PhD. Thesis

Estimation and Tracking of Atmospheric Delay Noise in a Long-Baseline Optical Stellar Interferometer and Determination of the Expected Estimation Error

Jeffrey Meisner

Estimation and Tracking of Atmospheric Delay Noise in a Long-Baseline Optical Stellar Interferometer and Determination of the Expected Estimation Error

A Thesis Submitted to the Faculty
of the Graduate School of the
University of Minnesota by:

Jeffrey Meisner

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

March 1995

Revised Edition
April 19, 1995

Abstract

A long-baseline optical stellar interferometer is capable of performing superior measurements of optical correlation (or "fringe visibility") in much shorter observation times if the instantaneous atmospheric delay $\tau(t)$, can be accurately tracked to well within an optical wavelength. This permits **coherent** integration of the optical correlation. Additionally, this mode of data acquisition permits the extraction of object phase (even when only using a two-element interferometer) and precise astrometry, advantages not realized using incoherently averaged interferometry.

The conventional approach to "fringe tracking" involves a control system that servos a rapidly responding path-length compensator in real-time. At marginal signal levels, the reliability of such a real-time delay-tracking system suffers. Precise delay-tracking can be achieved at somewhat lower signal levels by employing an **off-line** delay-tracking system, in which the raw data measured by the interferometer is stored for subsequent analysis. Then it is possible to estimate the delay error at time t using raw data collected both before **and after** time t, resulting in a superior estimate.

As opposed to **point estimation** procedures based upon the estimation of τ at a **point** in time, the optimum estimation of τ is based upon the comparison of all possible **functions**, $\tau(t)$, over a time period. Such a **path estimation** procedure fully incorporates the *a priori* statistics of the atmospheric delay process, given by the familiar model of D. L. Fried [1], or more generally as a Gaussian colored noise process with a specified spectral density. The *a posteriori* probability density of any $\tau(t)$ function can then be determined, using Bayes' theorem, as a combination of the *a priori* probability of such a function occurring, and the likelihood of such a $\tau(t)$ function having produced the sequence of photons that had been detected.

Solutions of $\tau(t)$ are found using an iterative procedure to maximize that *a posteriori* probability. However there will be more than one local maximum of *a posteriori* probability. At lower

signal-to-noise ratios it becomes increasingly difficult to differentiate among these multiple solutions, and the resultant estimate contains ambiguities. However by properly evaluating the array of solutions, sufficient information can be retained for the purpose of integrating the measurement of correlation. Very acceptable results are obtained at signal-to-noise ratios as low as 3.0. For an unresolved object ($|V| = 1$), using low-noise photon counting in the visible, a signal-to-noise ratio of 3 corresponds to around 500–1000 detected photons per second, assuming good seeing conditions. At that signal-to-noise ratio, the r.m.s. estimation error of the correctly identified solution is .7 radians.

Contents

List of Figures

Introduction

The concept of measuring stellar diameters by interfering starlight received at two separate points on earth was first proposed by Fizeau around 1868. However it was A. A. Michelson who first implemented an optical interferometer capable of actually resolving the disk of a star in 1920.[1] Using the frame and optics of the Mount Wilson reflector augmented by a system of mirrors, Michelson was able to produce a fringe pattern formed by light received from a star at a baseline of up to 6 meters. Through visual detection of the resultant interference pattern, viewed through a spectral filter, he was able to adjust the length of the baseline of the interferometer to the point at which fringe visibility was reduce to zero. Thus, finding the position of the first null of the zero-th order Bessel function describing the Fourier transform of the star's disk, Michelson became the first person to actually measure the (angular) diameters of Betelgeuse and Antares!

A larger instrument of the same type built around 1930 by F. G. Pease was a technical failure. The field of optical stellar interferometry subsequently languished. It was not until the 1970's that interest was revived, at first by the work of A. Labeyrie who pioneered the filed of **speckle interferometry**. It was also Labeyrie who, in 1974, it is believed, performed the first demonstration of optical amplitude interference between starlight received by separate telescopes on independent mounts [2].[2] The

[1] Some twenty years earlier, Michelson had already employed interferometry for astronomical measurements. Using interference produced by sub-apertures of a smaller telescope, he had performed accurate measurements of the diameters of the 4 larger satellites of Jupiter. Although the disks of these objects, unlike any star, were visible through an eyepiece (their diameters are on the order of one second of arc), Michelson's determinations of their diameters using interferometry were superior to any available estimation technique based on direct imaging.

[2] Also in 1974, C. Townes successfully measured interference of infrared starlight at 10 μm using **heterodyne detection** using a CO_2 laser as a local oscillator. The measurement of correlation was identical to that of radio-astronomy. Although a great

availability of improved technical apparatus was undoubtedly the leading factor enabling the revival of the field. This certainly included both the development of optical detectors (Labeyrie used an image intensifier feeding a television camera), and modern electronics and information storage and processing technology.

Concurrent developments in related fields were also responsible for encouraging work in optical stellar interferometry. In the 1950's Hanbury Brown had invented the **intensity interferometer** and constructed a practical instrument which was successful in measuring the diameters of several bright stars. Hanbury Brown, himself appreciating the limitations of the intensity interferometer, became an early advocate for the development of a modern rendition of Michelson's stellar interferometer.

Another important influence was the rapid introduction of interferometric techniques in the burgeoning field of Radio Astronomy. Using antennas whose individual spatial resolution was very poor compared to optical telescopes, the radio astronomers had, through interferometry, achieved image resolutions comparable to (and ultimately, using VLBI, **exceeding**) that of optical telescopes, The previous and concurrent development of aperture synthesis RADAR also deserves mention in this context.

Finally, developments in the area of control systems and interest in **adaptive optics**, and research in the optical effects of atmospheric turbulence especially by D. L. Fried [1], propelled interest in optical stellar interferometry and the implementation of "fringe-tracking" (or "delay-tracking") systems essential to many configurations of an optical stellar interferometer.

For various reasons, the long-baseline optical interferometers which have since been demonstrated have not

[2]accomplishment, his apparatus has little in common with optical/infrared interferometry using **optical correlation** produced by interfering the optical signals with a photomixer, and is thus only remotely related to the subject of the present work.

achieved scientific results comparable to those achieved by radio interferometry. It is fair to say that long-baseline optical interferometers built to date are better described as experimental rather than practical instruments. Among the technical challenges facing optical interferometry, not similarly encountered by radio interferometers, is overcoming atmospheric delay noise. This problem will be addressed in the context of optimizing the performance of a two-element long-baseline optical stellar interferometer using low-noise photon counting detectors in a configuration to be described.

The point of interferometric techniques is, of course, high resolution imaging of astronomical objects, and precision astrometry. Interferometers, we know, provide data points of the **Fourier transform** of the image. In some interferometer configurations the phase angle of the transform data is lost, and without using an array of unrealistic complexity, full coverage of the Fourier plane will not be achieved. Interferometers built to date invariably report loss of optical correlation (or "fringe visibility") and require empirical calibration. Considering these technical obstacles, the non-existence of high-resolution pictures of astronomical objects derived through optical aperture synthesis is not terribly surprising.

The problem of reconstructing image data from observation points in the Fourier plane, suffering noise, calibration errors, and sparse coverage, has been dealt with in theory and practice, largely for the sake of radio interferometers. Therefore, the reconstruction problem will not be addressed at all in the present work. Among the problems facing the reconstruction algorithms are: 1) Lost phase angle information of the Fourier data; 2) Incomplete Fourier plane coverage; 3) Noise in the data; and 4) Best applying **a priori** assumptions concerning the expected image characteristics. On the other hand, it might be noted that in many cases the full recovery of an arbitrary image is not at all required. For instance, Michelson set about to measure the angular diameter of a single star based on the assumption of uniform brightness over a disk. Determining a single parameter such as this, in principle

requires but one significant data point in the Fourier plane. If a limb-darkening model were assumed containing, say, two additional unknown parameters, then at last 3 data points in the Fourier plane (but preferably more) would be required to determine a fit for this particular model. Likewise the observation of a binary star system requires a minimum of 3 data points in order to determine the position vector and brightness ratio. So in many cases, it can be seen, sparse coverage of the Fourier plane is quite acceptable; what becomes of paramount importance, however, is the accurate determination of this data.

Realm of Optical Interferometry

Classification of a telescopic imaging system as "interferometric" itself includes a wide variety of techniques. In fact **all** optical imaging systems rely on the fundamental principle of interferometry: that the resultant amplitude of an optical field at a point is determined by the linear superposition of possible paths of the light form a source. Thus, the classical formation of a real image of a distant object by a convex lens can be explained as the phased recombination of waves received over the aperture of the lens, adding in phase at the point of the resultant real image but suffering phase cancellation at other points. Detailed calculation of the integrated amplitude received at points around the center of the image yields the Fourier integral which predicts the Airy disc pattern in the case of a circular aperture, or the results of Fraunhoffer diffraction in general.

Equivalently, the action by a convex lens of performing a Fourier transform on the incident optical field, describes the response function of a refracting telescope **relative to the impinging optical field**. Since a distant point-source produces a spatially coherent field of constant amplitude and phase, the resultant image is the Fourier transform of a constant, namely a delta function. Taking into account the aperture function which multiplies this constant field, we must convolve the delta function with the Fourier transform of the aperture function which yields

the resultant Airy disc. Also multiplying the impinging field by the random phase function of the atmosphere over a large aperture determines a net image which is larger than the diffraction-limited image size, and which is covered with a "speckled" pattern, reflecting the Fourier transform of the stochastic phase function introduced by the atmosphere.

While the formation of this image can be properly described as a result of optical interference, classical techniques for recording the image information reduce the instrument to an **incoherent** system described by a point-spread function with an extent inversely proportional to r_0 (the **Fried Parameter**, see page 241) rather than the physical aperture of the instrument. This occurs because the film being exposed (for instance) averages the **power** of coherently formed images over a period of time during which the random atmospheric phase function varies wildly. Additionally, the film records the image over a broad range of wavelengths, incoherently combining dissimilar images formed at different wavelengths. On the other hand electronic recording of short term images over time periods in which there is little change in the random atmospheric phase function can **retain** samples of the actual speckle patterns created as a result of amplitude interference, reflecting the combined effects of the atmosphere and the original object function. Hence, **speckle interferometry**.

It might be added that there are other high-speed correction techniques for reducing the effect of atmospheric turbulence affecting a large aperture telescope. Real-time hardware correction techniques are classified under **adaptive optics**. Alternatively, there are algorithms which extract improved image information from a large number of electronically-recorded images, each resulting from a short-term exposure. Because these techniques are more based on geometric optics concepts, and are incoherent relative to the treatment of different wavelengths, they can not be properly described as "interferometric." So although the principle of interferometry describes the instantaneous behavior of any optical system, the description of a technique as "interferometric" implies that the apparatus is sensitive to, and

able to process measurements based upon, unknown or fluctuating phase shifts in portions of the incoming optical wave.

Having so defined the scope of interferometry, it is not surprising that a multitude of optical configurations could be included in this category. Let us narrow down the range of configurations first by excluding from further discussion **speckle interferometry** and related techniques distinguished by their combining the light received over a **single aperture**, generally that of the objective of an existing large telescope. The results to be derived for the case of long-baseline interferometry are in fact somewhat applicable to the case of short baselines, among configurations in which light combination at any point on the detector is a combination of exactly two areas of the aperture. However a proper discussion of the attributes of such an instrument would still require a distinct analysis not included in the scope of the present work.

Long-baseline interferometers are designed to measure much higher spatial frequencies than the above instruments, and pose somewhat different technical challenges. Non-coherent imaging (long-term exposure) attains a resolution on the order of 1 arc second. Speckle interferometry and other single-aperture interferometric instruments can hope to extend this resolution by an order of magnitude or so. A long-baseline interferometer, on the other hand, obtains a resolution at least two to three orders of magnitude greater than the 1 arc second value but **does not** measure spatial frequencies correspond to the intermediate region of resolution. Such an instrument would generally be observing objects having no apparent extend using conventional imaging. The detector therefore need not image an interference pattern at the plane of the image produced according to geometric optics (image-plane interferometry) but can just as well produce an interference pattern at the re-imaged aperture of the two telescopes (pupil-plane interferometry). Then, instantaneous spatial variations in the atmospheric path delay over the apertures will cause shifts in the interference over the detector surface.

A special case of long-baseline interferometry occurs when the size of the apertures is on the order of r_0 or smaller. This configuration is termed the **Michelson Stellar Interferometer**, and is the model upon which the remainder of this work is based. As the apertures size is thus decreased, the distinction between pupil-plane and image-plane interferometry vanishes. The correlation between the light received at two points becomes characterized by a single complex number (at any instant in time), which may be resolved as the magnitude of visibility (the sought parameter) with a phase angle combining the atmospheric delay (the unwanted parameter) and the object phase. Conceptually, we have created a single-mode (**spatial** mode) system, in which each beam has an unchanging transverse variation and can be represented as a single field amplitude $A(t)$ varying in time. Unlike a multi-mode system, the two optical signals can now be transmitted by single-mode optical fibers or other guided-wave technology, if this is desired.

A pupil-plane-interference long-baseline interferometer using **larger** apertures can be viewed as the equivalent of several **Michelson** stellar interferometers operating in parallel. Since apertures separated by distances greater than r_0 receive **waves** which are essentially uncorrelated, but have random **delay functions** that are **partially** correlated (especially when looking at the lower frequency components), there **is** an advantage in delay-tracking employing information from adjacent sub-apertures. However it can be shown that the improvement is not terribly substantial in comparison to the increased expense of a much larger objective and complexity of the interferometer. Consequently the primary effect of an N-fold increase in the aperture area (also requiring a similar increase in the complexity of the subsequent optics, detector and electronics) is an N-fold decrease in required observation time to attain a given level of accuracy. Because delay-tracking interferometry is already capable of achieving a high net signal-to-noise ratio in a fairly short observation, this advantage would appear to not be worth the expense. However, there **will** be a certain range of magnitudes of objects for which the mutual delay information between adjacent sub-apertures would permit delay-tracking not achievable by an r_0 sized instrument.

This might justify the possible construction of such an instrument, despite the additional expense. However in the following pages we will restrict our analysis to the optimization of a **Michelson** (single-mode) stellar interferometer.

Delay-Tracking in a Michelson Stellar Interferometer

There are two ways in which estimation and correction of the random atmospheric differential delay function, is important to the operation of an interferometer. **Coarse delay-tracking** is necessary for interference to occur coherently with respect to wavelength over the operating bandwidth of the interferometer (or of any single detector, in a multi-spectral detection system). Secondly, more precise delay-tracking to within a fraction of a wavelength (or **"fringe-tracking"**) is required in order to **coherently** integrate the measured correlation over periods of time in which the phase of interference varies due to atmospheric turbulence. Let us briefly look at each of these problems.

Coarse delay-tracking in an absolute requirement of the observation of interference over a non-zero bandwidth. As is well known, the permissible bandwidth of light to be successfully interfered is inversely proportional to the differential delay affecting two arms of an interferometer. For instance, the familiar Michelson interferometer found in the physics classroom, illuminated by a filtered mercury vapor lamp, requires equalization of the two arms to within about a millimeter. This number reflects the bandwidth of the spectral line of mercury. Likewise, we seek to minimize delay discrepancies in the arms of an astronomical interferometer in order to relax the bandwidth requirements, thus increasing the total optical power that can be utilized.[3] One

[3]One way of relaxing this requirement, to be discussed on page 38, is the spectral dispersion of the interfered light and detection of different wavelengths in N separate channels. This is the equivalent of operating N narrowband interferometers in parallel. Since the bandwidth of each resulting system has been reduced by a factor of N, the allowable delay errors has increased

special case of coarse delay-tracking is **open loop delay compensation**, in which there is **no tracking** per se, but only correction for the expected differential delay base solely on geometric considerations. Clearly this entails a serious bandwidth constraint which grows more severe as the interferometer's baseline is lengthened, increasing the magnitude of differential delay.

The second sense in which delay-tracking is discussed, refers specifically to the estimation of delays to well within a wavelength of light. Also refereed to as **phase-tracking** or **fringe-tracking**, in this context the estimation of delay is specifically for the sake of coherently averaging the measured optical correlation. We will discuss both ways in which the estimation **and optical path length correction** can occur in real-time (page 45), and alternatively, in which the estimation process can occur **following the observation** (page 48), providing information that can then be used to implement coherent integration of the raw data in which the phase of measured correlation was randomly changing. We will also touch on **incoherently averaged** interferometry (page 42), in which delay-tracking, in this sense, is not required. However a delay-tracking instrument can achieve superior results using much shorter observation times.

Survey of Current Delay-Tracking Implementations

The following survey is intended to put the discussion of delay estimation and tracking in perspective in relation to interferometers that are presently operational.

The **Mark III interferometer** on Mt. Wilson, operated by the U.S. Naval Observatory, incorporates precision control of the delay-line monitored by a laser interferometer with a positional resolution of .01 μm [13]. Claimed **open-loop** delay errors are

[3]by the same factor. In any case, however, there remains an outer limit to the allowable delay error beyond which the measurement of interference is compromised.

within 10 μm, allowing operation in the **incoherently averaged** mode (page 42) for faint sources. However the primary mode of operation (especially due to its ostensible mission as an **astrometric** instrument), is in the **real-time fringe-tracking** mode, exactly as described on page 45. Closed-loop tracking errors of .05 to .1 μm are claimed. Although only medium-wide bandwidths are used in two separate spectral channels (only one of which is involved in delay-tracking), identification of and tracking on the central fringe has apparently been assured (which is, again, a requirement to insure the most precise astrometric measurements). The **Big Optical Array**, the U.S. Naval Observatory's follow-up project, is slated to include truly wideband detection using the **spectrally-dispersed** detection system (see page 38).

The **Infrared-Optical Telescope Array** (IOTA), a U.S. based interferometer consortium, planned in 1990 to also implement a real-time delay-tracking system using spectrally-dispersed detection [8]. Simulation results showed estimation errors approximately ranging from .06 to .3 wavelengths. Their estimation procedure follows the mathematics to be included under "Point Estimation of τ using Time Bins," (page 65) with a serious calculation of the optimum bin widths, and proper inclusion of model parameters (of both source and atmosphere). For the implementation of the spectrographic element, a **grism** device had been devised which exhibited a liner dispersion with respect to optical **frequency**. However as of 1990, their delay-tracking system had only been proved by way of simulation.

As of 1993, the **Sydney University Stellar Interferometer** (SUSI) had reported [14] the successful estimation of differential delay using artificial light only, and not using the actual optics of the interferometer (which is, presumably operating without any real-time delay estimation hardware). Based on hardware comparable to that advocated in the present work (the spectrally-dispersed detection system), the delay is estimated using the group delay method to a reported resolution of about 2 μm, quite coarse compared to an optical wavelength. Although termed "fringe tracking," it is not clear whether only **coarse** delay-tracking is

being attempted, or if different parameters would ultimately be used to allow coherent integration of optical correlation. Based on what will be termed **point estimation** of differential path delay, the trial results presented in [14] show errors characteristic of point estimation techniques: at various points in time wild mis-estimations of the delay occur, producing a net delay **function** which would have an *a priori* probability near zero.

The **Infrared Michelson Array** (IRMA) was a prototype project of the University of Wyoming an the National Optical Astronomy Observatories [12] (its hardware has since been incorporated into the IOTA projects). This instrument operated in the **delay-scanning** mode (as in [7]). In order to reduce costs, the designers compromised on a **passive delay-scanning** system, in which the delay line is moved well **past** the central fringe, using open-loop control. Then the earth's rotating is allowed to "scan" the delay pattern across the detector, after which the process may be repeated. While suffering from a very low duty-factor, the simplicity of the passive delay-scanning system clearly eliminates the need for a dynamically stable and well-controlled delay-line. The evaluation of the data reduction form a delay-scanning instrument not dealt with in the present work. However it should be noted that the use of such a system in which all of the light winds up in only two detectors is a favored choice for the detector-noise-dominated wavelength (2.2 μm) for which this instrument was designed. It should also be noted that the magnitude of atmospheric phase fluctuations at this wavelength is considerably reduced below that of the visible. However we would expect that delay-tracing instrument could unquestionably achieve higher performance.

Finally we should mention A. Labeyrie's **Grand Interferomètre à 2 Télescopes** (GI2T) [15] which is **not a Michelson** stellar interferometry per se, inasmuch as its apertures (1.5 m) are much larger than r_0. Thus real-time fringe-tracking in the present context is not even an option since different parts of the image would exhibit significantly different delays (of well over a wavelength). The output of the interference pattern, however, is

sliced and spectrally-dispersed, potentially implementing some 100 **Michelson** stellar interferometer operating in parallel. In any case, reasonably accurate adjustment of the delay-line is certainly required to reduce the **maximum** delay over the interfered apertures. **Part** of the interfered image is spectrally-dispersed and evaluated **visually** for fine adjustment of the delay-line (an automated version of spectrally-dispersed delay estimation was reported to be in the "testing" stage). After that point, an open-loop control system continues the motion of the delay-line with a reported accuracy of a few microns. Emphasis was placed on plans for a laser metrology system to aid in open-loop setting of the delay line, particularly important since these telescopes are movable on tracks.

Beam Recombination and Detection

Having restricted ourselves to the case of single-spatial-mode stellar interferometer, we now wish to construct a model for the optical circuit, and in particular, the beam recombination (photomixing) optics. Different photomixer configurations and their characteristics will be explored. For simplicity, we will start by assuming that the light received is narrowband. The application off these configurations to a wideband (white light) interferometer will be treated towards the end of the chapter.

Model

The general model for the optical train is depicted in Figure 1 where A_1 and A_2 represent the amplitudes of the starlight **as would be seen in outer space, prior to the effects of the atmosphere** at two points separated by (the projection of) the baseline of the interferometer. The quantity we ultimately seek to measure using interferometry is the normalized cross-correlation of these two amplitudes, generally referred to as **fringe visibility**, given by

$$V_{12} \triangleq \frac{\langle A_1 A_2^* \rangle}{\sqrt{\langle |A_1|^2 \rangle \langle |A_2|^2 \rangle}} \tag{1}$$

In a wideband instrument, V_{12} will generally be a function of optical frequency ν.

The light actually collected by the telescopes suffers a differential delay consisting of two terms. The first is a deterministic variation, $\tau_g(t)$, the geometrical delay encountered due to the physical placement of the telescopes and the (changing) position of the celestial object in the sky. The second is a random delay term simply denoted $\tau(t)$, attributable to the effects of atmospheric turbulence (there could also be random delay components introduced by imperfect instrumentation). We will

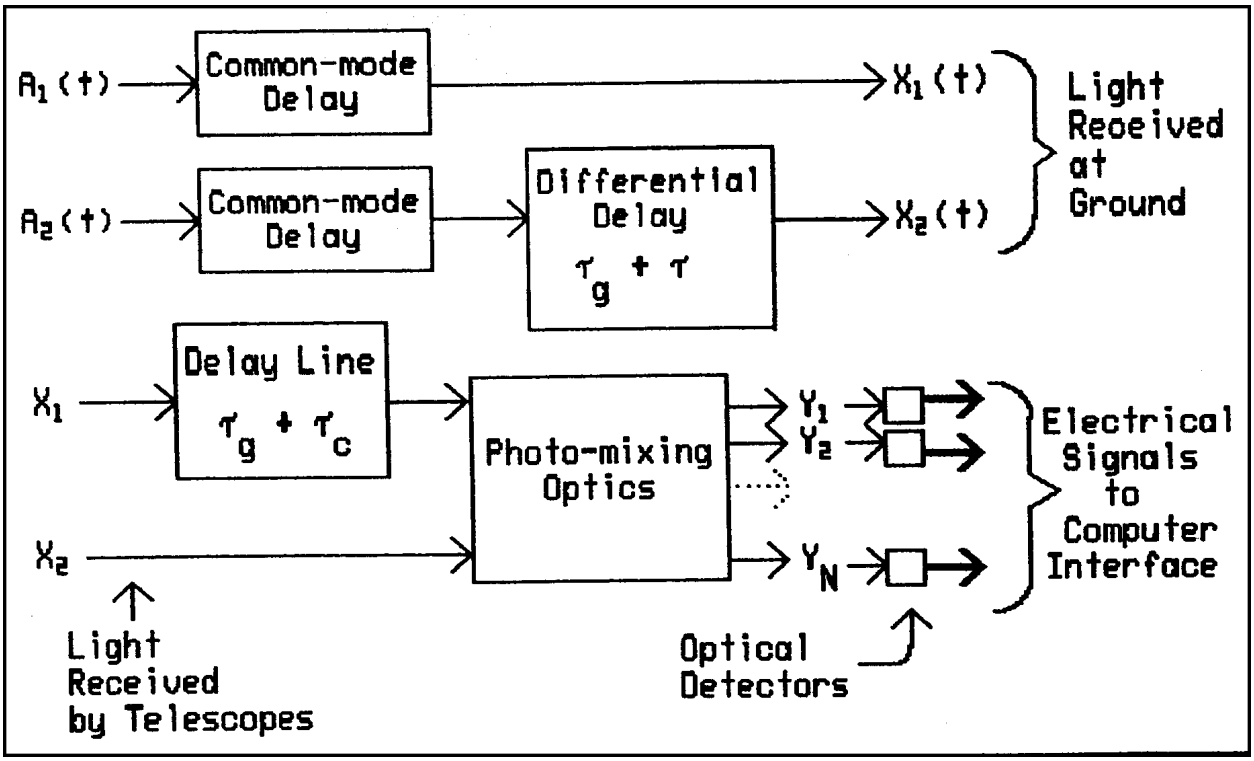

Figure 1: Model of optical train through atmosphere and interferometry, up to the optical detectors.

ignore any **amplitude** variations that may be introduced by the combination of the collection optics and the effects of atmospheric seeing.

The light collected at telescope 2 is labelled X_2. The light collected by telescope 1, labelled X_1, will pass through a delay line intended to compensate for the additional delay encountered by the light received at telescope 2. The delay line introduces a changing delay, τ_g, to compensate for the deterministic geometric delay calculated on the basis of the celestial coordinates of the object, the placement of the telescopes, and the exact time. Using a closed-loop system, the delay line may **additionally** introduce a "correlation delay" τ_c, intended to compensate for random or unanticipated delays measured by the interferometer during an observation. Depending on the requirements of the subsequent optical system, τ_c may try to follow the best estimate of $\tau(t)$ (the actual atmospheric delay function) in real time, or may only be required to follow delay offsets in a slow and/or coarse manner in order to keep the uncorrected delay offset within an allowable range.

For certain configurations it is also desirable to have the delay line introduce a rapid dither of peak-to-peak amplitude approximately $\lambda/4$ (or more in some configurations). This allows one to recover both sine and cosine components of interference when using only a two phase photomixer. This auxiliary dither would not generally be required if using photomixing optics which yield three or more phases of interference, as noted in the discussion below.

The field amplitudes entering the photomixer optics are denoted as (lower case) $x_1(t)$ and $x_2(t)$, where clearly:

$$\begin{aligned} x_1(t) &= X_1(t - \tau_g - \tau_c) \\ x_2(t) &= X_2(t) \end{aligned} \tag{2}$$

These signals can be seen to have a cross-correlation given by $V_{1,2}$ — the correlation prior to entering the atmosphere — modified by the uncorrected delay factor of $\Delta\tau = \tau - \tau_c$. As measured at the optical frequency ν, then, the **measured correlation** seen by the photomixer and optical detectors would be V_m.

$$V_m \triangleq \frac{\langle x_1 x_2^* \rangle}{|x_1||x_2|} = V_{12} e^{j2\pi\nu\Delta\tau} \tag{3}$$

Note that, in this narrowband context, we are treating the signals x_1 and x_2 as **phasors** at the optical frequency ν, even though the above results is wholly applicable to the response of a wideband instrument **as measured at** the optical frequency ν.

Photomixer Output Considerations

The output of the photomixer consists of N channels where $N \geq 2$. In the case of certain photomixing configurations, the output will consist not of discrete beams but rather of a **continuum** of non-identical modes. This case can be accommodated within the present

model by dividing the continuous spatial field into N regions, and making N sufficiently large.

It is assumed that each of the N output channels is fed into a separate photodetector of a distinct element (pixel) of a photodetector array. At this point we should address the relationship between N, the number of detector degrees of freedom, and detector noise performance. The following familiar result applies. In the case of low-noise detector in which the dominant noise term is due to signal quantum noise (Poisson photon statistics), **there is no penalty for increasing** N. However in the case of detectors that are dominated by **internal noise** sources (such as dark current or electronic pre-amp noise) then **the signal-to-noise ratio suffers according to the square root of increasing** N.

The first result is easily shown. Consider the case in which a single detector receives p photons in a time period. In the case of quantum limited detection there would be an r.m.s. noise amplitude of $\sqrt{p}$. Suppose we (needlessly) chose to split the light that had been received by that detector into two equal parts which are fed into **two** detectors. Now each of the two detectors would receive $p/2$ photons with an r.m.s. noise of $\sqrt{(p/2)}$. If we simply **add** the output of the two detectors then we would have a signal of strength $p/2 + p/2 = p$. However the noise contributions **add in quadrature** resulting in an r.m.s. noise level of $\sqrt{(p/2 + p/2)} = \sqrt{p}$. Thus we have produced a signal with the same signal and noise levels that would have been obtained by collecting all of the light in a single detector.

On the other hand, if the detector noise had been dominated by a dark count of d, where $d \gg p$, then the r.m.s. noise suffered using a single detector would have been $\sqrt{d}$. In this case, following the above example, if **two** detectors had been used to each detect half the light, then each detected signal of strength $p/2$ would have an r.m.s. noise of strength $\sqrt{d}$. Again we add the two signals together to obtain the original signal strength p. This time, however, we

would obtain an r.m.s. noise amplitude of $\sqrt{(2d)}$, cutting the signal-to-noise ratio by $\sqrt{2}$.

The upshot of this result is that when employing noisy detectors, as is inevitable in the infrared, concentrating the available light in the fewest number of detectors (and, practically speaking, in the smallest area detectors) will improve the signal-to-noise ratio, everything else being equal. However in the visible where photon-counting detection can be employed with a relatively small dark count rate, it is permissible to divide the optical power among detector elements, especially if there is any information-gathering advantage in doing so. The point beyond which increasing N will incur a noise penalty can be calculated as follows. Suppose that the total number of photons to be counted is about I_0 photons per second, and that each detector element has d_0 dark counts per second. The detectors will continue to be quantum noise limited as long as $I_0/N \gg d_0$. So as long as one sets $N \ll I_0/d_0$ there will be no significant penalty paid for the size of N. Of course, this calculation should be performed using the **smallest** value for I_0 in which operation is anticipated.

One more issue regarding the coverage of detector elements occurs when using noisy detectors (so that reducing N is desirable) but in which the output of the photomixer consists of a continuum of non-identical modes. It was noted above that such a photomixer can be characterized by arbitrarily dividing the output space into N regions and setting N to be such a large number that each region is essentially a pure mode. On the other hand, if we choose to route a number of such pure modes into a single detector we have now created an output channel which is **impure**, that is, it is not single-mode. The detector now is sensitive to the sum of the **power** of non-identical modes. This entails a certain loss of sensitivity insofar as the phase of interference detected varies spatially across the output beam.

Let us calculate this penalty. Suppose that the center of an output channel detects the interference between x_1 and x_2 at zero phase, but across the beam of uniform amplitude there is a linear

phase variation ranging from $-\Delta\theta$ to $+\Delta\theta$. Suppose that a portion, a^2, of the power from each input channel is routed to the output channel in question. Now the total power of light, I_d, falling on the photodetector is found by integrating over the width of the field over which the phase varies between these limits. Recall that the amplitudes entering the photomixer are given by x_1 and x_2, which have a normalized cross-correlation given by V_m. Then we can write the total beam power incident on the photodetector as an integral of the power over the beam width:

$$I_d = \frac{1}{2\Delta\theta}\int_{-\Delta\theta}^{\Delta\theta} d\theta\langle|ax_1 + ax_2e^{j\theta}|^2\rangle \tag{4}$$

$$= \frac{1}{2\Delta\theta}\int_{-\Delta\theta}^{\Delta\theta} d\theta(|ax_1|^2 + |ax_2|^2 + 2\,\mathrm{Re}\langle a^2x_1x_2^*e^{-j\theta}\rangle) \tag{5}$$

Let us call the total power entering the photomixer I_0 and assume that half this power is from x_1 and half from x_2. Thus:

$$|x_1|^2 = |x_2|^2 = \frac{I_0}{2} \tag{6}$$

We can take the exponential outside the expectation and use (3) to obtain:

$$I_d = \frac{1}{2\Delta\theta}\int_{-\Delta\theta}^{\Delta\theta} d\theta a^2 I_0(1 + \mathrm{Re}\{e^{-j\theta}V_m\}) \tag{7}$$

Setting the integral of the real part to be the real part of the integral, we can evaluate the result.

$$I_d = a^2 I_0 \left(1 + \frac{1}{2\Delta\theta} \mathrm{Re}\left\{\int_{-\Delta\theta}^{\Delta\theta} d\theta e^{-j\theta} V_m\right\}\right) \quad (8)$$

$$= a^2 I_0 \left(1 + \mathrm{Re}\{V_m\} \cdot \frac{\sin(\Delta\theta)}{\Delta\theta}\right) \quad (9)$$

We thus see that the sensitivity of the detection system to the real part of V_m (this channel measures the **real part** of V_m since we specified the center mixing phase as 0°) has been reduced by a factor we recognize as the **sinc function** evaluated at $\Delta\theta/\pi$. The average noise power is unchanged. Therefore, this is also the factor by which the signal-to-noise ratio has been reduced.

To give a numerical example, consider an output channel combining interference phases over a range of 90°, to that $\Delta\theta = 45°$. Then we find that the signal-to-noise ratio has been cut by a factor of .9. Since signal-to-noise ratio varies as the square root of collected light power (assuming quantum noise is dominant), we would require a 23% increase in signal power to compensate for the penalty imposed by mode impurity in the detection channel. On the other hand, in the case of detectors dominated by **internally generated noise**, the .9 factor affecting signal-to-noise ratio due to mode impurity compares **favorably** to the previously discussed .71 factor that would be imposed by dividing the light into two detectors, each with only a 45° interference range (in which case there would be a mode impurity penalty of only .97).

Characterization of the Photomixer

The fields a the photomixer output are denoted $y_1, y_2, \ldots y_N$. The relationship of these output modes to the input modes of the photomixer can be described by a scattering matrix S consisting of 2 columns and N rows.

$$\begin{Bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{Bmatrix} = \begin{Bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \\ . & . \\ . & . \\ S_{N1} & S_{N2} \end{Bmatrix} \times \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} \tag{10}$$

Clearly we wish the photomixer to be a **lossless** optical system, and we shall model it as such. For modelling purposes, the optical losses occurring throughout the optical train, including the effect of the detector quantum efficiency, can be lumped into an equivalent attenuating term reducing the magnitude of A_1 and A_2 relative to the actual starlight received by the telescopes.

Given that the photomixer is modelled as a lossless system, we can view it as a lossless system with N input modes and N output modes, in which $N - 2$ of the input modes are unused. A lossless system with N inputs and N outputs is described by a **unitary matrix**. Thus the scattering matrix S, above, must consist of two columns of an $N \times N$ unitary matrix. That implies two conditions. First, the inner product of either column with itself must be unity. That means that if only one input is excited then the power entering the system must equal all the power exiting the system. Accordingly, the first column satisfies:

$$\{S_{11}S_{21}\ldots S_{N1}\}^* \times \begin{Bmatrix} S_{11} \\ S_{21} \\ . \\ . \\ . \\ S_{N1} \end{Bmatrix} = \sum_{i=1}^{N} |S_{i1}|^2 = 1 \tag{11}$$

and likewise for column 2. The second condition for a unitary matrix is that the inner product between two **different** columns is zero. That expresses the fact that the orthogonal input vectors $(1, 0)$ and $(0, 1)$, must produce orthogonal outputs. Thus:

$$\{S_{11}S_{21}\ldots S_{N1}\}^* \times \left\{ \begin{array}{c} S_{12} \\ S_{22} \\ . \\ . \\ . \\ S_{N2} \end{array} \right\} = \sum_{i=1}^{N} S_{i1}^* \cdot S_{i2} = 0 \qquad (12)$$

In addition to these requirements on the characterization of a lossless system, we would almost certainly want a photomixer in which any output combines an equal portion of the power of each input signal, thus satisfying:

$$|S_{i1}| = |S_{i2}| \qquad (13)$$

We assume that the signals entering the photomixer, x_1 and x_2, have equal power levels, at least on the average. Then (13) will maximize the sensitivity of correlation detection. Clearly if the power reaching an output mode contributed by each input is **not** equal, then complete cancellation of anti-phase inputs will be impossible as will full doubling of the power of in-phase inputs. The noise level, however, continues to be governed by the total photon level (and/or detector noise). Thus a significant loss of signal-to-noise performance occurs for large violations of (13).

Assuming a photomixer satisfying (13) in which the power from either input mode is routed to output mode #i in the proportion of p_i, we can rewrite S as:

$$S = \left\{ \begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22} \\ . & . \\ . & . \\ S_{N1} & S_{N2} \end{array} \right\} = \left\{ \begin{array}{cc} \sqrt{p_1}e^{j\Phi_{11}} & \sqrt{p_1}e^{j\Phi_{12}} \\ \sqrt{p_2}e^{j\Phi_{21}} & \sqrt{p_2}e^{j\Phi_{22}} \\ . & . \\ . & . \\ \sqrt{p_N}e^{j\Phi_{N1}} & \sqrt{p_N}e^{j\Phi_{N2}} \end{array} \right\} \qquad (14)$$

where the power splitting coefficients must satisfy (11) so that:

$$\sum_{i=1}^{N} p_i = 1 \tag{15}$$

Now, the power **leaving** the photomixer is detected **incoherently** and any phase delays **following** the photomixer are undetectable and unimportant. Thus we can, without loss of generality, multiply (14) **on the left** by a **diagonal phase matrix** in order to express the essence of the photomixer's scattering matrix in the following canonical form:

$$\begin{aligned} S' &= \begin{Bmatrix} e^{-j\Phi_{11}} & 0 & 0 & . & 0 \\ 0 & e^{-j\Phi_{21}} & 0 & . & 0 \\ 0 & 0 & . & . & . \\ 0 & 0 & . & . & . \\ 0 & 0 & 0 & . & e^{-j\Phi_{N1}} \end{Bmatrix} \times S \\ &= \begin{Bmatrix} \sqrt{p_1} & \sqrt{p_1}e^{j\theta_1} \\ \sqrt{p_2} & \sqrt{p_2}e^{j\theta_2} \\ . & . \\ . & . \\ \sqrt{p_N} & \sqrt{p_N}e^{j\theta_N} \end{Bmatrix} \end{aligned} \tag{16}$$

We have now characterized each output channel by a power coefficient p_i and a mixing phase θ_i. In addition to the condition (15), the condition (12) must be met, which now becomes:

$$\sum_{i=1}^{N} p_i e^{j\theta_i} = 0 \tag{17}$$

One simple class of discrete beam photomixers uses (16) setting $p_i = 1/N$ and $\theta_i = 2\pi i/N$, resulting in:

$$S = \sqrt{1/N} \begin{Bmatrix} 1 & e^{-j\frac{2\pi}{N}} \\ 1 & e^{-j2\frac{2\pi}{N}} \\ . & . \\ . & . \\ 1 & 1 \end{Bmatrix} \tag{18}$$

The Two Phase Photomixer

Applying (18) with $N = 2$, we get the simple balanced two-phase photomixer:

$$S = \frac{\sqrt{2}}{2} \begin{Bmatrix} 1 & 1 \\ 1 & -1 \end{Bmatrix} \tag{19}$$

This network can be implemented by a single 50/50 beamsplitter as shown in Figure 2. We assume an ideal beamsplitter with (amplitude) reflection and transmission coefficients:

$$r_1 = \frac{\sqrt{2}}{2}$$
$$r_2 = -\frac{\sqrt{2}}{2} \qquad (20)$$
$$t_{12} = t_{21} = \frac{\sqrt{2}}{2}$$

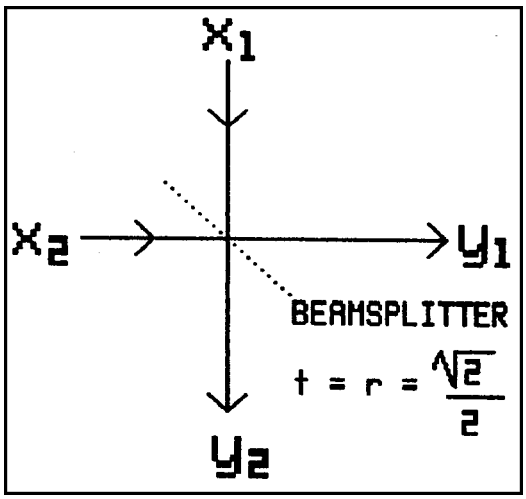

Figure 2: Two-phase photomixer mixes x_1 and x_2 at 0° and at 180°.

Not explicitly shown in the diagram are a number of essential details which will be mentioned for the sake of completeness. For a system that is not extremely narrowband, there should be a "compensating plate" introduced in the input beam which does **not** pass through the substrate of the beamsplitter, in order to cancel chromatic delay effects resulting from passage of light through that substrate (equivalently, the beam splitting surface may be sandwiched in between identical glass plates with an index-matching oil or cement). All of the air-glass interfaces should be anti-reflection coated. Also, for wideband interferometry, the beamsplitter's transmission and reflection coefficients (20) must not vary widely over the range of optical frequency.

Now, in order for y_1 and y_2 to truly be single spatial modes producing full interference between the input beams, optical alignment must be precise so that the wavefronts received at y_1 due to either x_1 or x_2 alone are indistinguishable. Even for circular (or Gaussian) anastigmatic beams, that requires matching 6 parameters characterizing the beam positions, angles, and focussing. If using beams that are guaranteed single mode (such as light transmitted down optical fibers), this can be accomplished by careful alignment of the optics. However in the case of light re-imaged from the apertures or image planes of two telescopes, there will be random fluctuations in these parameters. The magnitude of these fluctuations is nil for collecting optics much smaller than r_0 (the Fried parameter), and is significant but tolerable for objective diameters about equal to r_0, especially if a tip-tilt

active optical correction system is employed. Apertures much larger than r_0 do not fall under the classification of **Michelson** stellar interferometry, precisely because the received light **is** multimode. For apertures not much smaller than r_0, it would be necessary to continually monitor r_0 as it changes during an observation, and use the expected degradation in measured correlation to calibrate the results of that measurement.

Finally, interference will only be complete if the **polarizations** of the interfering beams are identical. Thus x_1 and x_2 must both be either vertically polarized or horizontally polarized (or any other identical polarization state, in the case of a beamsplitter which is non-birefringent). It is only necessary that the characterization of the beamsplitter (19) apply to the polarization used. In the very special case of a beamsplitter which is truly non-birefringent (which can be approximately implemented by reducing the angle of incidence) so that (19) applies to all polarizations and wavelengths, we realize the ability to run both polarizations of x_1 and x_2 into the same beamsplitter and simultaneously integrate the interference of each polarization, which would be identical in the (usual) case of non-polarized starlight. This would be advantageous when using detectors which are dominated by internally generated noise. When using low-noise photon counting detectors dominated by source quantum noise, this advantage disappears; in that case processing the two polarization components **in parallel systems** may be preferable.

The two-phase photomixer characterized by (19) has an important drawback. Its response is only sensitive to the **real part** of the measured correlation given by (3). Thus it is absolutely incapable of distinguishing between positive and negative uncorrected delays.

Quadrature Detection

This drawback of the two-phase photomixer may be eliminated by using the **quadrature** detection system in which half the optical power is used to detect the real part of V_m and the other half is used to detect the imaginary component of correlation (the splitting ratio need not necessarily be 50/50). Such a quadrature photomixing system would be characterized according to:

$$S = \frac{1}{2}\begin{Bmatrix} 1 & 1 \\ 1 & j \\ 1 & -1 \\ 1 & -j \end{Bmatrix} \tag{21}$$

A possible implementation of this network is shown in Figure 3. The light at input x_1 is specified to be in the vertical polarization, as is the input x_2. The latter signal is passed through a quarter-wave plate oriented at 45° to the horizontal. The beams are first mixed using a (non-birefringent achromatic) 50/50 beamsplitter as in the previous implementation of a two-phase photomixer. Each output beam is then split again using a polarizing beamsplitter oriented at 45° to the horizontal. It can easily be shown that the four output beams are characterized according to (21). There is also an implementation of (21) using only polarizing beamsplitters, which may be easier to realize than a broadband 50/50 beamsplitter.

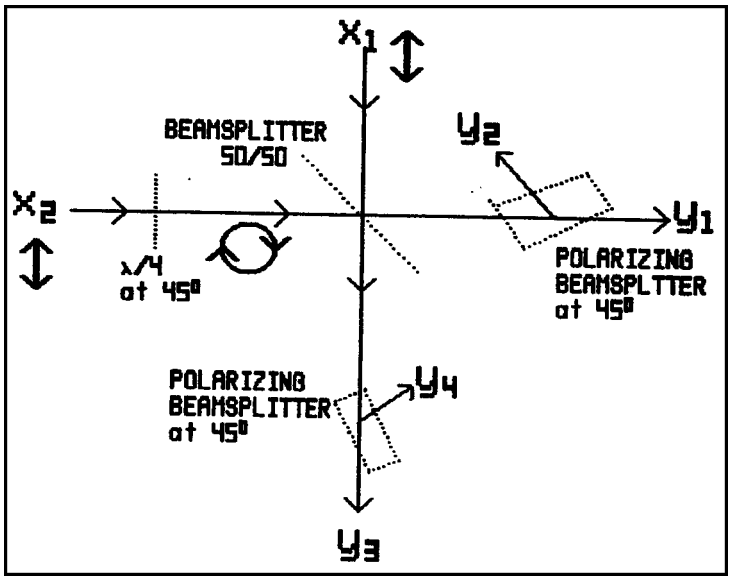

Figure 3: Four-phase photomixer mixes x_1 and x_2 at 0°, 90°, 180°, and 270°. Polarizations at various stages are indicated by double arrows.

Such a quadrature photomixing system successfully resolves the sign of phase delay, information that was lost using the two-phase interference configuration. It is also possible to build a photomixer with $N = 3$, the smallest value of N which maintains this property. Recall that reducing N is advantageous when the detectors are not quantum noise limited, since this allows each detector to receive more light.

A balanced three-phase photomixer would implement the scattering matrix S given by:

$$S = \frac{\sqrt{3}}{3}\begin{Bmatrix} 1 & 1 \\ 1 & e^{j120^\circ} \\ 1 & e^{j240^\circ} \end{Bmatrix} \qquad (22)$$

A possible implementation is shown in Figure 4. As in the previous example, the light at x_2 is first converted to circular polarization using a quarter-wave plate oriented at 45°. Then the beams are combined using a beamsplitter designed to completely transmit the horizontal component but reflect two-thirds of the power of the vertical polarization. One of the output beams is then split using a polarizing beamsplitter at 45°; the other is used directly. While less symmetric than the quadrature system, implementation of (22) is fully achieved.

