Algorithms for Ambiguity Function Processing

SEYMOUR STEIN, FELLOW, IEEE

Abstract—**Calculation of the complex ambiguity function is viewed as the basis for joint estimation of the differential delay and differential frequency offset between two waveforms that contain a common component plus additive noise. In many applications, the required accuracy leads to a need for integration over long data sets that can become a challenge for near real-time digital processing. The nature of the ambiguity processing is interpreted, and an algorithm approach is shown that minimizes the processing burden over a broad category of applications without affecting performance.**

I. Introduction

A WELL-KNOWN application for *correlation processing* is the estimation of *time delay* between replicas of an arbitrary (unknown) continuous waveform when these replicas are contaminated by additive noise [1]. Flexible correlation testing of long data segments at arbitrary time lags has become practical with the advent of inexpensive storage of digitized data. Digital calculation at the same time provides high precision estimation of the time delay.

In many radio applications, the problem is compounded by the existence of a frequency difference between the replicas as a result of receiver tuning offsets or because of differential Doppler shifts due to relative motion between the terminals.[1] For reasons explained later, it is frequently required in these cases to estimate jointly the differential time offset (DTO) and *differential frequency* offset (DFO), in order to measure either one adequately.[2] For joint DTO/DFO estimation, the natural generalization of the correlation process is the complex *ambiguity function,*

$$A(\tau, f) = \int_0^T s_1(t)\, s_2^*(t+\tau) \exp(-j2\pi f t)\, dt. \tag{1}$$

In this expression, $s_1(t)$ and $s_2(t)$ are complex envelopes of two waveforms that contain a common component, while τ and f are, respectively, the time lag and frequency offset parameters to be searched simultaneously for the values that cause $|A(\tau, f)|$ to peak. For $f = 0$, the complex ambiguity function reduces to the traditional complex correlation function. For $f \neq 0$, the computation in (1) can be viewed as a correlation carried out after first translating all the spectrum in $s_1(t)$ by an amount f,[3] with peak values occurring when the value for f exactly compensates for the frequency offset from the corresponding component in $s_2(t)$. Section II discusses the character of $A(\tau, f)$ in more detail.

For given input signal-to-noise ratios (SNR's) and given input bandwidths, the integration time T determines the accuracy with which DTO or DFO can be measured. When the operation in (1) is a cross-ambiguity, the additive noises associated with $s_1(t)$ and $s_2(t)$ are usually independent and do not correlate at any values of τ, f. Then (1) provides ambiguity lobes characteristic of only the "signal" component, along with an additive random component that limits the accuracy of estimation of the τ, f values at which the lobe peaks. When the operation is an autoambiguity, so that $s_1(t)$ and $s_2(t+\tau)$ represent output from a single receiver, both the signal and noise components have an autoambiguity lobe that peaks at $\tau = 0, f = 0$. In application, however, interest in such a case lies in regions of τ, f sufficiently away from the origin to avoid the central lobe; for example, propagation research on signals of convenience might use autoambiguity calculation to detect multipath components in the signal, represented by presence of lobes at such other values of τ, f. At such other values, the additive noise has zero mean autocorrelation, and contributes only a random additive component in the result of the ambiguity calculation. The noise limitations on *accuracy* of location of the peak (Section III) then are formally identical to those for the cross-ambiguity case (with the second SNR taken to be that associated with the echo). Hence, we do not refer further to distinctions between auto- and cross-ambiguity calculation.

As we will show by example, many applications require integration time T and input bandwidth B that are orders of magnitude greater than the domains of uncertainty of τ(DTO) or f(DFO) that need to be searched; $N = 10^4$ to 10^6 data samples are often involved. When near real-time calculation is desired, brute-force calculation even over the limited domains of τ and f is usually totally impractical. Even "fast" techniques like using an FFT convolver for correlation are efficient only when results are wanted for all possible values of time lag τ within the integration time T. This paper describes useful algorithms for large N.

In Section II, we present an intuitive characterization of the ambiguity calculation process that points the way to efficient processing *algorithms.* Sections III and IV discuss typical processing parameters and related aspects of performance. Then,

Manuscript received February 8, 1980; revised October 30, 1980.

The author is with SCPE, Inc., Newton Centre, MA 02159.

[1]We consider here only the narrow-band radio case where all signal spectral components can be assumed to be shifted by the same value as the carrier or center frequency.

[2]In various applications, other common names for DTO are DTOA (differential time of arrival) or TDOA (time difference of arrival), and DFO is often called FDOA (frequency difference of arrival) or simply DD (differential Doppler).

[3]Note that both positive and negative frequency lines of the complex envelope are translated in the same direction by multiplication by $\exp(-j2\pi ft)$.

0096-3518/81/0600-0588$00.75

Section V describes the algorithmic approach that preserves all the desirable features, with minimal processing burden.

II. Interpretation as Filtering of a Mixing Product

We observe that for each τ, ambiguity function calculation can be regarded as operating on a product waveform (with τ as a parameter),

$$r(t;\tau) = s_1(t)\, s_2^*(t+\tau). \tag{2}$$

We call $r(t;\tau)$ a *mixing product;* it is the complex envelope that results if the real bandpass signals represented by $s_1(t)$ and $s_2(t)$ are offset in frequency and combined in a mixer, with the difference frequency component extracted. To preface the more general statement, consider first an autocorrelation function,

$$\psi_u(\tau) = \frac{1}{T}\int_0^T u(t)\, u^*(t+\tau)\, dt.$$

For typical waveforms that contain no hidden periodicities, $\psi_u(\tau)$ has a single "correlation lobe" peaking at $\tau = 0$, whose width (versus τ) is of order $1/B$ where B is the nominal width of the power spectrum of $u(t)$. We can interpret this result in terms of the mixing product

$$r_u(t;\tau) = u(t)\, u^*(t+\tau) \tag{3}$$

as follows. For values of τ around $\tau = 0$, $r_u(t;\tau)$ contains a nonzero dc value. The simplest example is that at $\tau = 0$, the product is $|u(t)|^2$, whose dc (mean) value is in fact the mean power in $u(t)$. The product continues to have a dc value for other values of τ, although the value decreases as τ increases away from 0, and eventually vanishes for τ well outside of $(-1/B,\ 1/B)$. The mixing product also contains a broad-band term, with generally continuous spectrum covering the range $(-2B,\ 2B)$. The operation $(1/T)\int_0^T [\ ]\ dt$ which forms $\psi_u(\tau)$ from the mixing product can be viewed simply as a low-pass filter, with unity gain at dc and bandwidth of order $1/T$. Larger T means a narrower filter and less background energy passed. The background energy is just the statistical fluctuation component in the measurement of $\psi_u(\tau)$. As is well known, the longer one integrates, the closer the result is to the "true" value of the autocorrelation function.[4]

Returning to (1) and (2), since $s_1(t)$ and $s_2(t)$ include noise, $r(t;\tau)$ includes broad-band signal $\times$ noise and noise $\times$ noise products whose level is reduced by the low-pass filter. Moreover, consider the case when $s_1(t)$ and $s_2(t)$ contain two differently delayed and frequency-shifted versions of some $u(t)$, say

$$\begin{aligned} s_1(t) &= u(t)\exp(j2\pi f_1 t) + n_1(t) \\ s_2(t) &= u(t-D)\exp[j2\pi f_2(t-D)] + n_2(t). \end{aligned} \tag{4}$$

[4] This reference to the "true" mean correlation over an ensemble average for a random signal should not be confused with our application objective. In DTO/DFO applications, one is usually interested only in the fact that the longer the integration, the better the discrimination against background additive noise. The values of τ and f will be correctly determined from the peak of the correlation of whatever waveform is present over the observation period.

Then, the desired term in the mixing product has the form

$$u(t)\, u^*(t+\tau-D)\exp[j2\pi(f_1-f_2)t]\exp[-j2\pi f_2(\tau-D)]. \tag{5}$$

When $f_1 - f_2 \neq 0$, the dc term in the product is replaced by (or "gives rise to") a sinusoid at the difference frequency $(f_1 - f_2)$. The ambiguity operation on the mixing product, $1/T\int_0^T [\ ]\exp(-j2\pi ft)\, dt$, is correspondingly a bandpass filter centered on frequency f, or equivalently a heterodyning of $r(t;\tau)$ downwards by an amount f, followed by low-pass filtering. When f matches $f_1 - f_2$ to within $1/T$, the filter has significant output, peaking when $f = f_1 - f_2$. Note that τ must also match the differential delay D to well within the reciprocal of the signal bandwidth $1/B$ in order that the mixing product itself have any significant magnitude. The relationships are pictured in Fig. 1.

