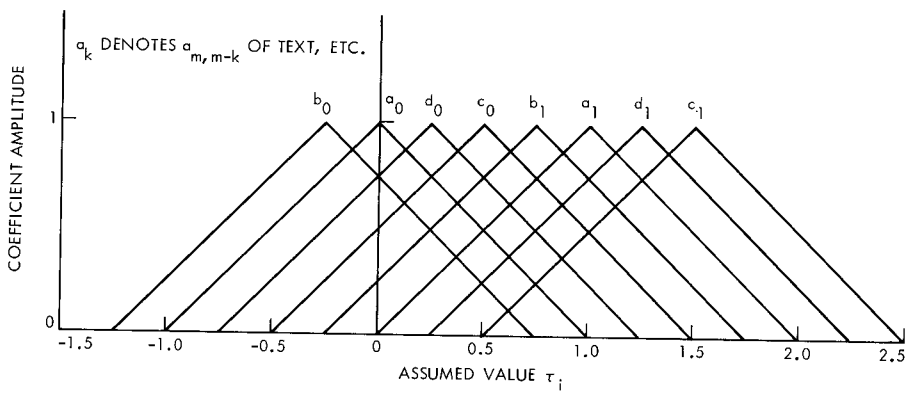

Fig. 2. Weighting coefficients for optimum sampling