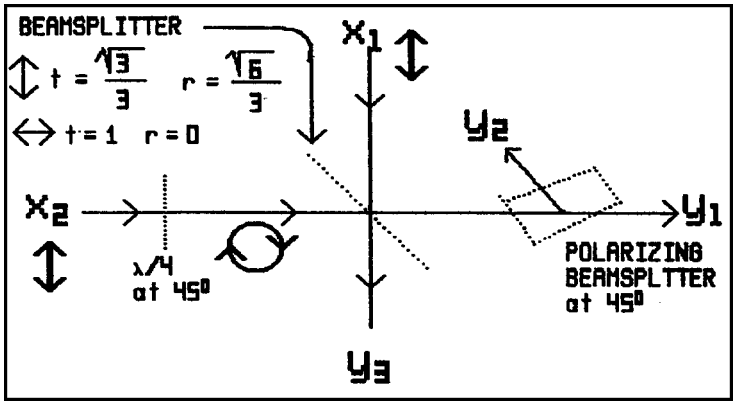

Figure 4: Three-phase photomixer using a custom beamsplitter, mixes x_1 and x_2 at 0°, 120°, and 240°. Polarizations at various stages are indicated by double arrows.

The Delay-Dispersed Photomixer

As earlier suggested, a photomixer may, instead of outputting **discrete** beams, produce an output in which the phase of interference varies over a **spatial continuum**. A notable example of such a configuration, we shall term the **delay-dispersed** photomixer. This has been the most common photomixing configuration used by stellar interferometers and is basically a generalization of Thomas Young's original two-slit interference experiment. The two optical sources are focussed onto two closely separated points and the then-diverging beams are allowed to fall upon the same imaging detector, at which point we observe (finally!) interference **fringes**, in the classical sense.

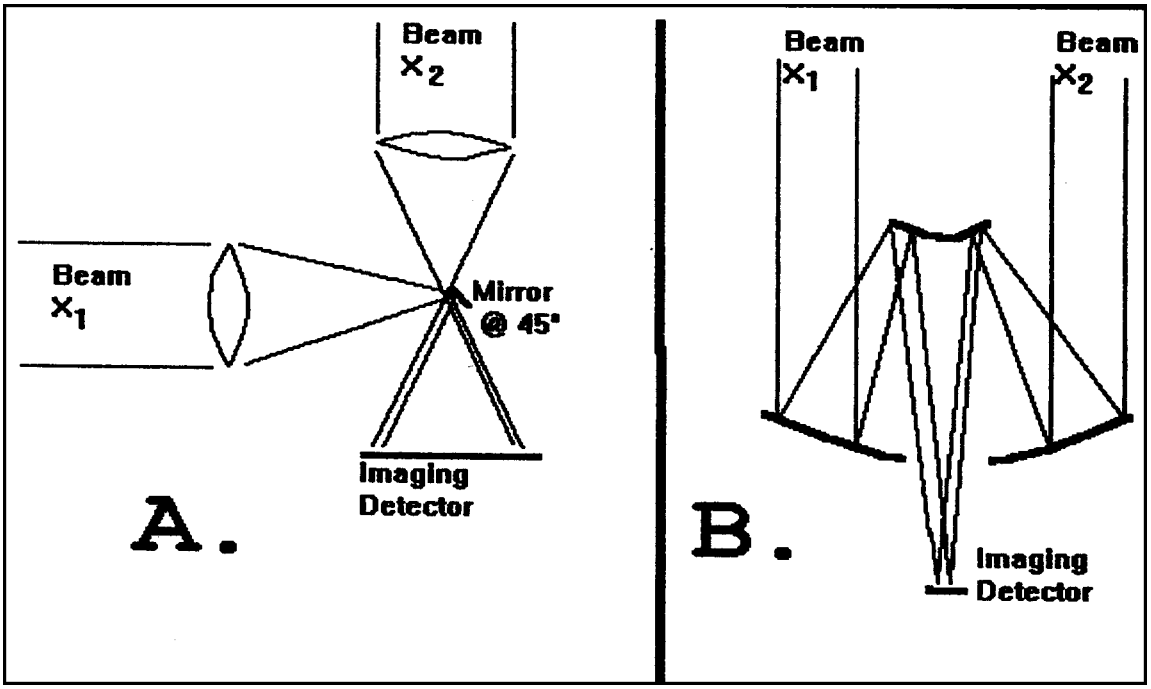

Figure 5: **Delay-dispersed photomixer**, two implementations.

One possible implementation is shown in Figure 5A. A popular scheme involves directing collimated input beams onto subapertures of a Cassegrain telescope (Figure 5B). The Airy disc corresponding to the subaperture used is then observed at the Cassegrain focus; correlation between the two input beams will superimpose a fringe pattern on the Airy disc.

In any such scheme, the fringe pattern, as it ultimately reaches the imaging detector, can be arranged to be of an arbitrary size. Using cylindrical optics, the extent of the image in the transverse dimension normal to the fringes can always be reduced (to the limit imposed by diffraction). The qualitative behavior of the delay-dispersed photomixer is ultimately determined by a single parameter, κ. If a two-slit interferometer consists of slits of width d_1 which are separated by d_2, then we define $\kappa = d_2/d_1$. In the case of Figure 5B, d_1 is the width (or diameter) of each input beam and d_2 is the separation between the input beams. κ is sometimes referred to as "the number of fringes" (although the **total** number of fringes in between the zeros of the single slit diffraction pattern is actually 2κ). Note that the minimum possible value of κ is 1, since a smaller value would require a physical impossibility. Also, it is possible to make κ **adjustable**, for instance, by finely adjusting the position of the small right-angle mirror in Figure 5A.

As previously noted, a photomixer with a continuum output can be viewed as one with discrete mode outputs by dividing this continuum into N regions, where N is a **large** number, in which case each region will essentially be a pure mode. In the case of noisy detectors, however, we wish to reduce N, thus invoking a penalty associated with **mode impurity** in the detected output. For simplicity, let us assume that the input modes to the delay-dispersed photomixer are **rectangular** in shape, rather than circular (like the image of a telescope objective) or Gaussian.

Having modelled each source as a slit of width d_1, we can write the power going to the detector at the deflection angle ψ from either source as:

$$p(\psi) = \mathrm{sinc}^2\left(\frac{\psi d_1}{\lambda}\right) = \left(\frac{\sin(\pi\psi d_1/\lambda)}{\pi\psi d_1/\lambda}\right)^2 \tag{23}$$

λ is the optical wavelength. The phase of interference due to two sources separated by d_2 is a linear function of ψ:

$$\theta(\psi) = \frac{2\pi\psi d_2}{\lambda} \tag{24}$$

Using the normalization:

$$x \triangleq \frac{\psi d_1}{\lambda} \tag{25}$$

we can express the power and interference phase in the following form:

$$\begin{aligned} p &= \operatorname{sinc}^2(x) = \left(\frac{\sin(\pi x)}{\pi x}\right)^2 \\ \theta &= 2\pi\kappa x \end{aligned} \tag{26}$$

thus revealing the significance of κ. We can define a function, $Z(\psi)$ which combines the power and phase. In order to determine the sensitivity of a detector element covering a segment of the fringe pattern, we can write an integral similar to (7) where the integrand is now replaced by:

$$p(\psi)(1 + \operatorname{Re}\left\{e^{j\theta(\psi)}V_m\right\}) = |Z| + \operatorname{Re}\{Z(\psi)V_m\} \tag{27}$$

where

$$Z(\psi) \triangleq p(\psi)e^{j\theta(\psi)} \tag{28}$$

Z is plotted in Figure 6A for $\kappa = 1$, and in Figure 6B for $\kappa = 2$. Note that $Z(x)$ is **not** a plot of a physical interference pattern; it is the **response function** of the photomixer at positions on the detector plane. Integrating its **magnitude** over the extent of a detector element yields the total power incident on that detector. Integrating its **complex value** over the detector yields a complex number indicating the phase and amplitude of interference.

Dividing the magnitude of **that** number by the total power obtained by integrating the magnitude of Z, yields the penalty due to **mode impurity**. The previous result obtained in the integral (8) can be seen to be a special case of such an integral where the optical power level (magnitude of Z) is constant over the detector field.

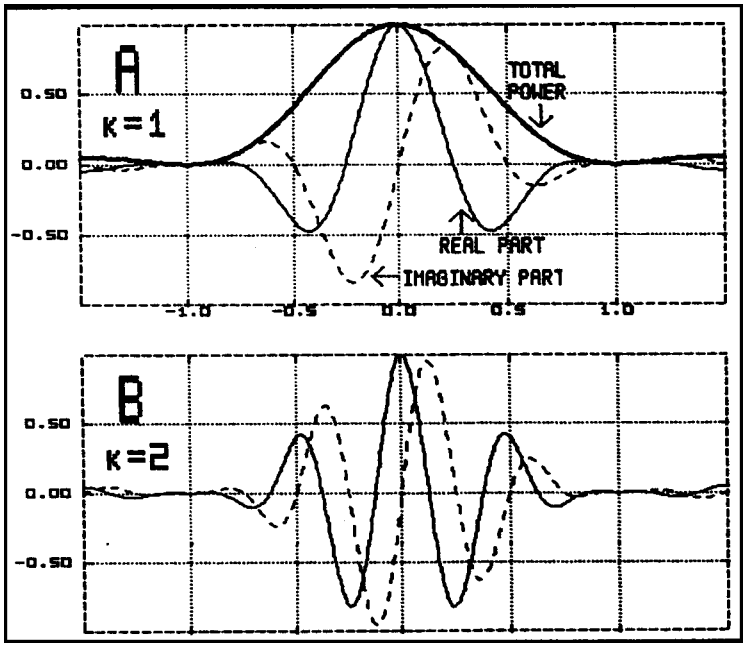

Figure 6: $Z(x)$, the response function of the delay-dispersed photomixer, plotted for two values of κ.

Looking at Figure 6, it can be seen that more than a few detector elements would be necessary to integrate most of the power without a significant mode impurity penalty.

Suppose we have chosen to only measure the 90% of the total power which falls in between $x = -.8$ and $x = +.8$. Let $\kappa = 1$ (the most favorable case). Then we could achieve mediocre performance using only 4 detectors of equal size:

Range of x		Net Power	Phase	Efficiency
-.80	-.40	.110	169°	.819
-.40	.00	.338	295°	.769
.00	.40	.338	65°	.769
.40	.80	.110	191°	.819

Note that the phase separation between outputs is about 130°. If instead we employ **6 detectors** the mode impurity penalty is relaxed to about .9. The phase separation between output channels is now close to 90 °:

Range of x		Net Power	Phase	Efficiency
-.80	-.53	.049	134°	.907
-.53	-.27	.152	222°	.893
-.27	.00	.247	314°	.889
.00	.27	.247	46°	.889
.27	.53	.152	138°	.893
.53	.80	.049	226°	.907

If we tried to apply the same configuration to a larger κ, however, we again suffer degraded performance. For $\kappa = 1.5$ the same detector coverage would yield a much poorer efficiency:

Range of x		Net Power	Phase	Efficiency
-.80	-.53	.049	21°	.801
-.53	-.27	.152	154°	.770
-.27	.00	.247	291°	.761
.00	.27	.247	69°	.761
.27	.53	.152	206°	.770
.53	.80	.049	339°	.801

While the actual integral over Z over the detector width yields an exact result, the mode impurity penalties we have calculated would be reasonably well approximated by the previous result (9) applied to the range of phase angle of the detectors. Overall, it can be seen that this photomixer configuration is handicapped when using noisy detectors (when minimizing N is advisable). Advantages of the configuration include its simplicity of construction (not requiring beamsplitters using elaborate coatings) especially for wide bandwidth operation, and its insensitivity to polarization (assuming a symmetric configuration such as Figure 5B). And unlike all of the discrete photomixers, this configuration can be constructed of **all reflecting optics**. These advantages are of most interest in the infrared, where, unfortunately, detector noise is dominant and the mode impurity penalty thus takes effect. If the delay-dispersed photomixer is to be used in such a case, one might want to lower the value of κ to be close to unity so as to minimize that penalty.

Bandwidth Considerations

Up until now we have assumed that the light entering the photomixer is monochromatic. Let us now relax that assumption and examine the effect that optical bandwidth has on the detection of interference. Refer to the model of the optical signal train in Figure 1. The discrete output photomixers had the characteristic that the response of any output channel is a function of the **phase** of the uncorrected delay, but that phase is itself a function of wavelength. Thus light of different wavelengths is not treated uniformly and a given output channel integrates different wavelengths **incoherently**. For instance, the quadrature photomixer described by (21) has a power in the first output beam given by:

$$\begin{aligned} |y_1|^2 &= |\frac{1}{2}(x_1 + x_2)|^2 \\ &= \frac{1}{4}E\left\{|e^{-j2\pi\nu(\tau_g+\tau)}A_1 + e^{-j2\pi\nu(\tau_g+\tau)}A_2|^2\right\} \\ &= \frac{1}{4}E\left\{A_1^2 + A_2^2 + 2\,\mathrm{Re}\left\{A_1 A_2^* e^{j2\pi\nu\Delta\tau}\right\}\right\} \\ &= \frac{A^2}{2}\left(1 + |V|\cos(2\pi\nu\Delta\tau + \arg(V))\right) \end{aligned} \tag{29}$$

Recall that $\Delta\tau$ is the **uncorrected** component of atmospheric delay given by $\tau - tau_c$. Assume for simplicity that $\arg(V) = 0$ for all ν. The this output channel would clearly exhibit a peak at $\Delta\tau = 0$. It would also reach the same peak for $\Delta\tau = 1/\nu$, $2/\nu$, etc. **However** if the light were not of a single wavelength, then the argument of the cosine in (29) would vary over the wavelength range. Suppose that the light's spectrum were centered at optical frequency ν_0 with a uniform power over a bandwidth of $\Delta\nu$. If the uncorrected delay $\Delta\tau$ were exactly N wavelengths at ν_0, then, assuming that V is independent of wavelength, we would measure a response integrated over wavelength of:

$$
\begin{aligned}
&= \frac{1}{\Delta\nu}\int_{\nu_0-\frac{\Delta\nu}{2}}^{\nu_0+\frac{\Delta\nu}{2}} d\nu \frac{A^2}{2}\left(1+|V|\cos(2\pi\nu\Delta\tau)\right) \\
&= \frac{A^2}{2}\left(1+V\frac{1}{\Delta\nu}\int\cos\left(2\pi\nu\frac{N}{\nu_0}\right)\right) \qquad (30)\\
&= \frac{A^2}{2}\left(1+V\frac{2\sin(2\pi\frac{\nu}{2}\frac{N}{\nu_0})}{(2\pi N/\nu_0)\Delta\nu}\right) \\
&= \frac{A^2}{2}\left(1+V\cdot\mathrm{sinc}(\Delta\nu\Delta\tau)\right) \qquad (31)
\end{aligned}
$$

So we find that for a time-bandwidth product, $\Delta\nu\Delta\tau$, which is not small compared to 1, the **measured** correlation is degraded by the sinc function evaluated at $\nu\Delta\tau$, in exactly the same way that measured correlation was degraded by a photodetector receiving light of spatial modes whose interference phase varied over a range, calculated in (9). There are only two ways to avoid this predicament. Either the maximum uncorrected delay must be reduced, or the fractional bandwidth must be reduced.[4]

Sufficiently reducing the delay is possible only with real-time delay-tracking ("fringe-tracking," **not** coarse delay-tracking),

[4]The delay-dispersed photomixer presents an **apparent** exception to this assertion. In this case, different wavelengths undergo the same time delay, rather than the same phase shift, so peaks in the output field (notably the central fringe) are shifted according to the uncorrected delay, but not **smeared** by different wavelengths.

However the hidden trade-off presents the same dilemma. On one hand, κ can be made small, in which case the tolerable delay error is small. Limiting τ to this range would require a suitable **real-time** delay-tracking system. On the other hand κ can be made arbitrarily large allowing the instrument to perform equally well over a large range of uncorrected delay. The, however, the amount of power routed to the central fringe is proportionally reduced. Since most of the delay estimation information is contained in the immediate vicinity of the central fringe, the overall delay-tracking performance is reduced by an amount similar to the reduction that would result from limiting the bandwidth of an instrument to achieve the same end.

and requires operation such that the excursions of uncorrected delay error are quite limited, generally within half an optical wavelength. An occasional delay excursion causing a loss of the central fringe would require "restarting" the delay-tracking process, beginning with a search procedure. This method, does in fact have certain advantages in the infrared where detector noise rules out the spectrally-dispersed system to be presented below.

Limiting the instrument's bandwidth, on the other hand, is the classical approach to avoiding loss of correlation predicted by (31). The required bandlimiting is inversely proportional to the maximum expected differential delay in order to keep the argument of the sinc function in (31) from approaching ±1. For constant atmospheric conditions, the maximum extent of differential delay is expected to vary as the 5/6 power of baseline distance (up to the "outer layer of turbulence") according to (42). For a long-baseline interferometer, at visible wavelengths, this may be in excess of 100 wavelengths. Thus a bandwidth of less than 1% would be required to maintain coherence, entailing a corresponding decrease in the amount of light received by the interferometer contributing to the estimation of the random atmospheric delay. We shall find that there is a minimum photon flux necessary for successful delay-tracking, and that photons of different wavelengths more or less equally contribute to this function. Consider the effect, for instance, of the 1% bandwidth suggested above. The associated 99% loss of optical power would entail a decrease in sensitivity fully amounting to 5 stellar magnitudes!

In addition to conserving the received optical power for the sake of delay-tracking, there are two further reasons that **white-light interferometry** is desirable. One concern is that V, the optical correlation or **fringe visibility** that we seek to measure, is itself a function of wavelength. In order to obtain the most information from an observation, we would like to receive different wavelengths simultaneously in order to measure V as a function of ν. However simply introducing broadband light into a photomixer as previously described, would measure some sort of "average" correlation.

Secondly, it is possible for a white-light interferometer to measure not only the amplitude, but also the **phase** of the visibility function $V(\nu)$. This is of extreme importance for the sake of image recovery which requires performing the Fourier transform on a **complex** visibility function to retrieve the **real** image brightness profile in the case of a non-symmetric image[5]. In the case of narrowband interferometer, the phase of $V(\nu)$ is intermingled with the phase delay caused by the random atmospheric delay, preventing both from being determined with one measurement[6]. The inclusion of simultaneous interferometric

[5]Measurement of the object phase is not required in the case of objects that can be safely assumed to be symmetric, since the Fourier transform of a symmetric function has zero phase. This would include common observations such as measurements of the diameter and brightness profile of solitary stars. For the case of non-symmetric objects, it has been shown that mathematical redundancy in the Fourier transform of two-dimensional objects **can**, given low noise measurements and good Fourier plane coverage, make possible the recovery of lost phase information.

There have also been techniques described under the category of **triple correlation interferometry** or **phase closure**, which permit extraction of the phase of the visibility function. This requires light received from at least **three telescopes** undergoing simultaneous interference.

Each of these approaches is dependent on dense coverage of the Fourier plane, thus requiring measurements on the same object using a wide range of baselines. For that reason they have been exclusively applied to **speckle interferometry** and are unlikely to prove useful for long-baseline interferometry

[6]More specifically, the **measured** complex correlation, V_m (see (3)), incorporates the phase not only of V, but also of $\Delta\tau$, the uncorrected delay error. Not only does $\Delta\tau$ vary wildly over a range of many wavelengths, but it includes *a priori* errors in the knowledge of the position of the object, which determines τ_g (see page 13 and Figure 1). Clearly that position could not be known to within a small fraction of a wavelength over a long baseline!

On the other hand, **simultaneous** measurement of V_m at N different wavelengths, yields instantaneous determinations of measured correlation at these wavelengths **all of which are subject to the same** $\Delta\tau$, thus permitting $N - 1$ comparisons of phases of $V(\nu)$ at N different ν. The remaining free parameter corresponds to the definition of the "center" of the image, which may be arbitrarily defined.

For this procedure to work in practice, it is first necessary that the instrumental dispersion of the interferometer optics be calibrated (for instance, by observation of a known symmetric object, or by observing an object at a baseline near zero), and that the dispersion function be stable over the period of an observation.

measurements using N different wavelengths, on the other hand, provides information combining both the atmospheric delay, **and** $V(\nu)$ at N different wavelengths, with one free parameter undetermined.

Assume that the object being observed is **uniformly colored**, so that there is no wavelength dependence to the normalized Fourier transform of the object function. The key to the recovery of the phase of interference, rests upon the fact that for a given **physical** baseline L, the measurement of interference at the optical frequency ν, determines the component of spatial frequency ν/L of the object. Thus in a wideband observation, a **range** of points on the Fourier plane will be determined in magnitude **and phase**, with one free parameter to be determined. Additional observations at different baselines can be used to match overlapping data, eventually resulting in a complete description of the Fourier transform of the object function in one direction. The possibility of recovering phase information in this manner has been examined by and [10], and simulations have been performed demonstrating the technique.

White Light Interferometry using the Delay-Dispersed Photomixer

The previous discussion of the delay-dispersed photomixer has been discussed relative to narrowband light. Unlike the previous discrete output photomixers, the detection response of this system is **wavelength dependent**. That can be seen from the definition of x (25); at a given detector position ψ the photomixing phase θ is characterized by a **time delay** $\Delta\tau$ (hence the name "delay-dispersed") so that the photomixing phase at the optical frequency ν is $2\pi\nu\Delta\tau$. One ramification of this characteristic in a wideband system, especially for larger κ, is the formation of a classical **white light fringe** at the center of the fringe pattern (for $\Delta\tau = 0$). As

the uncompensated delay error between the photomixer inputs wanders, the white light fringe shifts across the detector field in direct relationship to the delay error. While greatly simplifying the computational complexity for delay tracking, such a system has few additional merits. For large κ, an inordinately small amount of power is devoted to the channels where the white light fringe is most likely to be detected. For small κ, it would take but a small untracked delay discrepancy to move the white light fringe outside the range of detection.

Neglecting the issue of delay tracking, it can be observed that the superposition of different optical frequencies at different **spatial** frequencies, will preserve wavelength dependent information in the **Fourier transform** of the fringe pattern, permitting, in theory, its recovery. The wider the fringe pattern, (large κ), the finer the spatial frequency resolution determined by the fringe pattern, and therefore the finer the resolution in **optical frequency** (wavelength). Using such a system, implemented by **scanning** the delay line, [7] were able to reconstruct correlation as a function of wavelength. Amazingly they achieved useful results without even maintaining close delay tracking over the period of a scan.

Spectrally Dispersed Detection

When using low noise detectors, there is an alternative configuration that implements white light interferometry **without** either compromising the amount of light usable for delay-tracking, or the breadth of information received. All of the received light can be efficiently employed simultaneously as long as different wavelengths are independently detected. Imagine that the light received by each telescope is separated into N spectral channels, and each pair of optical signals is processed using a separate interferometer. Rather than constructing N interferometers to be operated in parallel, however, it is equivalent to route the broadband light through a single delay line and photomixer, and

then to disperse the detected light beams into spectra, so that different wavelengths get detected separately.

The **spectrally dispersed detection** system is diagramed in Figure 7. The output beams of the photomixer are each shown to be routed through a dispersive prism (or other spectrometer implementation) and each spectrum falls upon a linear detector array which detects that spectrum in N wavelength channels. In the case of a photomixer with one dimensional continuum output (page 28), the spectrometer would be implemented to preserve the image in that dimension, but spectrally disperse the light in the opposite two dimensional detector array.

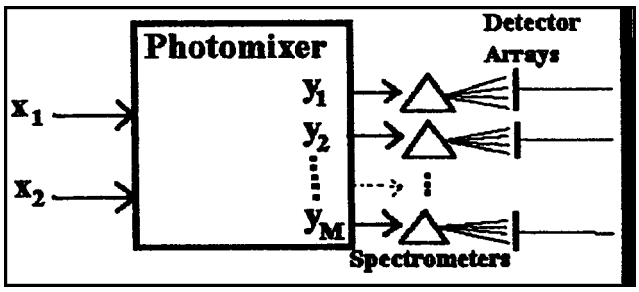

Figure 7: The **Spectrally-dispersed** detection system creates spectra of photomixer output beams for simultaneous detection of photon wavelength and interference phase.

Again, considering only the detector elements receiving a particular wavelength, we are presented with a system which is indistinguishable from a narrowband interferometer. All characteristics of a narrowband instrument apply to this subset of the total hardware. Now considering **two** wavelengths, we could just as well be looking at two interferometers side-by-side, each independently measuring the optical correlation at its optical frequency. From a delay-tracking standpoint, however, the situation is far improved over having two side-by-side instruments. That is because the optical delay affecting the two wavelengths is then guaranteed to be **identical**[7], unlike the case for physically

[7]At least this will be **assumed** for the discussion of delay-tracking. Specifically, the assumption states that a delay error at any one time can be characterized by a single number, τ, such that the phase difference measured by an interferometer at optical frequency ν, is exactly $2\pi\nu\tau$. A somewhat more general assumption — but one that is equally suitable for the purposes of delay-tracking — is that the delay process at any one time is described by a single number, x, and that the phase difference at optical

distinct optical collectors. Therefore their measurements of delay error could be combined, yielding an improved estimate. So using, say, 100 wavelength channels, we can reap the benefit of 100 simultaneously operating narrowband interferometers, but require only the hardware of a single instrument, up to the point of the prism in Figure 7. The **total** optical power received, then, can be used to estimate the single delay error characterizing the receiving light.

[7]frequency ν is given by $\Phi_0(\nu)x$, where $\Phi_0(\nu)$ is a fixed function of optical frequency describing the chromatic dispersion of a unit quantity of the turbulent medium. The more general model that we will **not** consider, is to suppose that the function of phase over optical frequency contains **two** (or more) degrees of freedom, and is thus given by $\Phi_1(\nu)x + \Phi_2(\nu)y$ where x and y represent the delay contributions from two different species whose chromatic dispersions are dissimilar.

The near-field approximation that is valid in describing the effect of the atmosphere on the light received at the earth's surface (except under poor seeing conditions), would rule out the possibility of geometric effects leading to anachromaticity in the atmospheric delay. The other possibility that might dispute the assumption would be the existence of two different material sources of atmospheric delay. It is generally assumed that in the visible and near-infrared the dominant source of differential delay is the turbulent mixing of air masses of different **temperatures**. A **secondary** source of differential delay would be the effect of air masses having different water-vapor content. While unquestionably of little significance in the visible, the differential effect of water vapor would be exaggerated in the infrared. Conceivably, an interferometer extending well into the infrared might require implementing a delay model with **two degrees of freedom** as suggested above.

Overview of Delay-Tracking Methods

Recall that there were two contexts in which delay-tracking applied to an optical stellar interferometer. We first identified **coarse** delay-tracking as a requirement to insure that interference was not partially or totally canceled due to the finite bandwidth of light falling upon a detector. The required accuracy for this aspect of delay-tracking was dependent upon the bandwidth of detector elements, and on the permissible delay-bandwidth penalty we were willing to tolerate in the detection of interference (complete cancellation of detected interference would occur for a delay-bandwidth product $\Delta\tau\Delta\nu = 1$).

The second way in which we discussed delay-tracking was also referred to as **phase-tracking** or **fringe-tracking**. This referred to the estimation, and possible optical compensation, of the differential atmospheric delay, **to well within a wavelength**, specifically for the purpose of interpreting the raw data in order to accumulate statistics yielding the magnitude of optical correlation that we seek to measure.

While delay-tracking in the second sense requires an error of much less than a wavelength, the tolerable error of coarse delay-tracking in the first sense may be similar (in the case of wideband detection) or much greater (in the sense of spectrally-dispersed detection with off-line delay-tracking). Also "coarse" delay-tracking necessarily refers to adjustment of the hardware (optical) delay-line, whereas "fringe-tracking," as we shall see, may occur in hardware **or** software. There **may** or may not be a great deal of overlap between the two tasks. However, in the following pages we will be solely addressing the problem of "fringe-tracking," or delay-tracking **for the purpose of integrating the raw data to measure optical correlation**.

We will first note, however, that there are three general cases of "coarse" delay-tracking. In the looses case, only open-loop delay compensation is implemented in order to correct for the expected geometrical delay, with no additional compensation for random delay components. At the other extreme, wideband detectors

will require real-time delay-tracking to well within a wavelength, in which case the two senses of delay-tracking may coincide. In the intermediate case, the detection bandwidth is such that a delay error of several or many wavelengths may be tolerated, but a feedback system is used which **reduces** the overall magnitude of delay error below that of the random atmospheric delay process that would remain uncorrected in the first case. Such techniques include those that are termed "group-delay tracking" in the literature ([4], [6]), and also could include tracking the envelope of delay error determined running the global path-estimation algorithm (page 172) in real-time. Henceforth, such a system will be termed **medium-coarse** (real-time) delay-tracking.

Now, relative to the tracking of differential path delay for the purpose of integrating the raw data produced by an interferometer, there are three general approaches we can identify. These three approaches to delay-tracking apply to somewhat different optical power levels (i.e. stellar magnitudes). These are termed **absolute** interferometry (which uses **no** delay-tracking), **real-time** delay-tracking, and **off-line** delay-tracking. Let us look at each.

Incoherently Averaged Interferometry

An **absolute** or **incoherently averaged** interferometer uses only *a priori* delay compensation, that is, **there is no feedback or attempt to estimate or compensate for atmospheric delay noise.** The measurement of magnitude of optical correlation or "fringe visibility" is obtained from an averaging scheme which assumes a random phase, and integrates the **square** of the magnitude of the optical correlation. The success of this scheme rests upon the assumption that within a certain time interval the differential phase delay does not change appreciably. Then, it can be shown that within a very short time interval the probability of **two** photons being found in the same interference phase is enhanced by a factor of $(1 + |V|^2/2)$ where V is the fringe visibility or normalized cross-correlation between the two received signals (we

have assumed a low photon arrival rate, so that the expected photon count in such a time interval is much less than 1).

Such a system has the advantage that it will work at **any** light level. Unfortunately the required integration time increases rapidly with decreasing signal level, since the base probability of detecting two photons in a short time interval falls according to the square of received optical power level. What's more, the detection is of the correlation **squared**, further reducing the sensitivity to smaller correlations. Also, as we have noted, the instrument responds only to the **magnitude** of correlation and loses all phase information (this limitation **can be** overcome with a "triple correlation" system, whose sensitivity falls even more rapidly with signal level). Despite these drawbacks, this is the only possible mode of operation in the case of very low light levels, in which case delay-tracking would be impossible.

Figure 8 shows the block diagram of an incoherently averaged system, for illustrative purposes based on analog electronics (rather than discrete photon processing). The two optical signals, after receiving **coarse** (open loop) delay compensation, are interfered in two opposite phases, producing optical outputs I_1 and I_2 which are detected to produce analogous electrical signals. Assuming that the total input is I_0 (which may be varying) and is equally divided between x_1 and x_2, we know that the intensities expected at I_2 and I_2 are governed by:

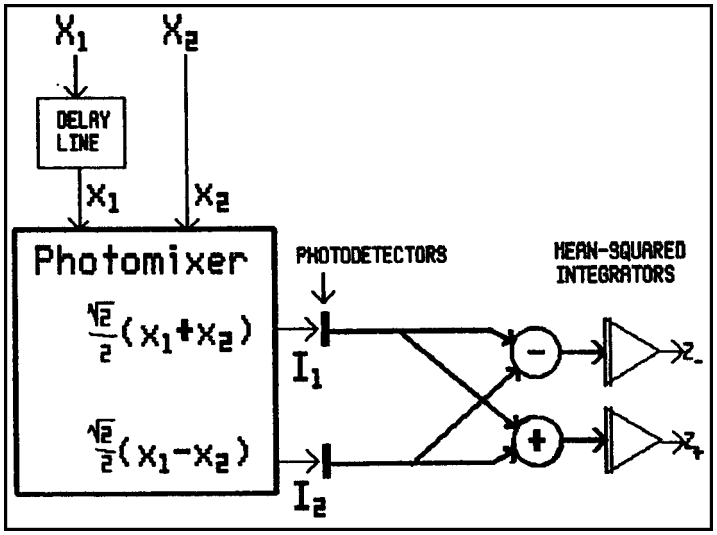

Figure 8: An **incoherently averaged** interferometer uses open-loop control of the delay line. Optical correlation is manifest as an increase in the "noise level" of $I_1 - I_2$.

$$I_1 = \frac{I_0}{2}(1 + |V|\cos\Phi) \tag{32}$$

$$I_2 = \frac{I_0}{2}(1 - |V|\cos\Phi) \tag{33}$$

where Φ must be regarded as a totally random phase (since no delay-tracking is employed). We assume that the response of the photodetector and preamp combination is (whether intentionally or not) low-pass filtered, but still has a rapid response in relation to the changing phase of atmospheric delay. We also assume that any dark current has been subtracted. Then the sum and differences of I_1 and I_2 are formed and **squared**. The results are averaged over a period of time to reduce the effects of noise, resulting in the accumulated statistics Z_- and Z_+. Using (33) and considering Φ to be random with a uniform distribution, the expectations of the resultant statistics are evaluated.

$$Z_- = \langle (I_1 - I_2)^2 \rangle = \frac{1}{2}\overline{I_0^2}|V|^2 + \overline{n_1^2} + \overline{n_2^2} \tag{34}$$

$$Z_+ = \langle (I_1 + I_2)^2 \rangle = \overline{I_0^2} + \overline{n_1^2} + \overline{n_2^2} \tag{35}$$

n_1^2 and n_2^2 refer to the mean squared noise introduced by detectors 1 and 2 respectively. By subtracting these noise terms, we form the following estimator for the **squared magnitude** of optical correlation.

$$|V|^2 \approx 2\frac{\langle (I_1 - I_2)^2 \rangle - \overline{n_1^2} - \overline{n_2^2}}{\langle (I_1 + I_2)^2 \rangle - \overline{n_1^2} - \overline{n_2^2}} = 2\frac{Z_- - Z_{-0}}{Z_+ - Z_{-0}} \quad (36)$$

In the final expression, Z_{-0} refers to the observed value of Z_- **under the condition of no optical correlation**, which determines a baseline value for the detector noise power. Z_{-0} could be measured, for instance, by introducing an optical misalignment or path length discrepancy which destroys the interference without affecting the net optical power falling upon the detectors.

Real-Time Delay-Tracking

A **real-time** delay-tracking system is workable and desirable under the condition of a sufficiently strong signal level. In such a system the **imaginary part** of measured correlation is monitored in real-time and comprises the error signal of a control system. The output of the control system servos a high speed path length compensator in order to cancel differential delay errors. The net delay error is kept to a small fraction of a wavelength enabling **coherent integration** of the real part of measured correlation. Unlike the incoherently averaged mode described above, integration is now of fringe visibility V, not $|V|^2$, removing the penalty for observing well resolved objects (except insofar as the delay-tracker **itself** will eventually fail as the magnitude of V is reduced).

The block diagram of a possible real-time delay-tracking system based on analog electronics is shown in Figure 9. The received optical signal X_1 passes through a path-length compensator (delay line) which rapidly responds to the output of a feedback system. The delayed light, x_1 is fed into a quadrature (4-phase)

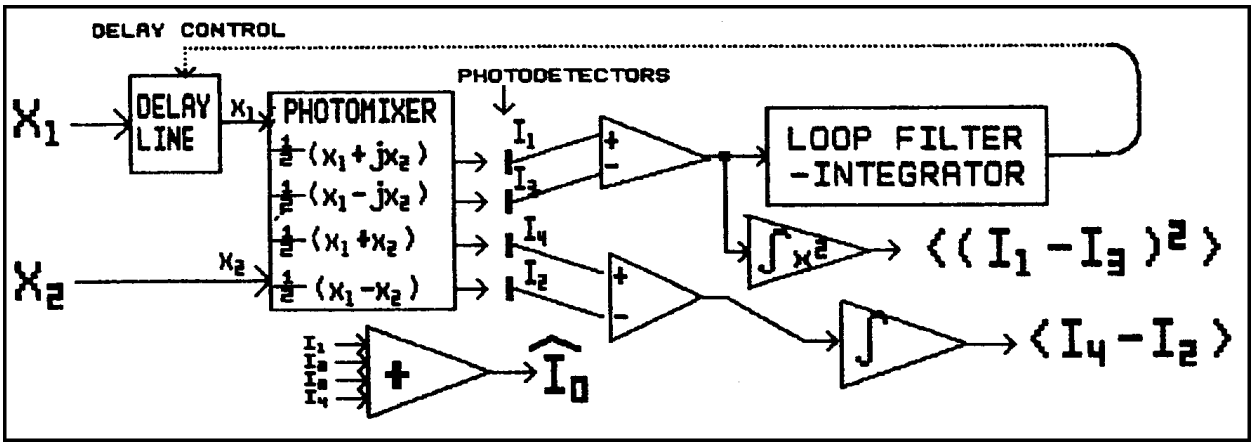

Figure 9: A **real-time** delay-tracking system servos the delay line to maintain a delay error of much less than a wavelength, using the imaginary component of correlation as an error signal.

---

photomixer, the outputs of which are detected and subtracted to obtain instantaneous estimates of the real and imaginary parts of the correlation between x_1 and x_2. Of course these measurements are inevitably corrupted by noise.

The difference between the signals labelled I_1 and I_3 is sensitive to the imaginary part of correlation.

$$\begin{aligned} I_1 - I_3 &= \left(\frac{1}{2}|x_1 + jx_2|\right)^2 - \left(\frac{1}{2}|x_1 - jx_2|\right)^2 \\ &= -\mathrm{Re}\{jx_1x_2^*\} = \mathrm{Im}\{x_1x_2^*\} \end{aligned} \tag{37}$$

This signal is then filtered and controls the delay line in order to reduce the imaginary component of correlation. The filter **might** consist of simply an integrator with a certain gain, in which case we have implemented a basic first-order linear control system (we have assumed a small phase error so that the imaginary part of correlation, given by the sine of phase error, is proportional to the phase error itself). Increasing the loop gain would increase the speed of the control loop which, not considering noise, would lower the net phase error. Unfortunately, detector noise (not considered in (37)) will also be amplified with increased gain. The optimum gain and frequency response of the feedback loop will

thus depend on the level and spectral characteristics of the random delay process and detector noise.

The **real part** of the correlation is measured in a like manner using the other two phases of interference.

$$\begin{aligned} I_4 - I_2 &= \left(\frac{1}{2}|x_1 + x_2|\right)^2 - \left(\frac{1}{2}|x_1 - x_2|\right)^2 \\ &= \mathrm{Re}\{x_1 x_2^*\} \end{aligned} \tag{38}$$

The result can be converted to the magnitude of "fringe visibility" (normalized correlation) V. We may use an adder, as shown in Figure 9, to sum the four detected intensities (we assume that dark current has been subtracted out) to determine the **total** light power received, labelled I_0. If we can assume that the two input beams are of the same power, then we can write:

$$|x_1|^2 = |x_2|^2 = \frac{I_0}{2} \tag{39}$$

Using I_0 to normalize $I_4 - I_2$ we find:

$$\frac{I_4 - I_2}{I_0/2} = \frac{\mathrm{Re}\left\{x_1 x_2^*\right\}}{I_0/2} = \mathrm{Re}\left\{\frac{x_1 x_2^*}{|x_1||x_2|}\right\} = |V|\cos(\Phi) \tag{40}$$

where Φ represents the instantaneous phase error between x_1 and x_2, the phase that the control loop is attempting to cancel. Integrating the voltage will yield an estimate of $|V|$ that is degraded by the effect of non-zero Φ:

$$\left\langle \frac{I_4 - I_2}{I_0/2} \right\rangle = |V| E\{\cos(\Phi)\} = |V| e^{-\frac{1}{2}\overline{\Phi^2}} \tag{41}$$

This result, based on Φ having a Gaussian distribution, is thus seen to depend on the mean squared phase error, which might be estimated by integrating the square of $I_1 - I_3$ (as shown in

Figure 9), which (neglecting noise) we have shown is proportional to the imaginary part of correlation.

While further details of hardware implementation of a real-time delay-tracking system would require an extensive discussion, its **performance** can be estimated using the results that will be obtained for an **off-line** delay-tracking instrument by imagining that the data processing unit has been made infinitely fast (and in fact real-time delay-tracking **can be implemented as part of** what will be termed the off-line delay-tracking system, precisely by employing a fast computer). The expected estimation error that would thus result from employing only past photon detections, is discussed starting on page 149. We will not further dwell upon the characteristics of a real-time delay-tracking instrument, except to emphasize its practicality under the condition of sufficiently strong signals.

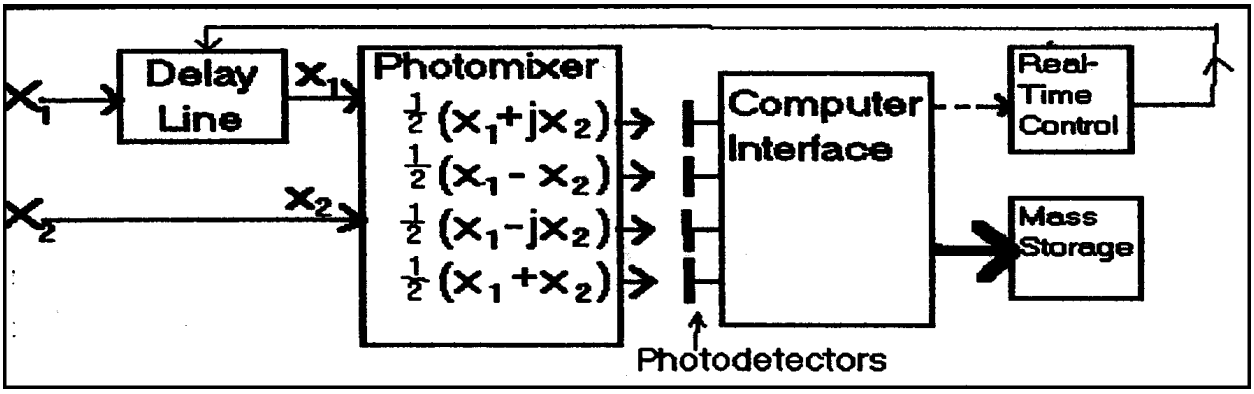

Figure 10: In an **off-line** delay-tracking system, real-time control of the delay-line is **optional**. Estimation of the delay function is performed by subsequent processing of stored photon events.

Off-Line Delay-Tracking

Finally, the **off-line** delay-tracking system, on which the remainder of this work is based, will be briefly introduced. **Off-line** delay-tracking has a sensitivity similar to, but extending substantially beyond that of **real-time** delay-tracking. A block diagram of such an interferometer is shown in Figure 10. As in the previous configurations, the signal X_1 is first passed through a delay line. As in the incoherently averaged system, only **coarse** delay equalization is absolutely required. The control of the

delay line is derived from a controller which may **optionally** receive feedback from the detection system. The light, as shown, is then fed into a quadrature photomixer[8]. The detector outputs are accepted by a computer interface for recording onto mass storage for **subsequent** analysis.

Of course a data processing system of sufficient speed **may** be able to carry out this analysis at the same rate that the data is collected. We shall still refer to this as an **off-line** system for the reason that the algorithms employed to estimate delay error at time t will employ data collected before **and after** time t. Thus even if a very fast computer is used to servo the hardware delay line in real-time (an **optional** configuration), **that** estimate of optical delay will necessarily be inferior to the ultimate estimate employing what amounts to a **non-causal** filter. The beauty of the off-line process is that the **improved** estimate of optical delay error can be employed to better estimate V, the optical correlation we are ultimately trying to determine. What's more, the interferometer hardware required to implement the off-line delay-tracking system does not go beyond that described for the previous two cases; the only additions involve data processing hardware.

Now, consider the hardware depicted in Figure 10, assuming that we **have** implemented the optional feedback loop to control a rapidly responding delay line. Then it is apparent that we have implemented all of the hardware required to also perform either **incoherent averaging** or **real-time** delay-tracking, as they were previously described. All we require is a modest computational power in the real-time delay-tracking control unit shown in Figure 10. Thus, by adopting this general hardware configuration, we have subsumed the other two methods.

[8]Since close control of the uncorrected delay is not expected, the implementation of a wideband system would require using the spectrally-dispersed detection system, not explicitly shown in Figure 10.

Estimation of the differential delay function using the stored data, is the primary issue addressed throughout the following pages. Having created that estimate of the random atmospheric delay function, the procedure for integrating the accumulated off-line data in order to determine $V(\nu)$ is described in Appendix II. First, however, we will make a few general observations concerning the relationship between delay-tracking systems and the choice of photomixing and detection hardware.

Photomixing and Bandwidth Considerations

In the preceding descriptions of the three approaches to delay-tracking, there was no reference to detection systems and bandwidth considerations in particular. Note that each photomixing configuration shown in the examples was included for illustrative purposes, and does not represent the only possible configuration for the respective delay-tracking system. We chose to illustrate the real-time and off-line delay-tracking systems using a quadrature photomixer since, as we have previously noted, a two-phase photomixer is incapable of resolving the sign of delay error. In the case of incoherently averaged interferometry, this was not a concern, and the two-phase photomixer shown in Figure 8 would therefore suffice.

Let us recall the characteristics distinguishing different photomixing and detection configurations. One important characteristic was the degree to which the received light was divided among a number of detectors. This was of particular importance in the infrared where detector noise would be dominant, but also could become important if the number of low-noise detectors were to be increased to the point that a faint optical signal would no longer be quantum noise dominated in each one. The second important characteristic was the fractional bandwidth received by any one detector element, since this placed a limit on the maximum optical delay that could be tolerated without compromising the measurement of interference as determined in (31). Third, there was a penalty to be paid for detectors which, in quest

of the first criterion, combined light from spatial modes of differing phases, reducing the detection of interference according to (9). Recall that there were trade-offs possible between these three goals.

Now, let us see how these characteristics relate to the three approaches to delay-tracking. The most striking relationship pertains to bandwidth reduction. In either the **incoherently averaged** system or the **off-line** delay-tracking system **not using real-time feedback**, the net excursion of uncorrected delay error is fairly large, and increases with longer baselines. The preferred solution to bandwidth reduction is, as we have previously discussed, implementation of spectrally-dispersed detection. The repercussions of dividing the light into N spectral channels are far more serious in the case of the incoherently averaged system for two reasons. First, because that is a system that would need to be used at the lowest light levels, at which point detector dark counts would become dominant for a much smaller value of N. On top of that, because this configuration does nothing to measure either the instantaneous value or statistics of the atmospheric delay process, the effect of a non-negligible value of $\Delta\nu\Delta\tau$ is much more pernicious. The degradation of measured correlation according to (31) would affect the interferometer's estimation of $|V|$ **without any ability to correct for this degradation**, due to the lack of any independent measurement of the delay error in effect during the observation. In the case of the off-line delay-tracking system this is not problematic precisely because the delay function **will** eventually be estimated, thus supplying a figure for correcting degraded measurements of correlation, as described in Appendix II.

However, even using off-line delay-tracking with a spectrally-dispersed detection system, there will still be an eventual point at which spectral diversity will compromise the signal-to-noise ratio, due to the optical level at each detector element becoming insufficient for quantum limited operation. This becomes particularly true as we venture into the infrared wavelengths. A decrease in detector signal-to-noise ratio will translate into an increased optical level required in order for the off-line delay

tracking algorithm to become operational. One useful feature in a spectrally-dispersed detection system might be a (cylindrical) "zoom lens" in between the spectrograph and the detector, so that N, the number of detector elements covering the spectrum, may be adjusted. Then N could be reduced to optimize the instrument for observation of the very dimmest objects, especially under the best seeing conditions and/or using shorter baselines. N would otherwise be increased to accommodate the larger expected excursions of atmospheric delay anticipated in the case of longer baseline and poorer seeing, and also to obtain superior spectral resolution.

A strategy that could be of assistance in reducing N in the operation of the off-line delay-tracking system, is to simultaneously employ, in real-time, what was previously referred to as **medium-coarse** delay-tracking (page 42). In this case, by rapidly processing the received interference signal to track the **envelope** of possible atmospheric delay estimates in a coarse manner, the required allowable range of uncorrected delay error may be greatly reduced (especially relative to that necessary for processing light received from a very long baseline configuration). That would allow the reduction of N, reducing the effect of detector noise. Of course, at higher signal levels, the same hardware would just become an implementation of the real-time **phase**-tracking system already discussed.