The filters we have shown are bandpass or low-pass integrate-and-dump (I&D) filters. While an I&D filter precisely matches the specification of the calculation of an ambiguity function, obviously many other filters will give a useful result. Indeed, if it is only desired to determine D and $f_1 - f_2$ by finding the values of τ and f that cause the filter output to peak, and the actual value of the correlation function is secondary, any reasonable filter can be chosen to smooth on the mixing product. In particular, when the signal bandwidth B is much wider than the final filter bandwidth $(1/T)$ of the ambiguity calculation, the digital processing burden can be greatly reduced by realizing the filter as a cascade of successively narrower filters, with reduction in the sampling rate between stages commensurate with the filtering that has already been accomplished. However, any sample rate reduction must avoid aliasing (foldover) of any significant nonvanishing spectral components.

To illustrate this filtering concept, consider the example detailed later in Section V. Assume two 1 MHz bandwidth received signals plus noise represented by complex samples at a 1.25 MHz rate, and assume that the differential frequency offset can never exceed ±1 kHz. For any τ of interest, the mixing product

$$r(t;\tau) = s_1(t)\, s_2^*(t+\tau) \tag{6}$$

can first be processed by a low-pass I&D filter that integrates over 128 samples at a time. This filter will have a (two-sided) −3 dB width of approximately 10 kHz (hence certainly passing up to ±1 kHz efficiently), with nulls at multiples of 10 kHz. The natural output rate (no overlapping of I&D operations) is 10 kHz, and there is relatively little out-of-band energy folded onto the (−1, 1) kHz band of interest. Any further ambiguity function processing for that value of τ uses data corresponding only to the 10 kHz sampling rate, a 128-fold reduction in storage and processing requirements from the original 1.25 MHz rate.

Note that while referring to *down-sampling*, we are describing intermediate data in the computation of $A(\tau, f)$ at one value of τ, and over a limited range of f. We are *not* describing how closely spaced the values of τ are to be taken. Generally, computation will be carried out at discrete values of τ defined by the input sample interval. This guarantees that at least three equally spaced values of τ are well within the width of

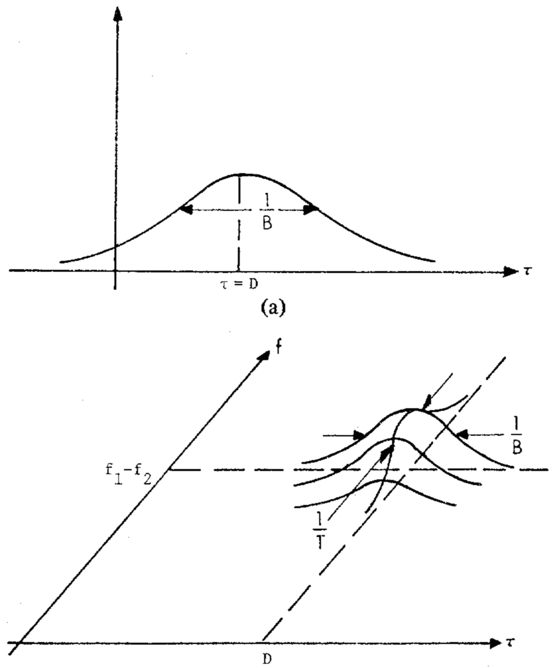

Fig. 1. Correlation and ambiguity functions. (a) Cross correlation. (b) Cross ambiguity.

the correlation lobe, and from these one can easily interpolate for even more precision in the DTO estimate. For example, with three values, a parabolic fit may be used to determine the location of the peak. With calculations at more than three values of τ, more elaborate interpolation filters can be designed.

The filter interpretation also makes evident another useful and exploitable characteristic. Suppose that at some lag value τ, the differential frequency offset that maximizes $A(\tau, f)$ is $\Delta f = f_1 - f_2$. When we define a filter centered on some f that is sufficiently close (within $\pm 1/T$) to assure a detectable output, the continuous filter output will represent complex samples of a sine wave at frequency $f - \Delta f$. This residual frequency difference can be estimated by using a digital discriminator to measure the frequency of this sine wave (digitally, by measuring the phase difference between successive samples). If Δf is changing slowly, this measurement may be more convenient (and more accurate) than finding the peak by interpolation between the outputs of filters computed at several different values of f.

III. Signal-to-Noise Ratio and Accuracy

For each τ, f, the output resulting from an ambiguity function calculation contains a desired signal × signal term, and an output noise consisting of unwanted noise × signal and noise × noise. In order for the desired lobe peak to be uniquely identified (very low probability of spurious noise lobes exceeding a detection threshold), the SNR in the output has to exceed about 10 dB. Within this restriction, the errors in locating the lobe peak are due to output noise perturbations. The estimates are unbiased and have a variance that achieves the Cramér–Rao bounds when any of several reasonable techniques is used for identifying the apparent location of the peak. The accuracies (standard deviation) are given by

$$\sigma_{\mathrm{DTO}} = \frac{1}{\beta} \frac{1}{\sqrt{BT\gamma}} \tag{7}$$

$$\sigma_{\mathrm{FDO}} = \frac{1}{T_e} \frac{1}{\sqrt{BT\gamma}} \tag{8}$$

where

B = noise bandwidth at receiver input, assumed same for both receivers;
β = "rms radian frequency" in the received signal spectrum (9);
T_e = "rms integration time" (10);
γ = effective input signal noise ratio (13).

By definition

$$\beta = 2\pi \left[\frac{\int_{-\infty}^{\infty} f^2 W_s(f)\, df}{\int W_s(f)\, df} \right]^{1/2} \tag{9}$$

where $W_s(f)$ is the signal power density spectrum, as shaped by the receiver and defined to have zero centroid, and

$$T_e = 2\pi \left[\frac{\int_{-\infty}^{\infty} t^2 |u(t)|^2\, dt}{\int_{-\infty}^{\infty} |u(t)|^2\, dt} \right]^{1/2} \tag{10}$$

where again $|u(t)|^2$ is defined to have zero centroid. For a rectangular spectrum, for example,

$$\beta = \frac{\pi}{\sqrt{3}} B_s \approx 1.8 B_s \tag{11a}$$

where B_s is the signal RF bandwidth, so that

$$\sigma_{\mathrm{DTO}} \approx \frac{0.55}{B_s} \frac{1}{\sqrt{BT\gamma}}. \tag{11b}$$

Similarly, for a constant energy signal with ambiguity processing carried out over only a segment of length T (a "gated segment"),

$$T_e = 1.8T \tag{12a}$$

$$\sigma_{\mathrm{DFO}} \approx \frac{0.55}{T} \frac{1}{\sqrt{BT\gamma}}. \tag{12b}$$

The effective input SNR is defined by

$$\frac{1}{\gamma} = \frac{1}{2} \left[\frac{1}{\gamma_1} + \frac{1}{\gamma_2} + \frac{1}{\gamma_1 \gamma_2} \right] \tag{13}$$

where γ_1 and γ_2 are the SNR's in the respective receivers (in the noise bandwidth B).[5]

[5]Note that γ appears in the accuracy formulas always multiplied by B, so that the results are the same whether B represents the equivalent noise bandwidth defined by the signal or some larger bandwidth, so long as the noise term in γ always refers to the noise level in the bandwidth B.

The quantity $BT\gamma$ can be viewed as the effective output SNR resulting from the ambiguity processing, with γ improved by the BT product of the processing. As stated earlier, the qualification in the stated formulas is that no matter how small the value of γ, $BT\gamma$ is well above unity. While the results are derived for the case where the input is perturbed only by additive Gaussian noise, because of the large amount of smoothing, the results apply equally well when SNR's are computed with respect to any other independent additive disturbances. While the results above are actually lower bounds on the standard deviation, it has been shown in repeated experiments that they can actually be achieved in reasonable implementation,[6] even when γ itself is well below 0 dB.