Now let us consider the question of bandwidth in the use of the **real-time** delay-tracking system. A real-time delay-tracking system using a feedback system based on the imaginary part of measured correlation, such as shown in Figure 9, either maintains a delay error of less than half a wavelength **or it "loses lock."** So, as long as it works at all, we can see that allowing detectors to receive a very wide bandwidth **is allowed** in this case. This is therefore the optimum configuration to be used in the infrared, at wavelengths where low-noise detectors are unavailable.

Even though we may thus choose to use wideband detection for optimizing the delay-tracking performance of an infrared

interferometer, we need not simultaneously abandon any hope of measuring $V(\nu)$ with a higher degree of spectral resolution. By using **most** of the received light in a wideband detection channel for delay-tracking, one can use a **narrowband pick-off** to simultaneously measure the correlation in a narrow spectral band. Since the amount of received light in such a narrow bandwidth would be small, with the detector contributing no less noise, the result of this detection would be a weak signal **buried in noise**. However since the delay-tracking information is obtained from the wideband channel, the assurance of a reasonably small delay error is maintained. By taking advantage of that delay-tracking, coherent integration of the (noisy) narrowband correlation channel can be performed over time, eventually producing an accurate estimate of $V(\nu)$.

That would be a reasonable strategy for delay-tracking and coherent integration of optical correlation using an interferometer having only infrared capabilities. It would also be appropriate for observing objects whose emission is almost totally in the infrared. On the other hand, when observing objects with a more or less "white" spectrum (even including stars that are described as "red"), the most sensitive mode of delay-tracking would probably be based on use of the visible radiation. Even though, in interferometric detection, there is a strong advantage realized with increasing wavelength, a sharp decline in sensitivity occurs as we exceed the maximum wavelength (about 1 μm) at which efficient photon counting with low noise detection is technologically possible. However a delay-tracking instrument in which most (or all) of the delay-tracking information is obtained from visible photon detections, can simultaneously perform infrared detection, using the successful estimation of delay using the visible radiation to enable coherent integration of the infrared correlation.

Model of Atmospheric Path Delay

We wish to characterize the random process affecting the phase of interference observed by a stellar interferometer due to atmospheric turbulence. Consider the case of two telescopes at positions r_1 and r_2, receiving light from the same star, as shown in Figure 11. Let us denote the net delay a light wave encounters in its trip trough the atmosphere, when observed at a point r_1 on the earth, as $\tau_0(r_1)$. According to the widely accepted Fried model [1], the random field τ_0 is Gaussian and characterized by the structure function $D_{\tau 0}$ given by:

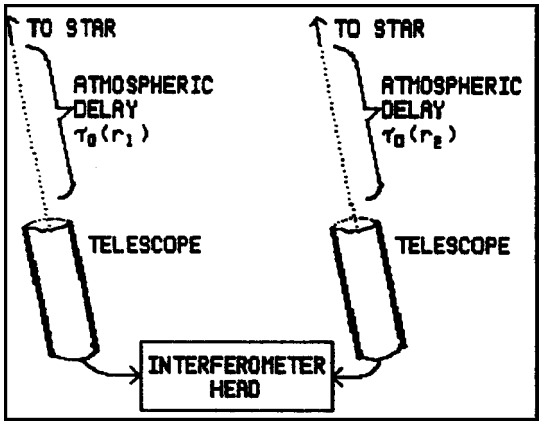

Figure 11: Starlight in each of two paths is subject to atmospheric delay denoted τ_0. The interferometer sees the differential delay, denoted, simply, $\tau(t)$.

$$D_{\tau_0}(\vec{r_1}, \vec{r_2}) \stackrel{\Delta}{=} E\left\{\left(\tau_0(\vec{r_1}) - \tau_0(\vec{r_2})\right)^2\right\} = \frac{6.88}{G_\tau} L^{5/3}$$

where: (42)

$$L \stackrel{\Delta}{=} |\vec{r_1} - \vec{r_2}|$$

L is the distance separating two observation points (or **baseline**) and G_τ is a parameter expressing the magnitude of atmospheric turbulence having dimensions of length$^{5/3}$*time^{-2}. In terms of the more familiar **Fried Parameter**, r_0, applicable to the optical frequency ν_0, G_τ is given by:

$$G_\tau = r_0^{5/3} (2\pi\nu_0)^2 \tag{43}$$

Consider an interferometer over a baseline of length L. The **differential** delay is simply denoted τ.

$$\tau \stackrel{\Delta}{=} \tau_0(\vec{r_1}) - \tau_0(\vec{r_2}) \tag{44}$$

τ will thus be a zero-mean Gaussian random process of variance:

$$E\left\{\tau^2\right\} = \frac{6.88}{G_\tau} L^{5/3} \tag{45}$$

Dynamics of $\tau(t)$

To evaluate the temporal characteristics of this process, we shall accept the **Taylor hypothesis** which assumes that the turbulence of the atmosphere is "frozen in" and is simply swept across our field of view by a wind of velocity v_0. Then it can be shown that $\tau(t)$ is a zero-mean stationary Gaussian process characterized by an auto-correlation function $R_{\tau\tau}(\Delta t)$ given by:

$$R_{\tau\tau}(\Delta t) = \frac{3.44}{G_\tau} v_0^{5/3} \Delta t^{5/3} \quad \left[\left(1 + L \frac{L + 2v_0 \Delta t \cos\theta}{(v_0 \Delta t)^2} \right)^{5/6} + \left(1 + L \frac{L - 2v_0 \Delta t \cos\theta}{(v_0 \Delta t)^2} \right)^{5/6} - 2 \right] \tag{46}$$

where θ is the angle between the wind velocity vector and the baseline vector. We are mainly concerned with the small Δt behavior of this function, in which case $R_{\tau\tau}(\Delta t)$ is well approximated by:

$$R_{\tau\tau} \approx \frac{6.88}{G_\tau} L^{5/3} \left[1 - (\frac{v_0}{L} \Delta t)^{5/3} \left(1 - \frac{5}{6} (\frac{v_0}{L} \Delta t)^{1/3} \right) \right] \tag{47}$$

For long baselines and short time intervals, $L \gg v_0 \Delta t$, we can neglect the final factor and further approximate $R_{\tau\tau}$ as:

$$R_{\tau\tau} \approx \frac{6.88}{G_\tau} L^{5/3} \left[1 - (\frac{v_0}{L} \Delta t)^{5/3} \right] \tag{48}$$

For a random process $x(t)$ with a finite variance, the structure function is related to the auto-correlation by:

$$D_x(\Delta t) = 2\,[R_{xx}(0) - R_{xx}(\Delta t)] \tag{49}$$

In this case, the **temporal** structure function of τ, defined as the mean-squared change in τ over a time interval Δt, would be approximately:

$$D_\tau(\Delta t) \approx \frac{13.76}{G_\tau} v_0^{5/3} \Delta t^{5/3} = T_\tau^{1/3} \Delta t^{5/3} \tag{50}$$

where we now define the single controlling parameter, T_τ, as:

$$T_\tau \triangleq \frac{2605}{G_\tau^3} v_0^5 = \frac{v_0^5}{r_0^5} \frac{2605}{(2\pi\nu_0)^6} \tag{51}$$

Note that under this assumption (large L) the structure function for τ is no longer dependent on the baseline L, and is, in fact, exactly twice the value of the temporal structure function of τ_0, the one-way atmospheric delay. This is what we would expect from the addition (or in this case, subtraction) of two identically distributed independent random processes. Thus for large L, this equation treats $\tau_0(r_1)$ and $\tau_0(r_2)$ as independent. Given typical separations for the elements of a long-baseline interferometer, we have formed a reasonable approximation for the dynamics of τ using the statistics of τ_0 alone. (50) is no longer contaminated by the assumption of the atmospheric delay structure propagating as a monolith over a long range, which accounted for the complexity of the more "exact" expression, (46).

A simple parameter reflecting the time basis on which the phase of interference changes, is often termed the "fringe coherence time" or "speckle lifetime" parameter T_0, defined as the Fried parameter r_0 divided by the "wind velocity" v_0 (different

authors have used definitions of this quantity that may vary by a constant factor[9]).

$$T_0 \triangleq \frac{r_0}{v_0} \tag{52}$$

It can be verified using (50) that in the time period T_0, the r.m.s. change in optical phase is 3.71 radians (which would suffice as an alternate definition). Unfortunately T_0 is defined in terms of r_0 which varies with wavelength. Thus we will, in most instances, employ non-narrowband parameters such as T_τ and G_τ, even if their definitions are relatively opaque.

We have now simplified the characterization of $\tau(t)$ using (50) in which only one parameter, T_τ, need be specified. However, one word of caution is in order. According to this simplified model, $\tau(t)$ has ever increasing energy at low frequencies. In fact, our approximations had assumed a small Δt and were thus never meant to be accurate at low frequencies. Therefore at low temporal frequencies, or equivalently, at large Δt, this formula grossly overestimates the magnitude of differential path delay. The eventual result of this discrepancy will thus be a pessimistic evaluation of an interferometer's performance. It will be seen, however, that in tracking $\tau(t)$, the low frequency components are the least problematic, and the resulting degradation of estimated performance is small.

It might be pointed out, however, that for an interferometer **not performing** tracking of the interference phase (i.e. using open-loop control of the delay lines), the penalty for low frequency components of $\tau(t)$ differs, and the low frequency components are thus of much greater significance. Since the low frequency cut-off

[9]For instance [8] uses τ_0 corresponding to the period in which the expected r.m.s. phase change due to τ would be exactly 1 radian. In terms of our T_0, his $\tau_0 = .207T_0$. [4] uses a time period, t_0, at which the temporal structure function of the **one-way** atmospheric delay τ_0 (see page 54), would be 1 radian^2. Thus, in terms of our T_0, his $t_0 = .314T_0$.

below which (50) breaks down, increases with the interferometer's baseline L (at least up to the size of the so-called "outer layer of turbulence"), it can be seen that an open loop non-delay-tracking scheme suffers with increasing baseline while a delay-tracking instrument maintains its performance up to the point where an object begins to become resolved.

Given the desirability of a model specified by only one parameter, we shall continue using (50). We must be prepared to consider an alternative description, however, in addressing issues which are dependent on low-frequency components of $\tau(t)$, such as establishing an *a priori* search space for initial acquisition of interference, and certain other problems not addressed in the present work.

The Power Spectrum of $\tau(t)$

Due to the increasing energy at low frequencies, given the above model, we face a problem when we try to find the power spectrum of $\tau(t)$. Normally the power spectrum would be found by obtaining the Fourier transform of the auto-correlation. The auto-correlation of $\tau(t)$, however, according to the present approximation (50) is not well behaved (it is not described as a positive-definite function) and the Fourier integral does not converge.

Therefore we shall use an indirect method of determining the power spectrum of this process. We shall, at the same time, be introducing and characterizing the cosine transform of $\tau(t)$ which will be employed in the theory to be later presented concerning **path estimation**. Le us consider a sample function of $\tau(t)$ over the interval $= 0, T$, which is Gaussian and governed by (50). We can expand over this interval in a cosine transform:

$$\tau(t) = \sum_{i=0}^{\infty} U_i \cos(\frac{\pi}{T} it) \quad 0 < t < T \tag{53}$$

Over the ensemble of realizations of $\tau(t)$, the U_i are then distributed as zero-mean independent Gaussian random variables whose variance we shall denote as:

$$\text{Var}\{U_i\} = E\{U_i^2\} = \sigma_i^2 \tag{54}$$

Our task, then, is to find the values of the σ_i^2 consistent with (50).

Recalling the definition of the structure function D_τ:

$$D_\tau(\Delta t) \triangleq E\left\{[\tau(t) - \tau(t + \Delta t)]^2\right\} \tag{55}$$

and substituting in the cosine transform expansion (53) for τ, we obtain:

$$D_\tau(\Delta t) = E\left\{\left[\sum_{i=0}^{\infty} U_i[\cos(\frac{\pi}{T} it) - \cos(\frac{\pi}{T} i(t + \Delta t))]\right]^2\right\} \tag{56}$$

While the expected value is stipulated to be taken over the ensemble of τ functions, we could additionally let it run uniformly over all t in the interval $0, T$. This will allow us to evaluate the right hand side of (56). Expanding the square inside the summation, and replacing the expected value of the sum with the sum of the expected value we find:

$$D_\tau(\Delta t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} E\Big\{U_i U_j \quad [\cos(\frac{\pi}{T} it) - \cos(\frac{\pi}{T} i(t + \Delta t))] \cdot [\cos(\frac{\pi}{T} jt) - \cos(\frac{\pi}{T} j(t + \Delta t))]\Big\} \tag{57}$$

For terms in which $i \neq j$, the U_i and U_j are uncorrelated, so the expected value of such terms is necessarily zero. Retaining the terms in which $i = j$ we find:

$$D_\tau(\Delta t) = \sum_{i=0}^{\infty} E\Big\{U_i^2 \quad [\cos^2(\frac{\pi}{T}it) + \cos^2(\frac{\pi}{T}i(t+\Delta t)) \\ -2\cos(\frac{\pi}{T}it)\cos(\frac{\pi}{T}i(t+\Delta t))]\Big\} \tag{58}$$

Noting that the U_i are not functions of t or Δt, we may express the expectation of the product as the product of the expectations. We then substitute in σ_i^2 for the mean squared value of the U_i as defined in (54), and take the expected value **over** t of the cosine terms. We can immediately assign $1/2$ to the expected value (over time) of terms of the form $\cos^2(kt)$ (except for the zero frequency term, $i = 0$, resulting in an in accuracy of negligible significance as T grows large). Then expanding the last cosine factor, we obtain:

$$D_\tau(\Delta t) = \sum_{i=0}^{\infty} \sigma_i^2\Big[1 - 2E\Big\{\cos(\frac{\pi}{T}it) \quad [\cos(\frac{\pi}{T}it)\cos(\frac{\pi}{T}i\Delta t) \\ -\sin(\frac{\pi}{T}it)\sin(\frac{\pi}{T}i\Delta t)]\Big\}\Big] \tag{59}$$

Again taking the expectation over t, the first term inside the brackets is seen to produce a $\cos^2(t)$ term while the second term results in zero expectation. We find:

$$D_\tau(\Delta t) = \sum_{i=0}^{\infty} \sigma_i^2 \left[1 - \cos(\frac{\pi}{T}it)\right] \tag{60}$$

Now equating this expression for D_τ with the formula (50) derived from the dynamics of atmospheric turbulence, we shall determine the values of the σ_i^2.

$$\sum_{i=0}^{\infty} \sigma_i^2 \left[1 - \cos(\frac{\pi}{T} i\Delta t)\right] = D_\tau(\Delta t) = T_\tau^{1/3} \Delta t^{5/3} \tag{61}$$

For large T there is no loss in accuracy if we transform the discrete summation into an integral over i. If we also divide both sides by $T\Delta t^{5/3}$ and make the substitution:

$$x \equiv \frac{i\Delta t}{T} \tag{62}$$

and consider σ_i^2 to now be a function of a continuous argument, we obtain:

$$\int_0^{\infty} \frac{\sigma^2(Tx/\Delta t)}{\Delta t^{8/3}} [1 - \cos(\pi x)] dx = \frac{T_\tau^{1/3}}{T} \tag{63}$$

Since this must hold equally well for any Δt, and Δt only appears in the fraction of the integrand, this fraction must therefore not by a function of Δt:

$$\frac{\sigma^2(Tx/\Delta t)}{\Delta t^{8/3}} = A \tag{64}$$

where A is an as-of-yet undetermined function of only x and T. Let us hold x and T constant, so that i is inversely proportional to Δt according to (62). Then substituting the frequency index i back into (64) in order to again express σ^2 in terms of i, we obtain:

$$\sigma^2(i) = A \frac{T^{8/3} x^{8/3}}{i^{8/3}} = \frac{\alpha}{i^{8/3}} \tag{65}$$

where α is now to be determined.

Substituting (65) into the integral of (63) we find:

$$\int_0^\infty \frac{\alpha(\frac{Tx}{\Delta t})^{-8/3}}{\Delta t^{8/3}}[1-\cos(\pi x)]dx = \frac{T_\tau^{1/3}}{T} \tag{66}$$

All parameters can now be taken out of the integral:

$$\alpha \int_0^\infty x^{-8/3}[1-\cos(\pi x)]dx = T_\tau^{1/3}T^{5/3} \tag{67}$$

The definite integral has been evaluated numerically and is approximately 14.06. Thus

$$\alpha \approx \frac{T_\tau^{1/3}T^{5/3}}{14.06} \tag{68}$$

yielding our final expression for the variances of the cosine transform coefficients:

$$\sigma_i^2 = E\{U_i^2\} = \alpha i^{-8/3} = \frac{T_\tau^{1/3}T^{5/3}}{14.06} i^{-8/3} \tag{69}$$

The power spectrum $S_{\tau\tau}(f)$ can now be calculated. By convention, the total power contained in a small frequency interval $f, f+\Delta f$ is

$$\Delta P = \int_f^{f+\Delta f} 2S_{\tau\tau}(f')df' \approx 2S_{\tau\tau}(f)\Delta f \tag{70}$$

But this frequency interval corresponds to frequency components of the cosine transform used above (53) running from i to $i+\Delta i$ where:

$$\begin{aligned} i &= 2Tf \\ \Delta i &= 2T\Delta f \end{aligned} \tag{71}$$

in which there is therefore an expected power of:

$$\Delta P = \sum_{j=i}^{i+\Delta i-1} \frac{1}{2}\sigma_j^2 \approx \Delta i \frac{1}{2}\sigma_i^2 = T\Delta f \sigma_{2Tf}^2 \tag{72}$$

Substituting the value we have found for σ_i^2 in (69) we find:

$$\begin{aligned} \Delta P &= T\Delta f \alpha i^{-8/3} \\ &= T\Delta f \frac{T_\tau^{1/3} T^{5/3}}{14.06} (2Tf)^{-8/3} = \Delta f \frac{2^{-8/3}}{14.06} T_\tau^{1/3} f^{-8/3} \end{aligned} \tag{73}$$

Equating (70) with (73) we solve for the spectral density $S_{\tau\tau}(f)$:

$$S_{\tau\tau}(f) = .0056 T_\tau^{1/3} f^{-8/3} \tag{74}$$

The above expression for the spectral density function is included for reference only. Future calculations requiring knowledge of the spectral characteristics of $\tau(t)$, will be based on the cosine transform expansion(53) and our determination of the variances (69) of the cosine transform coefficients.

It should be pointed out that the Fried model [1] for the spatial structure function of τ_0, on which the determination of the temporal structure function of τ was based, are theoretical predictions derived from the Kolmogorov theory of turbulence. While there have been empirical studies producing results consistent with this model (for instance [5]), the validity of the $-5/3$ power law for the structure function of τ_0 (42) is not a given. Some authors have preferred to model the power spectrum as a general power law in which the exponent remains undetermined. Additionally, there is unquestionable decrease in the magnitude of the low-frequency components of τ relative to this simplified

model, due to the finite baseline distance L as discussed earlier, and also due to the existence of an "outer layer of turbulence," the size of which is subject to considerable disagreement. Although results in the following pages are strictly based upon the above model, it should be apparent that the mathematical procedures employed could be generalized to accommodate any characterization of the spectrum of τ.

Point Estimation of τ using Time Bins

In this chapter and the next chapter we shall introduce relatively simple approaches to the $\tau(t)$ estimation problem. In so doing, we introduce the tools necessary for the development of the ultimate solution to be presented in a later chapter. We will, in this journey, present sub-optimum, but nevertheless, reasonable solutions to the estimation problem which are more closely related to conventional approaches than is the **path estimation** procedure, to be presented later.

One simplification present in the following solutions is that the entity we are estimating is $\tau(t_1)$, the differential atmospheric delay **at a given time** t_1, rather than a function, $\tau(t)$, over a **range** of t. We face a far simpler problem in the one dimensional estimation problem of τ at **one** time t_1 (or **"point estimation"**, as opposed to the infinite dimensional problem of estimating **the path of the function** $\tau(t)$ over an **interval** (or **"path estimation"**. The point at which the degrading approximations, (75) and (144), are invoked in order to thus limit the scope of the estimation problem, will be noted.

Now let us initially restrict ourselves to the **narrowband** case. Under the narrowband assumption, all received photons are of approximately the same optical frequency ν. This simplifies the solutions we shall obtain. The application of the point estimation methods to **wideband** (white) light is depicted in the simulation results of Figure 15 and Figure 16.

Simplified Model

A casualty of point-estimation procedures is the lack of detailed incorporation of the *a priori* statistics of $\tau(t)$, that is, properly rewarding solutions which display the smoothness in $\tau(t)$ favored by its power spectrum, which, as we have shown, is dominated by low-frequency components. In the first part of the discussion of point-estimation procedures we will even **further**

ignore the *a priori* characteristics of $\tau(t)$. The **very** simplified characterization of $\tau(t)$ is essentially to assert that over a short time interval its value must be almost constant. In our case "constant" means changing much less than an optical wave period $1/\nu$. Our assumption is then:

$$\tau(t) \approx \tau(t_1) \qquad t_1 - \frac{T}{2} < t < t_1 + \frac{T}{2} \tag{75}$$

which, as we have noted, must be true for a sufficiently small T given that $\tau(t)$ is a smooth function. We shall thus proceed taking (75) as an **equality**.

This approach is explicitly employed in classical treatments of the τ estimation problem. Generally a criterion is first set for the "approximately equals" in (75), and then a maximum limit is arrived at for T, the width of the time "bins," so that (75) is satisfied. This maximum T is then on the order of, but somewhat smaller than T_0, the atmospheric coherence time parameter (52), during which interference "fringes" or "speckles" formed by the interference of starlight, are relatively stationary. In [8], using a time binning procedure equivalent to that described below, there is a proposed solution to the optimization of T taking into account the expected delay errors both due to the approximation of (75), and due to the lack of available photons as T is decreased.

The upshot of the assumption (75) is that photons received during the interval $(t' - T/2) < t < (t' + T/2)$ are employed, with equal weight, to estimate the point $\tau(t')$, whereas photons outside this interval will be considered irrelevant to the estimation of $\tau(t')$. One thus forms N time bins: $0 < t_1 < T$; $T < t_2 < 2T$;... $(N-1)T < t_N < NT$; so that the problem of estimating $\tau(t)$ is reduced to independent estimations of the N points $\tau(t_i)$, each of which is based only on photons received in its particular time bin[10]. We

[10]One could, of course, gain additional time resolution in the estimate of τ by creating **overlapping** time bins at a spacing, Δt, which is smaller than T, the width of the time bins. In the

simplistically ignore any correlation between $\tau(t_i)$ and $\tau(t_j)$ for $i \neq j$. We will assume the *a priori* probability density of $\tau(t_1)$ to be uniform over some indefinite (or, perhaps, infinite) interval.

Following the present discussion we will discover how better results can be obtained by abandoning the time-binning strategy and instead taking into account the *a priori* spectrum of $\tau(t)$ in order to best weight the received photons in order to estimate the point $\tau(t')$. Following that, we will further improve our results by going to **path** estimation procedures, in which the entire function $\tau(t)$ is estimated over an extended **interval** of time.

So let us reiterate the very simplified model to which we have reduced our problem. We wish to estimate the value of $\tau(t)$ at $t = t'$ using K received photons acquired during an interval of width T centered about t', during which $\tau(t)$ does not appreciably change according to (75). All photons are of the same optical frequency ν. Let us assume, for the sake of the following presentation, that the interferometer employs a balanced **quadrature** photomixer feeding noiseless detectors[11]. Also, for the sake of simplicity, let us assume that the "correction delay" τ_c included in the model discussed on page 14, is zero. Each photon, numbered k for $k = 1$ to K is found in a quadrature channel q_k, where $q_k = 0, 1, 2,$ or 3. Assume that we are employing a photomixer characterized by (21). Now, **given** that the underlying $\tau(t)$ around time t' is equal to $\tau^{(\mathrm{ACT})}$, with $\tau_c = 0$, then the field amplitude seen at the photomixer output channel #i at optical frequency ν would be given by the generalization of (29).

---

[10]extreme case, we would compute a "moving average" so that each point in time is based upon a slightly different set of photons, as in [8]. In any case, the results that will be presented are not affected by the choice of Δt.

[11]The choice of a balanced quadrature detector (see page 26) will add to the symmetry of the following calculations. However **any** photomixer configuration will achieve essentially identical results. The only exception is the **two-phase** photomixer which, as we have previously noted, is capable of measuring only the **cosine** and not the sine of phase angle, resulting in an expected error which is dependent on τ itself, and which is incapable of resolving the sign of any phase error.

$$|y_i|^2 = |\frac{1}{2}(x_1 + j^i x_2)|^2$$
$$= \frac{|A|^2}{2}\left(1 + \mathrm{Re}\left\{j^i V e^{j2\pi\nu\tau^{(\mathrm{ACT})}}\right\}\right) \quad (76)$$

Note that we have changed the index of y to run from 0 to 3, rather than 1 to 4 as previously shown. Assume that photon #k was detected at optical frequency ν. Then the conditional probability of having found photon #k in quadrature channel q_k would clearly be given by:

$$P(q_k|\tau^{(\mathrm{ACT})}) = \frac{1}{4}\left(1 + \mathrm{Re}\left\{j^{q_k} V e^{j2\pi\nu\tau^{(\mathrm{ACT})}}\right\}\right) \quad (77)$$

where, again, V is the complex visibility for photons of optical frequency ν. Given $\tau^{(\mathrm{ACT})}$, these events are independent, thus the joint probability of finding K photons in a series of quadrature channels q_1, q_2, $\ldots q_k$ would simply be the product of K factors, each given by (77).

Finding the A Posteriori Density of τ

We wish to find the *a posteriori* probability density for $\tau^{(\mathrm{ACT})}$ under these assumptions given a received data set q_1, q_2, $\ldots q_k$. Bayes' theorem says that the *a posteriori* probability of an underlying random variable x given an observation y is given by:

$$P(x|y) = k_1 P(y|x)P(x) \quad (78)$$

where k_1 is a constant that normalizes the total probability of $P(x|y)$ such that its integral over all x is unity. In other words, the *a posteriori* probability is proportional to the *likelihood* of x given y (given by $P(y|x)$ above) times the *a priori* probability of x. In the case of a **uniform** *a priori* probability for x, then, we conclude that the *a posteriori* probability of x given an

observation is just proportional to the likelihood function[12] $P(y|x)$. Applying this to the estimation of $\tau(t')$ given the q_k from K photons, we find using (78):

$$P(\tau|\vec{q}) = k_1 \prod_{k=1}^{K} \left(1 + \mathrm{Re}\left\{j^{q_k} V e^{j2\pi\nu\tau}\right\}\right) \tag{79}$$

From now on we shall write expressions not for the probability density in (79) but for the **logarithm** of this probability. This will conveniently transform **products** (such as in (79)) into **summations**. The normalization factor k_1 will now become an **additive** term C_1, which, incidentally, we will really never need to evaluate or further consider[13]. Now, the problem of maximizing (in either a global or a local sense) the *a posteriori* probability density is clearly achieved by maximizing its logarithm, knowledge of C_1 being totally unnecessary.

One further ramification of looking at the **logarithm** of probability will be mentioned. Let Λ be a constant plus the logarithm of probability. Then, if Λ is at all smooth, it can naturally be approximated around its maximum point in a **second-order** polynomial. But a second-order polynomial is exactly the logarithm of a normal (Gaussian) density! This procedure, therefore, lends itself to the evaluation of Gaussian solutions. The variances of these Gaussians can be directly obtained from the curvature (second derivatives) of Λ around the maxima.

---

[12]In other words the **Maximum A Posteriori Probability** (MAP) estimate, in the case of a **uniform** *a priori* characterization of the unknown, is identical to the **Maximum Likelihood** estimate.

[13]It should be pointed out that k_1 or C_1 is only a constant **relative to** the quantity being estimated, τ. It will generally **not constant** for different data sets!

The Complex Likelihood Function

Let us introduce the following notation. The **Complex Likelihood** L_k, for photon #k, will be defined as:

$$L_k(\tau') \stackrel{\wedge}{=} j^{q_k} V e^{j2\pi\nu\tau'} \tag{80}$$

L_k is thus a function which embraces the information derived from the reception of photon #k at time t_k with optical frequency ν_k and detected in quadrature channel q_k. Given these parameter, L_k is then a function of a single argument: τ'. When the argument of L_k is not explicitly specified, we will take it to be $\tau(t_k)$, the **independent variable** of **estimated** $\tau(t)$, that is, the estimated τ function evaluated at time t_k, the time that particular photon was received (there will only be a few cases when the argument of L_k needs to be specified differently.) Since, in the present case, we are taking τ to be constant over a time bin, we will just use the estimated τ for time bin #1, or τ_1 to be the argument of L_k for all photons received during the interval of length T centered about $t_1 = T/2$. Thus:

$$L_k = j^{q_k} V e^{j2\pi\nu\tau_1} \qquad (0 \le t_k \le T) \tag{81}$$

ν and V are the same for all photons, in the present case, since we are assuming narrowband reception. We shall continue to refer only to time bin #1 since all time bins are treated identically and independently of one another.

Now let us write the *a posteriori* probability density of $\tau(t_k)$ given only the detection of a **single photon**. As before we assume a uniform *a priori* distribution for $\tau(t_k)$. Then we can rewrite (79) for $K = 1$ (only one photon detected) as:

$$P(\tau|q_k) = k_1 \left(1 + \mathrm{Re}\{L_k\}\right) \tag{82}$$

Note that the physically significant quantity derived from L_k is always given by one plus its real part. This number is always

between 0 and 2, and yields the probability as seen in (82). Since we shall henceforth be dealing solely with the **logarithm** of probability density, we rewrite (82) as:

$$\log P(\tau|q_k) = C_1 + \log\left(1 + \mathrm{Re}\{L_k\}\right) \tag{83}$$

The Linear-Log Approximation

Before proceeding, we shall digress and introduce one important approximation which will be frequently employed throughout the rest of this work. The Linear-Log Approximation (LLA) says that:

$$\log(1 + x) \approx x \tag{84}$$

is a very good approximation under the condition:

$$|x| \ll 1 \tag{85}$$

Even for $x = +1$ it is not a terribly bad approximation, although as x approaches -1 it clearly fails. In the practical case to which we now apply this approximation it will be seen to be reasonably good (for $|V| \approx 1$) to excellent (for $|V| \ll 1$).

Applying the linear-log approximation to (83) results in the simplification:

$$\log P(\tau(t_k)|q_k) = C_1 + \mathrm{Re}\{L_k\} \qquad (LLA) \tag{86}$$

Equations such as (86) in which the linear-log approximation has been applied will **not** be written with an "approximately equals" sign (which will be reserved to indicate **further** levels of approximation) but will be punctuated with **"LLA"** to remind the reader that this approximation has been invoked. Similar equations

in which the LLA approximation has been **avoided** will be labeled with **"EL"** for "exact logarithm."

Let us see to what extend using the LLA approximation in (86) will distort the determination of probabilities. Let us call Λ a constant plus the logarithm of probability density. Then probability density is given by

$$P = k_1 e^{\Lambda} \tag{87}$$

But (for the case of observing a single photon) Λ is given by (83) or (86) depending on whether or not we are employing the LLA approximation. Thus:

$$\Lambda = \begin{cases} \log(1 + |V|\cos(\Phi)) \\ |V|\cos(\Phi) \end{cases} \tag{88}$$

where we have now rewritten L_k as:

$$\begin{aligned} L_k &= j^{q_k} V e^{j2\pi\nu\tau} = |V| e^{j\Phi} \\ \therefore \mathrm{Re}\{L_k\} &= |V|\cos(\Phi) \end{aligned} \tag{89}$$

We have assembled all the phase factors of L_k in the $e^{j\Phi}$ factor, which then multiplies the **magnitude** of V, the "fringe visibility" or correlation being detected by the interferometer. $|V|$ **can** be as high as unity in the case of observation of a totally unresolved object (a virtual point source). In this case the LLA approximation will be at its worst. More typically, a source will be at least sightly resolved. Or there might be a significant level of dark counts in the photon detection system which, equivalently, reduced the **detected** correlation well below unity. As the correlation decreases, then x in (84) is bounded by:

$$-|V| \le x \equiv \mathrm{Re}\{L_k\} \le |V| \tag{90}$$

Now the relative probability densities computed using (87) using the exact and approximate expressions in (88) would be

$$\frac{P}{k_1} = e^{\Lambda} = \begin{cases} e^{\log(1+|V|\cos(\Phi))} = 1 + |V|\cos(\Phi) & (EL) \\ e^{|V|\cos(\Phi)} & (LLA) \end{cases} \tag{91}$$

In Figure 12 this computed probability is plotted for all possible values of x both with and without the LLA approximation. For correlations higher than .8 it can be seen that the LLA approximation can be off by a factor of 2 (when $x = -.8$). Such a discrepancy, however, is less serious than might appear at first sight. That is because the whole point of the estimation procedure is to find solutions which **maximize** the *a posteriori* probability. But the y axis of Figure 12 is exactly a plot of probability, and our solution, if it is any good, would favor points the upper right of the graph (whichever curve you choose), the points at the lower left being strongly avoided. The major difference, therefore, between an exact algorithm and one using the LLA approximation, is that the latter would consider solutions which find $\mathrm{Re}\{L_k\}$ near -1 to be **quite unlikely** whereas the exact algorithm would judge such solutions to be **totally unlikely**. However the reckoning of the probabilities of **likely** solutions, having higher values of $\mathrm{Re}\{L_k\}$, tends to be much less distorted by the LLA approximation.

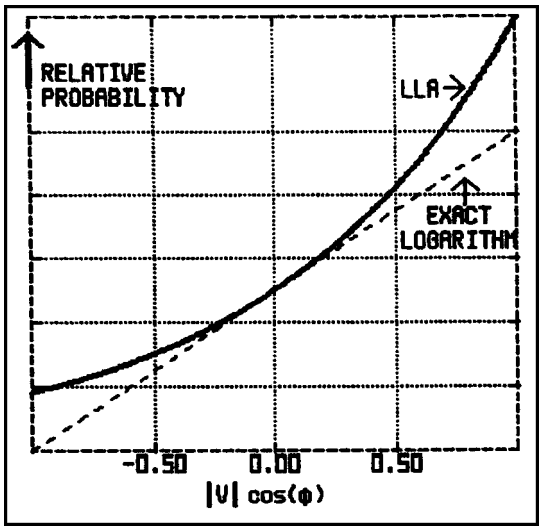

Figure 12: Distortion of probability reckoned using the LLA approximation compared to exact logarithm.

Now it **can** be seen that the LLA approximation consistently **overestimates** the Λ of any solution. In fact there is a very marked difference in the overall figures of merit produced using the two methods, as will be pointed out on page 159. However this **bias** is of no concern since we are computing **relative** not absolute

probabilities, and choosing solutions on the basis of these relative numbers.

The net degradation suffered due to using the LLA approximation will be quantified in a later section. For $|V|$ not approaching unity, this degradation will be found to be negligible. Even if not used, however, the increase in accuracy using the exact logarithm remains of academic interest since the enhancement gained is a specifically quantum effect which can not be accounted for using classical models in which the quantized nature of light is reduced to an equivalent "shot noise" level with an r.m.s. amplitude given by the square root of photon flux.

For the remainder of this chapter we shall apply the LLA approximation as in (86).

The Net Complex Likelihood Function

(86) was derived by applying the LLA approximation (84) to (82), representing the probability due to a single photon. Generalizing this result for K Photons, the **logarithm** of probability is expressed as a **summation** of logarithms approximated using (84).

$$\begin{aligned} \log P(\tau|\vec{q}) &= \log k_2 + \sum_{k=1}^{K} \mathrm{Re}\{L_k\} \qquad (LLA) \\ &= C_2 + \mathrm{Re}\{L_{\Sigma}^{(b)}\} \end{aligned} \tag{92}$$

where we hereby introduce the notation $L_{\Sigma}^{(b)}$ for the **net complex likelihood** for the time bin being considered. The superscript (b) reminds us that this complex likelihood is computed using **binning** rather than a more sophisticated weighting yet to be introduced.

The net complex likelihood is now defined simply as the summation of the photons' complex likelihoods for each photon in the time bin:

$$L_{\Sigma}^{(b)} \triangleq \sum_{k=1}^{K} L_k \tag{93}$$

For the narrowband case, this summation takes on a particularly simple form:

$$L_{\Sigma}^{(b)} = \sum_{k=1}^{K} j^{q_k} V e^{j2\pi\nu\tau} = V e^{j2\pi\nu\tau} \sum_{k=1}^{K} j^{q_k} \tag{94}$$

We are reminded that the summation of K phasors (with respect to τ) of the same frequency ν, yields a phasor of that frequency of some magnitude and phase. Since the real part of $L_{\Sigma}^{(b)}$ gives us the logarithm of *a posteriori* probability density, we can see that the plot of *a posteriori* probability will be periodic with peaks separated by $1/\nu$. Of course we already knew that a very narrowband interferometer is only capable of phase-tracking, with integer wave-period delay ambiguities being unavoidable. Let us suppose that we have overcome that ambiguity (or are willing to ignore it!) and calculate the expected error in phase determination obtained by maximizing the logarithm of *a posteriori* probability in (92), equal to the real part of (94) plus a constant.

The Resulting Estimator

The real part of (94) will clearly be maximized for phases in the exponential which cancel the phase angle of V times the complex summation, so that we would naturally form the estimator:

$$\hat{\Phi} = 2\pi(\nu\hat{\tau} + N) = -(\arg(V) + \arg(Z)) \tag{95}$$

where Z is a statistic derived from the received photon events:

$$Z \stackrel{\wedge}{=} \sum_{k=1}^{K} j^{q_k} \tag{96}$$

Of the K received photons, about half, K_R, will have $q = 0$ or 2 and contribute to the real part of Z. K_I will likewise contribute to the imaginary part.

$$\begin{aligned} Z &= \sum_{k_R=1}^{K_R} (-1)^{q_{k_R}/2} + j \sum_{k_I=1}^{K_I} (-1)^{(q_{k_I}-1)/2} \\ &= \sum_{k_R=1}^{K_R} z'_{k_R} + j \sum_{k_I=1}^{K_I} z''_{k_I} \\ &= Z_R + jZ_I \end{aligned} \tag{97}$$

We have now expressed the effect of each photon accumulated in (96) as either z'_{kr} or z''_{ki} depending on whether it contributes to Z_R or Z_I, the real and imaginary parts of the complex statistics Z. Using (77) we can see the K_R photons counted in the z' summation of (97) are each independently distributed according to:

$$P\{q = 0\} = 1 - P\{q = 2\} = \frac{1}{2}\left(1 + \mathrm{Re}\{Ve^{j\Phi^{(ACT)}}\}\right) \tag{98}$$

where

$$\Phi^{(ACT)} \stackrel{\wedge}{=} 2\pi\nu\tau^{(ACT)} \tag{99}$$

is the actual phase we are attempting to estimate. Note that the probabilities for $q = 0$ and $q = 2$ have been doubled since we are only selecting photons whose q is **even** and have omitted from consideration photons whose q is odd. Then the random variables z' in (97) each take on the value $+1$ or -1 with probabilities given by

(98) (A similar expression applies to the probabilities of the z'' taking on $+1$ or -1). Computing the mean and variance of each z' we find:

$$\begin{aligned} E\{z'\} &= (+1)\cdot P\{q=0\} + (-1)\cdot P\{q=2\} \\ &= \frac{1}{2}\left(1 + \mathrm{Re}\{Ve^{j\Phi^{(ACT)}}\}\right) - \frac{1}{2}\left(1 - \mathrm{Re}\{Ve^{j\Phi^{(ACT)}}\}\right) \\ &= \mathrm{Re}\{Ve^{j\Phi^{(ACT)}}\} \end{aligned} \tag{100}$$

$$\begin{aligned} \mathrm{Var}\{z'\} &= E\{(z')^2\} - (E\{z'\})^2 \\ &= 1 - (\mathrm{Re}\{Ve^{j\Phi^{(ACT)}}\})^2 \end{aligned} \tag{101}$$

Likewise for the z'':

$$E\{z''\} = \mathrm{Re}\{jVe^{j\Phi^{(ACT)}}\} \tag{102}$$

$$\mathrm{Var}\{z''\} = 1 - (\mathrm{Re}\{jVe^{j\Phi^{(ACT)}}\})^2 \tag{103}$$

The means and variances of the real and imaginary parts of Z are given by multiplying by K_R and K_I respectively. Calling:

$$\Phi' \triangleq \Phi^{(ACT)} + \arg(V) \tag{104}$$

then:

$$E\{\mathrm{Re}\{Z\}\} = K_R|V|\cos(\Phi') \tag{105}$$

$$\mathrm{Var}\{\mathrm{Re}\{Z\}\} = K_R(1 - |V|^2\cos^2(\Phi')) \tag{106}$$

$$E\{\mathrm{Im}\{Z\}\} = -K_I|V|\sin(\Phi') \tag{107}$$

$$\mathrm{Var}\{\mathrm{Im}\{Z\}\} = K_I(1 - |V|^2\sin^2(\Phi')) \tag{108}$$

$$\therefore E\{Z\} = \frac{\bar{K}}{2}|V|e^{-j\Phi'} = \frac{\bar{K}}{2}V^* e^{-j\Phi^{(ACT)}} \tag{109}$$

where $\bar{K}$ is the expected value of K, so that K_R and K_I each have an expectation of half this value. Substituting the expected phase angle of Z according to (109) into the estimation equation (95) we indeed find that the estimator works and is unbiased:

$$\begin{aligned} E\{\hat{\Phi}\} &= -E\{\arg(V) + \arg(Z)\} \\ &= -(\arg(V) + \arg(V^*) - \Phi^{(ACT)}) \\ &= \Phi^{(ACT)} \end{aligned} \tag{110}$$

(We have assumed that the expected value of a phase angle is the phase angle of the expected value, always true for complex Gaussians.)

Determination of the Estimator's Performance

For a sufficiently strong signal we can determine the r.m.s. error in the resultant estimation $\Phi^{(ACT)}$ using the following procedure. By "strong signal" we mean that

$$K|V|^2 \gg 1 \tag{111}$$

Under this condition, the approximations invoked in various of the following steps become justified. Since we expect (and shall find!) that the expected estimation error is significantly less than a radian, we can proceed to determine the mean squared estimation error.

$$
\begin{aligned}
E\{(\hat{\Phi} - \Phi^{(ACT)})^2\} &\approx E\{\sin^2(\hat{\Phi} - \Phi^{(ACT)})\} \\
&= E\{(\mathrm{Im}\{e^{j\hat{\Phi}} e^{-j\Phi^{(ACT)}}\})^2\} \\
&= E\{(\mathrm{Im}\{e^{j(\hat{\Phi}+\arg(V))} e^{-j(\Phi^{(ACT)}+\arg(V))}\})^2\}
\end{aligned} \tag{112}
$$

Substituting (94) and (104) into (112) yields

$$
\begin{aligned}
&= E\{(\mathrm{Im}\{e^{j\arg(Z)} e^{-j\Phi'}\})^2\} \\
&= E\{(\mathrm{Im}\{\frac{Z^*}{|Z|}(\cos\Phi' - j\sin\Phi')\})^2\} \\
&= E\{\frac{(-\mathrm{Im}\{Z\}\cos\Phi' - \mathrm{Re}\{Z\}\sin\Phi')^2}{|Z|^2}\}
\end{aligned} \tag{113}
$$

We shall crudely evaluate the expected values of the numerator and denominator separately, and then perform the division. Working on the denominator:

$$
\begin{aligned}
E\{|Z|^2\} &= E\{Z_R^2 + Z_I^2\} \\
&= (E\{Z_R\})^2 + \mathrm{Var}\{Z_R\} + (E\{Z_I\})^2 + \mathrm{Var}\{Z_i\} \\
&= E\{K_R + K_I + K_R^2|V|^2\cos^2\Phi' + K_I^2|V|^2\sin^2\Phi' \\
&\qquad -K_R|V|^2\cos^2\Phi' - K_I|V|^2\sin^2\Phi'\}
\end{aligned} \tag{114}
$$

By (111) the dominant terms can be identified and retained. Loosely applying (119) we then obtain

$$
E\{|Z|^2\} \approx \frac{|V|^2\bar{K}^2}{4} \tag{115}
$$

The numerator can now be expanded as:

$$\begin{aligned} \text{numerator} &= (\cos^2\Phi')Z_I^2 + (\sin^2\Phi')Z_R^2 \\ &\quad +2\sin\Phi'\cos\Phi' Z_R Z_I \end{aligned} \tag{116}$$

Taking the expected value of each term, and expressing the expected value of the square of a random variable as its variance plus the square of its mean, all given by (105)–(108) we find:

$$\begin{aligned} E\{\text{numerator}\} &= E\{(\cos^2\Phi')(K_I(1+(K_I-1)|V|^2\sin^2\Phi')) \\ &\quad +(\sin^2\Phi')(K_R(1+(K_R-1)|V|^2\cos^2\Phi')) \\ &\quad -2\sin\Phi'\cos\Phi' K_I|V|\sin\Phi' K_R|V|\cos\Phi'\} \end{aligned} \tag{117}$$

We have evaluated the expectations over z, but retain the expected value notation around the expression since K_I and K_R remain (slightly) random. Continuing:

$$\begin{aligned} &= E\{(K_I\cos^2\Phi' + K_R\sin^2\Phi') \\ &\quad +\cos^2\Phi'\sin^2\Phi'|V|^2(K_I^2 - K_I + K_R^2 - K_R) \\ &\quad -2\sin^2\Phi'\cos^2\Phi'|V|^2 K_I K_R\} \end{aligned} \tag{118}$$

Next we substitute in the expectations for terms containing K_I and K_R, given by:

$$E\{K_I\} = E\{K_R\} = \frac{\bar{K}}{2} \tag{119}$$

$$E\{K_I^2\} = E\{K_R^2\} = \left(\frac{\bar{K}}{2}\right)^2 + \frac{\bar{K}}{2} \tag{120}$$

$$E\{K_I K_R\} = \left(\frac{\bar{K}}{2}\right)^2 \tag{121}$$

Then we find that the last two major terms in (118) cancel exactly, yielding the expected value for the numerator of (113):

$$E\{\text{numerator}\} = \frac{\bar{K}}{2} \tag{122}$$

Dividing this number by the expected value of the denominator of (113) given by (115) we approximate the mean squared estimation error:

$$E\{(\hat{\Phi} - \Phi^{(ACT)})^2\} \approx \frac{\bar{K}/2}{(|V|^2\bar{K}^2/4)} = \frac{2}{\bar{K}|V|^2} \tag{123}$$

It is only natural to define the **bin** signal-to-noise ratio as:

$$SNR^{(b)} \stackrel{\wedge}{=} \bar{K}^{1/2}|V| \tag{124}$$

If photons are received at an average rate of I_0 (photons per second), then, using time bins of duration T, we would find the signal-to-noise ratio to be:

$$SNR^{(b)} = \sqrt{I_0 T}|V| \tag{125}$$

Now using this definition for signal-to-noise ratio, the r.m.s. estimation error is simply given by:

$$\tilde{\Phi} = \sqrt{E\{(\hat{\Phi} - \Phi^{act})^2\}} \approx \frac{\sqrt{2}}{SNR^{(b)}} \tag{126}$$

or in terms of delay error:

$$\tilde{\tau} = \frac{\tilde{\Phi}}{2\pi\nu} \approx \frac{\sqrt{2}}{2\pi\nu SNR^{(b)}} \tag{127}$$

Note that the condition, (111), required for the validity of the approximation used in the derivation of (126), can be re-written:

$$SNR^{(b)} \gg 1 \tag{128}$$

As can be seen from (126) the signal-to-noise ratio would need to exceed 2.8 just to attain an estimation error of 1/2 radian.

Point Estimation of τ using Optimum Weighting

The estimation of delay at a particular time by considering only photons in a certain time "bin" suffers a certain shortcoming. We wish to raise the signal-to-noise ratio as high as possible in oder to improve delay-tracking performance. Recall that the signal-to-noise ratio in (124) is proportional to the square root of K, the expected number of received photons in a time bin. K and the bin signal-to-noise ratio can be increased by lengthening T, the interval of time in which photons are accumulated into one bin. As earlier discussed on page 66 however, we have required that T be sufficiently small so that $\tau(t)$ does not change appreciably over this time period in order for the assumption (75) on which we have formulated the binning procedure to retain its validity (the derivation of (126) rests on this assumption). This dichotomy between the need to increase T and to limit T illustrates the need to go beyond "binning." We shall now develop a more sophisticated procedure which can optimize the use of all photons which may contain some information concerning $\tau(t')$ without simply allowing an overly wide bin size which, using the previous procedure, would be counter-productive.

To lead into this discussion, let's recall the definition of the "net complex likelihood" function (93) of $\tau(t')$, that is, the function related to the logarithm of *a posteriori* probability of τ evaluated at the particular time t'. Rather than summing over the K photons **in the bin**, we can rewrite (93) as:

$$L_{\sum}(\tau(t')) = \sum_{k=1}^{K} L_K(\tau(t'))w(t_k - t') \tag{129}$$

where the summation is now taken over a much larger set of photons, perhaps all photons received in an observation, but weighted according to w, which, consistent with (93) would be given by:

$$w^{(b)}(t_k - t') = \begin{cases} 1 & |t_k - t'| \le T/2 \\ 0 & |t_k - t'| > T/2 \end{cases} \tag{130}$$

In other words, photons which were assigned to the bin estimating $\tau(t')$ according to (75) (see the discussion on page 66) will have a weight of unity, and photons outside the specified interval will have zero weight. So with the choice of weighting function given by (130), then the giant sum in (129) clearly is identical to the binned sum in (93). The conclusion of the following analysis, however, yields a more natural weighting function which drops off from 1 to 0 in a smooth manner as shown in Figure 13.

Incorporating *A Priori* Statistics of $\tau(t)$

What we wish to do, therefore, is to incorporate knowledge regarding the **dynamics** of atmospheric path delay $\tau(t)$, into the problem of estimating $\tau' = \tau(t')$ (partially) from a photon $\#k$ detected at time $t_k = t' + \Delta t$ for a substantial time difference Δt during which τ is expected to vary at least a significant fraction of a wave period $1/\nu$. As before, the computation of the *a posteriori* probability of τ' given a series q_1, q_2,$\ldots q_K$, requires computation of the likelihood, defined as the *a priori* probability of having received the **set** of q_k given only that the actual τ at time t' takes on the value τ'. Let us proceed to determine this probability. The problem is as follows. **Given** that $\tau(t)$ evaluated at the particular time t' is exactly τ', then what is the **conditional probability** of a photon detected at some **other** time t_k being found in quadrature channel q_k of an interferometer?

Solving for that probability first requires determining the probability density of τ at one time t_k given τ at a **different** time t'. That quantity is dependent upon the structure function D_τ for the stochastic process $\tau(t)$, which we have determined in (50) based upon the widely-accepted Fried model of atmospheric turbulence. Recall the definition of the structure function:

$$D_\tau(\Delta t) \stackrel{\wedge}{=} E\{[\tau(t) - \tau(t + \Delta t)]^2\} \tag{131}$$

τ being a stationary stochastic process, the structure function D_τ is independent of t. The particular form for D_τ in (50) may be substituted into the results obtained in the following analysis. Regardless of the exact D_τ used, we assume that $\tau(t)$ is a Gaussian process. IN that case (131) implies that knowing only the value of $\tau' = t(\tau')$, then if τ_k represents τ **at some different time** $t_k = t' + \Delta T$, then τ_k will have an expected value of τ' and variance given by $D_\tau(\Delta t)$ as can be seen by its definition (131). Thus:

$$P(\tau_k|\tau') = \frac{1}{\sqrt{2\pi D_\tau(\Delta t)}} e^{-\frac{(\tau_k - \tau')^2}{2D_\tau(\Delta t)}} \tag{132}$$

Now we already have an expression for the *a priori* probability of finding a photon in quadrature channel q_k given τ_k:

$$P(q_k|\tau_k) = \frac{1}{4}\left(1 + \mathrm{Re}\{j^{q_k} V e^{j2\pi\nu\tau_k}\}\right) \tag{133}$$

We can combine (133) and (132) to form a joint probability, and then integrate over τ_k to find the marginal probability of q_k, conditional on τ' **only**.

$$P(q_k|\tau') = \int_{-\infty}^{\infty} d\tau_k P(q_k|\tau_k) P(\tau_k|\tau') \tag{134}$$

$$= \int_{-\infty}^{\infty} d\tau_k \frac{\frac{1}{4}\left(1 + \mathrm{Re}\{j^{q_k} V e^{j2\pi\nu\tau_k}\}\right)}{\sqrt{2\pi D_\tau(\Delta t)}} e^{-\frac{(\tau_k - \tau')^2}{2D_\tau(\Delta t)}} \tag{135}$$

Calling:

$$\Delta\tau = \tau_k - \tau' \tag{136}$$

and noting that the integral of a real part is the real part of the integral:

$$= \frac{1}{4}\left(1 + \mathrm{Re}\{\int_{-\infty}^{\infty} d\Delta\tau \frac{j^{q_k}V}{\sqrt{2\pi D_\tau}} e^{-\frac{\Delta\tau^2}{2D_\tau} + j2\pi\nu\Delta\tau} e^{j2\pi\nu\tau'}\}\right) \tag{137}$$

$$= \frac{1}{4}\left(1 + \mathrm{Re}\{\frac{j^{q_k}V}{\sqrt{2\pi D_\tau}} e^{j2\pi\nu\tau'} \int_{-\infty}^{\infty} d\Delta\tau e^{-\frac{\Delta\tau^2}{2D_\tau} + j2\pi\nu\Delta\tau}\}\right) \tag{138}$$

Evaluation of the definite integral yields:

$$= \frac{1}{4}\left(1 + \mathrm{Re}\{\frac{j^{q_k}V}{\sqrt{2\pi D_\tau}} e^{j2\pi\nu\tau'} \left(\sqrt{2\pi D_\tau} e^{-\frac{4\pi^2\nu^2}{4/2D_\tau}}\right)\}\right) \tag{139}$$

$$= \frac{1}{4}\left(1 + \mathrm{Re}\{j^{q_k}V e^{j2\pi\nu\tau'} e^{-2\pi^2\nu^2 D_\tau(\Delta t)}\}\right) \tag{140}$$

If we now, in anticipation, jump ahead of ourselves and set the weighting function to be

$$w(\Delta t) = e^{-2\pi^2\nu^2 D_\tau(\Delta t)} \tag{141}$$

then we can substitute the simplified notation of (141) and (80) into (140) yielding the solution of (134):

$$P(q_k|\tau') = \frac{1}{4}\left(1 + \mathrm{Re}\left\{w(t_k - t')L_k(\tau')\right\}\right) \tag{142}$$

Note that, in this important exception, we are using the complex likelihood L_k evaluated for an argument **other** than $\tau(t_k)$. L_k in

this case has an argument τ', the τ we are estimating for the time τ', separated from t_k by Δt.

What we really need, however, is not the probability of q_k given τ' as given by (142), but rather the **joint** probability of the **set** of q_k's for all k, given τ'. It would be tempting to write this joint probability as the product of individual factors each given by (142) And it is true that, because photon events occur independently, the joint probability of a q_k sequence **given a** $\tau(t)$ **function over all time** is given exactly by the product of the individual probabilities:

$$P(q_1, q_2, \ldots q_k)|\tau(t)(\infty < t < \infty)) = \\ P(q_1|\tau(t)) \cdot P(q_2|\tau(t)) \cdot \ldots \cdot P(q_K|\tau(t)) \tag{143}$$

But **given only** $\tau' = \tau(t')$, then, $\tau(t_1)$ and $\tau(t_2)$ at two times other than t' are still each random variables obeying (132). However $\tau(t_1)$ and $\tau(t_2)$ are not independent of **each other** (especially for small $|t_2 - t_1|$. Therefore q_1 and q_2, which are conditional on $\tau(t_1)$ and $\tau(t_2)$ respectively, are not exactly independent. Thus we can only write **as an approximation**:

$$P(q_1, q_2, \ldots q_k)|\tau') \approx \\ P(q_1|\tau') \cdot P(q_2|\tau') \cdot \ldots \cdot P(q_K|\tau') \tag{144}$$

We shall accept this approximation for the time being, however; any attempt to further quantification of the expression for joint probabilities would, out of necessity, duplicate work to be presented in a subsequent chapter in which **path estimation** is considered. The following results using the above approximation approach the performance limit attainable within the realm of **point estimation**, that is estimating τ at a particular **point** in time t', in terms of photons received at various times **around** t'. Only when we consider **path estimation**, evaluation of the likelihood of a **function** $\tau(t)$ in time, can we employ the exact expression of independence (143).

Nevertheless, (144) is not a bad approximation at all, and will lead to surprisingly close estimates of τ'.

Evaluation of the *A Posteriori* Density of $\tau(t)$

To evaluate the *a posteriori* probability density of τ' based upon the data set q_1, q_2, $\ldots q_K$, we again employ Bayes' theorem using a uniform *a priori* probability density for τ':

$$P(\tau'|\vec{q}) = k_1 P(\vec{q}|\tau')P(\tau') = k_2 P(\vec{q}|\tau') \tag{145}$$

Again we wish to write expressions for the **logarithm** of this probability density. The LLA approximation (84) is again applied in order to simplify the evaluation of the logarithm of (142). In the case of binning we noted that the LLA approximation held well when the magnitude of x, bounded by $|V|$ according to (90) was not close to unity. In the case of (142), application of (84) shows x to be bounded by:

$$-w(\Delta t)|V| \le x \le w(\Delta t)|V| \tag{146}$$

Since the weight function $w(\Delta t)$ varies between 0 and 1, this places us even more toward the (desirable) center of Figure 12, especially for photons separated appreciably in time from t'. Thus even for $|V| = 1$, the LLA approximation will hold well for most terms in the following equation.

Using the approximation of independence (144), the result (142), and applying the LLA approximation (84), then (145) becomes:

$$\begin{aligned} \log P(\tau'|\vec{q}) &= c_1 + \mathrm{Re}\{L_{\Sigma}^{(w)}(\tau')\} \quad (LLA) \\ &= c_1 + \mathrm{Re}\{\sum_{k=1}^{K} w(t_k - t')L_k(\tau')\} \end{aligned} \tag{147}$$

$L_{\Sigma}^{(w)}$ is the net complex likelihood **using the weighting function** $w(\Delta t)$. The maximization of the real part of **that** function, now, maximizes the *a posteriori* probability of $\tau(t')$ according to the above model, yielding the MAP estimate of τ. Note that we have achieved a summation of the generalized form proposed in (129) in which the classical binning procedure corresponds to a weighting function (130) which abruptly fell from 1 to 0 at $|\Delta t| = T/2$. Now we have derived a weighting function (141) which smoothly falls from 1 to 0 in a manner dictated by the dynamic characteristics of $\tau(t)$. Given the previous expression (50) of the structure function for $\tau(t)$, we can write out the weighting function in (141) as:

$$w(\Delta t) = e^{-2\pi^2 \nu^2 T_\tau^{1/3} |\Delta t|^{5/3}} \tag{148}$$

Note that this weighting is barely distinguishable from a Gaussian weighting function, in which case Δt in the exponent would be raised to the power of 2 instead of 5/3. This weighting function is plotted in Figure 13 for the case of the Fried parameter $r_0 = .1$ meter and the effective wind velocity (see page 55) $v_0 = 10$ m/sec. It can be seen that the weighting function has virtually fallen to zero at 10 milliseconds, the "speckle coherence time" parameter $T_0 = r_0/v_0$. Photons outside the central 10 millisecond wide window are seen to have little influence on the estimation of τ' using (147).

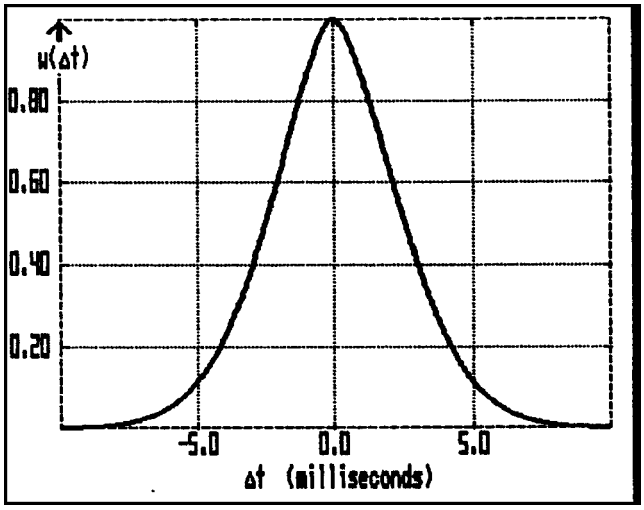

Figure 13: Weighting function for the contribution toward the net complex likelihood function by photons separated from t' by Δt.

In the narrowband case, we can, again, write the MAP solution for the estimate of τ or Φ in closed form:

$$\hat{\Phi}^{(w)} = 2\pi(\nu\hat{\tau}^{(w)} + N) = -(\arg(V) + \arg(Z^{(w)})) \tag{149}$$

where $Z^{(w)}$ is the **weighted** version of Z, which had been defined in (96).

$$Z^{(w)} \stackrel{\wedge}{=} \sum_{k=1}^{K} w(t' - t_k)\, j^{q_k} \tag{150}$$

Performance of the Estimator

It is possible to compute the r.m.s. phase estimation error as we did for the case of binning in (126). The mathematical procedure followed is similar to (112)–(127) but somewhat more involved. Heuristically, however, the result to be presented can be understood as follows.

If we think of each photon as providing an independent "estimate" of τ', then the expected "accuracy" (reciprocal of r.m.s. error) for a photon at time $t_k = t' + \Delta t$ is proportional to $w(\Delta t)$ (photons whose $t_k = t'$ are the most "accurate" for estimating $\tau(t')$ as $w(0) = 1$). Then we take the contribution of each photon and weight it by w to manufacture our estimate of τ'. The accuracy of the net estimate is governed by the familiar result which says that an estimate based upon the optimally-weighted average of independent measurements has an accuracy equal to the accuracy of each individual measurement added in **quadrature**. Thus the photon at time t_k has a relative contribution of $(w(t_k - t'))^2$ toward the **square** of the net "accuracy." In the discussion of error using the binning procedure we found the square of the "accuracy" to be proportional to the number of photons considered, as seen in (123). Retaining this relationship, we would find the contribution of each photon to be the denominator of the mean-squared phase error in (123) to also be proportional to the square of its weight as given by (148).

We are thus, in effect, producing an average base upon an **effective** number of photons given by:

$$K_{eff} = \sum_{k=1}^{K} (w(t_k - t'))^2 \tag{151}$$

where this summation is taken over all possible photons, with only a finite number really contributing to the sum. Given an average photon arrival rate I_0, the expected value of (151) can easily be seen to be:

$$E\{K_{\text{eff}}\} = I_0 \int_{-\infty}^{\infty} d\Delta\tau (w(\Delta t))^2 = I_0 T_{\text{eff}} \tag{152}$$

where we hereby define an effective time interval analogous to (but better than!) the time **bin** interval T in (125). Using this analogy the resultant estimation error takes on exactly the same form as (123).

$$E\{(\hat{\Phi} - \Phi^{act})^2\} \approx \frac{2}{\bar{K}_{\text{eff}} |V|^2} = \frac{2}{I_0 T_{\text{eff}} |V|^2} \tag{153}$$

Given the assumed structure function (50) for τ, and resultant weighting function (148) shown in Figure 13, numerical integration has determined the effective time window to be:

$$T_{\text{eff}} = .37 T_0 \tag{154}$$

for a "speckle coherence time" parameter $T_0 = r_0 / v_0$.

Henceforth, we would like to refer to the signal-to-noise ratio in a way (unlike (124)) which is not dependent upon the estimation algorithm being considered. We thus define the **intrinsic signal-to-noise ratio** as the square root of the number of photons received in the period T_0 (52) times the magnitude of correlation $|V|$.

$$SNR \triangleq \sqrt{I_0 T_0}|V| \tag{155}$$

Applying (154) to (153), and expressing the result in terms of the intrinsic signal-to-noise ratio (155), we find the expected r.m.s. phase or delay estimation error to be given by:

$$\begin{aligned}\tilde{\Phi} &\triangleq \sqrt{E\{(\hat{\Phi} - \Phi^{act})^2\}} \approx \frac{\sqrt{2}}{\sqrt{I_0 T_{\text{eff}}}|V|} \\ &= \frac{\sqrt{2}}{\sqrt{\frac{T_{\text{eff}}}{T_0}}SNR} = \frac{2.3}{SNR}\end{aligned} \tag{156}$$

$$\tilde{\tau} = \frac{\tilde{\Phi}}{2\pi\nu} = \frac{2.3}{2\pi\nu SNR} \tag{157}$$

Simulation Results

Wideband simulations were performed in which the real part of the net complex likelihood (129) using the optimum weighting given by (141), was calculated as a function of time and τ. Note that in the wideband case, the weighting function (141) becomes a function not only of Δt, but also of ν_k, the optical frequency of the detected photon to which it is being applied. Photons were randomly generated over a 2:1 wavelength range, based on a randomly generated $\tau(t)$ function shown in Figure 14. Plots of the real part of the net complex likelihood are shown in Figure 15 for an

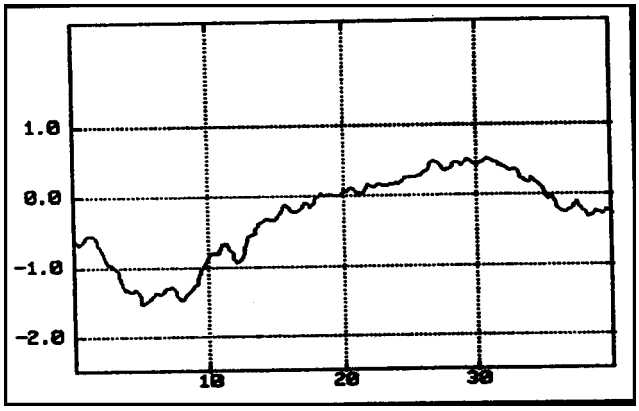

Figure 14: The **actual** τ function estimated in Figure 15 and Figure 16 using optimum weighting, and also in Figure 44 using a more sophisticated algorithm.

intrinsic signal-to-noise ratio of 2.5 (a marginal signal level) and in Figure 16 for a signal-to-noise ratio of 5, corresponding to a four-fold increase in the photon flux. The underlying $\tau(t)$ function is also indicated with a thick line on the top of the surfaces plotted in Figure 15 and Figure 16.

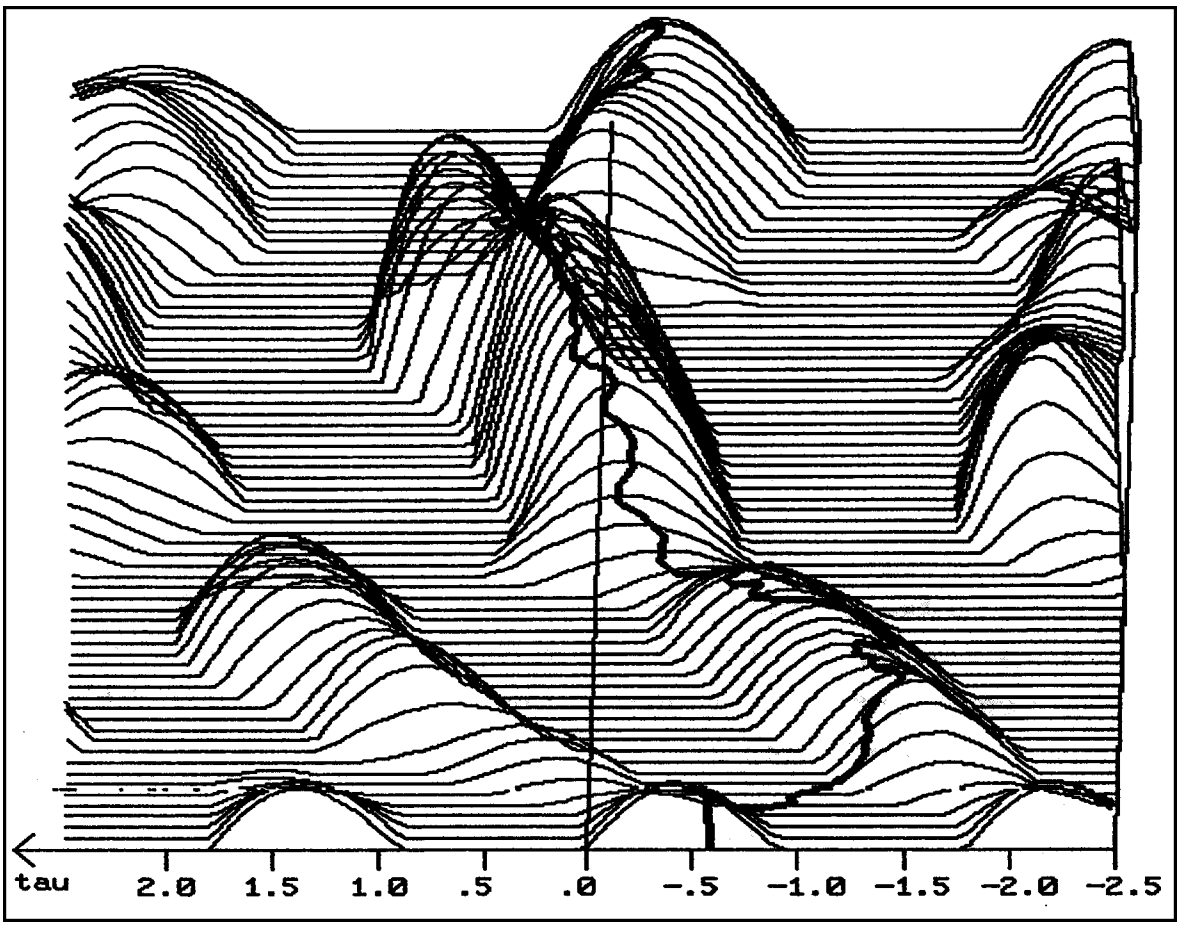

Figure 15: Plot of real part of the net complex likelihood at SNR=2.5, with the underlying τ function shown in Figure 14. Negative values have been clipped at zero, to aid in visualization.

In Figure 16 with the higher signal-to-noise ratio, identification of the dominant peaks as close estimates of τ is evident. However with the marginal signal level in Figure 15, examination of the contour of the complex likelihood alone would hardly be the basis for an estimate of τ at most times. Part of the actual $\tau(t)$ function even crosses an area of the complex likelihood where the real part is **negative** (negative real parts have been plotted as zero, to aid in visualization of the contour). This outcome can be compared to the results of the **global path estimation algorithm**, using the same set of photons, in

Figure 44, in which the multiple path estimate of τ at least includes the actual τ function itself, albeit with a reported probability of only 15%. Compared to the path estimation procedure, maximization of the real part of the complex likelihood using optimum weighting produces acceptable performance at higher signal levels. At the marginal signal level used in Figure 15 and Figure 44, however, the path estimation procedure produces estimates which are both **closer** to $\tau^{(ACT)}$, and which are better quantified in terms of their **confidences**, as plotted numerically in Figure 43.

Computer simulations were run to test the prediction of expected estimation error given by (156). In each case, a random $\tau(t)$ function was generated, and for every point in time the position of the peak of the real part of the net complex likelihood **closest to** $\tau^{(ACT)}$ was ascertained. The resultant errors over a series of time were squared and averaged to find the r.m.s. error in estimating $\tau(t)$, which was multiplied by $2\pi\nu_0$ to determine the r.m.s. **phase** error. ν_0 is the r.m.s. optical frequency over all detected photons, which were simulated as being uniformly distributed over a 2:1 range of optical frequency. The results are plotted versus intrinsic signal-to-noise ratio (155) (also based on the r.m.s. optical frequency ν_0) in Figure 17.

While the simulation data points are seen to roughly follow the direction of the theoretical curve in Figure 17, the error levels can be seen to exceed the theoretical prediction by about 30% at the lower signal-to-nose ratios, to about 60% at the higher signal levels. This discrepancy cannot come as a great surprise. The above analysis of estimation error accepted the validity of the approximation (144) which ignored the interdependence of the received photon events beyond their information concerning τ'. This interdependence becomes more pronounced at higher signal-to-noise ratios. The validity of the approximation $\sin(x) \approx x$, originally employed in (112) and propagated into the present analysis, deteriorates for low signal-to-noise ratios not yielding estimation errors much smaller than 1 radian. Given these limitations, however, the performance of delay-tracking using

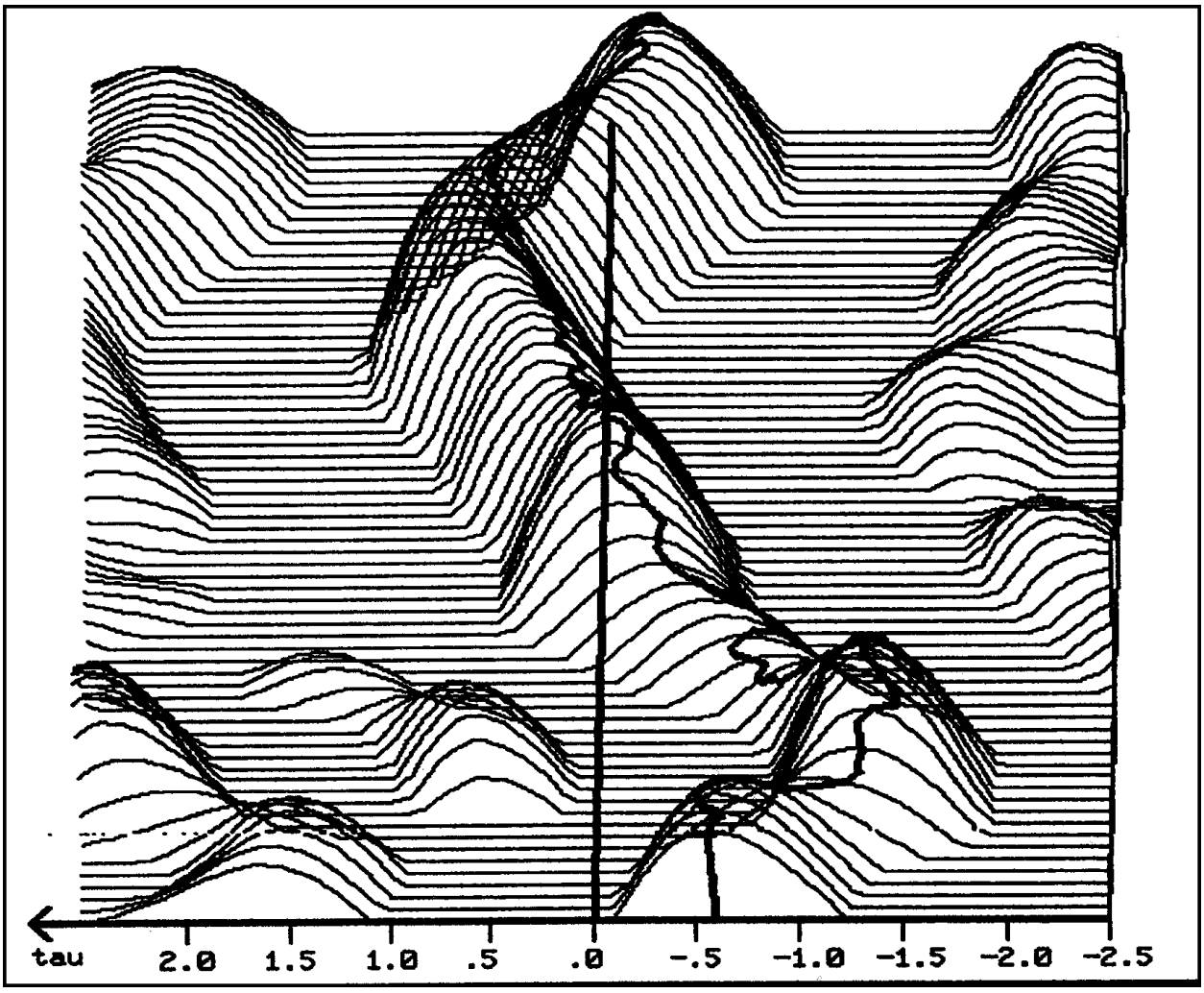

Figure 16: Plot of real part of the net complex likelihood at SNR=5.0. The **actual** τ function is plotted in a thick line on the surface of the contour.

conventional point estimation approaches has been quantified with reasonable accuracy.

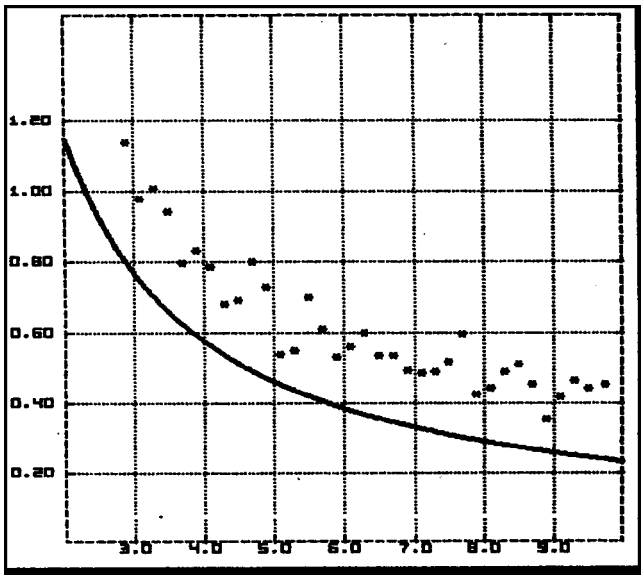

Figure 17: Simulation results of r.m.s. estimation error (in radians) vs. SNR, using the optimum weighting estimator. Theoretical value given by (156) graphed for comparison.

Theory of Path Estimation

The previous two chapters considered estimation of τ at a particular time t', given a collection of received photons. The estimate of $\tau(t')$ was the value which maximized the *a posteriori* probability density given by the result of Bayes' theorem. In each of the previous two chapters, the determination of the *a posteriori* probability density was achieved with the aid of **approximations**, (75) and (144). These approximations compromised the r.m.s. error of the resultant estimates. In both cases, we calculated the estimator's expected error on the basis of assumptions whose validity was tenuous at best.

More importantly, there was no concerted effort to produce estimates of τ at slightly different times which were necessarily well connected. In other words, there will be occasions (especially using wideband reception) when the τ value which maximizes (147) at different times causes a **discontinuity** in the net τ function. So even through **each point** of the reconstructed function **maximizes** the *a posteriori* probability of τ **for that one point**, the probability of the **net** τ function is near zero, due to the total improbability of a jump in the value of τ at one point in time.

We now turn from **point estimation** to **path estimation**. As opposed to the point estimation procedure, path estimation relies on **no approximations** in the determination of the *a posteriori* probability density (except for the previously discussed LLA approximation, which may be applied **optionally**). Thus we will be able to express the estimate as the $\tau(t)$ function which maximizes the r.m.s. error of that estimate by observing the **variance** of the *a posteriori* density. While it is unrealistically optimistic to imagine that, in a practical system, all parameters and elements of the model would be accurately known, we will nevertheless present a self-consistent analysis in which accurate prior knowledge of the model is **assumed** in order to find the attainable performance limits of the estimator. In the present work, we will not address the

sensitivity of the estimator to model parameter discrepancies, or the issue of model parameter re-estimation.

In the case of point estimation, we found the relative *a posteriori* probability for τ', and found the τ' for which that function peaks. In other words we consider the function over all possible values of τ'. Recall that τ' represents the differential path delay function τ, evaluated at the time t'. Now in the case of **path estimation** we consider not simply all possible values of $\tau(t')$, but rather **all possible paths** (functions) of $\tau(t)$ over an **interval of time**, say $0 \leq t \leq T$. This infinite dimensional problem can be somewhat simplified by considering the path of τ over that interval to be piece-wise approximated by specifying its values at $M + 1$ equally spaced points between $t = 0$ and $t = T$ inclusive. Thus we could picture a function of $M + 1$ variables, $\tau(0)$, $\tau(T/M)$, $\tau(2T/M)$, $\ldots \tau(T)$.

Over this $M + 1$ dimensional space, the *a posteriori* probability consists of various peaks and valleys. The function around each peak can be (approximately) described as a Gaussian density with an expected value given by the point of the peak and a covariance structure specified by an $M + 1$ square matrix. The τ functions corresponding to these peaks represent distinct paths obtained according to the MAP (Maximum *A Posteriori* Probability) estimation procedure. Under conditions of sufficient signal strength, there will be a dominant peak near which the underlying τ function almost surely lies. Under conditions of much lower signal strength, the bulk of probability mass can only be accounted for by considering the union of several such peaks, thus revealing an ambiguous result. For still lower signal levels, the probability mass is so dispersed throughout the $M + 1$ dimensional space that accounting for it is pointless; in this regime the path estimation procedure is inapplicable.

There are two facets to the path estimation problem. Consider the strong signal case in which the *a posteriori* probability density is dominated by a single Gaussian function. Given an initial estimate of the path $\tau(t)$, expressed as a point in $M + 1$

dimensional space, which is in the tails of that dominant peak, then, **by what procedure** is the peak found which identifies the estimated path $\tau^{(MAP)}(t)$ which maximizes the *a posteriori* probability density? And when that point is found, what is the expected r.m.s. distance from the point of the peak to the **actual** τ function being estimated? That expected distance determines the r.m.s. estimation error. The discussion related to these questions will henceforth be refereed to as the **"local estimation problem,"** since it concerns the behavior of the *a posteriori* probability function in the **locality** of a dominant peak.

In contrast, the second aspect of the path estimation problem is termed the "**global** estimation problem." This study, to be presented in later chapters, considers the **totality** of the *a posteriori* probability density function. Now we will ask how to find the dominant peak **without** being supplied an initial estimate in the tails of that peak (information to which we would normally not be privy!). And, especially in the case of lower signal levels, how can we exhaustively search the $M + 1$ dimensional space for probability mass in order to account for all likely paths? Furthermore, we would like to know at what signal level the transition from a single dominant peak to multiple peaks is expected to occur. It should be noted that, unlike the local problem, we are no longer dealing with a mathematically tractable problem as the probability density can no longer be described as a Gaussian. Thus the answers to most of these questions are based on the results of computer simulations.

We shall now set up the ingredients for the determination of the *a posteriori* probability density of $\tau(t)$ given an observation of K received photons over a time period of duration T. In the case of path estimation, we will be taking into detailed account the *a priori* characterization of $\tau(t)$ as well as the probablistic response of the instrument. Let us recall the expression of *a posteriori* probability given by Bayes' Theorem. If y is an observation and x is an underlying process that we wish to estimate, then Bayes' Theorem says that the *a posteriori* probability for the process x given the observation y is:

$$P(x|y) = k_1 P(y|x) P(x) \tag{158}$$

$P(y|k)$ is the response function of the instrument, and $P(x)$ is the *a priori* probability for the process x. k_1 is a normalizing constant chosen to insure that the integral of $P(x|y)$ over all x is unity. For simply **comparing** *a posteriori* probabilities, k_1 need never be evaluated. We will henceforth be evaluating not probability density but the *logarithm* of probability density, which we shall denote as Λ.

$$\Lambda \stackrel{\wedge}{=} C_1 + \log P(\tau(t)|\text{receiveddata}) \tag{159}$$

C_1 is an additive constant which, again, may be set arbitrarily without effecting the maximization procedure. It should be noted that C_1 is a constant as we consider different $\tau(t)$ functions; if convenient, it is allowed to be a function of the received data set, since the maximization of Λ only occurs **after** the data has been received at which time C_1 would no longer be treated as variable.

To evaluate Λ using Bayes' Theorem, we need expressions for the *a priori* probability for an arbitrary $\tau(t)$ function, and the conditional probability of having received a particular data set **given** a sample function $\tau(t)$. The *a priori* characterization of $\tau(t)$ is best expressed in terms of its **Fourier** coefficients using the previously defined cosine transform (53). The Fourier coefficients of Gaussian colored noise consist of independent zero-mean Gaussian random variables. Let us truncate the Fourier cosine series to M terms, where M may be arbitrarily large. Then we can write the probability density of $\tau(t)$ in U space as:

$$P_\tau = k_2 e^{-\frac{1}{2}\sum_{i=0}^{M-1} \frac{U_i^2}{\sigma_i^2}} \tag{160}$$

We have previously evaluated the variance σ_i^2 of each Fourier component U_i given the assumed model of atmospheric turbulence in

(69). That model will eventually be inserted into the expressions containing σ_i^2. We can now write the logarithm of this probability density as:

$$\log P_\tau = C_2 - \frac{1}{2}\chi^2 \tag{161}$$

where:

$$\chi^2 \stackrel{\Delta}{=} \sum_{i=0}^{M-1} \frac{U_i^2}{\sigma_i^2} \tag{162}$$

is a random variable whose **a priori** probability density is χ^2 with M degrees of freedom (assuming the accuracy of our model).

Now to determine the instrumental response component to insert into Bayes' Theorem, we will refer back to the model describing the path delays in an interferometer shown in Figure 1. We would first like to determine the conditional probability of detecting a photon in photomixer output channel y_i at time t_k **given** an assumed atmospheric delay $\tau(t_k)$ for time t_k. We can find the field amplitude at y_i using the delay model and the model for a balanced photomixer given by (16).

$$\begin{aligned} y_i &= \sqrt{p_i}x_1 + \sqrt{p_i}e^{-j\theta_i}x_2 \\ &= \sqrt{p_i}\left(e^{-j2\pi\nu\tau_c}A_1 + e^{-j\theta_i}e^{-j2\pi\nu\tau}A_2\right) \end{aligned} \tag{163}$$

We will assume that the input fields are each of the same power. The expected optical power in y_i is given by the squared magnitude of the field amplitude.

$$\begin{aligned} \langle|y_i|^2\rangle &= p_i\left(|A_1|^2 + |A_2|^2 + 2\,\mathrm{Re}\left\{e^{j\theta_i}e^{j2\pi\nu(\tau-\tau_c)}A_1A_2^*\right\}\right) \\ &= 2p_i|A|^2\left(1 + \mathrm{Re}\left\{Ve^{j\theta_i}e^{j2\pi\nu(\tau-\tau_c)}\right\}\right) \end{aligned} \tag{164}$$

In a very short time interval, the probability of detecting a photon is proportional to the optical power. Thus in some short time interval around t', the probability of detecting a photon in photomixer channel i, at optical frequency ν, **given** the atmospheric delay function $\tau(t)$ which **may** be partially compensated by the "correction delay" τ_c, is:

$$\begin{aligned} P(D_i|\tau(t)) &= k_2\left(1 + \mathrm{Re}\left\{V e^{j\theta_i} e^{j2\pi\nu(\tau-\tau_c)}\right\}\right) \\ &= k_2\left(1 + \mathrm{Re}\{L_k\}\right) \end{aligned} \tag{165}$$

$P(D_i|\tau)$ in (165) denotes the probability of detecting a photon in a photomixer output channel of mixing phase θ_i, given the knowledge of $\tau(t)$ at that time. We will not need to specify the "short time interval" or evaluate k_2. We have again consolidated the combined phases and correlation magnitude in L_k, the **complex likelihood**, as we did in the case of point estimation. Given a photon, labelled $\#k$, detected at time t_k, in a photomixer output channel of mixing phase θ_k, at optical frequency ν_k, we can write L_k as:

$$L_k \stackrel{\wedge}{=} V(\nu_k) e^{j\theta_k} e^{j2\pi\nu_k(\tau(t_k)-\tau_c(t_k))} \tag{166}$$

Since we will be dealing with the **logarithm** of probability, we now define the logarithm of the likelihood function **for photon** $\#k$ as λ_k.

$$\lambda_k \stackrel{\wedge}{=} \log\left(\frac{1}{k_2} P(D_i|\tau(t))\right) = \log(1 + \mathrm{Re}\{L_k\}) \tag{167}$$

We **may** choose to simplify the evaluation of λ_k by using the Linear-Log Approximation (LLA) presented on page 71.

$$\lambda_k = \begin{cases} \mathrm{Re}\{L_k\} & (LLA) \\ \log(1 + \mathrm{Re}\{L_k\}) & (EL) \end{cases} \tag{168}$$

What we need to insert into Bayes' Theorem, is the joint probability of detecting the set of K photons received during the time interval $0 < t < T$. Given a $\tau(t)$ function over this time interval, the joint probability of these K photons being found in the photomixer channels $i_1, i_2, \ldots i_K$ are given by the **product** of the probabilities of each photon being detected, since given an optical power level, the occurrence of photons are independent.

$$\begin{aligned} &P(i_1, i_2, \ldots i_k | \tau(t)(0 < t < T)) = \\ &\quad P(i_1|\tau(t)) \cdot P(i_2|\tau(t)) \cdot \ldots \cdot P(i_K|\tau(t)) \end{aligned} \tag{169}$$

Note that this is the assumption that **we were not able to use** in the derivation of point estimation on page 87. Let us denote the logarithm of this joint conditional probability, plus a constant, as λ (**without** a subscript).

$$\begin{aligned} \lambda &\stackrel{\wedge}{=} \log P(i_1, i_2, \ldots i_K|\tau(t)) + C_2 \\ &= \log \prod_{k=1}^{K} P(i_k|\tau(t)) + C_2 = \sum_{k=1}^{K} \lambda_k \end{aligned} \tag{170}$$

Now, the logarithm of *a posteriori* probability given by Λ in (159) is evaluated using Bayes' Theorem (158). Then we can substitute in the value we have obtained for the *a priori* probability of $\tau(t)$ when expressed as a Fourier cosine series given by (161) and (162). We can also substitute the determination of the instrumental response function, the logarithm of which is expressed by λ in (170). We obtain:

$$\begin{aligned} \Lambda &= C_2 + \log P(i_1, i_2, \ldots i_K|\tau(t))P(\tau(t)) \\ &= \lambda - \frac{1}{2}\chi^2 \end{aligned} \tag{171}$$

Now, the solution to the path estimation problem amounts to finding one or more maxima of (171) in order to account for most of the probability mass given by e^{Λ}. The components of (171), again, are contained in (162) and (170); the relationship between the time

domain representation of $\tau(t)$ (used in (170)) and the frequency domain representation (used in (162)) is given in (53).

MAP Path Estimation of τ

Consider the set of all possible $\tau(t)$ functions over $0 < t < T$. Suppose we seek to find the **Maximum A Posteriori Probability** (MAP) estimate of τ, defined as the $\tau(t)$ which maximizes Λ. We start from an initial estimate $\tau^{(1)}(t)$. We might suppose that $\tau^{(1)}$ is in the tails of the **global maximum** of Λ. However the following discussion would equally apply to estimation of a local maximum of Λ given an initial guess in the tails of that local maximum. Call $\tau^{(\mathrm{MAP})}(t)$ the position of the maximum of Λ nearest the starting point $\tau^{(1)}(t)$. Let the Fourier cosine series of $\tau^{(MAP)}(t)$ be denoted[14] $U_i^{(MAP)}$. We shall consider the set of band-limited $\tau(t)$ functions whose Fourier series are truncated to M terms. M may be arbitrarily large.

Around $U^{(MAP)}$, Λ can be expanded to second order as:

$$\Lambda(\vec{U}) = \Lambda^{(MAP)} - \frac{1}{2}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} A_{ij}(U_i - U_i^{(MAP)})(U_j - U_j^{(MAP)}) \qquad (172)$$

where $A_{ij} = A_{ji}$. There are no first order terms in $(U_i - U_i^{(MAP)})$ since we are located at the maximum of Λ.

To put this equation in perspective, recall that Λ is by definition a constant plus the logarithm of the *a posteriori* probability density of U given the received photon detections. Therefore the probability density of U around $U^{(MAP)}$ would be given by:

$$P(\vec{U}) = ke^{-\frac{1}{2}(\vec{U}-\vec{U}^{(MAP)})^T A(\vec{U}-\vec{U}^{(MAP)})} \qquad (173)$$

where we have assembled the coefficients A_{ij} into a matrix A. The reader will recognize (173) as the Gaussian probability density whose covariance matrix is the inverse of A.

[14]Note that we are using **superscripts** to denote different sample functions, as the subscripts refer to different frequency components of U.

We wish to find the $U^{(MAP)}$ in (172) starting from an initial guess $\tau^{(1)}(t)$ whose cosine transform we shall call $U^{(1)}$. If $U^{(1)}$ is sufficiently close to $U^{(MAP)}$ then the second order expansion (172) around $U^{(MAP)}$) will still describe Λ around $U^{(1)}$. Not initially knowing $U^{(MAP)}$, we would find the expansion of Λ around $U^{(1)}$ to be:

$$\Lambda(\vec{U}) = \Lambda^{(1)} - \sum_{i=0}^{M-1} B_i(U_i - U_i^{(1)}) - \frac{1}{2}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} A_{ij}(U_i - U_i^{(1)})(U_j - U_j^{(1)}) \quad (174)$$

The A_{ij} in (174) and (172) are, under this assumption, the same. The B_i can be expressed in terms of $U^{(MAP)}$ and $U^{(1)}$ by equating the two expressions for Λ, and then equating the coefficients of first order terms in U_i, resulting in:

$$\sum_{j=0}^{M-1} A_{ij}U_j^{(MAP)} = -B_i + \sum_{j=0}^{M-1} A_{ij}U_j^{(1)} \quad (175)$$

Calling ΔU the vector from the initial estimate $U^{(1)}$ to the MAP estimate $U^{(MAP)}$:

$$\Delta\vec{U} \triangleq \vec{U}^{(MAP)} - \vec{U}^{(1)} \quad (176)$$

then (175) becomes

$$\sum_{j=0}^{M-1} A_{ij}\Delta U_j = -B_i \quad (177)$$

The set of M equations in M unknowns can be written as the matrix equation:

$$A\Delta\vec{U} = -\vec{B} \quad (178)$$

which has the standard solution for ΔU given by Cramer's rule, or equivalently, by multiplying by the inverse of the matrix A.