When both γ_1 and γ_2 are large (well above 0 dB), the third term in (13) is negligible; if also $\gamma_1 = \gamma_2$, then $\gamma = \gamma_1 = \gamma_2$, while if one is considerably smaller than the other, the effective γ is 3 dB above that smaller value (e.g., $\gamma = 2\gamma_1$ when $\gamma_1 << \gamma_2$ and $\gamma_2 >> 1$). When γ_1 and γ_2 are both much less than 0 dB, the effective input SNR is approximately

$$\gamma = 2\gamma_1\gamma_2. \tag{14}$$

For example, when both γ_1 and γ_2 equal -10 dB, $\gamma = 0.016$ or -18 dB. To cite some specific illustrative values, we consider two examples, both stated for 0 dB effective input SNR:

a) $\gamma = 0$ dB, $T = 1$ ms, $B = B_s = 1$ MHz

$\sigma_{\mathrm{DTO}} = 17.4$ ns

$\sigma_{\mathrm{DFO}} = 17.4$ Hz

b) $\gamma = 0$ dB, $T = 10$ s, $B = B_s = 25$ kHz

$\sigma_{\mathrm{DTO}} = 44$ ns

$\sigma_{\mathrm{DFO}} = 1.1(10)^{-4}$ Hz.

Note the major difference: in order to approach the same σ_{DTO}, a narrow bandwidth situation requires very long integration time, which in turn leads to a very low associated σ_{DFO}. For instance, consider an objective of $\sigma_{\mathrm{DTO}} = 1$ μs and $\sigma_{\mathrm{DFO}} = 1$ Hz, on a 1 MHz bandwidth signal at input SNR's of $\gamma_1 = \gamma_2 = -10$ dB ($\gamma = -18$ dB). The value $T = 1$ ms would result in accuracies 9 dB poorer than a),

$\sigma_{\mathrm{DTO}} = 140$ ns

$\sigma_{\mathrm{DFO}} = 140$ Hz.

Because of the wide bandwidth, the σ_{DTO} goal can still be achieved handily with $T = 1$ ms, but σ_{DFO} requires more processing. However, even the σ_{DTO} statement is too facile, since it is "achieved" at a value $BT\gamma = 12$ dB which may be marginal for spurious peak considerations as cited earlier. For the latter purpose, a more desirable effective processing SNR would be $BT\gamma = 20$ dB, achieved by increasing the processing time T to $T = 6$ ms. The value σ_{DTO} is then decreased to about 56 ns and σ_{DFO} is decreased to $\sigma_{\mathrm{DFO}} = 9$ Hz. To reduce σ_{DFO} to 1 Hz requires an even longer processing time, to the order of $T \sim 25$ ms.

A potential problem in processing over very long intervals, as in the second example b), is that neither the differential frequency nor the differential delay may remain sensibly constant over that long a period if moving platforms are involved. If either parameter changes during the interval, the ambiguity function lobe will be smeared. For example, a linear change in the differential frequency offset will result in a mixing product that can be described as a chirp (linear FM) tone over the interval, rather than a fixed frequency, and the filter responses will reflect this as line broadening. It can be shown [3] that for a linear change in differential delay over the integration interval, the peak in τ of the calculated lobe corresponds to a centroid value of the differential delay, where, if there is any change in signal energy level over the interval, the centroid is calculated by energy-weighting the changing delay versus time. Presumably, the same is true of the peak in the f-direction.

In many situations, it will be sufficient to use these centroid interpretations. However, more elaborate processing (Section V-C) can compensate for the effects of change. The appropriate generalization of the ambiguity calculations is 1) form mixing products over the observation interval with a changing value of τ, seeking a variation that maintains a fixed alignment between the desired signal components in the received versions of $s_1(t)$ and $s_2(t)$; and 2) consider matched filters in which the value of f used to define the filter center frequency is gradually changed, in effect making the filter into a more general matched filter (like a chirp filter). These generalized calculations recognize that the processing technique builds up coherent integration over an interval of length T by compensating for changes in τ and f, so that one is really finding a (τ, f) curve that best fits the observed data over the interval $(0, T)$. Indeed, the optimum approach when parameters are dynamic is to treat the *entire* problem (from RF through final estimates) as one of nonlinear parameter estimation over the total set of observed data. However, the problem often has to be partitioned because of the mass of data that would be involved, and the practical solution is a measurement system whose outputs can in turn be used for nonlinear parameter estimation over even longer intervals.

IV. Additional Characteristics of Ambiguity Function Calculations

We describe here some additional aspects of ambiguity function analysis that often arise in particular applications.

A. *Predetection versus Postdetection Processing*

Up to this point, we have been discussing so-called predetection processing. The ability to correlate is based on a perfect match-up (to within a constant phase) of two versions of the same RF or IF signal, and is totally independent of the detailed signal contents. In some applications, one might instead consider first demodulating the RF (or IF), followed by correlation of the results. This results in a loss in potential accuracy in most applications for several reasons.

[6] The accuracy formulas are very similar to familiar radar accuracy formulas (e.g., [2]) for matched-filter measurement of echo delay and Doppler shift. The effective SNR is 3 dB smaller than in the usual radar statements because here both inputs to the correlator are noisy. A direct derivation has been shown for the DTO problem [3].

1) At low SNR's, detection can further lower the SNR appreciably, particularly if FM demodulation must be used rather than AM demodulation. This problem would be particularly severe at below unity RF signal-noise ratios.

2) If there is any multipath associated with either received signal in a crossambiguity application, the demodulated signal likely will be distorted differently at the two receivers, leading to a significant reduction in correlation. In contrast, predetection processing may quite well result in well-separated and distinguishable ambiguity lobes. This concern is greater for continuous or high-duty cycle waveforms than it is for pulsed or low-duty cycle waveforms.

3) Postdetection processing of the signals would completely eliminate carrier frequency shift, eliminating the possibility of DFO measurements when these may be useful to system applications.[7] It would also eliminate the possibility of using the differential Doppler as a means of separating multiple ambiguity lobes (discussed below).

Nevertheless, under some circumstances, postdetection processing may be used. The processing reduces to correlation, but the basic formula for σ_{DTO} still applies, with all bandwidth and SNR parameters now describing the demodulated signal.

B. Interference

Since the DTO/DFO processing is independent of signal structure, an interferer present within the receiver bandwidth (in both receiver inputs, for cross-ambiguity processing) is as valid a signal for ambiguity processing as is the intended signal. Hence, it will give rise to its own ambiguity lobes. Frequently, these are quite distinct in DTO or DFO from those of the intended signal, and may be properly interpreted from refined estimates on each lobe; in the latter case each interferer present increases the processing burden proportionately.[8]

If the interferer is "close" to the desired emitter in both DTO and DFO, its ambiguity lobe may overlap and mask or confuse the lobe for the desired signal. Since closeness in both DTO and DFO usually implies the interferer being physically close to the desired signal transmitter as well as at the same frequency, this will often not be the case because of the interference they would cause to each other.

One common interference situation is that of narrow bandwidth interferers when one is trying to process wider bandwidth signals. The ambiguity lobe of a narrow bandwidth interferer may be quite extended as a function of τ. For example, 10 kHz bandwidth implies 100 μs nominal delay width of the lobe, and 2 kHz bandwidth implies 500 μs width. These values are large enough to bracket significant DTO ranges in which the desired signal lobe may fall, and imply that, most of the time, any use of lobe separation to discriminate against narrow-band interferers will have to be based on differential frequency resolution, as available through motion of either the emitters or receivers. For example, if the desired emitter is in high speed motion while most of the interfering emitters are fixed, the potential for discrimination is greatly enhanced.