Determination of the Matrix Elements

To find the B_i and A_{ij} at a point $U^{(1)}$ we need only evaluate the first and second partial derivatives of (174) at $U^{(1)}$:

$$\frac{\delta\Lambda}{\delta U_{i\,|\vec{U}^{(1)}}} = -B_i \tag{179}$$

$$\frac{\delta^2\Lambda}{\delta U_i \delta U_j} = -A_{ij} \tag{180}$$

We can find these partial derivatives of Λ given by (171) in terms of the received photon events as expressed in λ by (170), (168) and (166), and the *a priori* model for U expressed in χ^2 given by (162). For now, we will determine the results using the LLA approximation (see page 71). Evaluating the partial derivatives of (171) at $U^{(1)}$, we therefore find:

$$\begin{aligned}\frac{\delta\Lambda}{\delta U_i} = \quad & -\frac{U_i^{(1)}}{\sigma_i^2} + \sum_{k=1}^{K} 2\pi\nu_k \cos(\frac{\pi}{T} i t_k) \cdot \\ & Re\{jVe^{j\theta_k} e^{j2\pi\nu_k(\tau^{(1)}(t_k)-\tau_c(t_k))}\} \quad (LLA)\end{aligned} \tag{181}$$

$$\begin{aligned}\frac{\delta^2\Lambda}{\delta U_i^2} = \quad & -\frac{1}{\sigma_i^2} - \sum_{k=1}^{K} (2\pi\nu_k)^2 \cos^2(\frac{\pi}{T} i t_k) \cdot \\ & Re\{Ve^{\theta_k} e^{j2\pi\nu_k(\tau^{(1)}-\tau_c)}\} \quad (LLA)\end{aligned} \tag{182}$$

$$\begin{aligned}\frac{\delta^2\Lambda}{\delta U_i \delta U_j} = \quad & -\sum_{k=1}^{K} (2\pi\nu_k)^2 \cos(\frac{\pi}{T} i t_k) \cos(\frac{\pi}{T} j t_k) \cdot \\ & Re\{Ve^{j\theta_k} e^{j2\pi\nu_k(\tau^{(1)}-\tau_c)}\} \quad (LLA)\end{aligned} \tag{183}$$

By evaluating (181) (182) and (183), inverting the matrix A, and solving (178), we can obtain ΔU and thus $U^{(MAP)}$.

Practically speaking, there are a few drawbacks to this procedure. (181), (182) and (183) require tallying a large number of coefficients (about $M^2/2$). Then we need to invert a large $(M \times M)$ matrix to solve for ΔU. Finally, the solution $U^{(1)} + \Delta U$ will not reach $U^{(MAP)}$ **exactly** since we have ignored third and higher order terms in the expansion (172). Thus we would need to perform more than one iteration of this lengthy procedure in order for it to converge onto $U^{(MAP)}$.

A Computationally Practical Solution

The practical solution we have adopted is to perform several iterations involving a much simpler calculation. We shall take advantage of the fact that the matrix A in (178) is dominated by its diagonal terms. If we arbitrarily set all off-diagonal terms to zero:

$$A_{ij} = 0 \quad (i \neq j) \tag{184}$$

then we could easily solve for an **approximate** ΔU:

$$\Delta U_i \Leftarrow \frac{-B_i}{A_{ii}} = \frac{\frac{\delta \Lambda}{\delta U_i}}{-\frac{\delta^2 \Lambda}{\delta^2 U_i^2}} \tag{185}$$

Although we have eliminated most of the computation involved in solving (178) directly, we have nevertheless produced a procedure which will, after 5 or 10 iterations, converge to $U^{(MAP)}$. In practice, it has been found we obtain convergence at the same rate if instead of (185) we use

$$\Delta U_i \Leftarrow \frac{\frac{\delta \Lambda}{\delta U_i}}{\hat{A}_i} \tag{186}$$

where

$$\hat{A}_i \triangleq E\left\{-\frac{\delta^2\Lambda}{\delta^2 U_i^2}\right\} \tag{187}$$

is a pre-tabulated function of i. This further reduces the burden of computation.

Consider the iteration of U towards $U^{(MAP)}$ using (186). If the denominator of (186) were a constant, then this would simply amount to adding to U at every iteration a factor times the gradient of Λ with respect to U. To guarantee the quickest convergence, however, we prefer to scale the elements of the gradient non-uniformly according to the reciprocal of (187).

Continuously following the gradient of a function will always lead to a maximum of that function. IN this case we are taking discrete steps. The amplitude of each step is scaled according to (185) or (186) in order to, as nearly as possible, approach $U^{(MAP)}$ in that single step. That will not occur, as we have noted, due to mixed terms in (174) and higher order terms that have been omitted from (174). Eventual convergence is guaranteed, however, as long as we do not overshoot the correct solution by an amount greater than the initial value ΔU.

Such an overshoot could occur if the denominator of (186) underestimates the actual value of A_{ii} by a factor of 2 or more. While an uncommon occurrence, a practical algorithm must check for non-converging behavior and reduce the scaling factor as necessary.[15]

[15]A simple measure of over/undershooting is the inner product of the ΔU calculated in successive iterations:

$$[\Delta\vec{U}^{(N+1)}]^T[\Delta\vec{U}^{(N)}] = \sum_{i=0}^{M-1} \Delta U_i^{(N+1)} \Delta U_i^{(N)}$$

A negative quantity indicates overshooting, while a positive measure will occur when the step size has been reduced below its optimum level.

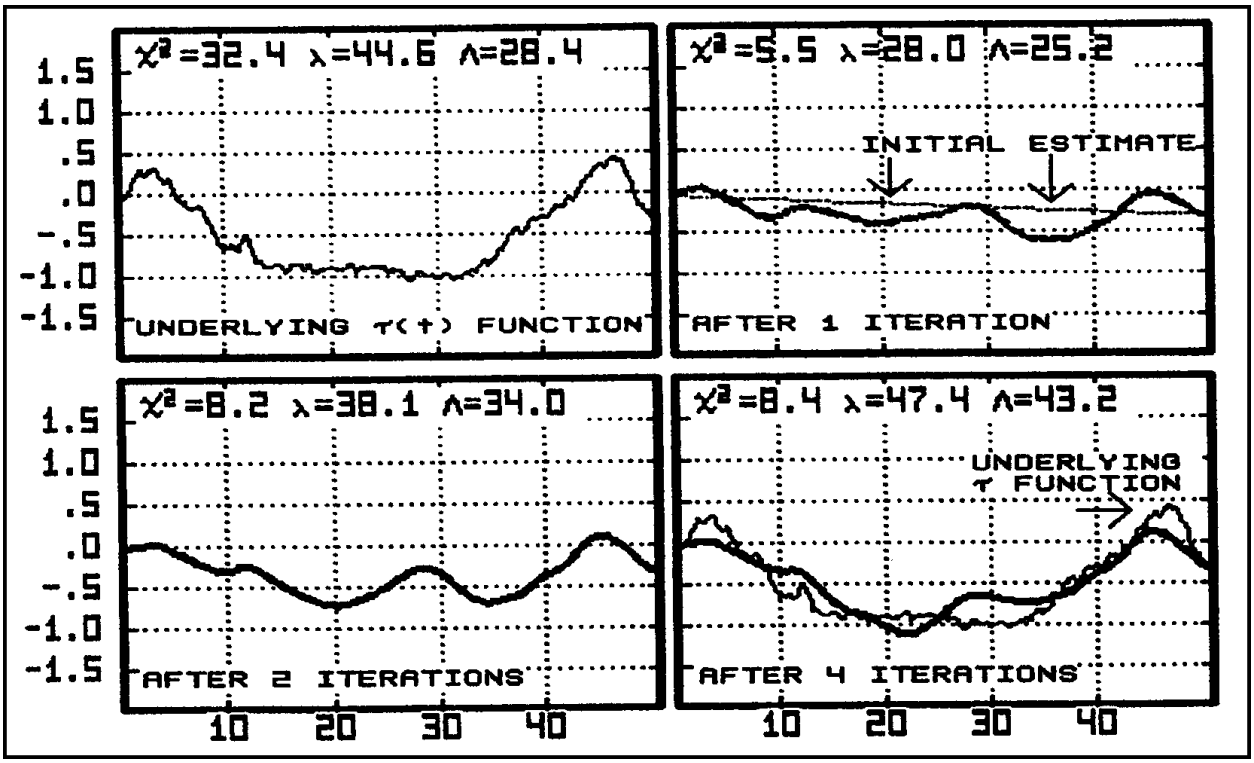

Figure 18: τ (in femtoseconds) vs. time (in milliseconds) of underlying function, and progression of iterative estimation procedure.

Simulation Results

An example of performing such a maximization procedure is shown in Figure 18. The underlying τ function, shown in the upper left, is typical for the atmosphere under good "seeing" conditions. The y-axis is τ in femtoseconds; the x-axis is time in milliseconds. The initial estimate, as shown in the upper right, is constructed as a line joining the endpoints of the underlying τ (which, in a real situation, would, of course, not be initially known). During the 50 milliseconds, 100 photons are received over

[15]A simpler test that has proved satisfactory operates on the τ function itself. The values of the starting function $\tau(t)^{(N)}$ at a number of points on a coarse scale, are stored initially. Following one iteration of the estimation procedure, yielding $\tau(t)^{(N+1)}$, the absolute deviations from the initial stored function are found and the maximum absolute deviation is noted. Following the **second** iteration yielding $\tau(t)^{(N+2)}$, the maximum absolute deviation from the **original** stored function is, again, found. If the **second** maximum absolute deviation from the initial stored function is found to be **smaller** than the determination of that number following the **first** iteration, that is deemed evidence of overshooting, and the gain will be reduced.

a 2 : 1 wavelength range. After the fourth iteration, shown in the lower right, Λ has essentially reached its peak value of 43.2.[16]

A few features of the resulting solution in the lower right of Figure 18 will be noted. The higher frequency components of the underlying $\tau(t)$ are clearly absent from the solution $\tau^{(MAP)}$. That is because the χ^2 penalty term in (171) is preferentially sensitive to high frequency components in τ according to (69) and the definition of χ^2 (162). The smoothing over of high frequency components contributes a significant but tolerable component of estimation error. In Figure 19 the MAP solution is similarly calculated using 50, 200 and 500 photons. The increased recovery of high frequency information with increased signal level is both apparent on visual inspection, and from the value of χ^2 printed at the top of each graph.

Some of the estimation error observed, however, is clearly not due to smoothing but simply results from statistical fluctuations (noise) involved in reconstructing τ from a finite number of randomly produced photons. In the top left and top right of Figure 19 **different** samples of 50 randomly produced photons are seen to generate dissimilar errors in the resultant MAP estimates of τ.

Finally, we must note that the success of the convergence to the global $\tau^{(MAP)}$ in Figure 18 was dependent upon the initial estimate $U^{(1)}$ having its probability density dominated by the tails of the global $U^{(MAP)}$ (even though it was only 10^{-16} of the peak probability density at the point of the initial estimate!) and not some other local maximum of *a posteriori* probability. In Figure 20, using the same set of 100 photons as in Figure 18, we

[16]The data in this example, although randomly generated, has been **selected** for presentation. **Not all** random τ functions and random sets of photons generated with these statistics, when initially estimated with a straight line in this manner, would converge to the global MAP solution. Refer to the discussion on page 100. Also, the algorithm used in these examples is virtually equivalent to that discussed in the text, but utilizes a modified **sine** transform in which the endpoints of the τ function, that is $\tau(0)$ and $\tau(50)$, are constrained.

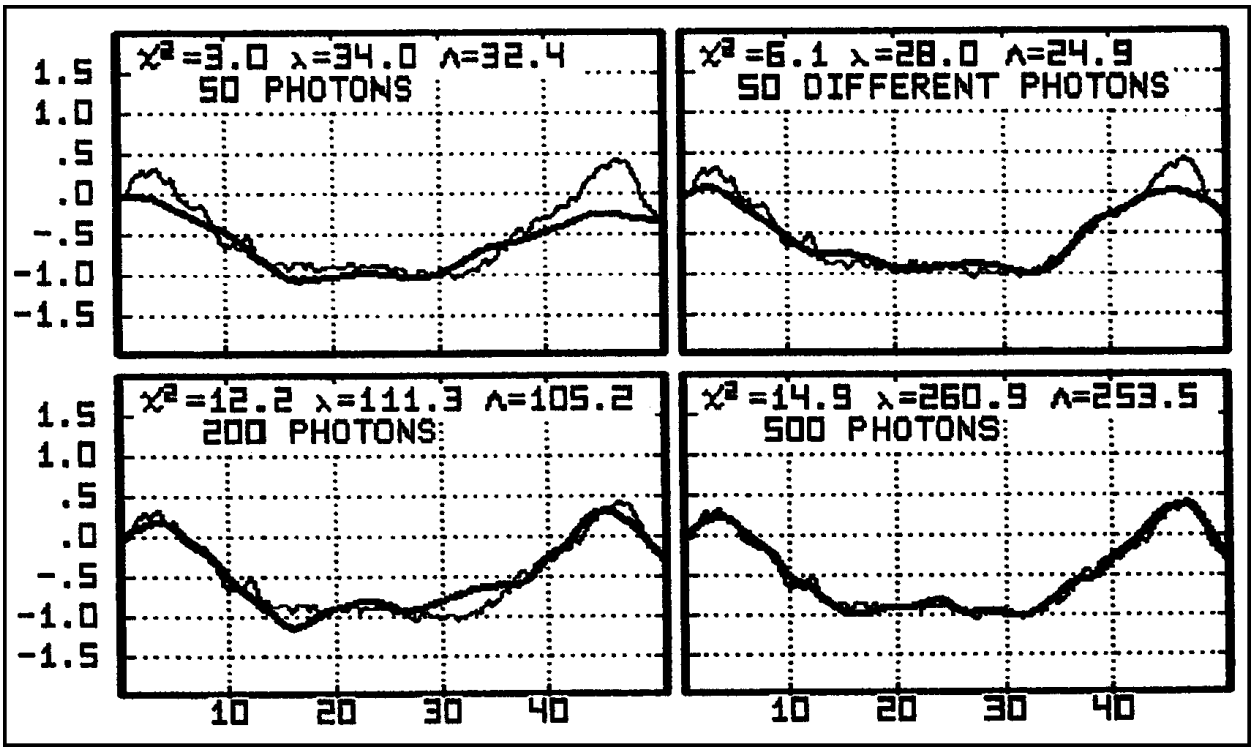

Figure 19: Results of estimation of $\tau(t)$ as in Figure 18, using different signal levels.

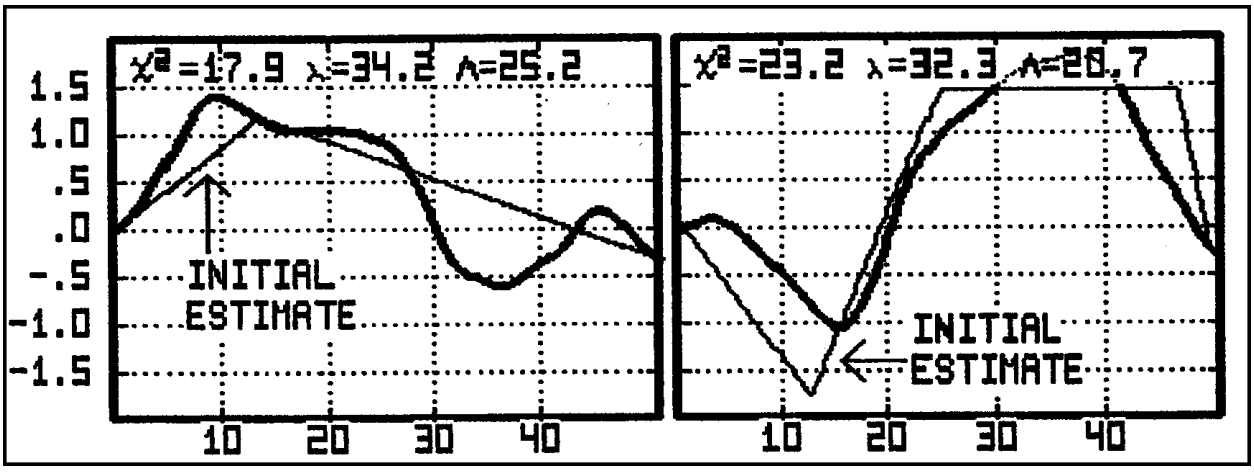

Figure 20: Estimation in which the initial estimates caused the iterative procedure to converge on **local** maxima of Λ, well below the figure of 43.2 obtained in Figure 18.

start in the tails of **different** local maxima and converge to erroneous solutions which have *a posteriori* probabilities of only 10^{-8} and 10^{-10} of that of the correct solution. Only if we had found the global MAP solution, however, would we have become aware of the improbability of the solutions shown in Figure 20. That is the crux of the global problem.

Solution using the Exact Logarithm

Suppose that instead of employing the LLA approximation for the determination of the derivatives of Λ, as in (181)–(183), we had employed the **exact logarithm** for the determination of λ written out in the lower expression of (168). We will then, for compactness, express the results in terms of the previously defined **complex likelihood**, for photon #k. Given a sample function $\tau^{(1)}$, then L_k is given by:

$$L_k \triangleq V(\nu_k)e^{j\theta_k}e^{j2\pi\nu_k(\tau^{(1)}(t_k)-\tau_c(t_k))} \tag{188}$$

Then using the **exact logarithm**, the derivatives of Λ with respect to the U_i are found to be:

$$\frac{\delta\Lambda}{\delta U_i} = \sum_{k=1}^{K} \frac{2\pi\nu_k \cos(\frac{\pi}{T}it_k)Re\{jL_k\}}{1+Re\{L_k\}} - \frac{U_i^{(1)}}{\sigma_i^2} \quad (EL) \tag{189}$$

$$\frac{\delta^2\Lambda}{\delta U_i^2} = -\frac{1}{\sigma_i^2} - \sum_{k=1}^{K}(2\pi\nu_k)^2\cos^2(\frac{\pi}{T}it_k)\frac{|V|^2+Re\{L_k\}}{(1+Re\{L_k\})^2} \quad (EL) \tag{190}$$

$$\frac{\delta^2\Lambda}{\delta U_i\delta U_j} = -\sum_{k=1}^{K}(2\pi\nu_k)^2\cos(\frac{\pi}{T}it_k)\cos(\frac{\pi}{T}jt_k)\frac{|V|^2+Re\{L_k\}}{(1+Re\{L_k\})^2} \quad (EL) \tag{191}$$

In the practical algorithm resulting from the previous analysis, only the substitution of (189) for (181) is required to perform estimation based on the exact logarithm. (190) and (191) are included only for future use in analyzing the **performance** of the algorithm.

It should be mentioned that in applying the algorithm using (189), especially for $|V| \approx 1$ (which is the only time that a significant improvement can result from using the exact logarithm), the **stability** of the iterative process is affected. This characteristic can be easily traced to the denominator of the terms summed in (189), which can, with an unfortunate selection of net

phase, find itself close to zero, causing the associated term to blow up. This is the result of a dilemma in which the algorithm is considering a sample function of τ which predicts nearly zero probability of a photon being detected **which actually had been detected**. A practical algorithm has alleviated this problem by multiplying the real part of the complex likelihood in the denominator of (189) by a factor which starts at 0, and gradually approaches 1 during the course of the iterations. Note that setting this supplementary factor to zero exactly implements the LLA estimation algorithm.

The Time-Domain Algorithm for Maximizing Λ

While the convergence toward $\tau^{(MAP)}$ using the path estimation procedure described above is no less than optimum, it is, in practice, computationally inefficient. This is particularly true as T, the interval over which τ is estimated, increases.

Consider that the iteration of each Fourier component of U_i in (186) requires the evaluation of K terms, one for each photon, as described by (181). This then be done for all M frequency components of $\tau(t)$ being considered, for a total of MK computations. Now, consider what happens when the interval T is doubled. The frequency spacing between adjacent Fourier components is halved, so that in order to make a comparable estimate, the number of Fourier components considered, M, must be doubled. Also the number of photons in that interval, K is approximately doubled. Thus, this interval requires **four times** as much computation in order to estimate only **twice** as much of the unknown $\tau(t)$ function.

This progressive inefficiency can be trace to the distributed information contained in the Fourier coefficients U_i, each of which affects the function $\tau(t)$ over the entire interval $0 < t < T$. On the other hand, the effect of a given photon is relatively local, affecting the resultant solution $\tau^{(MAP)}(t)$ only over a time span on the order of T_0, typically around 10 milliseconds. Thus by carrying out the computation of (181) we are, in effect, tediously computing the Fourier transform of a time-domain series of impulses corresponding to the received photons, only to subsequently (after filtering) use a Fourier transform to return to the time-domain.

The reader may correctly suspect that the transformation to and from the frequency-domain can be bypassed. The effect of individual photons on the maximization of Λ will then result from passing a time series through a properly designed filter whose impulse response duration need only be on the order of the time period over which a photon affects the change in $\tau(t)$ from iteration to iteration. What follows, now, is the development of such a procedure in the time-domain, which realizes an operation

identical to that previously implemented by applying (186) to the Fourier coefficients U_i, followed by transformation to the time-domain.

In the previously described path estimation procedure we update our estimate of $\tau(t)$ over an interval $0 < t < T$ by successively adding to the present estimate of the Fourier cosine series of τ contained in the vector U, a correction vector ΔU calculated using (186). The updated U is then transformed to the time-domain to yield the updated $\tau(t)$. If the present estimate of $\tau(t)$ is denoted $\tau^{(N)}$ then we denote the **new** estimate as $\tau^{(N+1)}$ which would thus be given by:

$$\tau(t)^{(N+1)} = \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T}t)(U_i^{(N)} + \Delta U_i) \tag{192}$$

Using (186) we can write ΔU_i as a gain parameter a_i for that frequency component, multiplying the partial derivative of Λ:

$$\Delta U_i = a_i \frac{\delta \Lambda}{\delta U_i} \tag{193}$$

It was previously pointed out that the gain a_i did not affect the ultimate solution, but rather the rate and stability of convergence toward that solution. The **expected** "optimum" value of a_i has been given as the reciprocal of (187) and will be discussed below.

Recalling (189), we see the derivative of Λ has two components:

$$\frac{\delta \Lambda}{\delta U_i} = -\frac{U_i}{\sigma_i^2} + \sum_{k=1}^{K} 2\pi \nu_k \cos(\frac{\pi}{T} i t_k) \frac{Re\{jL_k\}}{1 + Re\{L_k\}} \tag{194}$$

The first term is a relaxation term flowing from the present estimate of τ, in which the higher frequency components are emphasized and **subtracted** from the present estimate in order to

achieve a smoother function incurring a smaller χ^2 penalty in its value of Λ given by (171). The remaining summation in (194) is due to the K photons. Let us define P_i to be proportional to this entire summation according to[17]

$$\epsilon_i \pi T P_i = \sum_{k=1}^{K} 2\pi\nu_k \cos(\frac{\pi}{T} i t_k) \frac{Re\{jL_k\}}{1 + Re\{L_k\}} \tag{195}$$

Using this simplified notation for the summation of photon increments in (194), substituting that result into (193), and substituting **that** result for U_i into the inverse transform for $\tau^{(N+1)}$ given by (192), then yields:

$$\begin{aligned} \tau(t)^{(N+1)} &= \sum_{i=1}^{\infty} \cos(\frac{\pi}{T} i t_k) \left[U_i^{(N)} + a_i \left(-\frac{U_i^{(N)}}{\sigma_i^2} + \epsilon_i \pi T P_i \right) \right] \\ &= \sum_{i=0}^{\infty} \cos(\frac{\pi}{T} i t_k) \left(1 - \frac{a_i}{\sigma_i^2} \right) U_i^{(N)} + \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T} t)(a_i \epsilon_i \pi T) P_i \end{aligned} \tag{196}$$

Implementation using Time-Domain Filters

Each of the two summations in (196) has a special form, given generally by:

[17]Note that the time-domain procedure using (195) implements the previously described algorithm using the **exact logarithm**. The implementation of the estimation procedure in which the LLA approximation has been applied simply requires eliminating the denominator of (195). Likewise, unstable or chaotic behavior of the time domain estimation procedure due to the denominator of (195) approaching zero, can be alleviated by multiplying the real part of the complex likelihood in the denominator of (195) by a factor in between 0 and 1. During the course of the iterations toward the solution, that factor is allowed to approach unity, as described on page 113.

$$y(t) = \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T} t) Y_i = \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T} t) H_i X_i \tag{197}$$

What this general form expresses is that if we start with a function $x(t)$, with Fourier coefficients X_i, multiply each coefficient by H_i, and then take the inverse transform, calling the result $y(t)$, this will have implemented the temporal filtering of $x(t)$ by a filter whose frequency response $H(\omega)$ is given by:

$$H(\omega) = H_{i|i=\frac{\omega T}{\pi}} \tag{198}$$

Applying this observation to (196) we find the new estimate of $\tau(t)$ to consist of the addition of two filtered time series. The first term is a filtered version of the present estimate of $\tau(t)$ since $U^{(N)}$ is the Fourier cosine transform of $\tau^{(N)}$. Now we have only to ask what, when transformed, yields P_i as defined in (195).

To answer that, we form a time function consisting of K Dirac delta functions, one corresponding to each photon. Each delta function is set at the time t_k of the respective photon, and is assigned a strength according to the multiplicative factors in (195) (aside from the cosine factor).

$$p(t) = \sum_{k=1}^{K} \nu_k \frac{Re\{jL_k\}}{1 + Re\{L_k\}} \delta(t - t_k) \tag{199}$$

To verify that P_i is the cosine transform of $p(t)$ so defined, we need only insert (199) into the general formula for the determination of the coefficients of the Fourier cosine series for a function of time $x(t)$:

$$X_i = \frac{2}{\epsilon_i T} \int_0^T x(t) \cos(\frac{\pi i}{T} t) dt \tag{200}$$

Each delta function sifts out a factor of the cosine of the frequency times t_k producing a summation equivalent to the definition of P_i expressed in (195).

To go further, we must address the frequency-dependent gain parameter a_i introduced in (193). We have previously discussed the fact that insufficient gain will slow down the rate of convergence, whereas excessive gain will lead to overshooting. Gain of over twice the optimum value will cause absolute instability making convergence impossible. The optimum gain was specified in (185) as the negative reciprocal of the second derivative of Λ with respect to U_i. But Λ, given by (171) has two components. The second derivative of the χ^2 component is deterministic. The second derivative of the other part, λ, is random with an expected value which we shall call $-Q'$ (this accepted value is not a function of the frequency index i, except at $i = 0$ where it is doubled). So the **optimum** gain would be:

$$\begin{aligned} a_i(optimum) &= \left[-\frac{\delta^2 \Lambda}{\delta U_i^2} \right]^{-1} \\ &= \frac{1}{-\frac{\delta^2 \lambda}{\delta U_i^2} + \frac{1}{2} \frac{\delta^2 \chi^2}{\delta U_i^2}} = \frac{1}{\frac{1}{\sigma_i^2} - \frac{\delta^2 \lambda}{\delta U_i^2}} \\ \therefore \quad & E\left\{ \frac{1}{a_i(optimum)} \right\} = \frac{1}{\sigma_i^2} + \epsilon_i Q' \end{aligned} \tag{201}$$

where:

$$Q' \stackrel{\wedge}{=} -E\left\{\frac{\delta^2\lambda}{\delta U_i^2}\right\} \tag{202}$$

We shall leave the deterministic part alone, but apply an additional gain factor b, to the random (and essentially unknown) quantity whose expected value we call $-Q'$.

We shall therefore choose, for the gain of frequency component i, the value given by:

$$a_i = \frac{1}{\frac{1}{\sigma_i^2} + \frac{\epsilon_i Q'}{b}} \approx \begin{cases} \frac{b}{Q'} & small\, i \\ \sigma_i^2 & large\, i \end{cases} \tag{203}$$

We have inserted b into this expression in such a way that at low frequencies (where σ_i is large) the gain is equal to the expected optimum gain times b, but at high frequencies, where the effect of the first term in the reciprocal dominates, the gain is unaffected by b (since the optimum gain has already been closely determined). Let us now substitute this choice into (196) so that we can evaluate the frequency response of the filters to be applied to $\tau(t)^{(N)}$ and $p(t)$.

Using this choice for a_i, (196) becomes:

$$\begin{aligned} \tau(t)^{(N+1)} &= \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T}t) \left(1 - \frac{1}{1 + \frac{Q'\sigma_i^2\epsilon_i}{b}}\right) U_i \\ &\quad + \sum_{i=0}^{\infty} \cos(\frac{\pi i}{T}t) \frac{\epsilon_i \pi T}{\frac{1}{\sigma_i^2} + \frac{\epsilon_i Q'}{b}} P_i \end{aligned} \tag{204}$$

$$= \sum_{i=0}^{\infty} \cos\left(\frac{\pi i}{T} t\right) \frac{1}{1 + \frac{b}{Q' \sigma_i^2 \epsilon_i}} \left(U_i + \frac{\pi T b}{Q'} P_i\right) \tag{205}$$

What this equation expresses is that by having judiciously chosen the gain in (203) and making b independent of frequency, we have reduced the problem to a **single** time-domain filter acting on the **sum** of the current estimate $\tau^{(N)}$ and a scaled version of $p(t)$. We expect the essence of the actual system to not be dependent on T, the time interval over which it operates. Q' in (201), is given by (202) as the negated second derivative of λ which is itself the sum of λ_k for each photon in the time interval. Thus Q' will be proportional to T, so let us call:

$$Q \triangleq \frac{Q'}{\pi T} \tag{206}$$

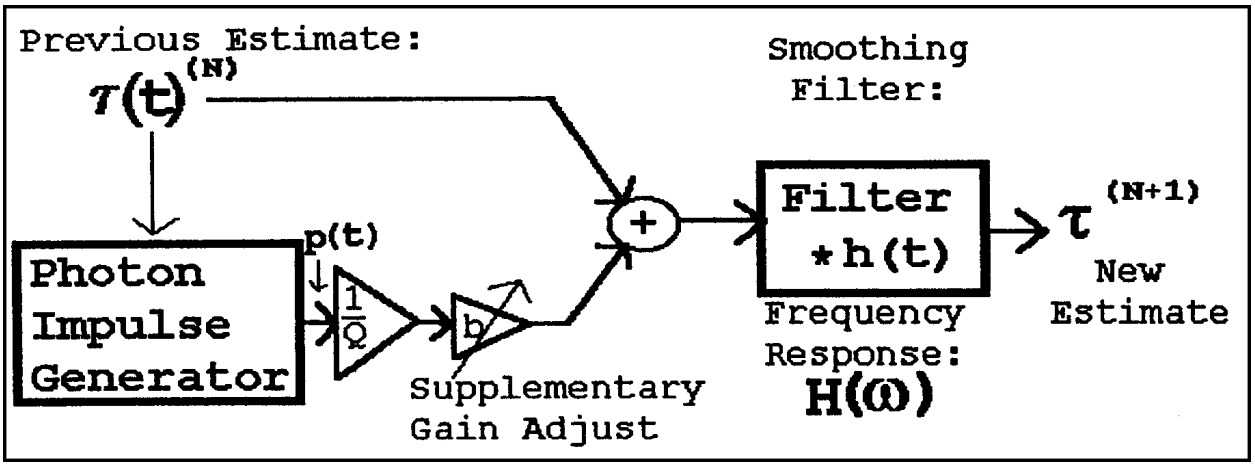

Figure 21: Signal flow diagram of one iteration of the time domain algorithm.

The resulting algorithm is depicted in Figure 21.

Determination of the Filter Transfer Function

Now using the value of σ_i derived for the atmosphere using our model in (69) we can determine the frequency response of the

smoothing filter whose impulse response is denoted $h(t)$ in the diagram. Its frequency response, $H(\omega)$, will be a real number, with no phase angle, at all frequencies. Thus its impulse response is **symmetric** about $t = 0$. Making the substitution for σ_i using (69) we find:

$$H(\omega) = \frac{1}{1 + \frac{b}{Q'\sigma_i^2\epsilon_i}} = \frac{1}{1 + \frac{b}{\alpha Q'} i^{8/3}} \tag{207}$$

Setting the frequency index i to correspond to the frequency ω, this reduces to:

$$H(\omega) = \frac{1}{1 + \frac{b}{\alpha Q \pi T} \left(\frac{\omega T}{\pi}\right)^{8/3}} = \frac{1}{1 + \left(\frac{\omega}{\omega_0}\right)^{8/3}} \tag{208}$$

where

$$\omega_0 \stackrel{\wedge}{=} \pi^{11/8} T^{-5/8} \alpha^{3/8} Q^{3/8} b^{-3/8}$$

ω_0 is the filter's -6 db cutoff frequency, which we shall proceed to determine. α has been determined in (68). The second derivative of λ is determined, using the LLA approximation as part of (236)–(238). Using the **exact logarithm** it is determined in (265). The result for Q, in either case is given by:

$$Q = \frac{E\left\{-\frac{\delta^2 \lambda}{\delta U_i^2}\right\}}{\pi T} = \frac{\pi^2 \bar{K} \overline{\nu^2 V^2} e_1}{\pi T} = \pi \overline{\nu^2 V^2} e_1 I_0 \tag{209}$$

Using the **exact logarithm**, e_1 is defined in (265). The appropriate result if using the LLA approximation, is obtained by setting $e_1 = 1$. Using these results for α and Q, we find the cutoff frequency ω_0 to be:

$$\begin{aligned}\omega_0 &= \pi^{11/8} T^{-5/8} \left(\frac{T_\tau^{1/3} T^{5/3}}{14.06} \right)^{3/8} (\pi \overline{\nu^2 V^2} e_1 I_0)^{3/8} b^{-3/8} \\ &= 2.75 T_\tau^{1/8} (\overline{\nu^2 V^2} e_1 I_0)^{3/8} b^{-3/8} \end{aligned} \tag{210}$$

As we require, T has dropped out of the expression, leaving a filter response whose shape is fixed, but frequency-scaled according to parameters expressing the signal level and the level of atmospheric delay noise. The cutoff frequency must also be adjusted slightly as we vary b, the supplemental gain factor applied to the photon impulses in the lower branch of Figure 21.

Substitution of the definitions for the intrinsic signal-to-noise ratio and T_0 (see (233)) results in the following simplified (and more transparent) form for ω_0:

$$\omega_0 = \frac{1.83}{T_0} \left(\frac{k_2 e_1}{b} \right)^{3/8} SNR^{3/4} \tag{211}$$

Note that the factors inside the parenthesis are on the order of unity. e_1 is defined in (265), and k_2 is defined in (257). The intrinsic signal-to-noise ratio is defined in (233). For a signal-to-noise ratio of 5, we see that the filter cutoff frequency occurs at a frequency of about 1 cycle per T_0, the atmospheric coherence time parameter. T_0 is evaluated at the nominal optical frequency, ν_0, defined in (231).

Similarly expressing the photon impulse gain, b/Q, in terms of the signal-to-noise ratio yields:

$$\frac{b}{Q} = \frac{T_0}{\pi \nu_0^2} \frac{b}{k_2 e_1 SNR^2} \tag{212}$$

Expected Error using the Path Estimation Procedure

Using the path estimation procedure described above to find the function $\tau(t)$ which maximizes Λ (and therefore the *a posteriori* probability of $\tau(t)$ given the set of received photon data), we would like to find the expected estimation error. We will use as an error measurement the r.m.s. difference between the τ function which maximizes Λ and the **actual** function, $\tau^{(ACT)}$, which governed the statistics of the received photons. We will, in this chapter, consider only the solution of the **local problem**, that is, we will be considering the **local maximum** of Λ nearest to $\tau^{(ACT)}$ **whether or not** that is also the global maximum of Λ, a concern that will be addressed in the following two chapters.

The derivation of the expected error will proceed on the basis of the LLA approximation (see page 71). The corresponding derivation using the **exact logarithm** is mathematically cumbersome and will be included at the end of this chapter.

We will define[18] $\Delta\tau$ as the difference between an arbitrary τ function and the MAP (Maximum *A Posteriori* Probability) estimator $\tau^{(MAP)}$. If that τ function is considered to be random and governed by the *a posteriori* probability density of U given the received data, then $\Delta\tau$ will be a random function describing the difference between the MAP estimate, and the real location of $\tau^{(\mathrm{ACT})}$. The r.m.s. average of $\Delta\tau$ may be performed over time ($0 < t < T$) and/or over the ensemble of realizations of the model. We choose to average over both. The mean squared value of $\Delta\tau$ **over the time interval** is calculated as:

$$\langle \Delta\tau^2 \rangle = \frac{1}{T} \int_0^T (\tau^{(ACT)}(t) - \tau^{(MAP)}(t))^2 dt \tag{213}$$

[18]Note that $\Delta\tau$ as used in this chapter is **not to be confused** with $\Delta\tau = \tau - \tau_c$ used in (3) and defined in page 15.

The result we seek will be the expected $\Delta\tau^2$ over the ensemble of realizations of the model. However first averaging over the time interval allows us to employ Parseval's theorem to express $\Delta\tau^2$ in terms of the corresponding cosine transform coefficients. Recall the form of the cosine transform of $\tau(t)$ over $0 < t < T$ that we introduced on page 59.

$$\tau(t) = \sum_{i=0}^{\infty} U_i \cos(\frac{\pi}{T} it) \quad 0 < t < T \tag{214}$$

Using this form, let us designate the transform coefficients of estimated $\tau(t)$ as U_i and designate the transform of $\tau^{(\mathrm{MAP})}$ as $U^{(\mathrm{MAP})}$. If ΔU_i is the difference between U_i and $U_i^{(\mathrm{MAP})}$, then ΔU is also the cosine transform of $\Delta\tau$. Parseval's theorem says that, for this transform pair, the power in the time function and the transform series must (with the proper normalization) be equal, according to:

$$\begin{aligned} \langle \Delta\tau^2 \rangle &= \frac{1}{T}\int_0^T (\Delta\tau(t))^2 dt = \sum_{i=0}^{\infty} \frac{\epsilon_i}{2} (\Delta U_i)^2 \\ &\textit{where} \\ \epsilon_i &= \begin{cases} 1 & i > 0 \\ 2 & i = 0 \end{cases} \end{aligned} \tag{215}$$

Using (215), we can determine the expected error in τ by finding the mean squared values of the ΔU_i. The ΔU_i themselves have zero mean, due to symmetry. Their mean **squared** values are then given by their variances, which comprise the diagonal elements of the covariance matrix describing ΔU. The covariance matrix for ΔU is identical to the covariance matrix of the *a posteriori* U itself, since they differ only by a constant, $U^{(\mathrm{MAP})}$. Given a set of received data, the *a posteriori* density of $\tau(t)$, expressed in terms of its Fourier coefficients U_i, is given by (173). Once $U^{(\mathrm{MAP})}$ has been evaluated by an algorithm, the ΔU separating it from τ is itself governed by (173), assuming the accuracy of our model.

Thus we shall undertake the following procedure to analyze the r.m.s. magnitude of estimation error. We will assume that the **actual** $\tau(t)$ function, expressed in U space, lies in the tails of a Gaussian probability density given by (173). The statistics of the components of the A matrix in the exponential of (173) will be estimated. Then, the covariance matrix of the vector ΔU, given by the **inverse** of A, will be considered. The statistics of the diagonal terms of the covariance matrix will be estimated, yielding the mean squared value of the ΔU_i. These values can then be inserted into Parseval's theorem (215), yielding and estimate for the man squared estimation error of $\tau(t)$ itself. Finally, the relevant mathematical steps will be summarily repeated using the **exact logarithm** form for the complex likelihood function.

The Statistics of the Complex Likelihood at $\tau^{(\mathrm{ACT})}$

Before proceeding, it will be convenient to have on hand the first and second order statistics of the real part of L_k, the complex likelihood function for photon #k, evaluated at $\tau^{(ACT)}$. The complex likelihood was defined in (166), which we repeat here:

$$\begin{aligned} L_k &\triangleq V(\nu_k)e^{j\theta_k}e^{j2\pi\nu_k(\tau(t_k)-\tau_c(t_k))} \\ &= |V(\nu_k)|e^{j\Phi} \end{aligned} \tag{216}$$

In the lower expression, we have consolidated all of the phase factors in Φ, where:

$$\Phi \triangleq \arg(V(\nu_k)) + \theta_k + 2\pi\nu_k(\tau(t_k) - \tau_c(t_k)) \tag{217}$$

Note that Φ is a combination of random and non-random components, thus causing it to be random variable τ in (217)) is a random variable governed by the *a posteriori* density of $\tau^{(ACT)}$ **given** θ_k). The probability density of Φ can be shown to be:

$$P(\Phi) = \frac{1 + |V|\cos(\Phi)}{2\pi} \tag{218}$$

Proof:

Although a simple result, the formal derivation requires several steps. Call:

$$\psi \stackrel{\wedge}{=} \arg(V) + 2\pi\nu_k(\tau - \tau_c) = \Phi - \theta_k$$

Given the observation of photon #k at mixing phase θ, then θ is known but ψ is unknown, thus **random**. **Not knowing** θ, the *a priori* density of ψ would clearly be uniform:

$$P(\psi) = \frac{1}{2\pi}$$

As previously determined, the probability of receiving a photon in mixing phase θ is enhanced by one plus the real part of the complex likelihood:

$$\begin{aligned} P(\theta_k|\psi) &= P(\theta)(1 + Re(\{L_k\}) \\ &= P(\theta)(1 + |V|\cos(\theta_k + \psi)) \end{aligned}$$

Applying Bayes' theorem to obtain the *a posteriori* density of ψ:

$$\begin{aligned} P(\theta_k|\psi) &= \frac{P(\theta_k|\psi)P(\psi)}{P(\theta_k)} = \frac{P(\theta_k)(1 + |V|\cos(\theta_k + \psi)) \cdot \frac{1}{2\pi}}{P(\theta_k)} \\ &= \frac{1}{2\pi}(1 + |V|\cos(\theta_k + \psi)) \end{aligned}$$

Now the random variable Φ is deterministic given ψ and θ_k. Since the derivative of Φ with respect to ψ is unity, there is no additional factor involved in computing the probability density of Φ from that of ψ. Thus:

$$\begin{aligned} P_\Phi(\Phi = \Phi|\theta_k) &= P_\psi(\psi = \Phi - \theta_k|\theta_k) = \frac{1}{2\pi}(1 + |V|\cos(\theta_k + (\Phi - \theta_k))) \\ &= \frac{1}{2\pi}(1 + |V|\cos(\Phi)) \end{aligned}$$

Since the probability of Φ given θ_k is no longer a function of θ, we can omit the conditional probability notation, proving (218).

Using this first result it is possible to evaluate a number of expectations. First, to evaluate the expected value of the real part of L_k **given** ν_k, we find:

$$\begin{aligned} E\{Re\{L_k\}|\nu_k\} &= \int_0^{2\pi} d\Phi\, P(\Phi)|V(\nu_k)|\cos(\Phi) \\ &= \int_0^{2\pi} d\Phi \frac{1+|V|\cos(\Phi)}{2\pi}|V|\cos(\Phi) \\ &= \frac{|V(\nu_k)|^2}{2} \end{aligned} \tag{224}$$

Similarly, the expected value of the **square** of the real part of L_k can be computed:

$$\begin{aligned} E\{(Re\{L_k\})^2|\nu_k\} &= \int_0^{2\pi} d\Phi\, P(\Phi)|\,(|V|\cos(\Phi))^2 \\ &= \int_0^{2\pi} d\Phi \frac{1+|V|\cos(\Phi)}{2\pi}|V|^2\cos^2(\Phi) \\ &= \frac{|V(\nu_k)|^2}{2} \end{aligned} \tag{225}$$

The expectations happen to be equal.

Now that we have the expectations **given** ν_k, we wish to also average **over all photons** and their respective optical frequencies. The likelihood of detecting a photon of optical frequency ν is proportional to $F(\nu)$, the **photon spectral density** function. Note that this differs from the actual spectral function describing the color of the observed object in several ways. In addition to the power spectral density function of the starlight, $F(\nu)$ takes into account the proportion of starlight that is accepted into a single mode by the collection optics (expected to vary as $\nu^{-5/3}$, for larger apertures), the throughput of the optical path, the quantum

efficiency of the detector, and the relation of photon count to optical power (obtained by dividing by the photon energy $h\nu$.).

Let us call the total detected photon count rate I_0. We will take $F(\nu)$ to be normalized so its integral is unity. Then the probability of detecting a photon of optical frequency ν in a bandwidth of $\Delta\nu$ in a time period Δt is $(\Delta\nu)(\Delta t)I_0F(\nu)$. As before, we will take $V(\nu)$ to be a "known" function of ν representing the (complex) correlation between the two light waves measured at optical frequency ν. Using these definitions, we can define several symbols representing "*a priori*" statistics that we shall consider to be "known." The mean squared visibility is denoted:

$$\overline{V^2} \triangleq E\{|V(\nu_k)|^2\} = \int_0^\infty d\nu\, F(\nu)|V(\nu_k)|^2 \tag{226}$$

Now using this notation, we can write the expectations **over all photons** of (224) and (225) as:

$$E\{Re\{L_k\}\} = E\left\{(Re\{L_k\})^2\right\} = \frac{\overline{V^2}}{2} \tag{227}$$

In a similar way, we wish to introduce the following notation for expectations that will arise in the course of the following pages.

$$\overline{\nu^2V^2} \triangleq E\left\{\nu^2|V(\nu_k)|^2\right\} = \int_0^\infty d\nu\, \nu^2F(\nu)|V(\nu_k)|^2 \tag{228}$$

$$\overline{\nu^4V^2} \triangleq E\left\{\nu^4|V(\nu_k)|^2\right\} = \int_0^\infty d\nu\, \nu^4F(\nu)|V(\nu_k)|^2 \tag{229}$$

$$\overline{\nu^2} \triangleq E\{\nu^2\} = \int_0^\infty d\nu\, \nu^2 F(\nu) \tag{230}$$

Let us call the square root of (230) the "average" or nominal optical frequency, ν_0.

$$\nu_0 \triangleq \sqrt{\overline{\nu^2}} \tag{231}$$

Let us again define the **intrinsic signal-to-noise ratio**, now that we have defined ν_0. T_0 was defined (52) as the time period that it would take an air mass moving at v_0, the hypothetical "wind velocity" according to the Taylor model, to traverse the distance of the Fried parameter r_0. We will consider T_0 relative to the optical frequency ν_0, in which case, using (52) and (51), T_0 can be shown to be given by:

$$T_0 = .53 T_\tau^{-1/5} \nu_0^{-6/5} \tag{232}$$

In the time period T_0 there are an expected number of photons given by $I_0 T_0$. The **noise**, assuming Poisson statistics, is given by the **square root** of that number. The **signal** is given by $I_0 T_0$ **multiplied by** the magnitude of correlation $|V|$. Using the root mean squared value of V, we would therefore define the intrinsic signal-to-noise ratio as:

$$SNR \triangleq \sqrt{I_0 T_0 \overline{V^2}} \tag{233}$$

Determination of the Statistics of the Elements of A

As expressed in (180), the elements of A can be found from the second (partial) derivatives of Λ, the logarithm of *a posteriori* probability density. The validity of (180) extends over the volume

of U in which Λ is dominated by the quadratic expression which peaks at $U^{(\mathrm{MAP})}$, that is, where the probability density of U is found to be dominated by the Gaussian probability density which peaks at $U^{(\mathrm{MAP})}$. We have assumed that the underlying function $\tau^{(\mathrm{ACT})}$ that we are trying to estimate lies in the tails of that Gaussian. Especially for small estimation errors, it can be said that the second derivatives of Λ evaluated at $U^{(\mathrm{ACT})}$ will be very close to the second derivatives of Λ evaluated at $U^{(\mathrm{MAP})}$. As the estimation error grows larger (due to decreasing signal-to-noise ratio) this approximation may become **less justified**, but still adequate for the purposes of the following analysis.