Even when lobe masking is not an issue, the energy level of the interferers always contributes to output noise. That is, an ambiguity lobe for a desired signal is built up in value by coherent addition over the processing time T. Any other signal present that does not have a significant lobe in that (τ, f) region nevertheless builds up a value noncoherently versus T, just as does noise. Hence, any interference energy present in either receiver input will show up in output noise, in interferer $\times$ interferer, interferer $\times$ noise, or interferer $\times$ signal terms. Each term represents a different kind of broad-band spectrum in the mixing product (e.g., a narrow-band interferer present in both receives contributes a broad-band spectrum of width double the interferer bandwidth). The BT product processing improvement applies to these interference terms just as it does to the additive wide-band noise. Thus, any unremoved interference power in the acceptance bandwidth of either receiver *should be included* in calculating the input SNR, γ_1 or γ_2. Realistically, the interference energy is significant only if it is at least comparable to the total additive noise energy in the bandwidth to which the signal can be filtered. If the latter is 1 MHz, for example, a 10 kHz bandwidth interferer will be significant if the interferer is 20 dB above the noise level in its 10 kHz bandwidth; and a 1 kHz bandwidth interferer only if it is 30 dB above noise in its local domain.[9] Since loss in γ can only be compensated by using greater processing time, which is usually undesirable, identifiable in-band interference should be eliminated to the greatest extent possible, by preprocessing prior to the DTO/DFO measurement system.

While excision of interferers by use of notch filters[10] may cause signal distortion, there is usually little effect on DTO/DFO accuracies (relative to the accuracies achievable in the absence of interferers). Excision will reduce the amount of useful signal energy. Moreover, if one receiver output is filtered differently from the other, some of the signal energy will no longer correlate and will effectively be converted to noise energy. Additionally, if differential Doppler shifts of the received signals are large, then even when the same filtering is applied to both received signals, it may cover noncorresponding regions of the signal spectrum and still increase the effective noise level.

As a practical matter, these considerations are important only when a large total fraction of the signal band is being excised to delete interferers. In fact, when the signal spectrum is roughly uniform over the band, the ratio of signal energy to

[7] In applications like sonar, with signal bandwidth of the order of the center frequency, Doppler effect will be observed as compression or expansion of the time base, and hence can still be observed on postdetection waveforms.

[8] At least until it is identified as an interferer, after which it can perhaps be rejected and only the desired signal lobe parameters investigated (or tracked, as discussed in Section V).

[9] For a very narrow-band interferer, the DFO of interest may lie outside the spectrum of the interferer $\times$ interferer term, in which case that term does not contribute to the output noise over the lobe of interest.

[10] Use of notch filters is intuitive. More sophisticated approaches, such as the Wiener filter to minimize the mean-square error relative to the desired signal, end up whitening the total undesired energy. When a narrow-band interferer is very strong in a localized spectral domain, the Wiener filter essentially places a notch over that localized region insofar as the signal is concerned.

additive wide-band noise energy is roughly the same in any spectral subband as in the whole band, and γ is actually unaffected. However, γ then refers to a smaller bandwidth, and hence the BT product of processing is smaller. For example, dropping 50 percent of the band deteriorates the processed SNR by about 3 dB (aside from eliminating interferers), and dropping 75 percent of the band causes 6 dB deterioration. In addition, the DTO suffers directly from diminution of the rms radian frequency β, if the band occupancy is significantly compressed. Interestingly, it is the rate of variations implied by the occupied signal band, not the solid occupied bandwidth, that governs this last effect. Thus, β will actually increase slightly in value if the interior portions of a rectangular signal spectrum are deleted, and only the extreme outer portions remain; on the other hand, it will certainly diminish if a solid half of the band is excised.

C. Use of Signal Subbands or Time Segments

Another approach to eliminating narrow-band interferers is to process several interference-free fractions of the signal band independently, and then combine the results. Use of subbands could also be attractive from an acquisition/tracking point of view, since the subbands have wider correlation lobes, making them easier to find and track. The time-domain analog is to process cross-ambiguity functions for time segments of some length $T_p = T/n$, and then to combine the results over n segments. We discuss both concepts here.

The time-domain version is perhaps easiest to picture. One problem is that to avoid identifying spurious lobes from examination of a time record of length T_p, $BT_p\gamma$ must be reasonably high, say exceeding 15 dB. If this is achieved, however, in principle the DTO estimates can be made from individual subrecords and averaged without loss of accuracy. Thus, for a subrecord of length T_p, the accuracy of a DTO estimate is

$$\sigma_{\mathrm{DTO}}(T_p) = \frac{1}{\beta} \frac{1}{\sqrt{BT_p\gamma}} \tag{15}$$

and averaging n such estimates, over a total duration $T = nT_p$, will reduce the standard deviation to

$$\sigma_{\mathrm{DTO}}(T) = \frac{\sigma_{\mathrm{DTO}}(T_p)}{\sqrt{n}} = \frac{1}{\beta} \frac{1}{\sqrt{BnT_p\gamma}} = \frac{1}{\beta\sqrt{BT\gamma}}. \tag{16}$$

Thus, noncoherent combining (averaging of the DTO estimates) gives precisely the same result as obtained from coherent integration over time T (but remember the caveat $BT_p\gamma \gg 1$). Unfortunately, the same is *not* true for DFO estimates, for which we would have

$$\sigma_{\mathrm{DFO}}(T_p) = \frac{1}{T_p} \frac{1}{\sqrt{BT_p\gamma}} \tag{17}$$

and

$$\sigma_{\mathrm{DFO}}(T) = \frac{1}{T_p} \frac{1}{\sqrt{BnT_p\gamma}} = \frac{1}{T_p} \frac{1}{\sqrt{BT\gamma}}. \tag{18}$$

Thus, for DFO, one would lose accuracy by the factor n by not coherently integrating over the full T-length interval.

Analogously, if the band is divided into n segments, each with noise bandwidth $B_p = B/n$ and rms radian frequency $\beta_p = \beta/n$, the averaging of DTO and DFO estimates fron n such subbands, each integrated over time T, would yield

$$\sigma_{\mathrm{DTO}} = \frac{1}{\sqrt{n}} \sigma_{\mathrm{DTO}}(B_p, \beta_p) = \frac{1}{\beta_p} \frac{1}{\sqrt{nB_pT\gamma}} = \frac{1}{\beta_p} \frac{1}{\sqrt{BT\gamma}} \tag{19}$$

and

$$\sigma_{\mathrm{DFO}} = \frac{1}{\sqrt{n}} \sigma_{\mathrm{DFO}}(B_p, \beta_p) = \frac{1}{T_e} \frac{1}{\sqrt{nB_pT\gamma}} = \frac{1}{T_e} \frac{1}{\sqrt{BT\gamma}}. \tag{20}$$

Here the DFO estimates would not deteriorate through the noncoherent averaging but DTO would. Incidentally, it should be recognized that coherent combining of the subband integrals will involve not only their complex envelope values but also the subband center frequency information (i.e., the values from the subband integrals must be added as data on spaced subcarriers).

V. Processing Algorithms

We now describe an algorithmic approach for efficient digital ambiguity processing in applications requiring long data segments. To be concrete, we will consider an example tailored to the wide-band example cited earlier:

signal bandwidth $B = 1$ MHz
sampling rate (complex) $f_s = 1.25$ MHz ($T_s = 0.8\ \mu$s)
accuracy requirements:
  DTO = 0.1 μs
  DFO = 1 Hz
maximum DFO uncertainty = ±1 kHz
integration time $T \approx 100$ ms.

Later, we shall comment as well on a typical narrow-band example such as cited earlier.

We will define two phases of processing that are generally needed (as in many estimation problems), which we term respectively the coarse (or acquisition) mode and the fine (or tracking) mode. The coarse mode generally involves short integration time, but a search over a wide domain in τ and f that typically can efficiently employ FFT convolution techniques. The burden is made tolerable by the short integration time, and by the limited use of this mode for initialization; its purpose is to narrow the domain in τ, f that has to be examined in the fine mode. In turn, the fine mode is the workhorse, containing the long integrations that result in DTO/DFO estimates with desired accuracies; its processing burden must be minimized, and it is here that we use the cascade filter concept of Section II with great effect. In an application where estimates are repeated or updated over longer intervals, the output parameter data can be entered into tracking (predictive) loops, so that it is self-sustaining (hence the name tracking mode), where reinitialization is required only to recover from a break in the data.

The fine mode is described first, under the assumption of constant DFO over each integration interval, and then the generalization for coping with varying DFO is described briefly. Subsequently, we describe the coarse mode.

A. Fine Mode (Constant DFO)

As stated, we take significant advantage here of the mixing product, cascade filter concept. We will define a two- or three-step cascade, the first step of which is a simple integrate-and-dump (I&D) filter operating on the mixing product for each value of τ that needs to be considered. Forming the mixing product (one complex multiply per input sample) and accumulating (one complex add) is about the minimum processing burden at the input data rate with which one can perform a useful function. More burdensome processing is then performed, but at the much lower I&D output rate. Clearly, the number of values of τ that must be considered must be small, and this is the function of the coarse mode or of tracking mode predictive operation.

1) Step 1–Partial Correlation with Bandwidth Reduction: The two input complex envelopes $s_1(t)$ and $s_2(t)$ are sampled at times $t = nT_s$, resulting in data streams $\{s_1(n)\}$ and $\{s_2(n)\}$. There is no loss in generality in assuming that the sets of samples are taken simultaneously for the two signals. The sampled-data version of the ambiguity computation defined in (1) is

$$A_R(m,p) = \sum_{n=RN}^{RN+N-1} s_1(n)\, s_2^*(n+m) \exp\left(-j2\pi \frac{np}{Q}\right). \tag{21}$$

In this representation, the ambiguity function is calculated at discrete time lags, $\tau = mT_s$, and at frequency offsets $f = p/QT_s$. The designation A_R denotes that this is the Rth successive block computation of the ambiguity function since the beginning of the data record. For our wide-band example, the size of each data block is $N = f_sT = 1.25(10)^5$. The values of the ambiguity function at values of τ, f other than the discrete grid set are available by standard interpolation, as we outline later, so that, in principle, use of a grid in no way hinders achieving the predicted accuracies. We assume use of frequency values spaced by 5–10 Hz, in order to simplify final interpolation to the desired 1 Hz range. This in turn implies a value $Q \approx 2^{17}$ or 2^{18}. Even if the exponentials do not themselves need to be represented to this 18-bit precision, indexing a table could be awkward for this large a value. Moreover, if implemented directly, (21) requires two full complex multiplies per input point for each value of p and m. Degrading the frequency resolution and assuming a greater interpolation burden (which is generally an unfavorable tradeoff, as we indicate later) would reduce the value of Q and the number of values of p over which the computation needs to be done, but would not reduce the number of operations for each value of p and m.

Instead, since the DFO uncertainty is so low compared to the sampling rate, we can simply form the mixing product for each m of interest (each τ), and apply as the simplest form of low-pass filter an unweighted integrate-and-dump (I&D) filter. For this purpose we accumulate (add) the mixing product over blocks of 128 samples each, representing $T_1 = 102.4\ \mu$s of real time (real time = clock time of the original analog data). This filter will have an output rate of about 10 kHz, a −3 dB bandwidth of about 10 kHz (±5 kHz) and first nulls at ±10 kHz. It will pass the desired difference frequency (no greater than ±1 kHz) with less than 0.2 dB loss (by contrast, integration over 256 points, resulting in a filter with −3 dB bandwidth of 5 kHz, would imply a possible loss up to 0.6 dB). Thus, in sampled data terms, for each m of interest we form the mixing product

$$r_1(n;m) = s_1(n)\, s_2^*(n+m) \tag{22}$$

and the I&D filter operation,

$$r_2(k;m) = \sum_{n=kN_1}^{kN_1+N_1-1} r_1(n;m) \tag{23}$$

where $N_1 = 128$ and k indexes the "output" sequence r_2. The values $r_2(k;m)$ represent sampled values every 102.4 μs of the complex envelope of a filter output that contains the desired difference frequency component within ±1 kHz around zero. Since the filter's first nulls are at ±10 kHz, the filter output sample rate of 10 kHz is below an effective Nyquist rate. If the sequence of samples is used for further filtering, some signal spectral foldover and aliasing will result. However, spectrum folded over into the interesting part of the frequency domain (from −1 kHz to +1 kHz) originates at multiples of ±10 kHz, where the I&D filter nulls occur. A straightforward calculation of the energy folded over from these regions, as weighted by the $[(\sin x)/x]^2$ filter power response, shows a negligible (<0.1 dB) increase for the additive noise components, while the self-noise resulting from the broad-band energy of the mixing product is more than 20 dB below the signal level.[11]

For each value of τ, the process just outlined requires calculation of a mixing product and a 102.4 μs partial-correlation sum (128 samples at 1.25 MHz). Assume, for reasons described later, that we need to do this in the fine mode for three values of τ that bracket the estimated correct value. If performed in real time, the processing burden is a 3.75 MHz complex multiply and add rate and the output data rate will be only 10 kHz (complex) for each value of τ, or a total of less than 30 kHz complex data rate for further processing.

2) Steps 2 & 3–Detailed Ambiguity Calculation: The remainder of the ambiguity calculation, at each delay $\tau = mT_s$, requires the computation of the forms

$$A_R(m,f) = \sum_{k=RK}^{RK+K-1} r_2(k;m) \exp(-j2\pi fkT_s') \tag{24}$$

for the selected values of f, where $T_s' = 102.4\ \mu$s and the index k represents samples occurring at a 10 kHz rate. Each value of f can be viewed as defining a bandpass I&D filter centered on that value of f. If we specifically select

$$K = 2^{10} = 1024$$

[11]With a significant increase in processing burden, overlapped windowed filtering of mixing products could be introduced to further reduce the foldover noise and, as well (because of filter bandwidth broadening), the attenuation of the desired spectral line.

we represent an integration interval of $T = KT_s' = 104.9$ ms. The basic resolution for a trigonometric look-up table would be $f_o = 1/T \approx 9.5$ Hz. The trigonometric coefficients would correspond to multiples of this value, $f = \nu f_o = \nu/T$, and have the form

$$\exp\left(-j2\pi \frac{k\nu}{K}\right).$$

Thus the look-up table would have 1024 entries (or 256 if sine and cosine for only one quadrant are represented in the table).

A number of interesting alternatives also present themselves here. For example, by increasing K in the trigonometric coefficients to 2048 or 4096 but keeping the same overall integration time (block length), we can define filters spaced 5 Hz or 2.5 Hz apart, respectively, although the basic resolution is still only 10 Hz. (This is the familiar step of zero-padding a DFT to give closer line spacing.) This could simplify the interpolation for a final estimate of DFO, at the cost of a few more filter calculations at the 30 kHz rate (for three values of τ). Another alternative would be an FFT over the entire 1024 point data block, thus giving for each τ all possible filters at 9.5 Hz spacing over a 10 kHz domain. The FFT will be computationally efficient as soon as more than about five such filters are wanted (say, to facilitate final interpolation). Again, spectrally overlapped filters can be obtained by zero-padding and use of larger FFT's. Three 1024-point or even 4096-point FFT's per 100 ms are well within the domain of readily available high-speed peripheral processors for real-time implementation.

Yet another approach is the equivalent of a Zoom FFT operation (but within the same overall resolution in the case at hand). Step 2 can be limited to being only another intermediate step in data reduction and in frequency resolution. For example, suppose the DFO is being tracked to within ±10 Hz. Then we can define a set of bandpass I&D filters centered roughly 10 Hz apart, with bandwidths around 100-200 Hz, and implement for Step 2 that filter centered closest to the last known DFO estimate. Again, the filter will pass the desired tone with less than 0.2 dB loss, and fold in a negligible amount of broad-band noise and signal products. Its processing is simply

$$r_3(q, f; m) = \sum_{k=qK}^{qK+K-1} r_2(k; m) \exp(-j2\pi f k T_s') \tag{25}$$

where the $\{r_3\}$ are values of the complex envelope of the filter output relative to its center frequency f. A convenient value is $K = 64$, giving -3 dB filter bandwidth of 153 Hz, and output sampling rate also 153 Hz, corresponding to integration over a total of 6.55 ms of real time. For this calculation, the discrete values of f can be selected as one of the set $f = \nu/8KT_s'$, which are separated by 19 Hz (or again, even finer spacings can be used). With ν representing the particular one of these frequencies that is closest to the estimated DFO, the calculated values would be the set

$$r_3(q, \nu; m) = \sum_{k=qK}^{qK+K-1} r_2(k; m) \exp\left(-j2\pi \frac{\nu k}{8K}\right) \tag{26}$$

where $K = 64$. The index q denotes the successive samples of the filter output complex envelope.