Thus we shall proceed to evaluate the statistics of the second derivatives of Λ at $U^{(\mathrm{ACT})}$ (there is no apparent method for finding these statistics at $U^{(\mathrm{MAP})}$ directly). Writing Λ as given by (171), and applying (170), (168), and (162), we previously found expressions determining the diagonal and off-diagonal elements of A in (182) and (183). We shall rewrite those expressions in terms of the complex likelihoods L_k.

$$\frac{\delta^2\Lambda}{\delta U_i^2} = -\frac{1}{\sigma_i^2} - \sum_{k=1}^{K}(2\pi\nu_k)^2\cos^2(\frac{\pi}{T}it_k)Re\{L_k\} \quad (LLA) \tag{234}$$

$$\frac{\delta^2\Lambda}{\delta U_i\delta U_j} = -\sum_{k=1}^{K}(2\pi\nu_k)^2\cos(\frac{\pi}{T}it_k)\cos(\frac{\pi}{T}jt_k)Re\{L_k\} \quad (LLA) \tag{235}$$

First, we will derive the expected values of the diagonal terms of A using (180) (with $i = j$) and (234).

$$\begin{aligned} E\{A_{ii}\} &= -E\left\{\frac{\delta^2\Lambda}{\delta U_i^2}\right\} \\ &= E\left\{\sum_{k=1}^{K}(2\pi\nu_k)^2\cos^2(\frac{\pi}{T}it_k)Re\{L_k\}\right\} + \frac{1}{\sigma_i^2} \end{aligned} \tag{236}$$

We will evaluate this expression at $\tau^{(\mathrm{ACT})}$ as discussed. The bottom expression contains three random variables, each of which are

uncorrelated. We can take the expectation of the $\cos^2$ term, for random t, to be 1/2. The expectation of the real part of L_k has been determined in (227). The expected number of photons received during the time interval T will be denoted:

$$\bar{K} \stackrel{\triangle}{=} E\{K\} = I_0 T_0 \tag{237}$$

The we can evaluate the expectation of the product in the summation as the product of the expectations, since the random variables are uncorrelated. We can also substitute the notation of (228) and (237).

$$\begin{aligned} E\{A_{ii}\} &= E\left\{\sum_{k=1}^{K}(2\pi\nu_k)^2\frac{1}{2}\frac{|V(\nu_k)|^2}{2}\right\} + \frac{1}{\sigma_i^2} \\ &= \bar{K}\cdot E\{\pi^2\nu_k^2|V(\nu_k)|^2\} + \frac{1}{\sigma_i^2} \\ &= \pi^2\bar{K}\overline{\nu^2V^2} + \frac{1}{\sigma_i^2} \end{aligned} \tag{238}$$

We will **not** go on to evaluate the **variance** of the diagonal elements. However it can be easily verified that as K, the number of photons in the time period being considered, increases, the magnitude of the variance in relation to the mean of (238) dwindles. Since we are free to set T to be arbitrarily large, we can cause the A_{ii} to become virtually deterministic.

No we can similarly look at the expectations of the **off-diagonal** elements of A using (235). We find:

$$\begin{aligned} E\{A_{ij}\} &= -E\{\frac{\delta^2\Lambda}{\delta U_i\delta U_j}\} \\ &= E\left\{\sum_{k=1}^{K}(2\pi\nu_k)^2\cos(\frac{\pi}{T}it_k)\cos(\frac{\pi}{T}jt_k)Re\{L_k\}\right\} \\ &= 0 \end{aligned} \tag{239}$$

The expectation is determined to be zero, given that the two cosine factors are uncorrelated for $i \neq j$.

Let us then find the expectation of the **square** of A_{ij}, identical to the **variance** of A_{ij} given that its mean is zero. Since the terms of the summation in (239) are uncorrelated, the expectation of the square of the sum is the sum of the expectations of the squares in the terms.

$$\begin{aligned} E\{A_{ij}^2\} &= E\left\{\sum_{k=1}^{K}\left[(2\pi\nu_k)^2\cos(\frac{\pi}{T}it_k)\cos(\frac{\pi}{T}jt_k)Re\{L_k\}\right]^2\right\} \\ &= \bar{K}E\left\{(2\pi\nu)^4\cos^2(\frac{\pi}{T}it_k)\cos^2(\frac{\pi}{T}jt_k)(Re\{L_k\})^2\right\} \\ &= \bar{K}E\left\{16\pi^4\nu^4\frac{1}{2}\frac{1}{2}\frac{|V|^2}{2}\right\} \\ &= 2\pi^4\bar{K}\overline{\nu^4V^2} \end{aligned} \tag{240}$$

Inversion of the A Matrix

Using the result in Appendix I, it will be possible to estimate the diagonal elements of the covariance matrix R applying to the vector U. The covariance matrix R, as can be seen by (173), is the inverse of the A matrix.

Application of the formula in Appendix I requires the evaluation of the diagonal terms of A (as we have found in (238)), and also determination of the variances of the off-diagonal elements and ability to express the variance of A_{ij} in the following form.

$$\frac{Var\{A_{ij}\}}{A_{ii}A_{jj}} = g_i g_j \tag{241}$$

Substituting in the results of (238) and (240), we would require the definition of the g_i to be consistent with:

$$g_i g_j = \frac{2\pi^4 \bar{K} \overline{\nu^4 V^2}}{\left(\frac{1}{\sigma_i^2} + \pi^2 \bar{K} \overline{\nu^2 V^2}\right)\left(\frac{1}{\sigma_j^2} + \pi^2 \bar{K} \overline{\nu^2 V^2}\right)} \tag{242}$$

That condition can be met using the following form for g_i:

$$g_i = \frac{\pi^2 \sqrt{2\bar{K} \overline{\nu^4 V^2}}}{\frac{1}{\sigma_i^2} + \pi^2 \bar{K} \overline{\nu^2 V^2}} \tag{243}$$

We now substitute in the previously determined value for the σ_i given in (69).

$$\begin{aligned} g_i &= \frac{\pi^2 \sqrt{2\bar{K} \overline{\nu^4 V^2}}}{\frac{i^{8/3}}{\alpha} + \pi^2 \bar{K} \overline{\nu^2 V^2}} \\ &= \frac{\pi^2 \sqrt{2\bar{K} \overline{\nu^4 V^2}}\alpha}{i^{8/3} + \beta} \end{aligned} \tag{244}$$

where:

$$\beta \triangleq \alpha \pi^2 \bar{K} \overline{\nu^2 V^2}$$

(The value of the parameter α was previously determined in (68).)

Now the estimation of the diagonal elements of R, the inverse of the matrix A, requires the evaluation of the quantity J, formed by the summation:

$$
\begin{aligned}
J &\triangleq \sum_{i=1}^{M} g_k \\
&= \sum_{i=1}^{M} \frac{\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha}{i^{8/3}+\beta}
\end{aligned} \tag{245}
$$

For large T, the magnitude of the terms of the summation varies slowly with i. Thus the summation can be safely transformed into an integral:

$$
\begin{aligned}
J &\approx \int_1^M di \frac{\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha}{i^{8/3}+\beta} \\
&\approx \pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha \int_0^\infty \beta^{3/8} dx \frac{1}{\beta x^{8/3}+\beta} \\
&= \pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha\beta^{-5/8} \int_0^\infty \frac{dx}{1+x^{8/3}} \\
&\textit{where} \\
x &\triangleq \beta^{-3/8} i
\end{aligned} \tag{246}
$$

Numerical evaluation of the normalized integral yields a value of 1.275. We thus determine the following estimate of the summation given by J.

$$
J \approx 1.275\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha\beta^{-5/8} \tag{247}
$$

Now, the formula derived in Appendix I for the expected value of the diagonal elements of the covariance matrix (equal to the inverse of A) is evaluated using (238), (244), and (247).

$$
\begin{aligned}
E\{R_{ii}\} = E\{[A^{-1}]_{ii}\} &= \frac{1 + g_i J}{A_{ii}} \\
&= \frac{1 + \frac{\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha}{i^{8/3}+\beta} 1.275\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2}}\alpha\beta^{-5/8}}{\pi^2 \bar{K}\overline{\nu^2 V^2} + \frac{i^{8/3}}{\alpha}} \\
&= \frac{\alpha/\beta}{1 + \frac{i^{8/3}}{\beta}} \left(1 + \frac{2.55\pi^4 \bar{K}\overline{\nu^4 V^2}\alpha^2\beta^{-13/8}}{1 + \frac{i^{8/3}}{\beta}} \right)
\end{aligned}
\tag{248}
$$

It should be pointed out that the fraction inside the parenthesis is usually much smaller than unit and therefore of little significance. It only becomes at all significant under low signal-to-noise ratios. However the term is proportional predicted by the analysis and will be retained and contribute to the final result below.

Determination of the Estimation Error

Finally, substitution of (248) into Parseval's theorem (215) produces a summation that will yield a mean squared expected difference (error) between $\tau^{(ACT)}$ (the underlying differential delay function) and $\tau^{(MAP)}$ (the best estimate of τ). We will ignore the doubling of the power contribution from the $i = 0$ (zero frequency) Fourier component towards the estimate of $\Delta\tau$; for larger T its relative significance dwindles.

$$\tilde{\tau}^2 \stackrel{\wedge}{=} E\{\Delta\tau^2\} = \frac{1}{2}\sum_{i=0}^{\infty} E\{R_{ii}\}$$
$$= \sum_{i=0}^{\infty} \frac{\alpha/2\beta}{1+\frac{i^{8/3}}{\beta}} + \sum_{i=0}^{\infty} \frac{1.275\pi^4 \bar{K}\overline{\nu^4 V^2}\alpha^3\beta^{-21/8}}{\left(1+\frac{i^{8/3}}{\beta}\right)^2} \tag{249}$$

Now we need only evaluate the summations. The factors in the numerators will be taken outside the summations. The first summation has already been performed above, with different factors. Repeating the derivation, a substitution for the index i allows transformation of the summation into a standard integral which has been evaluated numerically.

$$\sum_{i=0}^{\infty} \frac{1}{1+\frac{i^{8/3}}{\beta}} \approx \int_0^{\infty} \beta^{3/8} \frac{dx}{1+x^{8/3}} \approx 1.275\beta^{3/8}$$
$$where \tag{250}$$
$$x \stackrel{\wedge}{=} \beta^{-3/8} i$$

Similarly, the second summation is transformed into an integral which has been evaluated numerically.

$$\sum_{i=0}^{\infty} \frac{1}{\left(1+\frac{i^{8/3}}{\beta}\right)^2} \approx \int_0^{\infty} \frac{\beta^{3/8}\, dx}{(1+x^{8/3})^2} \approx .797\beta^{3/8}$$
$$x \stackrel{\wedge}{=} \beta^{-3/8} i \tag{251}$$

Now we can insert the values of the summations into (249) and arrive at the mean squared estimation error.

$$\begin{aligned}\tilde{\tau}^2 &= \frac{1.275}{2}\alpha\beta^{-5/8} + .797 \cdot 1.275\pi^4\bar{K}\overline{\nu^4V^2}\alpha^3\beta^{-9/4} \\ &= .637\alpha\beta^{-5/8} + 1.02\pi^2\frac{\overline{\nu^4V^2}}{\overline{\nu^2V^2}}\alpha^2\beta^{-5/4} \\ &= .637\alpha\beta^{-5/8}\left(1 + 1.6\pi^2\frac{\overline{\nu^4V^2}}{\overline{\nu^2V^2}}\alpha\beta^{-5/8}\right)\end{aligned} \tag{252}$$

The factor $\alpha\beta^{-5/8}$ occurs twice. We will proceed to simplify the $-8/5$ power of that factor, and express it in terms of the intrinsic signal-to-noise ratio.

$$\begin{aligned}\beta\alpha^{-8/5} &= \pi^2\bar{K}\overline{\nu^2V^2}\alpha^{-3/5} \\ &= \frac{4.88}{T_\tau^{1/5}T}\pi^2\bar{K}\overline{\nu^2V^2}\end{aligned} \tag{253}$$

Let's express T_τ in terms of T_0 using (232). Then we shall apply the definition of the intrinsic signal-to-noise ratio (233) yielding:

$$\begin{aligned}\beta\alpha^{-8/5} &= \frac{T_0(2\pi\nu_0)^{6/5}}{4.82}4.88\pi^2\frac{\bar{K}}{T}\overline{\nu^2V^2} \\ &= 1.88 \cdot 4.88\pi^2\nu_0^{6/5}I_0T_0\overline{\nu^2V^2} \\ &= 90.68\nu_0^{16/5}\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}SNR^2\end{aligned} \tag{254}$$

$$\therefore \quad \alpha\beta^{-5/8} = .0598\nu_0^{-2}\left(\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}\right)SNR^{-5/4}$$

Substituting this result into (252) produces our ultimate formula for the r.m.s. estimation error.

$$\begin{aligned}\tilde{\tau}^2 &= .038\nu_0^{-2}\left(\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}\right)^{-5/8} SNR^{-5/4}\cdot \\ &\quad \left(1+.94\frac{\overline{\nu^4V^2}}{\overline{\nu^2}\,\overline{\nu^2V^2}}\left(\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}\right)^{-5/8} SNR^{-5/4}\right) \qquad (255)\end{aligned}$$

$$\begin{aligned}\tilde{\tau} &= \frac{.195}{\nu_0}\left(\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}\right)^{-5/16} SNR^{-5/8}\cdot \\ &\quad \sqrt{1+.94\frac{\overline{\nu^4V^2}}{\overline{\nu^2}\,\overline{\nu^2V^2}}\left(\frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}}\right)^{-5/8} SNR^{-5/4}} \qquad (256)\\ &= \frac{.195}{\nu_0}k_2^{-5/16}SNR^{-5/8}\sqrt{1+.94k_4k_2^{-5/8}SNR^{-5/4}}\end{aligned}$$

where:

$$\begin{aligned}k_2 &\triangleq \frac{\overline{\nu^2V^2}}{\overline{\nu^2}\,\overline{V^2}} \\ k_4 &\triangleq \frac{\overline{\nu^4V^2}}{\overline{\nu^2}\,\overline{\nu^2V^2}} \qquad (257)\end{aligned}$$

Note that the combined statistics k_2 and k_4 are tailored to each be on the order of unity. In particular, for narrowband light they both approximate unity, and for observing an unresolved object, $|V| = 1$, k_2 is always equal to one. Thus we see that estimation error follows the $-5/8$ power of signal-to-noise ratio, or, for a fixed $V(\nu)$, follows the $-5/16$ power of collected optical power, assuming low-noise detection. At the "nominal" optical frequency ν_0 the above estimation error translates to a **phase** estimation error of:

$$\begin{aligned}\tilde{\Phi} &= 2\pi\nu_0\tilde{\tau} \\ &= 1.23k_2^{-5/16}SNR^{-5/8}\sqrt{1+.94k_4k_2^{-5/8}SNR^{-5/4}} \qquad (258)\end{aligned}$$

We shall find in the discussion of the **global** estimation problem that the lowest usable SNR is around 3.0. In what could therefore be termed the "worst case," (258) predicts a phase estimation error of about 40 degrees r.m.s. Simulations have been run at signal-to-noise ratios between 1 and 10 to verify (258). Using fully correlated light with photons uniformly distributed over a 2:1 range of optical frequency (not a very realistic spectrum), and employing the time-domain algorithm using the LLA approximation, the r.m.s. error was computed for a number of simulations which were averaged to reduce the uncertainty of the result. These results, converted to phase error by multiplying by $2\pi\nu_0$ are plotted in Figure 22 along with the theoretical value given by (258). Agreement between the data points and the theoretical curve is apparent, with the simulation results exhibiting a slightly lower than predicted by theory, except at very low signal-to-noise ratios where the simulations show a greater errorpredicted. In Figure 23 similar simulations

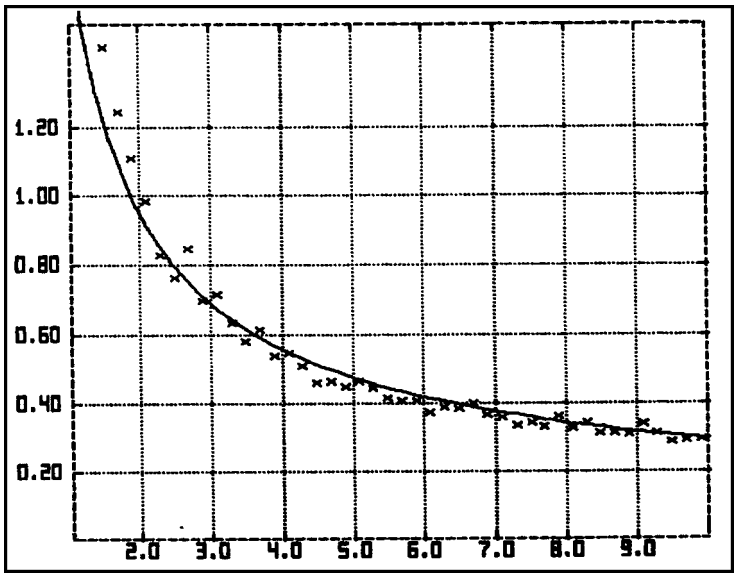

Figure 22: Average phase estimating error vs. signal-to-noise ratio, for simulations run with $|V| = 1$, using LLA, compared to theory (258).

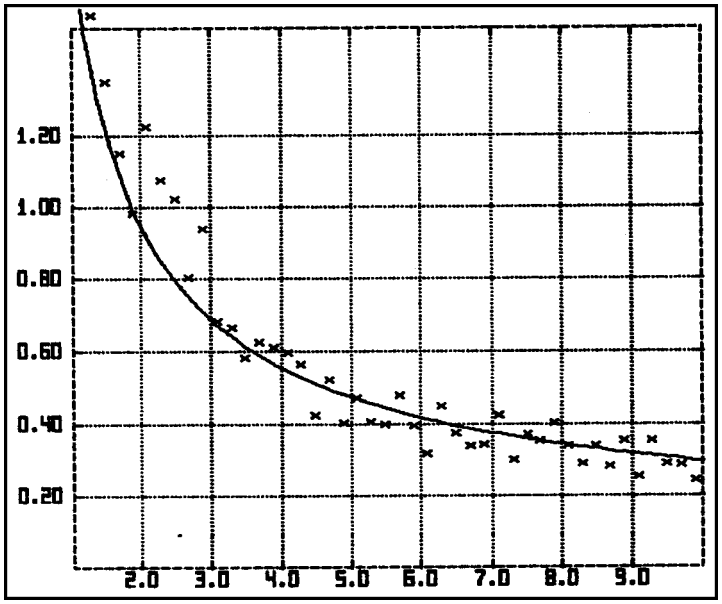

Figure 23: Average phase estimation error vs. signal-to-noise ratio, for simulations run with $|V| = .3$, using LLA, compared to theory (258).

were run with photons that were generated with reduced correlation: $V = .3$ at all wavelengths. Although the variance of the simulation results has greatly increased (even after averaging a larger number of simulation results to produce each plotted point), agreement with theory is again evident.

Derivation of the Estimation Error using the Exact Logarithm

The above derivation of the r.m.s. estimation error, employed the LLA approximation (page 71) in order to greatly reduce the complexity of the mathematics. We shall now repeat the same derivation using the **exact logarithm** in the estimation algorithm and in the analysis of the estimation error. The reader is reminded that the discrepancy involved in the LLA approximation is only significant for visibilities of magnitude approaching unity. In more typical cases of objects that are at least somewhat resolved, and/or photo-detectors that experience a dark count that is not very small compared to the photon count, the advantage of using the exact logarithm will be minuscule.

The use of the LLA approximation in the estimation procedure and analysis of estimation error first occurred in the use of the **upper** expression of (168) for the evaluation of λ_k, the logarithm of the conditional probability of receiving photon $\#k$ in the interference phase in which it eventually **was** detected. The resultant approximation was propagated into the determination of λ in (170), Λ in (171), and the derivatives of Λ in (181)–(183), results which were repeated above in (234) and (235). Using the **lower** (exact) expression of (168) for λ_k, we found the corresponding second derivatives of Λ to be given by (190) and (191), repeated below.

$$\frac{\delta^2\Lambda}{\delta U_i^2} = -\frac{1}{\sigma_i^2} - \sum_{k=1}^{K}(2\pi\nu_k)^2\cos^2(\frac{\pi}{T}it_k)\frac{|V|^2 + Re\{L_k\}}{(1 + Re\{L_k\})^2} \tag{259}$$

$$\frac{\delta^2\Lambda}{\delta U_i \delta U_j} = -\sum_{k=1}^{K}(2\pi\nu_k)^2\cos(\frac{\pi}{T}it_k)\cos(\frac{\pi}{T}jt_k)\frac{|V|^2 + Re\{L_K\}}{(1+Re\{L_k\})^2} \tag{260}$$

We can now go on to rework the mathematics in (236)–(258) using these more difficult expressions to determine the statistics of the elements of the A matrix. Again, the diagonal elements of A are found from the second derivative now given by (259).

$$\begin{aligned} E\{A_{ii}\} &= -E\{\frac{\delta^2\Lambda}{\delta U_i^2}\} \\ &= E\{\sum_{k=1}^{K}(2\pi\nu_k)^2\cos^2(\frac{\pi}{T}it_k)\frac{|V|^2 + Re\{L_k\}}{(1+Re\{L_k\})^2}\} + \frac{1}{\sigma_i^2} \end{aligned} \tag{261}$$

Evaluating this expression at $\tau^{(\mathrm{ACT})}$now requires the determination of the expected value of the fraction containing L_k at $\tau^{(\mathrm{ACT})}$. This will be done in two parts. Again we will express L_k as given by (216) in which ϕ is defined in (217) with a probability density given by (218). The first expectation to evaluate is:

$$\begin{aligned} E\left\{\frac{1}{1+Re\{L_k\}}\right\} &= \int_0^{2\pi} P(\phi)d\phi\frac{1}{1+Re\{L_k\}} \\ &= \int_0^{2\pi}\frac{1+|V|\cos\phi}{2\pi}d\phi\frac{1}{1+|V|\cos\phi} \\ &= 1 \end{aligned} \tag{262}$$

The evaluation of the second expectation is not quite as simple.

$$
\begin{aligned}
E\left\{\frac{1}{(1+Re\{L_k\})^2}\right\} &= \int_0^{2\pi} \frac{1+|V|\cos\phi}{2\pi} d\phi \frac{1}{(1+|V|\cos\phi)^2} \\
&= \frac{1}{2\pi}\int_0^{2\pi} \frac{d\phi}{1+|V|\cos\phi} \\
&= (1-|V|^2)^{-1/2}
\end{aligned}
\tag{263}
$$

(A table of integrals has been used in the final step.) The evaluation of the fraction in (261) can now proceed using these results.

$$
\begin{aligned}
&E\left\{\frac{|V|^2+Re\{L_k\}}{(1+Re\{L_k\})^2}\right\} \\
&\qquad = E\left\{\frac{1}{1+Re\{L_k\}}\right\} - (1-|V|^2)\cdot E\left\{\frac{1}{(1+Re\{L_k\})^2}\right\} \\
&\qquad = 1-\sqrt{1-|V|^2}
\end{aligned}
\tag{264}
$$

We have evaluated the expectation over the ensemble of realizations of photons **given the optical frequency** ν_k and the $V(\nu)$ at that wavelength. We have yet to average over all optical frequencies. Inserting this result into (261) we find:

$$E\{A_{ii}\} = \frac{1}{\sigma_i^2} + \bar{K}(2\pi)^2 \frac{1}{2} E\{\nu^2(1 - \sqrt{1 - |V(\nu_k)|^2})\}$$
$$= \frac{1}{\sigma_i^2} + \pi^2 \bar{K} \overline{\nu^2 V^2} e_1 \tag{265}$$

where

$$e_1 \triangleq \frac{E\{2\nu^2(1 - \sqrt{1 - |V(\nu_k)|^2})\}}{\overline{\nu^2 V^2}}$$
$$(1 \le e_1 \le 2)$$

We have created the exact logarithm **enhancement factor** e_1 as a dimensionless quantity which is close to unity for smaller magnitudes of V, but can attain a value of 2 in the case of $|V| = 1$ at all wavelengths. Recall that the advantage of using the **exact logarithm** only occurs for $|V|$ approaching unity. Note that, except for the inclusion of the enhancement factor, this expression for the value of the diagonal elements of A is identical to the previously determined value (238) using the LLA approximation.

The determination of the expectation of the **variance** of A_{ij}, using the **exact algorithm** undergoes a similar modification.

$$E\{A_{jj}^2\} = E\left\{\left(-\frac{\delta^2 \Lambda}{\delta U_i \delta U_j}\right)^2\right\}$$
$$= E\{\sum_{k=1}^{K} [(2\pi\nu_k)^2 \cos(\frac{\pi}{T} i t_k) \cos(\frac{\pi}{T} j t_k) \frac{|V|^2 + Re\{L_K\}}{(1 + Re\{L_k\})^2}]^2\} \tag{266}$$
$$= \bar{K} E\{(2\pi\nu_k)^4 \cos^2(\frac{\pi}{T} i t_k) \cos^2(\frac{\pi}{T} j t_k) \left(\frac{|V|^2 + Re\{L_K\}}{(1 + Re\{L_k\})^2}\right)^2\}$$

Now we are required to evaluate the expectation of the **square** of the same fraction that we addressed in (263). First, the squared fraction is algebraically decomposed into three terms, each of which will have its expectation calculated by brute force integration.

$$
\begin{aligned}
& E\left\{\left(\frac{|V|^2+Re\{L_k\}}{(1+Re\{L_k\})^2}\right)^2\right\} \\
& = E\left\{\frac{1}{(1+Re\{L_k\})^2}\right\} - (1-|V|^2)\cdot E\left\{\frac{1}{(1+Re\{L_k\})^4}\right\} \\
& \quad -2(1-|V|^2)\cdot E\left\{\frac{1}{(1+Re\{L_k\})^3}\right\}
\end{aligned} \tag{267}
$$

The first expectation has already been found in (263). The following two are evaluated by integration over ϕ (again a table of integrals has been employed).

$$
\begin{aligned}
& E\left\{\frac{1}{(1+Re\{L_k\})^4}\right\} = \int_0^{2\pi} \frac{1+|V|\cos\phi}{2\pi} d\phi \frac{1}{(1+|V|\cos\phi)^4} \\
& = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\phi}{(1+|V|\cos\phi)^3} \\
& = \frac{1}{4(1+|V|^2)^2\sqrt{1-|V|^2}}\left[\frac{3}{2}+\frac{1+|V|}{1-|V|}+\frac{3}{2}\left(\frac{1+|V|}{1-|V|}\right)^2\right] \\
& = \frac{1+\frac{|V|}{2}}{(1-|V|^2)^{5/2}}
\end{aligned} \tag{268}
$$

$$
\begin{aligned}
E\{\frac{1}{(1+Re\{L_k\})^3}\} &= \int_0^{2\pi} P(\phi)d\phi \frac{1}{(1+|V|\cos\phi)^3} \\
&= \frac{1}{2\pi}\int_0^{2\pi} \frac{d\phi}{(1+|V|\cos\phi)^2} \\
&= \frac{1}{(1-|V|^2)^{3/2}}
\end{aligned}
\tag{269}
$$

Substituting these results into (267) we obtain:

$$
\begin{aligned}
E\left\{(\frac{|V|^2+Re\{L_k\}}{(1+Re\{L_k\})^2})^2\right\} &= (1-|V|^2)^{-1/2} \\
&\quad -(1-|V|^2)^2\frac{1+\frac{|V|^2}{2}}{(1-|V|^2)^{5/2}} \\
&\quad -2(1-|V|^2)^2(1-|V|^2)^{-3/2} \\
&= \frac{|V|^2}{2\sqrt{1-|V|^2}}
\end{aligned}
\tag{270}
$$

Again, the final step involves averaging (270) over all optical frequency, ν, and substitution into (266) in which the expected

value of the product of uncorrelated factors is written as the product of the expected values.

$$E\{A_{ij}^2\} = \bar{K} \cdot 16\pi^4 \nu^4 \frac{1}{2}\frac{1}{2} E\left\{\frac{|V|^2}{2\sqrt{1-|V|^2}}\right\}$$
$$= 2\pi^4 \bar{K} \overline{\nu^4 V^2} e_2 \qquad (271)$$

where :

$$e_2 \triangleq \frac{E\left\{\frac{\nu^4 |V|^2}{\sqrt{1-|V|^2}}\right\}}{\overline{\nu^4 V^2}}$$

Again, we have formed an expression for the variance of the off-diagonal terms of A that is identical to the determination (240) using the LLA approximation, except for the addition of the factor e_2 which is approximately unity for small $|V|$. Both e_1 and e_2 are combined statistics of V and ν. For the (unrealistic) case of **constant** V over wavelength, the behaviors of e_1 and e_2 are plotted as functions of V in Figure 24.

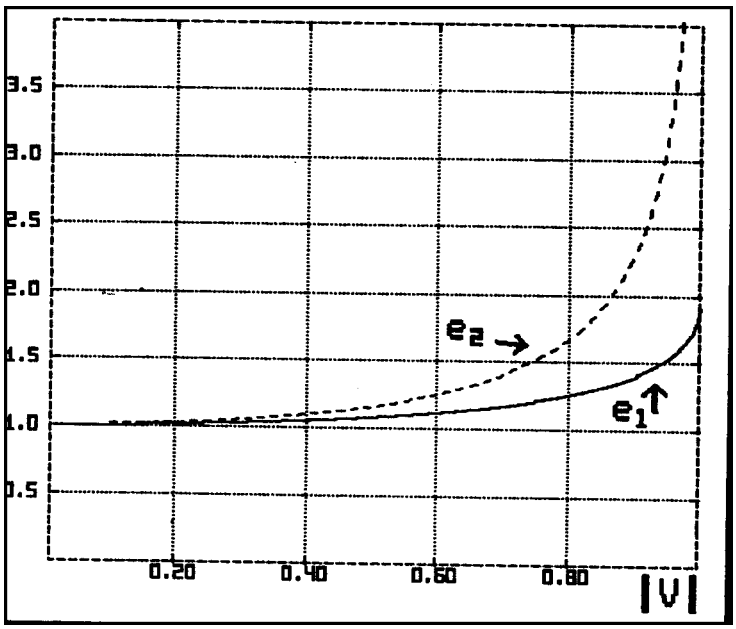

Figure 24: Enhancement factors e_1 and e_2 plotted vs. V, for the **special case** of constant V for all wavelengths.

We now use the results (265) and (271), derived using the **exact logarithm**, to modify the previous determination of the statistics of the elements of the matrix A. The determination of g_i consistent with (241) will now require the replacement of (244) with:

$$g_i = \frac{\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2} e_2}}{\frac{1}{\sigma_i^2} + \pi^2 \bar{K}\overline{\nu^2 V^2} e_1} = \frac{\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2} e_2 \alpha}}{i^{8/3} + \beta} \tag{272}$$

where

$$\beta \triangleq \alpha \pi^2 \bar{K}\overline{\nu^2 V^2} e_1$$

Note that we have absorbed the enhancement factor e_1 into the redefinition of β. We go on to form the summation of the g_i as in (245) with the inclusion of the enhancement factors. The result corresponding to (247) then becomes:

$$J \approx 1.275\pi^2 \sqrt{2\bar{K}\overline{\nu^4 V^2} e_2 \alpha} \beta^{-5/8} \tag{273}$$

Again, we are employing the **new** definition of β (272). We now substitute these results into the formula derived in Appendix I for the expected value of the diagonal elements of the covariance matrix, as in (248).

$$E\{R_{ii}\} = E\{[A^{-1}]_{ii}\} = \frac{1 + g_i J}{A_{ii}} = \frac{\alpha/\beta}{1 + \frac{i^{8/3}}{\beta}} \left(1 + \frac{2.55\pi^4 \bar{K}\overline{\nu^4 V^2} e_2 \alpha^2 \beta^{-13/8}}{1 + \frac{i^{8/3}}{\beta}} \right) \tag{274}$$

Again, the major modification due to e_1 is buried in the new definition of β. Propagating the effect of e_1 and e_2, we produce a slightly modified version of (256), the expected r.m.s. error in the estimation of $\tau(t)$.

$$\tilde{\tau} = \frac{.195}{\nu_0} k_2^{-5/16} e_1^{-5/16} SNR^{-5/8} \sqrt{1 + .94 k_4 k_2^{-5/8} e_1^{-5/8} e_2 SNR^{-5/4}} \qquad (275)$$

Setting e_1 and e_2 to unity will simply recover (256), the formula for estimation error using the estimation procedure in which the LLA approximation has been employed. The primary difference in (275) is the $e_1^{-5/16}$ factor which multiplies the entire expression. At its maximum, e_1 attains a value of 2 when $|V| = 1$ at all wavelengths, in which case a 20% decrease in the estimation error is predicted. This is equivalent to a $\sqrt{2}$ increase in the signal-to-noise ratio, corresponding to a doubling of the received optical power (in the case of a quantum noise dominated detector).

The effect of e_1 and e_2 under the radical in (275) is relatively minor. Curiously, the effect of the e_2 factor alone is to **worsen** the estimation error, according to (275). However its prediction of catastrophic estimation failure as $|V|$ approaches 1 is neither expected on general principles nor borne out by simulations. We can conjecture that the analysis that introduced the correction factor failed to observe the breakdown of approximations which had otherwise been valid. Specifically, taking the logarithm of a random variable **which is allowed to approach zero** (as can occur in the case of the logarithm of $1 + \mathrm{Re}\{L_k\}$ when $|V| = 1$) will yield a random variable with an unexpectedly wide and skewed distribution (as it takes on very negative values when $1 + \mathrm{Re}\{L_k\} \approx 0$). The Central Limit Theorem can no longer claim that the summation of several such random variables will be normally distributed.

As before, simulations have been run in which photons were randomly generated, with a uniform distribution over a 2:1 range of optical frequency. The optical correlation, V, was set to .9 at all wavelengths (setting V to unity was avoided, so that the anomalous behavior of e_2, discussed above, could be avoided, in plotting the theoretical curve). Again, there is good agreement between the simulations and theory.

However it should be noted that simulation results using the exact logarithm tend toward a slightly **higher** r.m.s. estimation error than predicted by theory, whereas the results of simulations using the LLA approximation showed a **lower** figure than theory (at SNR> 3). Taken together, we find that the **reduction** of estimation error using the exact logarithm seen in simulations, is only about half of the reduction predicted by theory. This only goes to make the usefulness of the exact logarithm (in the few cases that it even applies) even more questionable.

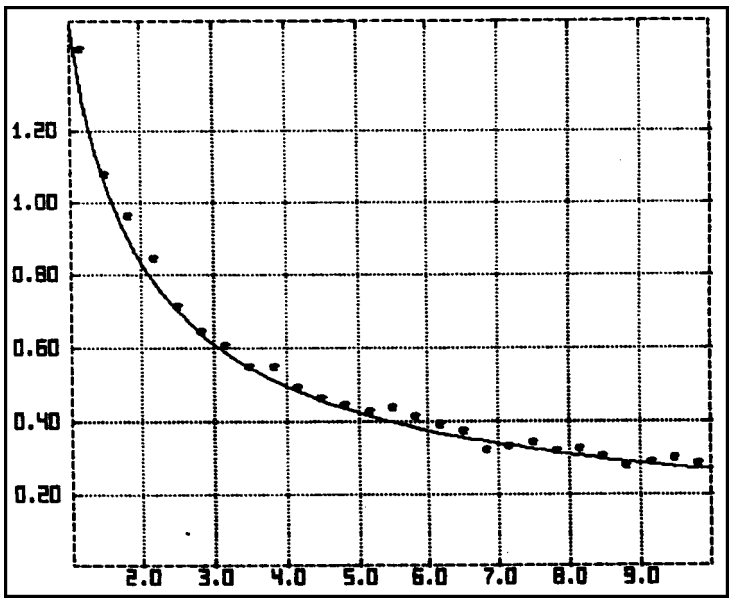

Figure 25: Phase error vs. SNR of simulations run with $|V| = .9$, using the **exact logarithm**, compared to theory (275) (multiplied by $2\pi\nu_0$).

The Estimation Error using only Prior Data

The determination of the mean squared estimation error incorporated the factor of $1/2$ resulting from Parseval's theorem (215). This factor entered into (249) in the determination of the mean squared estimation error **averaged over the interval** $0 < t < T$. However suppose we wish to determine the mean squared error at the endpoints, $t = 0$ and $t = T$. Then it can be seen at those endpoints, the expected value of the power contributed by each component of U to τ in (214) is no longer $\frac{1}{2}U_i^2$ but rather U_i^2. If we are willing to accept that there is zero average correlation between U_i and U_j, the we could write, for the endpoints of the time interval, an expression **similar** to Parseval's theorem:

$$\langle \Delta\tau^2 \rangle = \sum_{i=0}^{\infty} (\Delta U_i)^2 \qquad (276)$$

Substitution of (276) for (215) in the determination of the mean squared estimation error in (249) will **double** the value of that determination. This will result in a $\sqrt{2}$ increase in the r.m.s. estimation error found in (256).

Therefore a **real-time** delay-tracking system, using only photon detections **in the past** will be capable of a delay-tracking performance with an estimation error exceeding that of the **off-line** estimation system by the factor $\sqrt{2}$. More important, however, is the reduction in **confidence** (regarding the **global** estimation problem) that use of only prior data entails. This effect can be easily noticed in the character of the global solutions to be presented in Figure 32 through Figure 38, in which the confidence of the **interior** points is consistently superior to that of the estimates at $t = 100$ (or $t = 0$), at which point only photons **in the past** contributed to the estimation of $\tau(t)$.

Analysis of the Confidence of MAP Solutions

Using the MAP path estimation procedure, we can search for all the various paths which maximize Λ, the logarithm of *a posteriori* probability, in a local sense. After normalizing the probability of these paths, so that the sum of their probabilities is one, we consider the probability assigned to the path which describes the **actual** $\tau(t)$ function. We call that probability the **confidence** of the correct solution. We are focussing on that quantity because it can be used as a **quantitative** measure of the qualitative behavior of the global solution[19] In the following investigation we seek to analyze the theoretical relationship between signal-to-noise ratio and the expected confidence of the correct solution. The qualitative character of the global solution and the concrete **meaning** of confidence, will only be described in the following chapter (which the reader **may** choose to read first). At this point we will limit the discussion to a mathematical analysis whose quantitative results will apply to that problem.

In the ensuing analysis, some rough approximations and inequalities are invoked. The direction of the inequalities is so as to put an **upper limit** on the expected confidence of the correct solution. Thus we shall produce no guarantee of how high the confidence will be with a sufficient signal-to-noise ratio; such a guarantee might be inferred from simulation results. However we will determine a signal-to-noise ratio **below which** the confidence is predicted to rapidly deteriorate. Again, the specifics of that deterioration can be explored with numerical simulations. However it will be reassuring to have found analytically the approximate signal-to-noise ratio at which it occurs.

Consider two time intervals over which $\tau(t)$ is estimated. The first interval is from 0 to T; the second is from T to $2T$. If the confidence of the correct solution over the first interval is p_2

[19] A closely related quantitative measure of the uncertainty due to the multitude of local solutions would be the **entropy** of that uncertainty. That would be computed as the **expected** negative logarithm of the confidence of the correct solution.

and the confidence of the correct solution over the second interval is p_2, then, especially if T is sufficiently large, the confidence of the correct solution over the entire interval 0 to $2T$ will be **approximately** $p_1 p_2$. The approximation becomes more exact as the interaction is reduced between the reckoning of confidence of various paths leading to and starting from time T, the point that the two time segments are joined. For larger T, the relative effect of that joining diminishes. Therefore, it is expected that as T, the time interval considered, grows, the confidence of the correct solution over the growing interval should decay exponentially. Thus the **logarithm** of confidence should decrease **linearly** with T. Let us normalize T relative to T_0, the fringe coherence time parameter. Then if C is the confidence of the correction solution over the interval of time T, we will define the **confidence index**, C.I. as:

$$C.I. \stackrel{\wedge}{=} \lim_{T\to\infty} \frac{-\log C}{(\frac{T}{T_0})} \tag{277}$$

We shall, thus determine a **lower limit** on the confidence index (representing an upper limit in confidence) as a function of signal-to-noise ratio.

To begin the analysis, let us write out the expression given by Bayes' theorem for the *a posteriori* probability density of U given a set of K photons assembled into a data set D.

$$P(\vec{U}|D) = \frac{P(D|\vec{U})P(\vec{U})}{P(D)} = \frac{k_1 P(D|\vec{U})P(\vec{U})}{\int_{all\,\vec{U}} d\vec{U}\, k_1\, P(D|\vec{U})P(\vec{U})} \tag{278}$$

In this case, we have **normalized** the probability density by dividing the probability density in the numerator by the **total probability** of D. We have also multiplied the numerator and denominator by an arbitrary constant, k_1, to show that we can produce a normalized product by inserting the same **unnormalized**

probability into the numerator and denominator integral. That is useful, since we can identify e^{Λ} as the numerator, for some possible k_1. Λ was defined originally in (159). Thus:

$$P(\vec{U}|D) = \frac{e^{\Lambda}}{\int_{all\ \vec{U}} d\vec{U} e^{\Lambda}} \tag{279}$$

The direction of the subsequent analysis is as follows. First we shall place a lower limit on the denominator of (279). Then, using the second-order expansion for Λ around $U^{(\mathrm{MAP})}$ (172), we can integrate (279) to find (an upper limit on) the probability mass contained in the MAP solution, or the **confidence** of that solution. All ingredients to the solution will be based on use of the **exact logarithm** rather than the LLA approximation, although there is only one point in the following derivation where an appreciable difference is manifest, relative to the degree of approximation tolerated.

The numerator of (278), identical to the integrand of its denominator, is, as before, given by:

$$k_1 P(D, \vec{U}) = k_1 P(D|\vec{U}) P(\vec{U}) = e^{\Lambda} = e^{\lambda} e^{-\frac{\chi^2}{2}} \tag{280}$$

where we have previously defined λ in (170) using (167), and χ^2 in (162). Substituting these into the integral of the denominator of (279) using (280), we shall proceed to determine a lower limit on that integral.

$$denominator = \int_{all\ \vec{U}} k_1\, P(D,\vec{U}) = \int_{all\ \vec{U}} e^{\lambda} e^{-\frac{\chi^2}{2}}$$

$$= (2\pi)^{M/2} \left(\prod_{i=0}^{M-1} \sigma_i \right) \int d\vec{U} \frac{e^{-\frac{1}{2}\sum_{i=0}^{M-1} \frac{U_i^2}{\sigma_i^2}}}{(2\pi)^{M/2} \prod_{i=0}^{M-1} \sigma_i} e^{\lambda} \tag{281}$$

We have multiplied and divided by a factor so that the fraction inside the integrand is now simply the (normalized) *a priori* probability density of U. The value of the integral, is therefore just the expected value of e^{λ} over the *a priori* distribution of U or:

$$\int d\vec{U} \frac{e^{-\frac{1}{2}\sum_{i=0}^{M-1} \frac{U_i^2}{\sigma_i^2}}}{(2\pi)^{M/2} \prod_{i=0}^{M-1} \sigma_i} e^{\lambda} = \int d\vec{U}\, P(\vec{U}) e^{\lambda} = E\{e^{\lambda}\}$$

$$= E\left\{ \prod_{k=1}^{K} (1 + Re\{L_k\}) \right\} = E\left\{ \prod_{k=1}^{K} (1 + |V(\nu_k)| \cos(\phi_k)) \right\} \tag{282}$$

Using the **exact logarithm** form for λ, we have expanded e^{λ} into the product of K factors, each being 1 plus $|V|$ times the cosine of a random angle. The expected value of **each factor** is clearly unity.

If the factors were **truly** independent then the expected value of the product would be the product of their expected values, or in this case, 1. However there is some correlation between factors in-łwhich the t_k and the ν_k are close to each other. This is generally a positive correlation, since the photons were generated on the basis of the same underlying $\tau(t)$ function. Therefore the expectation of the product is necessarily **greater** than the product of the expected values:

$$E\{e^{\lambda}\} \geq 1 \tag{283}$$

leading to the lower limit on the value of the denominator of (279):

$$denominator \geq (2\pi)^{M/2} \prod_{i=0}^{M-1} \sigma_i \tag{284}$$

Now suppose we have found a solution to the maximization of Λ, which is the maximum closest to $U^{(\mathrm{ACT})}$, in which Λ can be closely approximated as a second order polynomial in the U_i as in (172) for a sufficient distance around $U^{(\mathrm{MAP})}$ so that the bulk of the probability mass expressed by (279) is included in this approximation, which therefore, constitutes a jointly Gaussian distribution in the U_i. The **confidence**, C, of our solution is the proportion of probability mass accounted for by (279) given the quadratic form for Λ around $U^{(\mathrm{MAP})}$. If we call Λ of that solution Λ_s and the probability density determined by that solution P_s, then the confidence would be:

$$C = \int_{all\ \vec{U}} d\vec{U}\, P_s(\vec{U}|D) = \int d\vec{U}\, \frac{e^{\Lambda_s(\vec{U})}}{denominator} \tag{285}$$

where the denominator in the integrand is the same as the denominator in (279), for which we now have an lower limit. Thus the confidence is bounded by:

$$C \leq \frac{\int d\vec{U} e^{\Lambda_s(\vec{U})}}{(2\pi)^{M/2} \prod_{i=0}^{M-1} \sigma_i} \tag{286}$$

The integral in the numerator is the integral of an **unnormalized** Gaussian density. Using the expression for the

exponent given by (172) we can easily evaluate this integral in terms of the determinant of the matrix A:

$$\begin{aligned}\int_{all\ \vec{U}} d\vec{U}\, e^{\Lambda_s(\vec{U})} &= e^{\Lambda^{(\mathrm{MAP})}} \int_{all\ \vec{U}} d\vec{U}\, e^{-\frac{1}{2}(\vec{U}-\vec{U}^{(\mathrm{MAP})})^t A(\vec{U}-\vec{U}^{(\mathrm{MAP})})} \\ &= e^{\Lambda^{(\mathrm{MAP})}} \frac{(2\pi)^{M/2}}{\sqrt{\det[A]}} \end{aligned} \tag{287}$$

Substituting (287) into (286), we find:

$$C \leq \frac{e^{\Lambda^{(\mathrm{MAP})}}}{\sqrt{\det[A]}\prod_{i=0}^{M-1}\sigma_i} \tag{288}$$

We shall estimate the value of the determinant of A using a result obtained in Appendix I. The result for the determinant parallels the matrix inversion procedure, using the same characterization of A on which (238) and (240) were based. According to the analysis of Appendix I, the expected value of the determinant of A is found from (315) and (325), yielding[20]:

$$E\{\det[A]\} \approx e^{-\frac{J^2}{2}}\left[1+\frac{H}{2}(1-J^2)\right]\prod_{i=0}^{M-1} E\{A_{ii}\} \tag{289}$$

The factor inside the brackets is of relatively small significance and will be ignored. For large T, the variance of A_{ii} and of the determinant is relatively small, so we will discontinue the "expected value" notation. Using the square root of (289), we find the following value for the denominator of (288).

[20]The index of rows and columns of the $M \times M$ determinant has been arranged to run from 0 to $M-1$, rather than the more standard 1 to M. This will assist in the identification of A_{ii} as the variance of U_i. The first element of the vector U has been designated U_0 for the zero-frequency component of the cosine transform.

$$\sqrt{\det[A]}\prod_{i=0}^{M-1}\sigma_i \approx e^{-\frac{J^2}{4}}\cdot\sqrt{\prod_{i=0}^{M-1}A_{ii}\sigma_i^2} \tag{290}$$

Now we will substitute in the expected values of the A_{ii} found in (265).

$$\begin{aligned}\sqrt{\det[A]}\prod_{i=0}^{M-1}\sigma_i &\approx e^{-\frac{J^2}{4}}\cdot\sqrt{\prod_{i=0}^{M-1}(1+\pi^2\bar{K}\overline{\nu^2V^2}\sigma_i^2)}\\ &= e^{-\frac{J^2}{4}}\cdot\sqrt{\prod_{i=0}^{M-1}(1+\beta i^{-8/3})}\end{aligned} \tag{291}$$

We used (69) for the determination of the σ_i^2. β was previously defined in (272).[21]

Taking the **logarithm** of (291) conveniently transforms the product into a summation, which can be closely approximated by a continuous integral.

$$\begin{aligned}\log\left(\sqrt{\det[A]}\prod_{i=0}^{M-1}\sigma_i\right) &\approx -\frac{J^2}{4}+\frac{1}{2}\sum_{i=0}^{M-1}\log(1+\beta i^{-8/3})\\ &\approx -\frac{J^2}{4}+\frac{1}{2}\int_0^{\infty}di\log(1+\beta i^{-8/3})\end{aligned} \tag{292}$$

[21]The lower expression of (291) clearly blows up for $i = 0$, making the product infinite. Recall, however, that σ^2 for $i = 0$ is certainly finite, however we did not wish to complicate our model with such details. Fortunately, the significant value to be derived presently is the **logarithm** of this expression, toward which any single factor in this product (which therefore becomes a **term** in a **summation**) is not of overwhelming significance (especially as T grows larger). In fact, when the summation (292) (which also contains an apparently infinite term) is evaluated using a continuous integral, the resulting "improper integral" converges without having to introduce an "exception" for the case of $i = 0$.

Using the substitution:

$$x \triangleq \beta^{-3/8} i \tag{293}$$

the integral is transformed into a standard form which has been evaluated numerically.

$$\begin{aligned} \int_0^\infty di\, \log(1+\beta i^{-8/3}) &= \beta^{3/8} \int_0^\infty dx\, \log(1+x^{-8/3}) \\ &= 3.40\beta^{3/8} \end{aligned} \tag{294}$$

J in the leading term in (292) has previously been evaluated in (273). Using the definitions for β in (272), k_2 and k_4 in (257), and the intrinsic signal-to-noise ratio in (233), this term can be shown to be equivalent to the following form.

$$\frac{J^2}{4} = \beta^{3/8} \left(.48 \frac{k_4}{k_2^{5/8}} \frac{e_2}{e_1^{5/8}} SNR^{-5/4} \right) \tag{295}$$

The quantity inside the parenthesis can be seen to be equal to $\frac{1}{2}$ of the second term under the radical of the correction factor in (275).

Substituting this result and (294) into (292) yields:

$$\log(\sqrt{\det[A]} \prod_{i=0}^{M-1} \sigma_i) \approx \beta^{3/8} \left(1.70 - .48 k_4 k_2^{-5/8} e_1^{-5/8} e_2\, SNR^{-5/4} \right) \tag{296}$$

The leading factor can be expressed in the following form, by applying the definitions of β (272), intrinsic signal-to-noise ratio (233), T_τ (51), T_0 (52), and k_2 (257).

$$\beta^{3/8} = .59 \frac{T}{T_0} k_2^{3/8} e_1^{3/8}\, SNR^{3/4} \tag{297}$$

Substituting this into (296) produces:

$$\log(\sqrt{\det[A]}\prod_{i=0}^{M-1}\sigma_i)$$

$$\approx .59\frac{T}{T_0}k_2^{3/8}e_1^{3/8}\,SNR^{3/4}(1.70-.48k_4k_2^{-5/8}e_1^{-5/8}e_2\,SNR^{-5/4}) \qquad (298)$$

Taking the logarithm of confidence in (288) and applying (298), leads to the following upper limit.

$$\log(C) \le \Lambda^{(\mathrm{MAP})} - .59\frac{T}{T_0}k_2^{3/8}e_1^{3/8}\,SNR^{3/4}(1.70-.48k_4k_2^{-5/8}e_1^{-5/8}e_2\,SNR^{-5/4}) \quad (299)$$

It can be shown that the expected value of $\Lambda^{(\mathrm{MAP})}$ is (especially for $|V|$ not close to zero) close to, but less than, the expected value of λ evaluated at $\tau^{(\mathrm{ACT})}$. The expected contribution of each λ_k at $\tau^{(\mathrm{ACT})}$ was determined in (224) in the case that the LLA approximation had been used. However using the **exact logarithm**, the expected value of λ at $\tau^{(\mathrm{ACT})}$ is approximately cut in half, and is given by:

$$E\{\lambda_k\} = \frac{|V(\nu_k)|^2}{4}e_3 \qquad (300)$$

where

$$e_3 = 1 + \frac{|V|^2}{8} + \frac{|V|^4}{24} + \ldots$$

The e_3 factor is appreciable only for $|V| >> 0$, and reaches a maximum value of 1.23 when $|V| = 1$. So for a time period T in which the expected number of photons is I_0T, we would expect the value of $\Lambda^{(\mathrm{MAP})}$ to be:

$$E\{\Lambda^{(\mathrm{MAP})}\} \;<\; E\left\{\lambda_{|\tau^{(\mathrm{ACT})}}\right\} = E\left\{\sum_{k=1}^{K}\lambda_k\right\}$$
$$\approx\; K\frac{\overline{V^2}}{4}e_3 = I_0 T\frac{\overline{V^2}}{4}e_3 = \frac{T}{T_0}\frac{SNR^2}{4}e_3 \qquad (301)$$

We have introduced the intrinsic signal-to-noise ratio according to its definition (233).

Substituting (301) into (299), we obtain:

$$\log(C) \le \frac{T}{T_0}SNR^{3/4}\left(\frac{SNR^{5/4}}{4}e_3 - k_2^{3/8}e_1^{3/8} + .28\,k_4 k_2^{-1/4}e_1^{-5/8}e_2\,SNR^{5/4}\right) \qquad (302)$$

The confidence C, being a probability, cannot be greater than one; $\log(C)$ must thus be negative. However the right hand side of the inequality (302) is **positive** for a signal-to-noise ratio higher than a particular value that we shall call $SNR^{(\mathrm{CRIT})}$. For signal-to-noise ratios **higher** than that number, therefore, the above inequality assures us of nothing. We **are** assured, however, that as the signal-to-noise ratio falls **below** that number, the confidence will certainly deteriorate; that is the qualitative result we have sought. We can solve for $SNR^{(\mathrm{CRIT})}$ by setting the expression inside the parenthesis equal to zero, resulting in:

$$SNR^{(\mathrm{CRIT})} = 1.74\,k_2^{3/10}\frac{e_1^{3/10}}{e_3^{4/5}}\left(1 + \sqrt{1 - .28\frac{k_4}{k_2}\frac{e_2 e_3}{e_1^{11/8}}}\right)^{4/5} \qquad (303)$$

k_2 and k_4 were defined in (257). They both take on the value of 1 in the narrowband case. e_1, e_2, and e_3 are all close to 1 except as $|V|$ approaches unity. For the case of $k_4 = k_2 = 1$, and $e_1 = e_2 = e_3 = 1$, we find the critical signal-to-noise ratio is approximately 3.0. This is consistent with the results of simulations which find a rapid disintegration of the *a posteriori* probability mass as the signal-to-noise ratio falls much below about 4.

To make a somewhat more quantitative statement we can look at the expected behaviour of C as the signal-to-noise ratio is decreased **below** $\mathrm{SNR}^{(\mathrm{CRIT})}$ according to (302). We shall write the result in terms of the **confidence index**, defined in (277) as the asymptotic ratio between the negative logarithm of confidence and T/T_0. Using the result of (302) we can write the following prediction for a lower limit on the confidence index for signal-to-noise ratios below $\mathrm{SNR}^{(\mathrm{CRIT})}$:

$$C.I. \geq SNR^{3/4}\left(k_2^{3/8}e_1^{3/8} - \frac{SNR^{5/4}}{4}e_3 - .28k_4k_2^{-1/4}e_1^{-5/8}e_2SNR^{-5/4}\right) \qquad (304)$$

In 20 simulations, using various signal-to-noise ratios between 2.8 and 6.5, there was a fairly exhaustive search for local maxima of Λ. Using the total probability found during that search to normalize the probability (confidence) of each individual path, and identifying the **correct** path among those found, the confidence of that correct path was ascertained. Dividing the negative logarithm of that confidence by T/T_0, the resulting estimates of the **confidence index** were plotted versus signal-to-noise ratio in Figure 26. Simulations in which $|V| = 1$ are each plotted with an "x" while simulations in which there was much less than full optical correlation are each plotted an "o." the theoretical lower limit (304) is plotted on the left of the graph for $e_1 = e_2 = e_3 = k_2 = 1$. k_4 has been set to 1.14 to reflect the model used in the simulations in which photons were uniformly distributed over a 2:1 range of optical frequency. The curve plotted with a **dashed** line reflects the modification to e_1, e_2, and

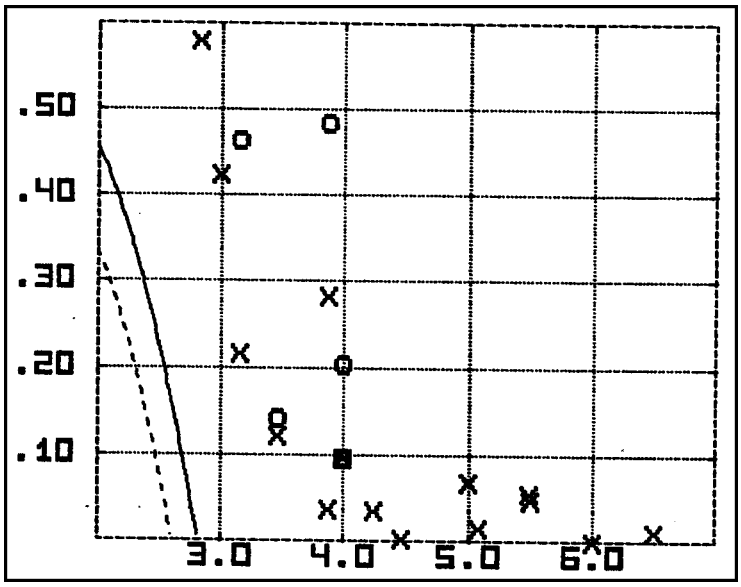

Figure 26: Plot of the lower limit on the confidence index according to (304) vs. SNR, and data points from 20 simulation runs.

e_3 appropriate for $|V| = .9$ ($|V| = 1$ was avoided because of the unacceptable behavior of e_2 as $|V|$ approaches unity, discussed on page 148). The dashed curve is thus more applicable to the points plotted as "x" while the solid curve applies to the points plotted as "o."

In considering Figure 26 it should be remembered that (304) only claims to be a **lower limit** on the confidence index. Any discrepancy between the observed behavior and theory can probably be traced to the weak inequality introduced in (283). One might conjecture, looking at Figure 26, that the slope of the data points appears similar to that of the theoretical curve; such a determination would be premature based upon limited data having such large variance. However it is gratifying that there is evidently a relationship between the **prediction** of a rapidly growing confidence index below a certain signal-to-noise ratio, and results showing such behavior relative to a signal-to-noise ratio which is reasonably close to the predicted SNR^{CRIT}.

Characteristics of Global MAP Solutions

We now wish to present the results of investigation into the nature of the **total** MAP solution, in which the global problem is addressed. Recall that the *a posteriori* probability density of solutions of $\tau(t)$ over the range of $0 < t < T$ is represented by Λ, the logarithm of (unnormalized) *a posteriori* probability density, approximated as a function in an M dimensional space. The value of M should be chosen to accommodate the highest frequency components in which there is appreciable energy; M must necessarily increase with T, the time length of the τ function being considered. The M dimensions have previously been referred to as the axes representing the amplitude of the M components of the discrete cosine transform U_0 through U_{M-1}. However the axes could just as well correspond to the values of τ at M points in time; the same function would result. Since the Fourier transform (if properly normalized) is a linear orthonormal transform, this corresponds to a simple **rotation** of the coordinate system. The shape of Λ remains unchanged.

Λ will be a smooth rolling function throughout the M dimensional space, with multiple peaks. Around each peak Λ can be well approximated as a quadratic function. However a quadratic function around a peak is exactly the logarithm of a Gaussian probability density whose mean is the position of that peak. The volume contained under any one peak corresponds to the net *a posteriori* probability (or **confidence**) of that Gaussian function containing the point corresponding to $\tau^{(\mathrm{ACT})}$, the actual function that we are attempting to estimate. Then, in order to insure, for instance, no more than a 1% change of estimating failure, we would need to account for 99% of the probability mass of e^{Λ} by identifying the N highest peaks of Λ. With a high signal-to-noise ratio, $N = 1$ might suffice, in which case the solution consists of a single path which, with 99% confidence, describes $\tau^{(\mathrm{ACT})}$. With a

lower signal-to-noise ratio, there will be a number of such peaks corresponding to multiple paths. The solution must then consist of the probabilistic **union** of N paths, each of which is characterized by a net probability or **confidence**, a mean path (at which Λ is maximized in a local sense), and a covariance structure (expected error) which we have already examined in detail (see (248)). We shall describe the net probability density as **poly-Gaussian**,, or the union of Gaussian solutions.

The Poly-Gaussian Probability Distribution

To illustrate the concept of a poly-Gaussian *a posteriori* probability density, consider the following one-dimensional estimation problem. We wish to estimate a temperature T whose *a priori* probability is considered to be uniform over $-\infty$ to $+\infty$ (to simplify the estimation problem). The temperature is measured using a thermometer that is known to have a Gaussian error with a standard deviation of $\sigma = 5$ degrees. Then a coin is tossed. The coin has a probability p of landing heads, and $1 - p$ of landing tails. If it lands heads then a particular number, ΔT is **added** to the measured temperature; if tails, it is subtracted. The result, Y, would then have a probability p of coming from the Gaussian distribution $\mu = T + \Delta T$, $\sigma = 5$, and a probability of $1 - p$ of coming from the Gaussian distribution $\mu = T - \Delta T$, $\sigma = 5$. The resulting probability density of Y would clearly be:

$$P(Y) = \frac{p}{5\sqrt{2\pi}} e^{-\frac{(Y-(T+\Delta T))^2}{50}} + \frac{1-p}{5\sqrt{2\pi}} e^{-\frac{(Y-(T-\Delta T))^2}{50}} \tag{305}$$

Such a density is what we call **poly-Gaussian**.

Now assume that an observer is aware of the above model, including the values of the parameters p and ΔT. Receiving the value of Y, the observer is required to estimate T. It can easily be seen (or proved using Bayes' Theorem) that, given Y, the *a posteriori* probability of T would also be poly-Gaussian, characterized as having a probability of p of belonging to the Gaussian distribution $\mu = Y - \Delta T$, $\sigma = 5$, and a probability of $1 - p$ of belonging to the Gaussian distribution $\mu = Y + \Delta T$, $\sigma = 5$.

$$\begin{aligned} P(T|Y) &= \frac{p}{5\sqrt{2\pi}} e^{-\frac{(T-(Y-\Delta T))^2}{50}} + \frac{1-p}{5\sqrt{2\pi}} e^{-\frac{(T-(Y+\Delta T))^2}{50}} \\ \therefore E\{T|Y\} &= p(Y - \Delta T) + (1-p)(Y + \Delta T) \\ &= Y - (2p-1)\Delta T \end{aligned} \tag{306}$$

If the object were to find a single estimate of T which minimized the mean squared estimation error, then the proper estimator would simply be the mean value of the *a posteriori* distribution, shown in (306).

However suppose we need to estimate T subject to a different criterion. For instance, suppose that it will snow if and only if $T < 0$. Now we would like to know the probability of it snowing given Y. In this case, the mean of the *a posteriori* distribution would be of little use. Instead, we would like to use the *a posteriori* **density** itself. Integrating $P(T|Y)$, above, from $-\infty$ to 0 would determine the probability of snow. Equivalently, the error function for each of the two Gaussians that comprise $P(T|Y)$ could be evaluated at the correct points, and combined by multiplying by p and $1 - p$, respectively. The point is that in such cases it is often **not** adequate to determine only the mean and variance of the *a posteriori* distribution. Rather, we would like to be supplied

with the probability density itself. In the case of a poly-Gaussian distribution, that density can be summarized by supplying, for each individual Gaussian component, the mean, variance, and probability of **that** Gaussian applying.

Plots of the *a posteriori* density (306) for $Y = 20$, are shown in Figure 27 for various values of p and ΔT. In A and B, the individual Gaussians comprising the probability density are clearly separated. However in C there is a large overlap between the tails of the two. In D, the overlap has reached the point that the resultant density is again unimodal. In the latter two cases, some likely values of T will not clearly be assigned to one Gaussian or the other. However, we shall see that this is a situation that will not occur, to any extent, in the global solution of the *a posteriori* probability of τ, except at signal-to-noise ratios so low that the results produced would be useless.

One feature of Figure 27 A and B that should be noted is that the area under each of the two distinct Gaussians is proportional to its peak value. That occurs only because we have set up the problem so that the two Gaussians have the **same variance**. That characteristic, it so happens, applies also to the τ estimation problem (approximately). Since the determinant of the covariance matrix of Gaussians corresponding

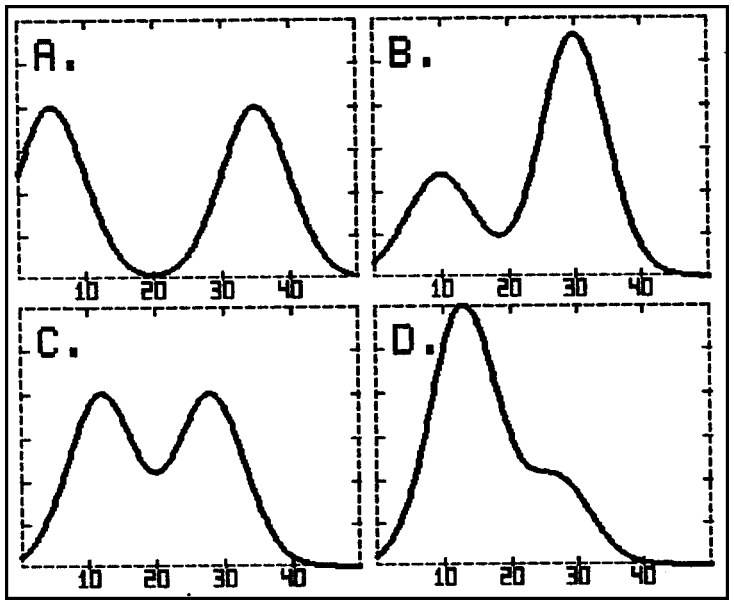

Figure 27: Poly-Gaussian *a posteriori* probability density for example in text. A: $\Delta T = 10$, $p = .5$. B: $\Delta T = 6$, $p = .3$; C: $\Delta T = 5$, $p = .5$; D: $\Delta T = 4$, $p = .7$.

to different peaks of Λ is approximately equal, we can say that the probability mass belonging to any individual Gaussian which is part of the total **poly-Gaussian** distribution of e^{Λ} will be approximately **proportional** to the exponential function of the local peak value of Λ. Using that fact, we can find the probabilities (or **confidence**) of each of N individual Gaussians of the *a posteriori* distribution using the following normalization procedure. If the **peak** Λ found for each of N local peaks is denoted $\Lambda^{(i)}$ for $i = 1$ to N, then the probability of Gaussian #i applying, denoted $P^{(i)}$, would be:

$$P^{(i)} = e^{\Lambda^{(i)} - \Lambda^{(\text{full})}}$$

where : (307)

$$\Lambda^{(\text{full})} \triangleq \log \sum_{i=1}^{N} e^{\Lambda^{(i)}}$$

For this procedure to work it is first necessary to find the N highest peaks of Λ which account for almost all of the total *a posteriori* probability of τ. In other words, N must be sufficiently large so that increasing it will not appreciably affect the value of $\Lambda^{(\text{full})}$ in (307). In that case, (307) will allow us to normalize the probabilities assigned to the N paths. This determination is one element of the global algorithm, the operation of which will be discussed starting on page 189. However we will immediately find useful the graphical output of that algorithm.

To begin the discussion of how a multiple path (or **poly-Gaussian**) solution to the global problem might present itself, let us consider the solution to the maximization of Λ in the case of **narrowband** reception. We recall that if ν_k is the same for all photons, then adding $1/\nu$ to all points of τ will produce a function which does not alter the phase angle supplied to every complex likelihood (80), thus producing the same λ. Using the strict $-8/3$ power low for the *a priori* statistics of the atmospheric path delay τ, there is no penalty whatsoever for the addition of a constant (d.c.) term, thus χ^2 will not be altered. Therefore in the narrowband case, there will be an infinite number of such solutions producing the same Λ. So even with a very high signal-to-nose ratio, we obtain an ambiguous result due to the lack of wavelength diversity. In Figure 28 the signal-to-noise ratio was set to 6.0, but due to the photons all being of the same wavelength, the algorithm has found repeated solutions separated by $1/\nu$ (in this case equal to 2.2 femtoseconds). The 4 solutions all reported value of Λ which were quite close (although not identical, due to idiosyncrasies of the algorithm). The remainder of the infinite number of possible solutions were not found by the algorithm which was based on a limited search space, roughly corresponding to the extent of the τ axis of Figure 28.

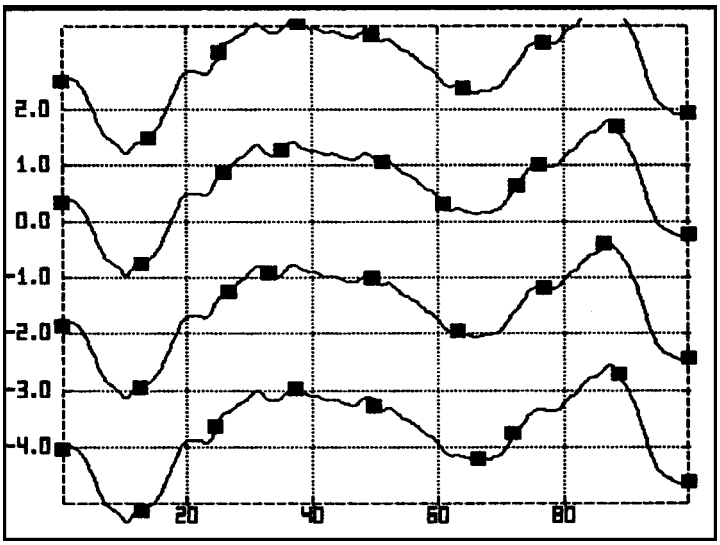

Figure 28: Result of global algorithm estimating paths based on **narrowband** reception at $SNR = .6$. Vertical axis is τ in femtoseconds; horizontal axis is time in milliseconds.

To illustrate the differentiating role that can be played by using a sizeable bandwidth, the algorithm was run using the same underlying τ function (shown in Figure 29), but with a set of photons whose wavelength varied over a 10% bandwidth. In Figure 30 the result of that simulation is plotted with *a posteriori* probability, in percent, printed at each nodal point. In this case it can be seen that the inclusion of multi-spectral information has enabled the algorithm to fairly well differentiate between the **correct path**, shown as having a 92% confidence, and the two first-order "sidelobe" solutions, whose probabilities are reported to be 8% and .2% (the very bottom solution from Figure 28 had such a poor Λ in the new simulation that it was eliminated by the algorithm according to the criterion that had been set). There also appear in Figure 30 a few, albeit improbable, paths **connecting** parts of the adjacent solutions. As we shall see, these are a natural part of the global solution, and were not included in Figure 28 only because that simulation had been performed with a tighter criterion on the basis of which paths were discarded (in Figure 30 it was necessary to relax that criterion so that the lower sidelobe, of *a posteriori* probability = .2%, would survive).

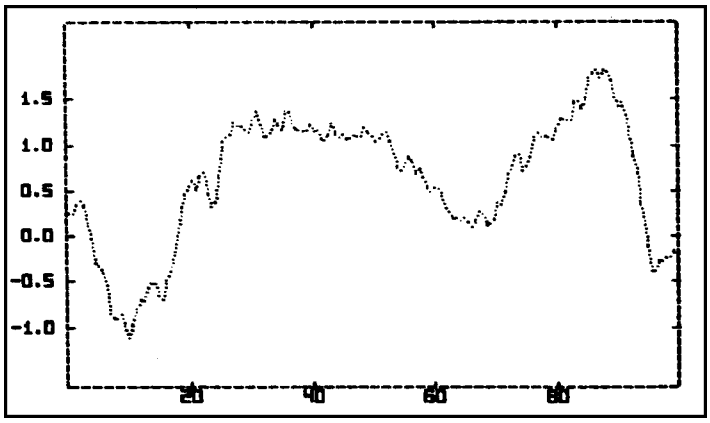

Figure 29: The actual $\tau(t)$ function used in the examples of the global solutions shown in Figure 28 through Figure 40.

The Effect of Signal-to-Noise Ratio

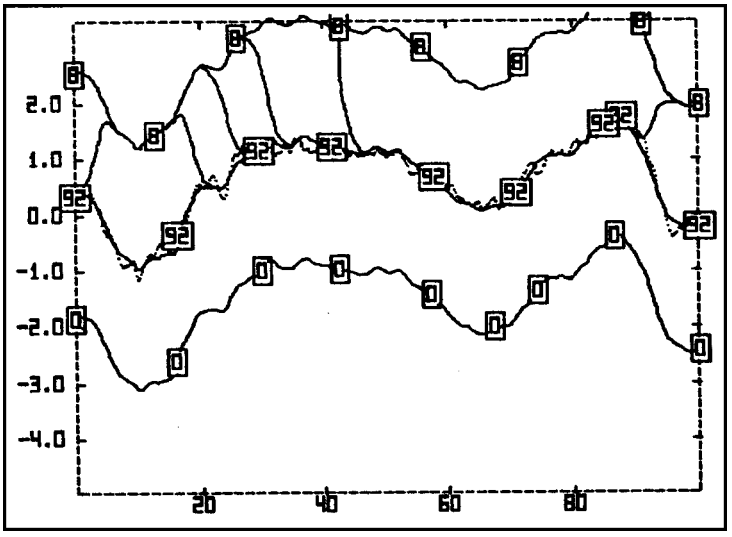

Figure 30: Simulation similar to Figure 28 except that the same number of photons have been produced over a 10% bandwidht. Numbers in boxes are the *a posteriori* probabilities in precent.

Such erroneous paths, however, become more prominent as the signal-to-noise ratio is reduced and the algorithm is required to evaluate the *a posteriori* probability of various paths on the basis of far fewer photons. In Figure 31 for instance, the photon level is 45% of that used in Figure 30, lowering the signal-to noise ratio to 4. The relative likelihood of several such paths has markedly increased. Another feature that might be noted is that with fewer photons, the probable identification of the correct path, achieved in Figure 30,has become quite tenuous. For instance in the region $t = 0 - 30$, the correct solution is only assigned a probability of 39%, with the first higher sidelobe receiving a probability of 46% form $t = 0 - 10$ and 15% from $t = 10 - 65$. It might be noted that **part** of the problem stems from the limited range of time t, over which photons were considered in order to determine the *a posteriori* probabilities of the paths shown. If the time series had extended further in both directions then it is likely that the (correct) 51% path a tthe left would have received a much stronger rating, substantially raising the confidence of the correct solution. However with a still relatively narrowband set of photons, it can be seen that differentiation between adjacent sidelobes remains problematic, especially at lower signal levels.

This can be seen clearly by contrasting Figure 31 with the result of a simulation of the same underlying τ function at the same signal-to-noise ratio, but in which the same number of photons now are spread over a 2:1 range of optical frequency (with the same r.m.s. optical frequency $\nu_0 = .458 \times 10^{15}$ Hz), shown in Figure 32. In this case it can be seen that the confidence of the correct path has risen to 99% (the algorithm has assigned a 1% probability of an erroneous path at $t = 0 - 5$, and 9% probability for the erroneous path at $t = 80 - 100$. The abundance of such erroneous paths, having nothing to do with the misidentification of **sidelobes** in the narrowband case, is rather typical (although somewhat better than average) for this signal-to-noise ratio (=4.0). Simply raising the signal-to-noise ratio to 5.0, as seen in Figure 33, has raised the confidence of the correct path to virtually 100% with no erroneous paths shown (paths of probability

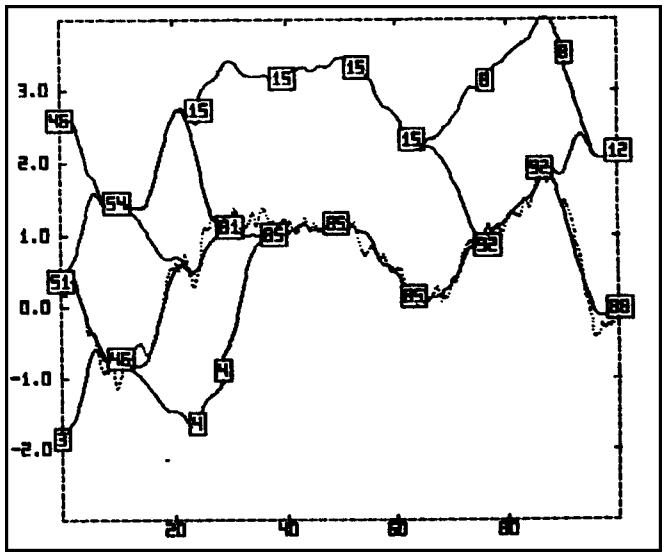

Figure 31: Simulation run identical to Figure 30, except that the signal-to-noise ratio has been cut to 4.

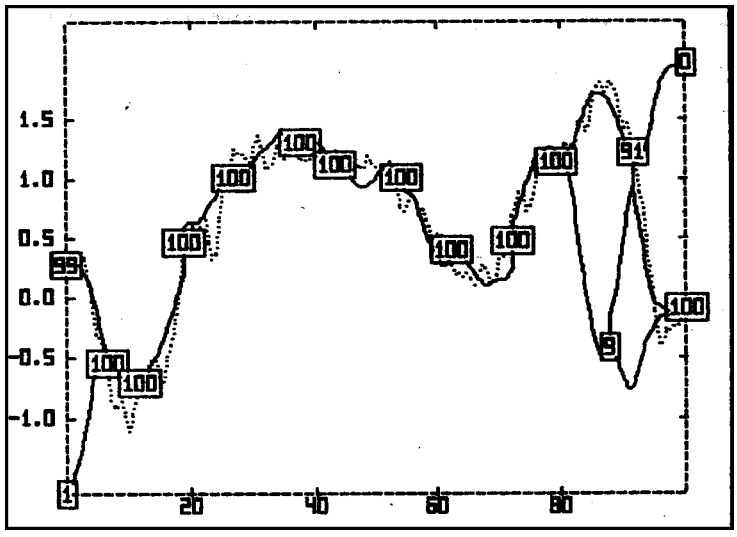

Figure 32: Simulation run on same underlying τ function using **wideband** reception with the same signal level (SNR=4.0) as in Figure 31.

less than about .001 would have been discarded by the algorithm.

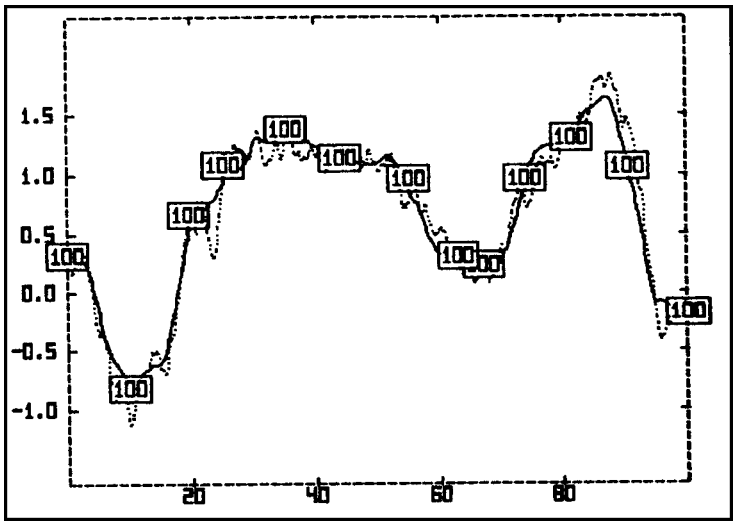

Figure 33: Simulation identical to Figure 32, but in which the signal-to-noise ratio has been **increased** to 5.

On the other hand, the rapid **disintegration** of the *a posteriori* probability mass, as the signal-to-noise ratio is reduced to 3.0 and below is clearly seen in Figure 34 through Figure 39. This is precisely the behavior predicted by (302) and depicted in Figure 26. As the signal level decreases, the resulting estimate becomes almost useless for the purpose of allowing the determination of optical correlation on the basis of the off-line estimate of τ. However it can still be seen that the concentration of paths still appears in the vicinity of $\tau^{(ACT)}$. Thus even at these very marginal signal-to-noise ratios, we maintain the ability to use the **envelope** of the resulting scattered estimation result, in order to perform some form of **medium-coarse** delay-tracking. In other words, if the algorithm shown were capable of operating in real-time, then we would be able to use the

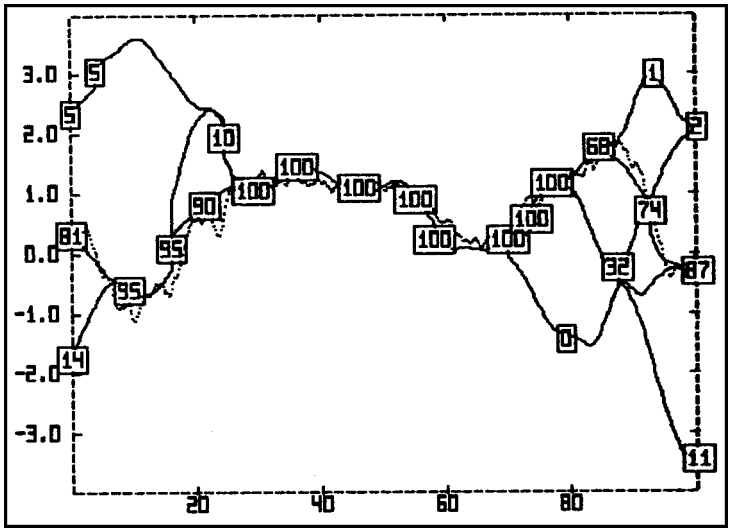

Figure 34: Simulation identical to Figure 32, but with reduced photon level. SNR=3.0.

envelope of the result to enable the delay-line to correct atmospheric delay to within a few wavelengths. Such a use of the global algorithm would probably duplicate the performance of the **group delay** method, which would require only a tiny fraction of the computational power employed to produce similarly useful results. However a performance comparison between the two has not been attempted.

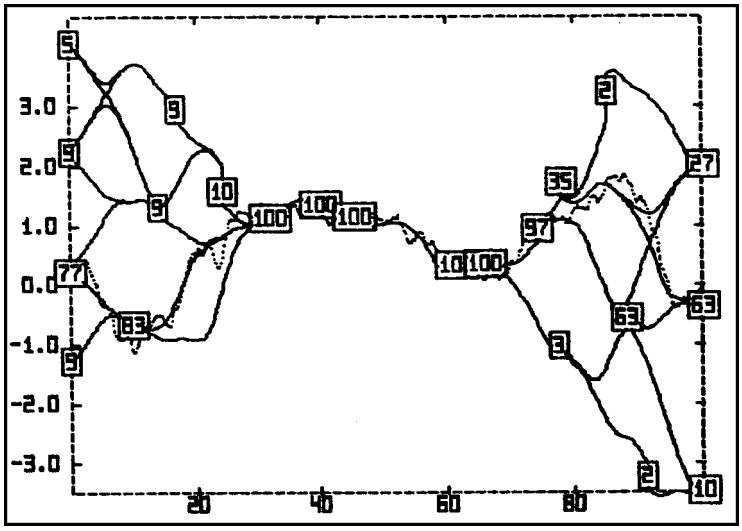

Figure 35: Simulation identical to Figure 32 but with reduced photon level. SNR=2.75.

In Figure 40 the simulation has been run using the **exact logarithm**. The set of photons used is identical to that of Figure 35, and the topology of the resulting solution is similar. However there is a significant increase in the confidence of the correct solution using the exact logarithm (in these examples, $V = 1$, providing the largest potential performance enhancement). Using the exact logarithm, we found, in the discussion of the **local problem**, that the decrease in the estimation error **of the**

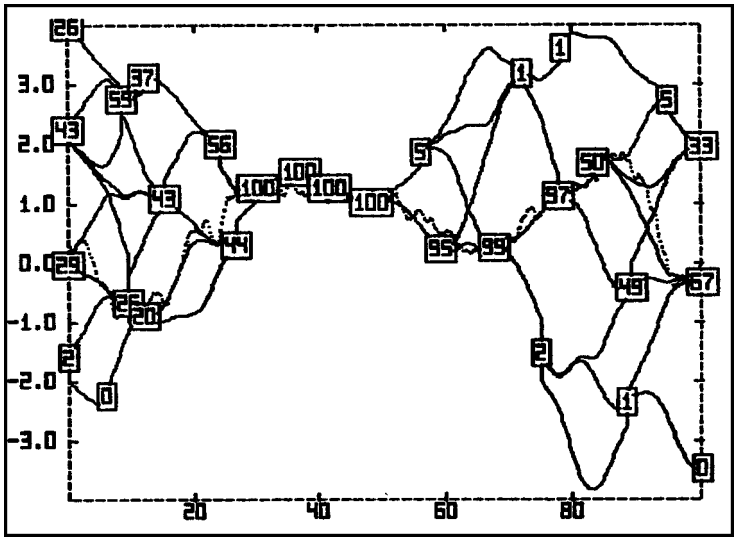

Figure 36: Simulation identical to Figure 32 but with reduced photon level. SNR=2.5.

correct solution was quite modest. However we can see that the increase in the **confidence** enabling **identification** of that correct solution, may make the use of the algorithm employing the exact logarithm advantageous after all.

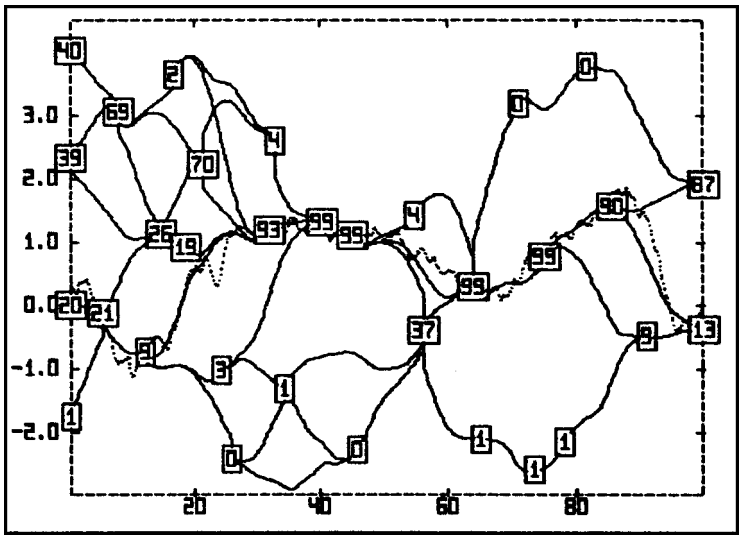

Figure 37: Simulation identical to Figure 32 but with reduced photon level. SNR=2.25.

Topology of Multiple Path Solutions

We would now, without specific reference to the **operation** of the global MAP algorithm itself, like to clarify the meaning and basis of the type of results depicted in Figure 31 – Figure 40. These figures consist of a number of **nodal points** (or "nodes") and a number of **partial paths** (or simply "paths"). The places where the computer drew a box around a number, are the positions of the nodes. These are not part of any path per se, but comprise the **endpoints** of the (partial) paths. The partial paths are just that: paths over a limited range of time.

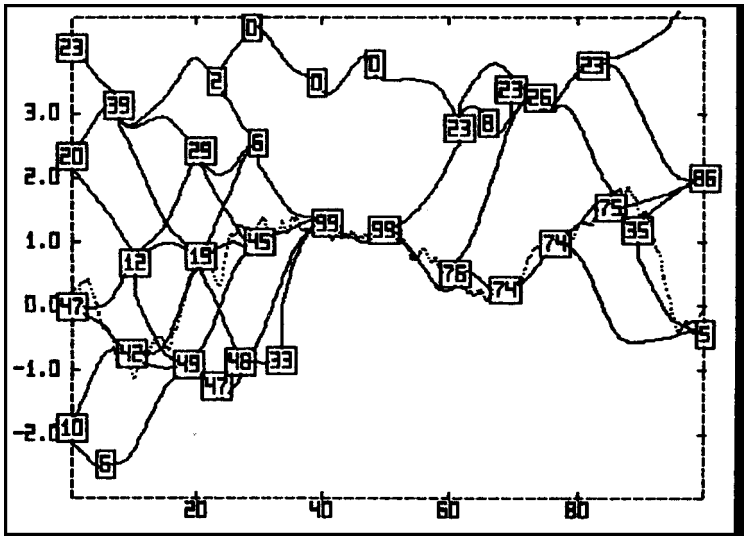

Figure 38: Simulation identical to Figure 32 but with reduced photon level. SNR=2..

That period of time is defined by the nodes to which the path is connected. The **shape** of the path is determined by the maximization of Λ over that interval, as we have discussed regarding the **local problem** (in some cases there may be more than one path connected between the same two nodes, each path being a local maximum of Λ). The values of τ at the endpoints of a path are required to match the τ value of the node, however that should more be viewed as a constraint on the position of the **node** rather than a constraint on the path (although, computationally, there is some element of each).

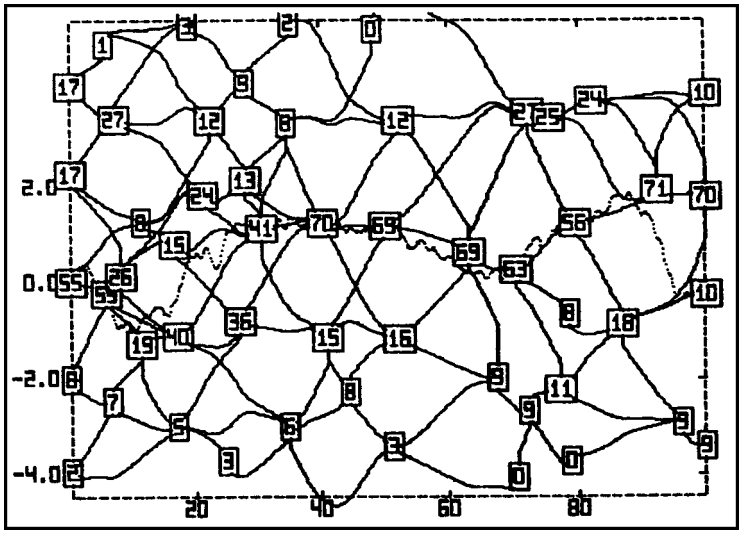

Figure 39: Simulation identical to Figure 32 but with reduced photon level. SNR=1.75.

Now, it is important to understand that these partial paths **do not constitute solutions** to the global problem themselves. In the first place, it's unlikely for two such paths to run over the same exact period of time, thus paths cannot be simply compared to each other on the basis of their individual Λ.

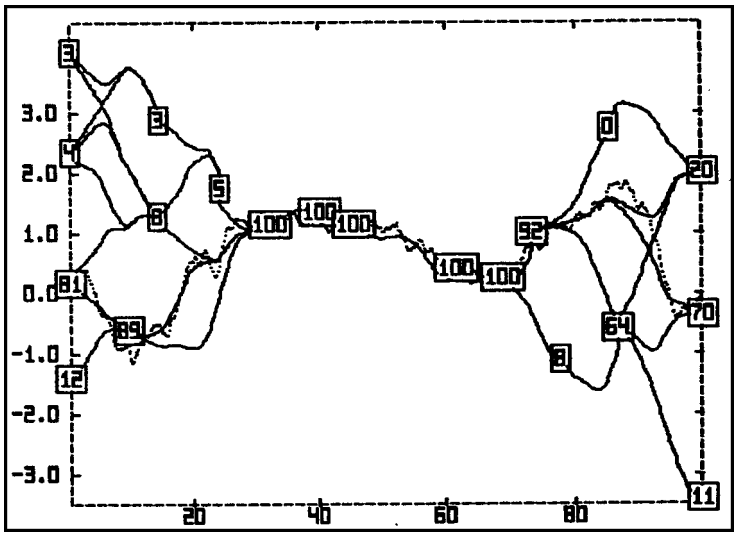

Figure 40: Simulation identical to Figure 35, also with SNR=2.75, but with the estimation algorithm employing the **exact logarithm**.

What's more, a partial path which (by pure chance) happens to have a particularly large Λ, but which is **not** part of a longer structure consisting of partial paths of sufficient Λ, is not part of a likely solution to the global problem. So let us see what constitutes a **total path** solution of the global problem, and how the determination of **partial paths** seen in the above figures, provide a way of describing that solution.

In Figure 41 a set of photons and a solution using the global algorithm has been generated to produce a figure of the type we have already seen, to be used for illustrative purposes[22]. Between the 100% node at $t = 0$ and the 100% node at $t = 27$, there are two partial paths we shall consider (actually, according to the computer, each consists of **two** partial paths with an intervening node; let us ignore the presence of that node). Then between $t = 27$ and $t = 45$, there are, again, two significant paths shown.

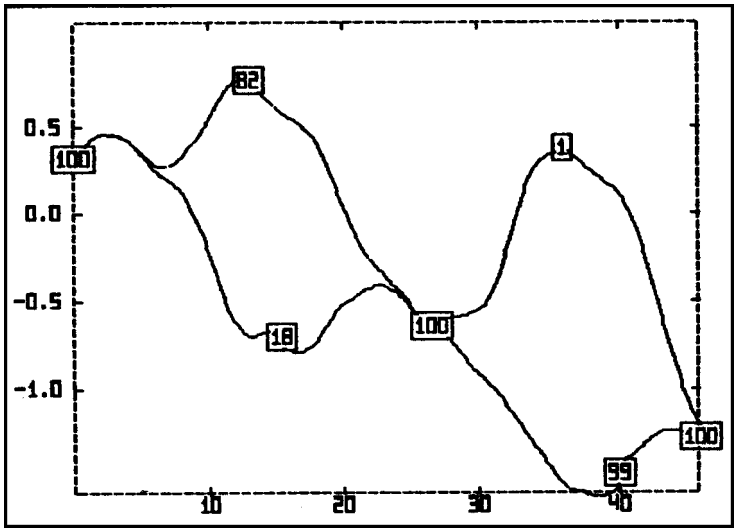

Figure 41: Illustrative example discussed in text regarding the topology of the multiple path solutions to the global problem.

The correct interpretation of such a diagram is as follows. There are altogether four **total paths** indicated by Figure 41.

[22]Figure 41 has been produced by a computer using a set of randomly produced photons, as in the previous examples. **However** the numbers shown in Figure 42 and quoted in the text are somewhat contrived. From the standpoint of the reader, however, the numbers are consistent and typical and serve to illustrate the principles for which this example has been created.

These four paths are shown individually in Figure 42, along with their *a posteriori* probabilities (or **confidences**. As can be seen, the set of **total paths** consists of every possible progression, starting at $t = 0$, from a node to a path starting at that node, to the node ending that path, and so on to another path, continued on until we reach a node at $t = 45$.

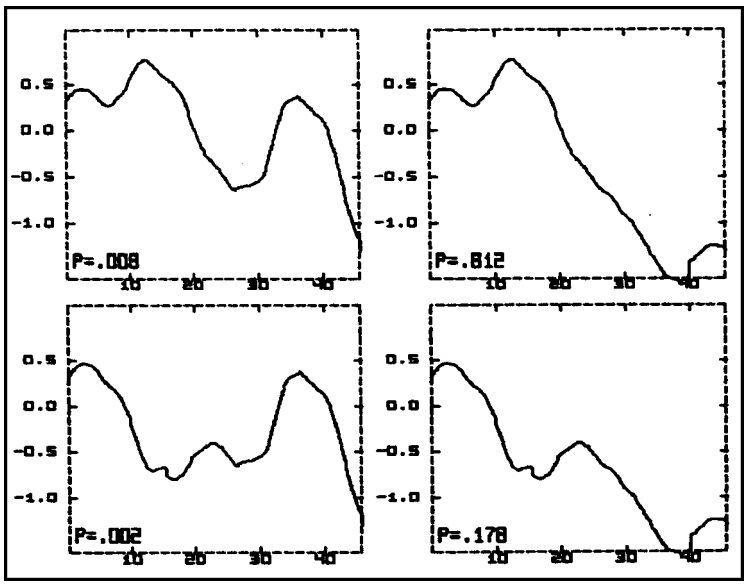

Figure 42: The four **total paths** implicit in the nodal diagram of Figure 41. The probability of each total path is shown, as might be computed using the figures shown in Figure 41.