In the third step, then, the ambiguity surface will be calculated by setting narrower band filters on the 153 Hz sequence $\{r_3(q, \nu, m)\}$. We can suppress the ν-dependence in r_3 henceforth. Letting $T_s'' = 6.55$ ms be the sampling interval at this rate, we have

$$A_R(m, f) = \sum_{q=RQ}^{RQ+Q-1} r_3(q; m) \exp(-j2\pi f q T_s'') \tag{27}$$

where $Q = 16$, so that $QT_s'' = 104.8$ ms. Now we select as values of f the set $f = \mu/\alpha QT_s''$ so that for $\alpha = 1$, integer values of μ define filters spaced by slightly under 5 Hz, and larger integer values of α correspond to finer-spaced filters (or again, an FFT can be used). The ambiguity calculation, for a given μ, becomes

$$A_R(m, \mu) = \sum_{q=RQ}^{RQ+Q-1} r_3(q, m) \exp\left(-j2\pi \frac{\mu q}{Q}\right). \tag{28}$$

It should be noted that the actual frequency at which the ambiguity function is being calculated from the total three-step sequence is the sum,

$$\frac{\mu}{\alpha QT_s''} + \frac{\nu}{8KT_s'} \text{ Hz} \tag{29}$$

for the sample parameters given.

The fractional increase in complex computation rate due to doing the overall computation in three rather than in two steps amounts to about 153/10 000 or only about 1.5 percent. It can actually lead to a saving when many points need to be calculated on the ambiguity plane to achieve accurate interpolation of the DTO/DFO estimates from points on the sampling grid.

For the particular parameters of these calculations, the breakdown into a Step 2 and a Step 3 calculation is probably an overkill. With not much change in parameters or in desired accuracy, the breakdown can lead to significant benefits; it can also be quite beneficial when "acceleration filters" are introduced into the processing, a topic discussed later.

B. Interpretation of Fine Mode Calculations

Next, we address the question of just how many points need to be computed on the ambiguity plane in the fine mode, and how DTO and DFO are extracted from the results. First, suppose the desired accuracies and precisions are no finer than the spacing between grid points on the ambiguity plane, as defined by integer values (m, μ) at Step 2. Then calculations are needed only at the point (m, μ) that is closest to the best prior estimate $(\hat{\tau}, \hat{f})$, and at the points on either side in the τ and f direction as diagrammed in Fig. 2 (we assume the rate of change of these parameters to be small enough that they will not change by more than one grid spacing from the previous estimate). Thus, one would compute A_R at the nine pairs defined by m, $m \pm 1$ combined with μ, $\mu \pm 1$. These discrete values of τ, f are already sufficiently precise so that it would suffice to select the coordinates corresponding to the largest magnitude of A_R as the parameter estimates. If the selected

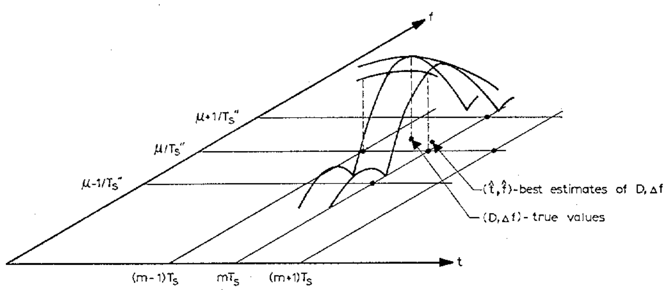

Fig. 2. Ambiguity plane–points of computation.

value of m or μ differs from the last previous estimate, in a continuing series of estimates, the next calculations would be made on a set of points shifted to center on the new values. A problem with this approach, although it is simple, is that the output DTO and DFO estimates would change stepwise, while the true values are changing continuously. Without implying any greater accuracy, it is desirable to have finer grained estimates, particularly if they are used to drive parameter tracking loops with the continuous loop outputs used as the system readout.

Finer grained data implies interpolation among the discretely calculated ambiguity values to determine location of the peak of the surface. However, we must recognize for these purposes a difference in the character of the ambiguity function surface in the τ- and f-directions. In the τ-direction, the lobe width is defined by the reciprocal bandwidth of the signal so that, with T_s smaller than the reciprocal bandwidth (as it must be, if Nyquist rate sampling rules have been observed), significant values are likely to be found over regions $\pm T_s$ from the peak. Hence, we can feel confident that some higher order curve can be used for interpolation to find the peak in the τ-direction; in particular, a curve fit as simple as a three-point polynomial (parabolic) fit may be adequate, since the ambiguity function has parabolic behavior around the peak (zero derivative) in both the τ and f directions (but for high accuracy, a higher order interpolation may be necessary). In the f-direction, the lobe is created by the integration and when the lobe happens to be almost centered in a particular filter, the adjacent f-direction values may be near zero (a null of the $\sin x/x$ filter response). Hence, in the f-direction, it may be preferable to use the adjacent values only to assure that of the three adjacent f-domain values, the value expected to be the largest is indeed the largest; and then to estimate the DFO by using the phase-change (discriminator interpretation) from successive ambiguity surface calculations, as outlined earlier in Section II. The discriminator calculation can be described as follows. Let $A_R(m, \mu)$ and $A_{R+1}(m, \mu)$ be the values calculated at (m, μ) for the Rth and $(R+1)$st intervals (each $T \approx 100$ ms long in our example). Then the frequency offset from the value denoted by μ is estimated by

$$\Delta\hat{f} = \frac{1}{2\pi} \frac{\arg A_{R+1}(m, \mu) - \arg A_R(m, \mu)}{T}. \tag{30}$$

When the value of m or μ for occurrence of the peak changes between blocks, this computation must still use both results for this last determined value of m, μ; our assumption that m or μ changes by no more than one increment from block to block guarantees that the appropriate sets of data will always be available. Another approach to f-domain interpolation is to use time weighting (windowing) as part of the calculation, thus sacrificing some resolution and some SNR in each calculation because of the widened filter bandwidth but assuring adjacent nonzero values that can be used in a standard interpolator.

When we estimated the Step 2-3 processing burden earlier, it was cited for three values of m, and for one value of μ for each m. Now, we have indicated two additional values of frequency $(\mu+1, \mu-1)$ at each m value. These additional values can clearly be only Step 3 calculations, which add negligible burden to that estimated earlier.

In a narrow-band example outlined later, the parameters are such as to require more elaborate interpolation.

C. Parameter Changes During Integration Interval

The basic ambiguity function calculation represented by (1) or (21) assumes that DTO and DFO remain constant over the total integration interval. Obviously (except for zero Doppler shift), these assumptions are incompatible in the strictest sense, since a nonzero differential Doppler shift implies a nonzero rate of change of DTO. We have already outlined the concept of using tracking loop predictions to provide more accurate compensation for changing delay or changing Doppler shift in a sequence of parameter estimates. Here we outline possible elaborations in the processing when such changes are rapid enough to possibly impact on the accuracy of the calculation.

First, consider differential delay. Assume there exists an estimate $\tau(0)$ at the beginning of a processing interval $(0, T)$, and an estimate of how τ changes over the interval. Then, instead of using fixed values of m in the ambiguity calculation, we can let m vary from block to block in the Step 1 calculation in (23), changing discretely at time blocks of 102.4 μs, but kept constant within each block. (One could for this purpose choose an even shorter time block for the Step 1 calculations, or even *interpolate* between values of m to obtain values of r_2 for time delays that more precisely match the predicted trend.)

Steps 2 and 3 also require compensation for the predicted variation in differential Doppler shift. If we confine discussion to at most a constant acceleration, the appropriate computation is to replace the previous "Doppler filters" by *acceleration filters* (really Doppler plus acceleration filters) of the form

$$\int [\ ]\, \exp(-j2\pi ft) \exp(-j\pi\alpha t^2)\, dt.$$

We assume the Step 2 and Step 3 mode of computation. In Step 2, in our example, we applied a 153 Hz bandwidth filter with center frequency selected from a set of values spaced by 19 Hz. Suppose that the differential Doppler will not vary over the total fractional second integration interval of Step 2 by more than a fraction of the 153 Hz. In this case, Step 2 could process the r_2 sequences exactly as prescribed in (26), since the varying DFO tone will be passed by the selected filter with essentially no distortion. The acceleration terms need only be introduced in the final narrow-band filter calculation represented by (28). This calculation would now be represented by a form such as

$$A_R(f) = \sum_{q=RQ}^{RQ+Q-1} r_3(q) \exp(-j2\pi q T_s'') \exp[-j\pi\alpha(qT_s'')^2]. \tag{31}$$

Because of the low processing burden, one might calculate for a number of different acceleration values for each "delay."

As represented here, there is no significant processing burden increase at Step 2, and Step 3 calculations are so negligible a burden that even a large factor increase is ignorable. However, to allow for extreme values of acceleration, one may need to consider adding them in at Step 2. This would increase the processing burden at that point to allow interpolating in frequency to follow a more continuous value than the discrete values in (28), in a manner analogous to the suggested treatment of the varying DTO. Note that the multiple-step approach to ambiguity calculation allows one to consider processes such as acceleration filtering without facing an exorbitant increase in system processing burden, because of the great downsampling after the first step.

When acceleration filters are used, the phase change between successive ambiguity function calculations represents the accumulated total phase differential over the integration time T between the actual phase of the mixing product tone produced by the Step 1 operations and that to which the filter is matched. It is specifically the phase differential accumulated at the *average* frequency differential between the filter and the mixing product. One would probably use this value to update the frequency estimate, and use only a fraction of it also to correct the estimated acceleration. A sudden acceleration change (a step function) would be followed by a tracking loop output only as some smoothed lagging response to a step.

As a summary on this aspect of the measurement problem, we are commenting on processing techniques that have not been adequately explored experimentally to the best of our knowledge. However, they appear to be on as firm a ground as most of the rest of the processing we have been discussing, where (unpublished) laboratory experimental results closely followed predictions for input SNR's as low as -20 dB.

D. Coarse Mode

The coarse or acquisition mode is needed to provide initial estimation of parameters with sufficient accuracy to enter the fine or tracking mode, or to reinitiate the tracking mode whenever tracking is lost for any reason. In either case, it is not generally a continuous operation. Hence, although coarse estimation or acquisition must be limited to a reasonable processing burden, it is defined so as to have little input on total processing system design.

The need is to search all possible positions in the ambiguity plane, with sufficient processed SNR to assure reliable detection (with low false alarm rate) of the correlation lobe. For example, it was indicated in Section III that for 1 MHz bandwidth signal, even with input SNR's of -10 dB, 5 ms integration intervals should provide adequate processed SNR, although not narrowing the lobe in the f-direction as finely as one would want for fine mode operation.

The range of DTO and DFO parameters to be searched can usually be bounded from prior geometric considerations. In our wide bandwidth example, we took the DFO to be reasonably bounded to ±1 kHz. The bandwidth corresponding to 5 ms integration is 200 Hz, so that in this case, up to ten Doppler cells would have to be searched for every time lag. The shorter the coherent integration time, the wider the Doppler bandwidth, and the fewer the number of cells that have to be searched. This makes attractive the use of shorter integration intervals in the coarse mode, with noncoherent summing to build up effective SNR, as discussed in Section IV-C.

We will actually outline a two-step acquisition procedure for parameter acquisition. The first step uses an FFT convolver to achieve DTO measurement with only coarse measurement of the differential Doppler shift. The second Step uses FFT spectral analysis of the mixing product at the determined time delay to provide precise measurement of the differential frequency (essentially identical to the Step 2 fine mode process).

The first step requires integration time long enough to provide an adequate processed SNR to assure reliable detection of the lobe, yet sufficiently short to be convenient from a processing viewpoint, and to require a small number of Doppler cells to be searched. In the convolver (to avoid circular convolution effects), the FFT's will use 50 percent overlap with one of the inputs zero-packed over 50 percent of the interval. For our example with 1.25 MHz sampling rate, a 2048-point FFT would correspond to using a data length of 819.2 μs for convolution. With this value, the frequency-direction width of an ambiguity lobe will be 1220 Hz (±610 Hz). If the initial differential Doppler uncertainty is larger than this value, the ambiguity will have to be calculated for several different values of frequency, spaced, say, by multiples of 610 Hz (to be consistent with the spacing of the FFT lines). Moreover, the differential time delay will be searched at all multiples of the equivalent complex sampling rate, in blocks of 819.2 μs. Again, if the initial differential delay uncertainty is larger than this value, many such blocks will have to be searched.

The specific calculations will be based on use of the now familiar ambiguity formula that we write in the sampled-data form, with its FFT convolution equivalent:

$$\begin{aligned} A_R(qN+m;\nu) &= \sum_{n=\frac{RN}{2}}^{R\frac{N}{2}+\frac{N}{2}-1} \hat{s}_1(n)\,\hat{s}_2^*(n+qN+m) \\ &\quad\cdot \exp\left(-j2\pi\frac{n\nu}{N}\right) \\ &= \sum_{k=0}^{N=1} S_1(k+\nu;R)\,S_2^*(k;R+q) \\ &\quad\cdot \exp\left(-j2\pi\frac{mk}{N}\right) \qquad m=0,\cdots,\frac{N}{2}. \end{aligned} \tag{32}$$

The notation $S_1(k; R)$, $S_2(k; R)$ refers to the Rth block of FFT's, one of which consists of N points (selected with 50 percent overlap), while the other consists of $N/2$ points (augmented by $N/2$ zeros). A_R refers to the ambiguity surface calculation which combines the Rth block taken on s_1 with the $(R + q)$th block of s_2 when the delay being tested is in some range other than (0, 819.2) μs; the delay at which A_R is being calculated is $\tau = (qN + m)\, T_s$ where T_s is the sampling interval for the complex envelope representation of the data. The frequency at which A_R is being calculated is $f = \nu/2T = 610\nu$ Hz since $T = 819.2$ μs. The basic computation will utilize the RHS of (32), with q and ν considered fixed (i.e., a separate set of computations for each q and ν). An FFT performed on the cross-spectral product will give A_R for all values of m (from 0 to $N/2$), for that given q and ν. Finally, the values of A_R will be calculated for the same set of parameters over a number of blocks of the order of 256 (to be compatible with the frequency search, as described below) and the magnitudes (or squared magnitudes) are averaged. The averaged values will be compared to a preset threshold, based on noise and signal level estimates, to determine whether any significant ambiguity lobes can be seen. The search would start with the most likely values of q and ν, and proceed with other values until such a lobe is found with some confidence or until all regions of interest have been searched.

For each lobe found, there is now an accurate estimate of time delay for entering the fine mode. To obtain a suitable estimate of DFO, one can form the mixing product at *that* time lag, and subject it to an FFT with about 5 Hz resolution. However, the intermediate data already included filtering the mixing product at that time delay and at the value of ν found most appropriate, with filter output represented by the sequence of complex values of A_R for that ν and delay. Therefore, simply, that sequence can be subjected to a final 256-point FFT to obtain the final desired frequency resolution, about 5 Hz, by determining the largest FFT value.

We must recognize that just as in the fine mode, the differential Doppler shift may change over time intervals of the order of 100 ms, and so may the differential delay. For purposes of acquisition, we believe it is satisfactory to assume that such changes are linear, and that the centroid interpretation mentioned earlier in Section III is sufficient for these purposes. To validate the initial coarse estimate, it would be desirable to carry out fine mode calculations on the *same* set of input data, at the acquired parameters. This can obviously be done in a multiple pass procedure, or by use of sizable buffer storage.

The major processing burden for the coarse mode is obviously in the first step, since the second step involves only one 256-point FFT. The first step involves FFT's on the individual signals, a spectral product for each block, and another FFT on the product. The assumption is that the convolutions will not have to be carried out in real time, unless the search range is minimal, and that either a multiple pass on the input data is available if overall operation is to be near real time (i.e., the coarse mode burden being negligible relative to a succession of many fine mode calculations), or that parameter changes are slow enough to allow a gap in the input data handled.

E. Narrow Bandwidth Example

To show another example of this kind of processing, with some different properties, we turn now to the narrow-band example suggested earlier, and add a different class of accuracy requirements. The parameters are

signal bandwidth = 25 kHz
sampling rate (complex) = 50 kHz (T_s = 20 μs)
accuracy requirements:
DTO: 40 ns
DFO: 0.001 Hz
maximum DFO uncertainty: ±100 Hz
integration time: 10 s.

The number of samples in an integration interval is $N = 5(10)^5$.

A major difference between this example and the previous wide-band case is the large discrepancy between the sampling interval (20 μs) and the precision (<40 ns) needed to be commensurate with the accuracy goal. Since in the fine mode (or in any calculation with the sampled data), the time lags will be 20 μs apart, a very fine interpolation will be necessary, locating the correlation lobe peak to 1/500 of the interval width. Similarly, if 0.1 Hz is the basic frequency resolution afforded by 10 s integration, the 0.001 Hz DFO requirement implies 1/100 splitting. In both cases, these are quite consistent with the high SNR built up by the processing; even for $\gamma = -10$ dB, $\sqrt{BT\gamma} = 158$. The interpolation of values over a cut along the τ-direction of an ambiguity function is basically that implied for the reconstruction of an analog waveform by the sampling theorem. In an unpublished study, it was found that roughly 11-15 sample values are needed with a raised-cosine filter interpolator to assure a 1/1000 interpolation with negligible added error. This implies fine mode calculations at probably 10 values of time lag for 1/500 interpolation. A similar statement applies for interpolation in the frequency direction.

The interpolation situation is pictured in Fig. 3, showing several samples and the reconstructed analog correlation function that they represent. It is, of course, only necessary to interpolate those values needed for testing to locate the peak. One simple algorithm is as follows. Select the three largest values (R_k, R_{k+1}, R_{k+2} in Fig. 3). The symmetry of the function guarantees that the middle one is on one side of the peak, and the other two on the other side. It can be shown [3] that the peak can be accurately located at the value of τ that is halfway between R_k and the point at the same level on the other side of the peak, as indicated by the dashed line in Fig. 3. Thus, it is only necessary to interpolate efficiently to find where that other point occurs. Knowing that it lies between the other two values, it suffices to interpolate the value at the middle of that sample interval ($T_s/2$ from each sample), test whether the interpolated value lies above or below the desired value, and accordingly define a new $T_s/4$ domain of uncertainty; then interpolate another value at the midpoint of that interval; etc. In nine steps of interpolation calculation, this locates the desired value to $\pm T_s/2^{10}$. The interpolation calculations themselves are linear sums over the uniformly spaced sample values.

To process ten time lag values in fine mode, with input data at 50 kHz, requires a 500 kHz rate of complex multiplies and

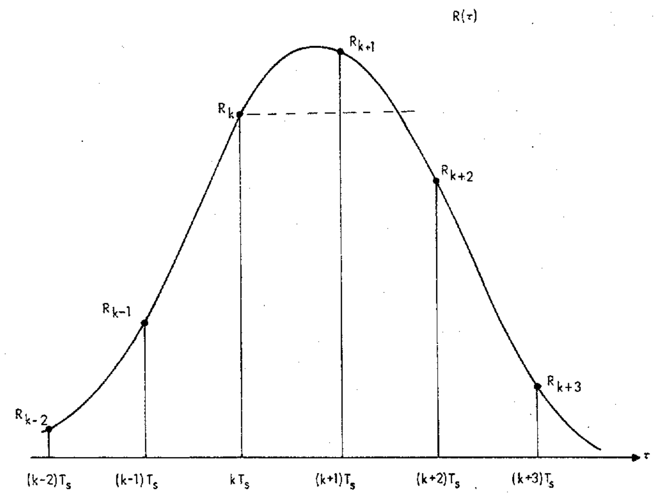

Fig. 3. Interpolation of a sampled ambiguity cut.

complex adds to form the mixing products and initial sums. With the maximum DFO uncertainty at ±100 Hz, the Step 1 I&D intervals could be taken over blocks of 256 samples (5.12 ms), corresponding to -3 dB bandwidths of 195 Hz, and an output sampling rate of 195 Hz for each of the ten time lags. The Step 2 processing could proceed directly to multiple filters of 0.1 Hz bandwidth and resolution, or could use a bandpass filter with bandwidth in the range 1-3 Hz (fraction of a second integration), with 0.1 Hz filtering (possibly by FFT) only in Step 3. Clearly, almost the entire significant processing burden here is in Step 1, and is well within the range of available high-speed peripherals.

To enter the fine mode, the coarse mode here would seek to isolate the DTO to ±10 μs by finding the discrete time lag with the largest ambiguity function magnitude, since this is adequate to define the set of time lags needed for fine mode calculations. Typically, ±10 μs accuracy (as compared to the ultimate 40 ns accuracy) can be accomplished with integration intervals of the order of 20-40 ms, governed more by the spurious lobe requirements than by the accuracy requirement, as discussed earlier in Section II. However, to estimate the DFO to within about 1 Hz so as to enter fine mode will require long integration times, of the order of 1 s. For this, one would in effect use the fine mode type of processing in its first step, operating on a mixing product formed at the determined time lag. Thus, the first step would use a low-pass I&D filter to reduce the bandwidth consistent with the known limited DFO uncertainty range (±100 Hz). If this first step results in a complex sample rate of 195 Hz, as suggested earlier, a 256-point FFT (operating therefore on about 1.25 s of input data) would provide the desired accuracy.

Fig. 4 indicates diagrammatically the processing steps just outlined for the narrow-band case.

VI. Conclusions

An approach to processing to obtain joint estimates of delay and frequency offset (DTO/DFO) for continuous waveforms has been outlined that is based on efficient computation of complex ambiguity functions. Generally, it comprises a two mode process, termed coarse and fine mode. The coarse mode is used to greatly reduce the domain of uncertainty of time delay and frequency offset, following which the fine mode calculations are performed. The fine mode uses a mixing product/filter interpretation to greatly reduce the processing burden.

Two illustrative examples were cited, one for wide bandwidth, one for narrow bandwidth, whose different characteristics (including different accuracy statements) lead to variations in details of the processing approach. In both cases, it was shown that the processing loads are consistent with real time implementation in existing high-speed peripherals or standalong systems. More recondite applications, such as requiring significantly greater accuracy in the wide bandwidth case, would probably result in slower than but "near" real time performance as a result of implementation cost considerations.

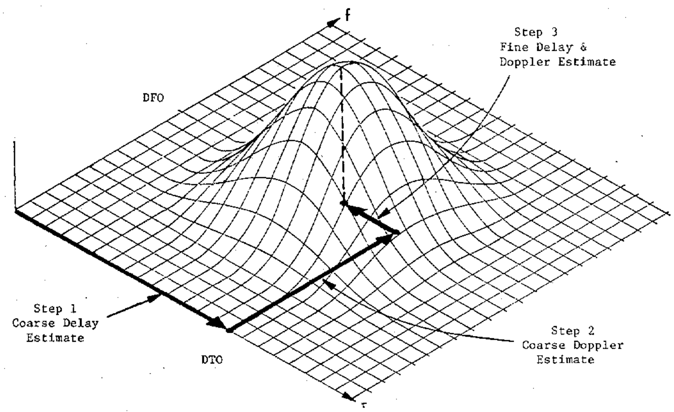

Fig. 4. Processing steps for narrow-band case.

References

[1] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 320-327, Aug. 1976.
[2] P. Swerling, "Radar measurement accuracy," in *Radar Handbook*, M. Skolnik, Ed. New York: McGraw-Hill, 1970, ch. 4.
[3] D. Gooding and R. Greenspan, Stein Associates, Waltham, MA, unpublished works.

Seymour Stein (S'48-A'54-SM'57-F'75) received the Ph.D. degree from Harvard University, Cambridge, MA, in 1955.

He was with Sylvania Electronic Systems from 1953 to 1956, Hermes Electronics from 1956 to 1959, and Sylvania again from 1959 to 1968, becoming Associate Director of its Applied Research Laboratory and then Director of a newly formed Communication Systems Laboratory. In 1969, he founded Stein Associates (which became a division of the Adams-Russell Company in 1970) and headed it until early 1979. Since 1979, he has been an independent consultant to government and industry, operating as SCPE, Inc., Newton, MA. Throughout his career, he has specialized in communication and signal processing theory and applications. He is coauthor of *Communication Systems and Techniques* (New York: McGraw-Hill, 1966) and *Modern Communication Principles* (New York: McGraw-Hill, 1967).