In the case of Figure 41 there were only two choices for the beginning path and two for the latter path, resulting in 2*2=4 total paths. In a solution of somewhat greater complexity, such multiplications will quickly produce a large number of total paths. For instance, the solution shown in Figure 34 indicates the existence of 55 total paths. Recall that this result is typical for 100 milliseconds of global estimation using typical parameters, for the indicated signal-to-noise ratio (3.0). Since the quantity of total paths derived from non-overlapping partial paths is **multiplicative**, it is not hard to see that the quantity of total paths present in global solutions grows exponentially with the time period considered. Thus it is not only for aesthetic reasons that one would seek to condense the solution in terms of partial paths and nodal points!

The Calculation of Λ for Multiple Path Solutions

There is a particular calculus for the determination of the *a posteriori* probabilities shown in Figure 30 through Figure 42 which is not obvious. What should be understood is that the fundamental determination of the *a posteriori* probability (or its logarithm, Λ) is on the basis of analysis of the **total path**. Such total paths are, for the time period $t = 0$–45, shown in Figure 42. Let us start by ignoring the numbers printed in Figure 41 and Figure 42, and the way in which these illustrations were obtained. Imagine that we had simply been seeking solutions of local maxima of Λ using the iterative procedure depicted in Figure 18. Imagine that we had run the algorithm starting with a variety of initial $\tau(t)$ functions, as suggested in Figure 20, and noted the Λ figures produced by each maximization. Suppose that we found the four most likely solutions to be those shown in Figure 42. These were reported as having: upper left, $\Lambda = 20.09$; upper right, $\Lambda = 24.69$; lower left, $\Lambda = 18.58$; lower right, $\Lambda = 23.18$. Then using (307) we would find $\Lambda^{(\text{full})} = 24.90$, providing the **normalization** necessary to form the *a posteriori* probabilities written at the lower left corner of each solution in Figure 42, as can be verified by applying (307). The sum of these probabilities is, of course, unity.

Now among these four possibilities, the top left and top right of Figure 42 share the same partial path from $t = 0$ to $t = 27$. Then we would say that the **marginal probability** of that partial path is given by the probability of the union of the total paths that include that partial path. Since the events shown in the top left and top right of Figure 42 are disjoint[23], the probability of

[23] At first glance, terming two paths whose values are identical over a long region ($t = 0$–27) as "disjoint" may seem odd. We have to remember, however, that the domain of events to which we

the event consisting of their **union** is given by the sum of the probabilities of the constituent events. In this case we would find the probability of the partial path from $t = 0$–27 to be $p = 0.008 + .812$ or 82%. This, we can see, is te humber printed on that partial path in Figure 41. **This is the meaning of all the probability percentages printed in Figure 30 through Figure 42: the marginal probability of that point, calculated as the sum of the probabilities of all total paths which include that point**.

At any one point in time, t', there will be one or more partial paths whose extend includes t'. The sum of the marginal probabilities of those paths must be unity. Let us take the value of τ for each of those paths at time t', use it as a mean for a Gaussian, and weight that Gaussian by the marginal probability of that path. The union of these weighted Gaussians will form a poly-Gaussian probability distribution for $\tau(t')$ as we created in Figure 27. **That is the resulting estimate for τ given the multiple path solutions such as depicted in Figure 30 through Figure 42.**

[23]are assigning probability consists of the set of all possible functions $\tau(t)$ running from 0 to T. The paths we are drawing on the graphs are each a **mean path**, around which a Gaussian distribution of possible paths is implied. Although a Gaussian density, technically, goes out to infinity, we know that 95% of its probability mass will occur within $\pm 2\sigma$ of the mean. In this case, σ, the r.m.s. distance of a probable path from the mean path, was computed in (275), and it can be seen that 2σ is small compared to the typical spacing between paths, except near to where they converge in on a node. So even if two total paths are coincident over a large stretch of time, their divergence in excess of 2σ over a finite period of time, qualifies them as being **disjoint**.

Another way of looking at this relies directly on the definition of "disjoint." Two events, A and B, are disjoint if the probability of the event "A and B" is zero. Clearly, the probability of τ following path A and following path B is zero (impossible) if A and B are not the same **total path** (even if they are coincident over **a period** of time), unless the mean paths A and B only differ by such a small amount that a sample path could very well belong to either distribution, as in Figure 27 C or D.

For instance, the multiple path solution shown in Figure 43 implies an *a posteriori* density of τ shown in Figure 44, plotted for the full range of t. Note that the probability function depicted as a function of **two** variables is not at all a **joint probability**, $p(\tau, t)$, but must be viewed as a probability in τ only: $p(\tau(t))$. After all, time is not a random variable but an **independent variable**. The integral over τ, at any time t, of the function depicted in Figure 44 is necessarily equal to one.

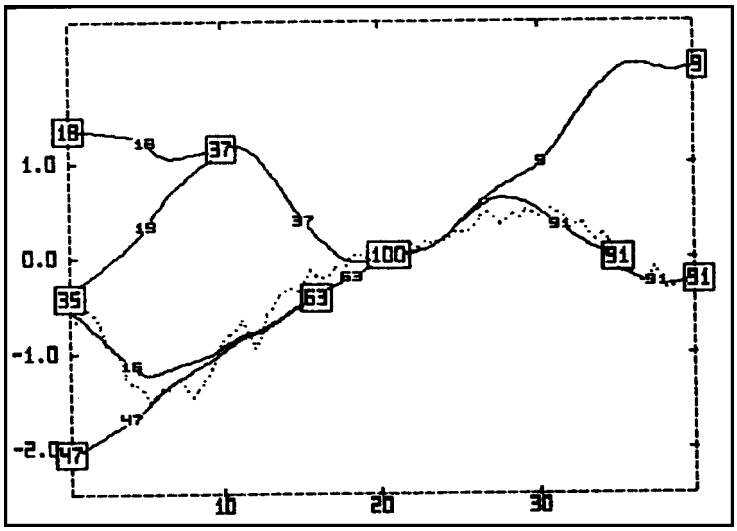

Figure 43: Multiple path plot obtained with SNR=2.5, the probability of which is plotted in Figure 44.

While we have explained the mathematical significance of the probabilities determined in the computer printouts above, the actual process by which the result have been obtained is really **opposite** that described. We conjectured that the marginal probabilities found in Figure 41 **may** have been obtained by a search for MAP solutions over the entire range of time, in which the four total path solutions shown in Figure 42 were then discovered. Such an approach, however, becomes less feasible considering the 55 total paths comprising Figure 34, not to mention the **thousands** of total paths present, for instant, in Figure 39. Remember, also, that these simulations only represent the solution over a time period of .1 seconds; the number of total paths grows exponentially with time. The truth is that only the partial paths were individually determined by the algorithm, and only **from** that set of nodal points and partial paths, was the existence of a larger number

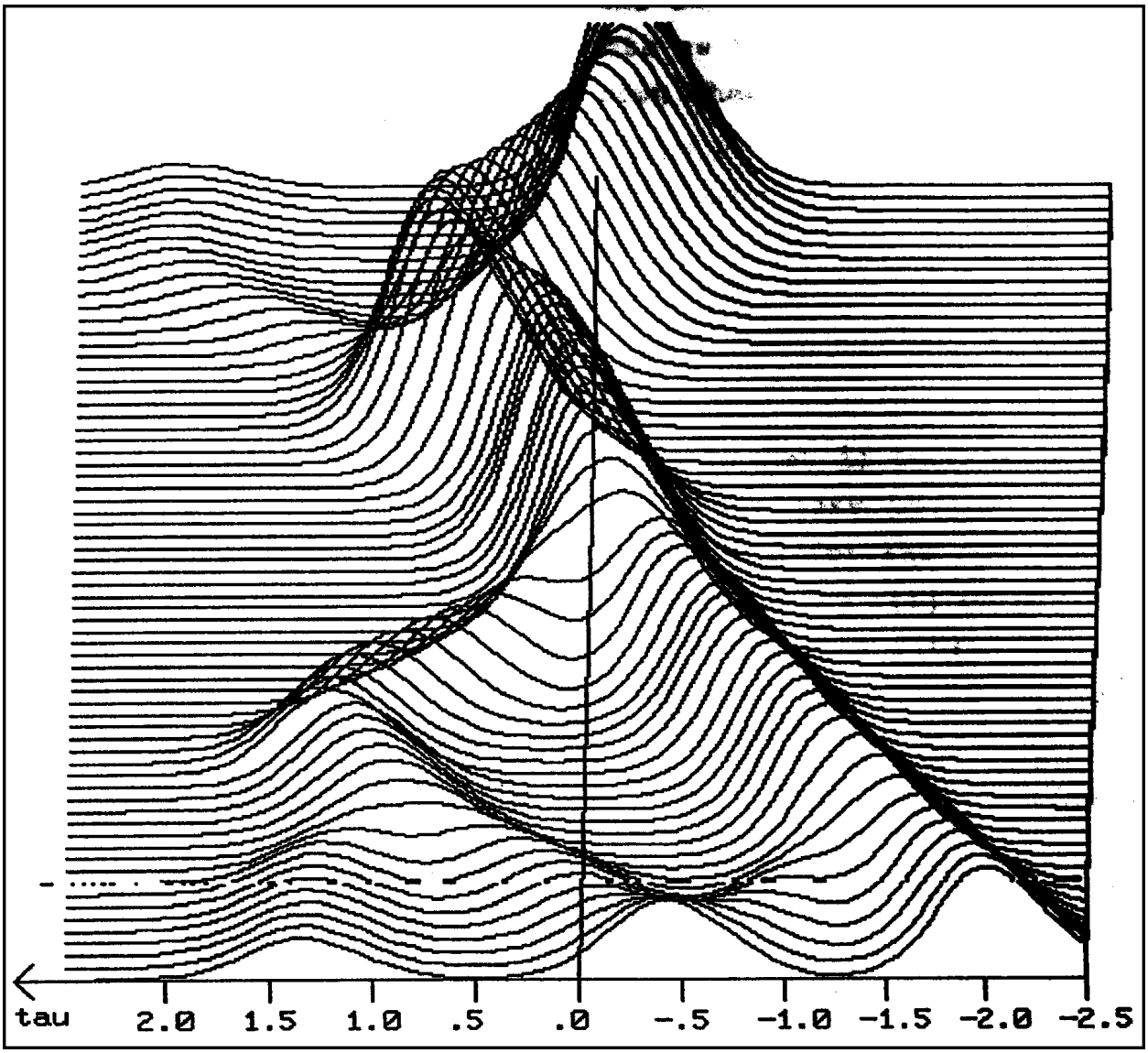

Figure 44: Plot of a posteriori probability vs. τ and t, implied by the multiple path solution shown in Figure 43. May be compared to Figure 15 (using optimum weighting).

of total paths **inferred**. Let us see how that works, and more importantly, why.

Consider the first part of Figure 41, from $t = 0$ to $t = 27$, in which 2 partial paths are depicted. The marginal probabilities of the upper and lower path are listed as 82% and 18%, respectively. In the **second** part of Figure 41, from $t = 27$ to $t = 45$, the marginal probabilities of the two partial paths found were 1% and 99%. If

we consider the choice of a path in the $t = 0$–27 region and the $t = 27$–45 region to be **independent** events, then, for instance, the **total path** shown in the upper left of Figure 42 which combines the upper path of $t = 0$–27 and the upper path of $t = 27$–45, would have a probability given by the **product** of the probabilities of the partial paths[24], or $.82 * .01 = .0082$. And, in fact, that is exactly how the probabilities printed in Figure 42 were arrived at.

The determination of net probability of a **product**, in this case, was predicted upon the characterization of **past** events and **future** events as being **independent**, given that $\tau(27) = -.7$ (which, in this example, was certainty: $p = 100\%$). Let us scrutinize this assumption. Two events are said to be independent if and only if their joint probability is equal to the product of their individual probabilities. In this case we are dealing with *a posteriori* probabilities, and the "intuitive" concept of independence hardly applies. Let us examine the assumption of independence based on the criterion of probability theory.

Consider a nodal point at time t. In the **past**, leading up to time t, there is path #1; in the future is path #2. We have calculated the Λ of each path using (171). We could similarly calculate the Λ of the **joint path**, that is, the path that follows

[24]More generally, the total probability, p_{12}, of a net path consisting of path 1 before time t, and path 2 after time t, both connected to node A at time t, is calculated as follows. We must assume in this case, that the choice of path 1 and the choice of path 2 are characterized as independent events **given** the passage of the solution through node A. Then if p_1 is the marginal probability of path 1 and p_2 is the marginal probability of path 2, then the probability of path 1 **given node A** is p_1/p_A, where p_A is the marginal probability of node A. Likewise the probability of path 2 **given node A** is p_2/p_A. Then, the probability of the **net path** *given node A* is the product $(p_1/p_A) * (p_2/p_A)$. The overall probability is **that** probability **times** the probability of node A itself, thus $p_{12} = p_1 p_2/p_A$. In the example discussed above, the connecting node had a probability of 1, thus simplifying the problem.

path #1 in the past and path #2 in the future. Let us call the Λ of path #1, path #2, and the joint path as $\Lambda(1)$, $\Lambda(2)$, and $\Lambda(1,2)$, respectively. Then we could break down these Λ as follows:

$$\begin{aligned} \Lambda(1) &= \lambda(1) - \frac{1}{2}\chi^2(1) \\ \Lambda(2) &= \lambda(1) - \frac{1}{2}\chi^2(1) \\ \Lambda(1,2) &= \lambda(1,2) - \frac{1}{2}\chi^2(1,2) \end{aligned} \tag{308}$$

Then the condition of independence will be met if we can express $\Lambda(1,2)$ as the sum of $\Lambda(1)$ and $\Lambda(2)$. It is easy to see, by the definition of λ in (170), that the value of λ for the net path is the sum of λ_k for all the photons in path #1 **using** $\tau(t)$ of path #1, **plus** the λ_k for each photon in path #2 using **its** $\tau(t)$. This is equal to the sum of the λ of paths #1 and #2:

$$\lambda(1,2) = \lambda(1) + \lambda(2) \tag{309}$$

This solves half the problem. Now if we could only make a similar statement concerning the χ^2 of paths #1 and #2 summing to that of the net path (plus a constant), then independence would be proved.

However recall that χ^2 is equal to the logarithm of the *a priori* probability of a $\tau(t)$ function (plus a constant). Therefore the condition that we seek, in order to establish the independence of path #1 and path #2, is identical to the determination of the *a priori* independence of a path in the past, and a path in the future, given that they both contain the nodal point $\tau(t)$. In probability theory, a stochastic process whose past and future are independent, **given** its present value, is said to be (first order) **Markov**. Technically, the characterization of $\tau(t)$ we have adopted is **not Markov**.

However it is "almost" Markov. Consider the range of stochastic processes, $x(t)$, whose power spectra are given by a power law of the form $S_{xx}(f) = f^{-\gamma}$. This is the spectrum of so-called $1/f$ noise, especially for γ between 1 and 2. Except for one special case, these processes are not strictly first order Markov. For instance, suppose that we know that $x(t)$ in the immediate past has been **increasing** in value. If $\gamma < 2$, then we would find it **more likely** than not for the **future** of $x(t)$ to reverse direction and decrease in value. However if $\gamma > 2$ (as it is for $\tau(t)$), then we would find it more likely to **continue increasing**. The limiting case, $\gamma = 2$, corresponds to the spectrum of the Weiner-Lévy process (random walk) which **is Markov**. However we can see that the exponent for the characterization of $\tau(t)$ we have used, $\gamma = 8/3$, does not differ too greatly from that of the Weiner-Lévy process, and introducing the **approximation** of independence between the past nd future given the present, will not lead to large numerical inaccuracies. Numerical simulations, in which the Λ of total paths (such as in Figure 42) was determined, and compared to the sum of the Λ of the constituent partial paths (as in Figure 41), finds the approximation of Markovian behavior to be well-founded.

The actual calculations of *a posteriori* probability that are performed are based on the addition of logarithms. The logarithm of the product of the probabilities of the two ostensibly independent events was formed by summing the logarithms of their probabilities. Recall that the logarithm of probability is given by Λ plus a normalizing constant. Thus we found the Λ for the solution in the upper left of Figure 42 by summing Λ for the 82% path in Figure 41 ($\Lambda = 15.20$) and that of the 1% path in Figure 41 ($\Lambda = 4.89$) to find the **total path** $\Lambda = 20.09$. Having found the normalization, $\Lambda^{(\text{full})} = 24.90$, using (307) we can convert this Λ to normalized probability: $e^{(20.09-24.90)} = .008$.

As suggested above, $\Lambda^{(full)}$ may be determined by summing the exponentials of the Λ of **each** total path. We shall show that enumeration of the possibly large number of total paths will not be necessary. In fact, the determination of $\Lambda^{(full)}$ will be shown to be a special case of the summation of disjoint probabilities.

The general rules on which our calculus is based are as follows:

1) The probability of the event including one or more paths in the pas **and** one or more paths in the future is obtained by **multiplying** the past probability and the future probability. Thus *the* Λ *of that event is determined by* **adding** *the past* Λ *and the future* Λ.

2) The probability of the event consisting of the union of several paths converging in on a node, all from the past, or all from the future, is given by the **sum** of their probabilities. Thus *the* Λ *of that inclusive event is found by* **logarithmically summing** *(as in (307)) the* Λ *of the disjoint events..*

3) The probability of the event that includes all possible paths over an interval of time, must be unity. Thus the Λ computed for an event that includes all possible paths over some interval of time, must equal $\Lambda^{(full)}$ for that time interval. Using the generalization of principle (2) above, this implies that the **logarithmic sum** of the Λ of N disjoint events whose union includes all possible paths, must equal $\Lambda^{(full)}$.

The application of those principles, and the basis on which the computer algorithm determined probabilities of the complex solutions we saw earlier, will now be illustrated with an example that is somewhat more general than that of Figure 41. A solution containing 15 total paths is shown in Figure 45 and Figure 46.

However we will never need to examine any such total path: the procedure to be described is based only on assigning numbers to the **nodes** and the **partial paths**.

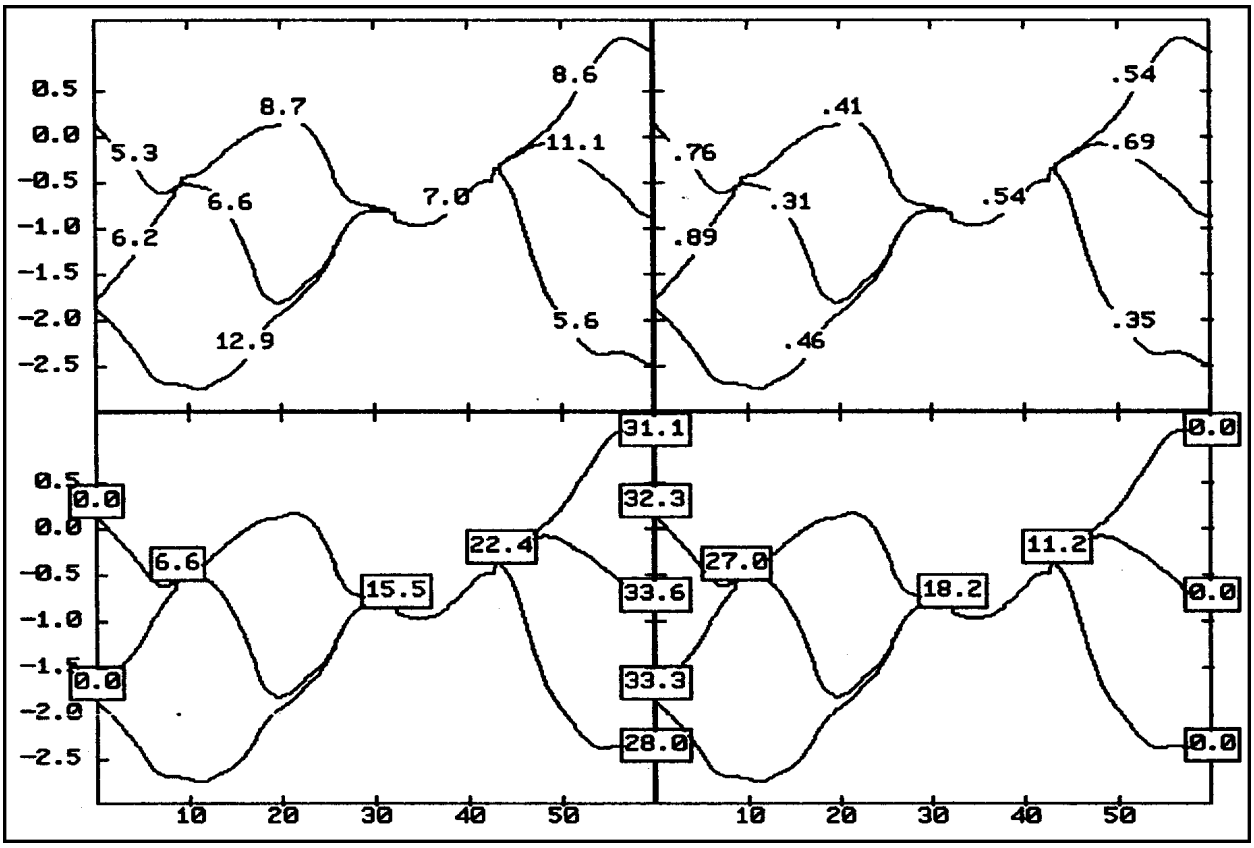

Figure 45: A multiple path solution showing: upper left, Λ of partial paths; upper right, Λ per photon; lower left, Λ^- of each node; lower right, Λ^+ of each node.

Each partial path has been "optimized" using the maximization procedure discussed in relation to the local problem. Over its extent in time, a number of photons have contributed to the λ component of its Λ. Its frequency spectrum has been computed to determine the χ^2 term **penalizing** the resultant value of Λ. That figure, hereafter denoted $\Lambda^{(PATH)}$, is printed, for each partial path, in the upper left of Figure 45. While properly being referred to as a **figure of merit** for each partial path, it is not possible, in most cases, to directly compare the values of Λ of different paths, since they generally contain a different number of

photons (there are two pairs and a triplet of paths in Figure 45 which **do** contain the same photons).

A slightly better comparison figure is the path's Λ divided by the number of photons in the path. In this case ($|V| = 1$) we expect an average value near .5. That figure, for each path, is plotted in the upper right of Figure 45. However it will still prove to be a very inadequate figure of merit. Note that in the time period $t = 10$–30, the top path (of three) shows a figure of $\Lambda = .41$ per photon, whereas that figure for the bottom path is .46. However, even with the lower figure of merit, the top path is over 10 times **more likely** than the bottom path. **That** determination, however, is only made on the basis of an integrated analysis in which the paths' Λ values (in the upper left of Figure 45) are considered in the context of the topology of nodes and partial paths on the basis of which 15 **total paths** are defined.

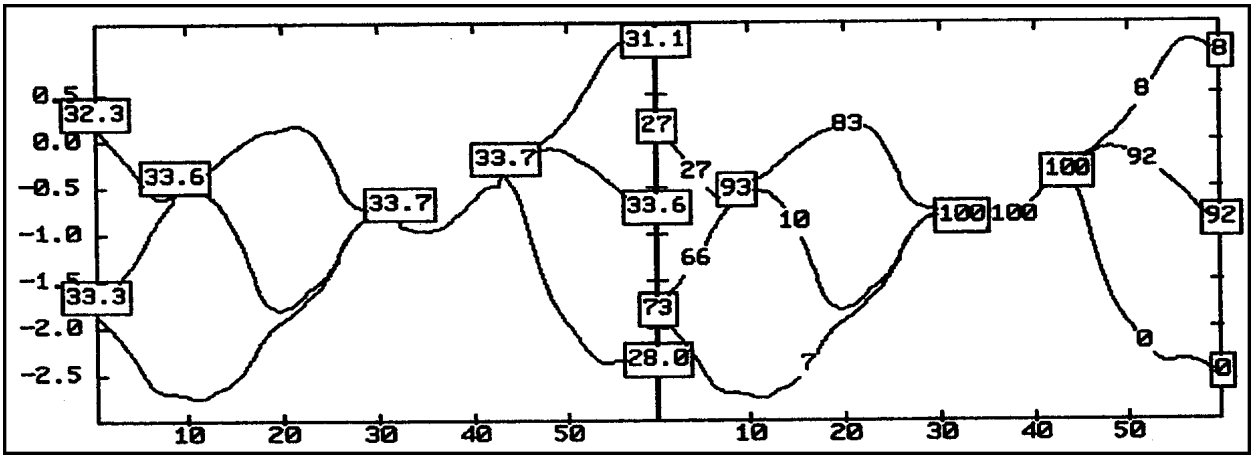

Figure 46: Graphs of the multiple path solution of Figure 45: left, **net** Λ of each node; right, the resultant net probabilities of each node and each partial path.

We shall now define two numbers for each node. Λ^+ is called the **future** Λ; Λ^- is the **past** Λ. Consider the two higher paths from $t = 0$ to $t = 9$, both of which converge on the same node at $t = 9$, $\tau = -.5$. The **past** of that node is the **logarithmic sum** of the past Λ of

each path connecting to the node from the past. Each of those is defined as the Λ of the path **plus** (multiplying the probability) the **past** Λ of the node in the past of the path. Since these paths both begin at $t = 0$, we arbitrarily label that past Λ as zero. Thus we form the logarithmic sum of these paths' Λ:

$$\Lambda^- = \log \sum_{\substack{all \\ past \\ paths}} e^{(\Lambda^{(\mathrm{PATH}} + \Lambda^-(\text{previous node}))}$$
$$= \log(e^{5.3} + e^{6.2}) = 6.6 \tag{310}$$

That figure is plotted, for that node, in the lower left of Figure 45. Likewise the two paths connecting **that** node to the node at $t = 32$, $\tau = -.8$, each contribute to the Λ^- of that node, $6.6 + 8.7 = 15.3$ and $6.6 + 6.6 = 13.2$. Additionally the very lowest path contributes $12.9 + 0$ toward the Λ^- of that node. Summing these number logarithmically yields a net $\Lambda^- = 15.5$ for that node, and so on. While none of these numbers indicates probability in any absolute sense, the following interpretation may be appreciated. **Given** that $\tau(t)$ passes through the node at $t = 32$ (which in this **particular** case, happens to be a certainty), then the probabilities of the past τ including the upper, middle, or lower paths leading to that node, are given by $e^{15.3-15.5}$, $e^{13.2-15.15}$, and $e^{12.9-15.5}$, respectively.

Now, the Λ^+ assigned to each node refers to the (logarithmic) sum of the net paths' Λ looking into the **future**. For instance, the three paths between $t = 42$ and $t = 50$, whose Λ are 8.6, 11.1, and 5.6 (as seen in the upper left of Figure 45) logarithmically sum to 11.2 as plotted in the lower right of Figure 45.

Now, the **net** Λ denoted $\Lambda^{(\mathrm{NODE})}$ for each node, equal to the logarithm of the (unnormalized) probability of $\tau(t)$ passing through that node, is proportional to the probability of all paths in the past (which end in that node) **times** the probability of all paths in the future (which start at that node). Thus we **add** Λ^- and Λ^+ for that node to form the net $\Lambda^{(\mathrm{NODE})}$, plotted in the left side of Figure 46. With the normalization $\Lambda^{(\mathrm{full})} = 33.65$, these numbers can be converted to probability using (307), as seen in the right side of Figure 46. This figure also shows probabilities reflecting the **net** Λ of the **paths** (not to be confused with the path Λ shown in the upper left of Figure 45). The net Λ of a path is computed as the path Λ **plus** the Λ^- of its past node **plus** the Λ^+ of its future node.

Note that $\Lambda^{(\mathrm{full})}$ can be found logarithmically summing the Λ^+ of all nodes at $t = 0$. The same figure would be obtained by logarithmically summing the Λ^- of all nodes at $t = 60$. It can also, in the case of Figure 46, be directly read as the value of $\Lambda^{(\mathrm{NODE})}$ at any node that includes all possible total paths, such as the node at $t = 31$ or the node at $t = 43$.

Description of the Algorithm

We will now briefly describe the operation of an algorithm which has been successful in searching for the multiple MAP solutions displayed in this chapter. No claim is made regarding the computational efficiency of the algorithm to be described; such concerns are not within the scope of this work, and there is little doubt that substantial improvements in the computational aspect of the τ estimation problem could be achieved in future work.

The key to the ability of the algorithm to find all MAP solutions, rests upon initially considering a dense combination of

partial paths, so that any ultimate solution is certain to have at least one such initial path in the tails of its Gaussian probability distribution. The initial set of partial paths consists of a dense grid pattern of nodes and connecting paths as shown in Figure 47. The **coverage** of the range of possible solutions achieved by the initial set of paths is critical. If the grid spacing in Figure 47 had been doubled, for instance, then it is likely that a partial path that **should** be contained in the final solution would never be found. Instead **weaker** MAP solutions would be found which, after the determination of $\Lambda^{(\text{full})}$, would of course have overall probabilities just as high (summing to unity!), but would be blissfully unaware of the potential Λ of the paths we **missed**.

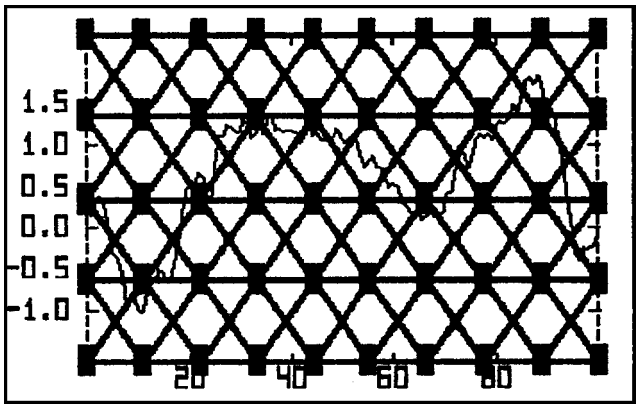

Figure 47: The initial set of nodes and paths from which the global algorithm proceeded to generate the solution stages of which are shown in Figure 48 – Figure 52.

We shall follow the operation of the algorithm in the determination of the global solution using the exact set of photons as in Figure 32. The underlying $\tau(t)$ function is plotted underneath the initial paths in Figure 47. The process begins with the "processing" of each initial path shown in Figure 47. That "processing" simply amounts to running several iterations of the Λ maximization algorithm discussed under the "local problem," using as a starting path each of the partial paths shown in Figure 47. Each path is tied at its initial and final points to two specific nodes. The initial and end point values of the path are **weakly** constrained by the τ values of these two nodes. Each resultant path is evaluated to determine its $\Lambda^{(\text{PATH})}$, and the path itself, $\tau(t)$, is stored. Also stored, is the extent to which the path

"wanted" the initial and final nodes, to which its value at those points was weakly constrained, to have higher or lower τ values.

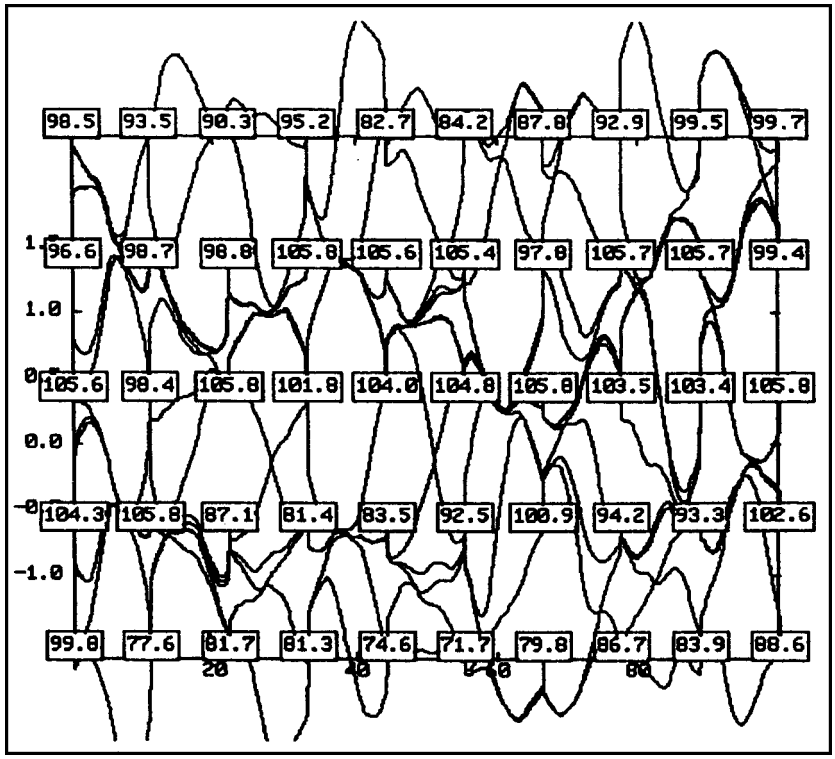

Figure 48: Result of processing each path shown in Figure 47 to maximize the Λ of each path. The resultant net Λ is written in the box of each node.

The set of paths following processing is shown in Figure 48. Because of the good initial coverage we established, there are often two or even three paths of the algorithm which have each found substantially the same path to maximize Λ, even though they are tied to different nodes. So although it may not be apparent at first sight, there is the same number of paths in Figure 48 as we had in Figure 47, tied to the same nodes to which they had been tied in Figure 47.

Following the processing of the paths to maximize each $\Lambda^{(\mathrm{PATH})}$, the algorithm used the determined values of $\Lambda^{(\mathrm{PATH})}$ to determine the Λ^{+} and Λ^{-} for each node, using the mathematics described earlier. Similarly, $\Lambda^{(\mathrm{full})}$ was calculated, and found to be 105.82. The **net** Λ for each node, determined as $\Lambda^{(\mathrm{NODE})} = \Lambda^{-} + \Lambda^{+}$, is plotted on each node in Figure 48.

Following the **processing** of the paths and the **calculation of** Λ, comes the process of "tugging on the nodes." Using data that was accumulated during the "processing" phase, the τ position of nodes is adjusted in response to the consensus (if any) of the "tugs" applied to it by the paths to which it is connected. In the not uncommon case that two future paths (or two past paths) are "tugging" in **opposite directions**, then the response is to shift the time t of that node **away** from the conflicting paths, thus giving them a chance to converge. Of course in the worst case there will be paths tugging τ up **and** down from both the past **and** the future. The tugs of these paths will have to go unanswered.

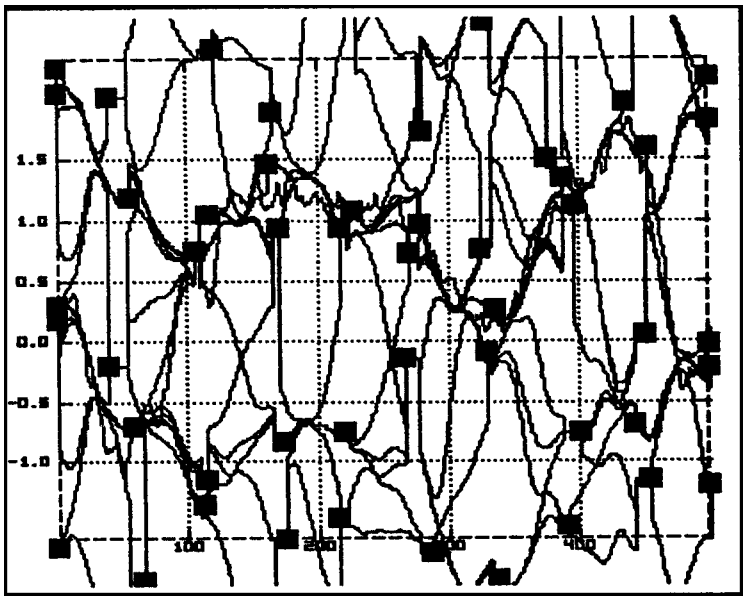

Figure 49: The same set of paths as in Figure 48 after the positions of the nodes have been tugged to the preferred positions. The underlying τ function is visible in the background.

The new positions of the nodes following the "tugging" process, are shown in Figure 49. Next, the large quantity of paths is sized down by eliminating paths whose **net** Λ fails to meet a criterion specified as $\Lambda^{(full)}$ minus 12. Why 12? After all, we will not be wishing to wind up with solutions of probability less than .001, which corresponds to $\Lambda^{(full)} - 6.9$. However **at this stage of the processing** we have to be very careful not to discard paths which **appear** to be improbable, but which may, after further processing, prove to have been worth keeping. Also, it should be noted that the $\Lambda^{(full)}$ that has been calculated, is based on a significant number of, what amounts to, **duplicate paths** (although

not connected to the same two nodes) which has exaggerated the calculated value of the total probability. For instance, two identical paths would, if undetected, increase the Λ of the total paths containing them by log(2) = .69. So we proceed cautiously in setting thresholds which will cause the discarding of paths that may never be recovered! Still, a great deal of clutter has been eliminated following the "trimming" of paths not meeting the above threshold, as seen in Figure 50.

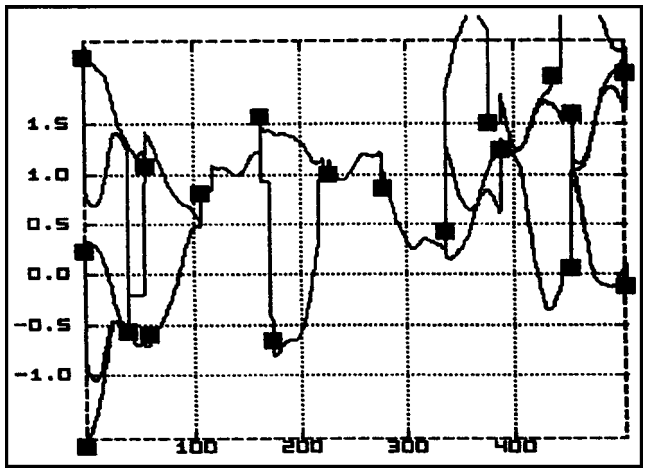

Figure 50: The result of "trimming" paths and nodes of Figure 49 not meeting a criterion described in the text. Nearby nodes have been merged and redundant paths have been eliminated.

Eliminating weak paths will often leave nodes which have no paths in the past or no paths in the future. Such nodes include no possible total paths; their $\Lambda^{(\mathrm{NODE})} = -\infty$ and are destroyed by the algorithm. Any paths connected to the **other** side of a such a node are **also** impossible, and are eliminated.

The next operation consists of the **merging** of nearby nodes. If two nodes have been "tugged" to positions very close to each other, then they are effectively redundant (except for the congregation of paths to which each is connected). One node is destroyed, and the paths connected to it are re-linked to the surviving node.

The merging of nodes invariably creates **redundant paths**. That is because a local MAP solution, especially a strong solution, would generally "attract" more than one nearby path. Those paths were unique only because they were not connected to the same two

nodes. After merging nearby nodes, many of these paths could be identified as redundant, due to the closeness of their $\tau(t)$ for **most** t. Based upon circumstantial criteria, it can be inferred that the maximization of Λ for each of these paths would almost surely lead to the same $\tau(t)$. Pairs of paths thus deemed redundant have one member destroyed.

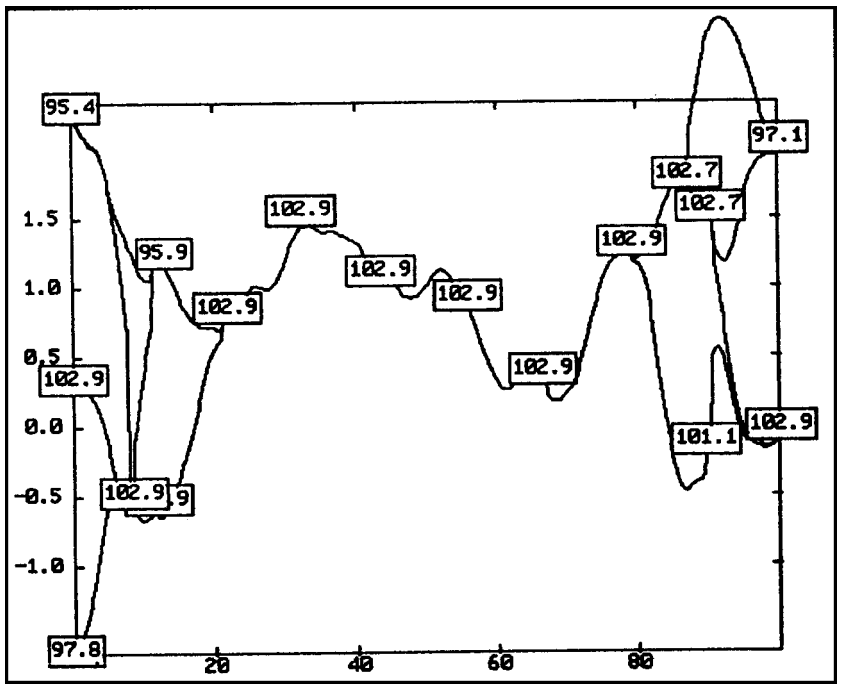

Figure 51: The set of surviving nodes and paths following the **second** overall iteration of the algorithm described in the test. Value of each net node Λ are shown inside the boxes at each node.

Finally, the entire process described above, is repeated a total of 3 to 6 times. The result of the second overall iteration for our example is shown in Figure 51 with the net Λ of each node shown in the boxes. With each new iteration, parameters are adjusted. For instance, the exactness to which the maximization of Λ for each local path proceeds, is increased each time. That reflects the need to more accurately quantify the likelihood of partial paths as we home in on the total solution. Also the criterion for eliminating "unlikely" paths is adjusted to destroy more paths. That is something that we can afford to do when we are nearer the final solution and the determination of the eventual values of Λ is more certain. In addition, the parameters allowing the "tugging" of nodes in τ or t are decreased so as to eventually **fix** the positions of all nodes for the final iterations. Otherwise

there could be interactions between the processing of paths and the tugging on nodes, causing a non-convergent "dithering" around the solution.

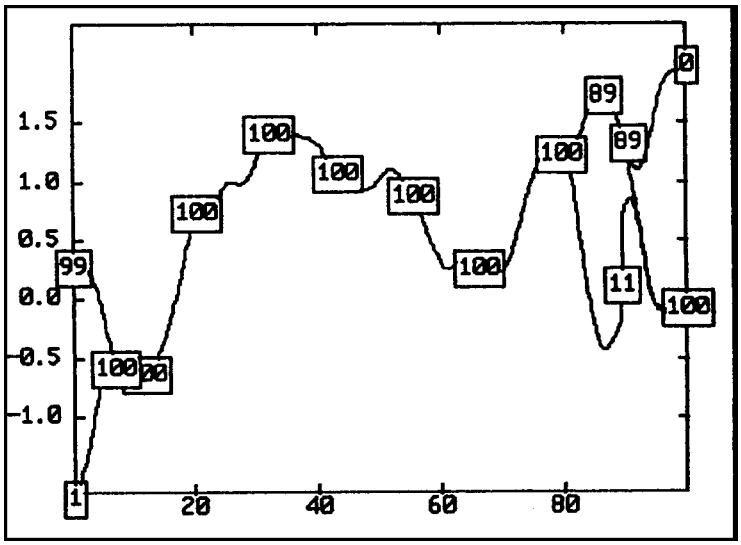

Figure 52: The final solution determined following 6 iterations of the global algorithm. Marginal probabilities in percent are written inside boxes at each node.

After 6 iterations of the total process described above, the final multiple path solution is shown in Figure 52, with net probabilities in percent plotted at each node. Comparison of Figure 52 to Figure 32 (which used the exact same set of photons) shows strong similarity but **not** exact replication of the resultant solution. The discrepancies illustrate certain sensitivities of the evaluation processes of the algorithm to, what should be, irrelevant considerations, such as the exact position in time of a node. Discrepancies in results from the same data set may also reflect sensitivity to changes in the initial path structure (Figure 47), propagating irrelevant attributes such as the number of nodes within a total path. However it is generally found that the identification of the **underlying** $\tau(t)$ function as the dominant solution (at sufficient signal-to-noise ratios) is relatively unaffected by such idiosyncracies of the present implementation of the algorithm.

Conclusion

We have seen that a long-baseline optical stellar interferometer may be operated in one of three modes as regards the estimation of differential atmospheric delay for the purpose of integrating measurements of optical correlation. An **incoherently averaged** interferometer requires only **coarse** delay-tracking in order to allow interference to be measured over a non-zero bandwidth. Reducing the maximum untracked delay in such a case will enable operation at wider bandwidths per detector. However even with light levels so low that no delay-tracking feedback can be produced, operation is still possible using detector bandwidths sufficiently small so that the maximum excursions of atmospheric delay will not destroy the detection of interference.

Real-time tracking of the atmospheric delay to well within an optical wavelength has a number of advantages over incoherently averaged interferometry. In this case, a feedback signal derived from the optical detection, controls a rapidly responding optical delay-line in order to cancel the differential delay as closely as possible. Keeping the physical path delay to within a fraction of a wavelength relaxes the bandwidth requirements on the optical detection system. This is a strong advantage in the infrared. Additionally, having provided a stabilized phase, the measurement of optical correlation, may now be integrated **coherently**, enabling the estimation of that quantity using much shorter observation times. Simultaneous coherent integration of interference using separate detection of **multiple wavelengths**, enables the measurement of the **phase** of the optical correlation function, information that would be lost in a narrowband system. Also, the identification of the white-light fringe in conjunction with a laser-monitored delay-line, makes possible astrometric measurements limited only by the

diffraction limit of the synthetic aperture defined by the instrument's baseline(s).

Unfortunately at signal levels below a certain point, it becomes more and more difficult for a real-time delay-tracker to follow the white-light fringe that defines zero delay error. There will inevitably be periods in which tracking fails, or in which the control system inadvertently locks into a sidelobe of the central fringe. In this tenuous regime there is still enough information in the correlation of the received light to enable some reasonable estimation of the atmospheric delay process. However that cannot necessarily be accomplished in real-time. Instead it is necessary to base the estimation of delay at a time *t*, on measurements taken not only before, but **after** time *t*. The detection hardware must also be capable of tolerating delay errors of at least several wavelengths, since the instantaneous untracked delay error cannot be guaranteed to stay within a small amount, as in the previous configuration. This then requires a **spectrally dispersed** detection system.

Operation as such, in which there is only **coarse** tracking of the hardware delay-line, but the ability to eventually produce a close estimate (within a fraction of a wavelength) of the delay function for the purpose of coherent integration of optical correlation (as would have been achieved using the real-time delay-tracking system), is termed **off-line delay-tracking**. Employment of such a system is advocated on the basis of its performance at light levels below which real-time tracking would be possible. The hardware requirements for the off-line delay-tracking system include the implementation of spectrally-dispersed detection, and the availability of a data processing system capable of performing thousands of arithmetic operations for every received photon (at the lower light levels). However if that computation cannot be performed at the rate that the data is acquired, it will still be

possible to employ the eventual results of that computation in order to extract information from the raw data stream which as been stored for off-line processing.

We analyzed the operation of **point estimation** techniques, in which the problem considers only the estimation of the differential path delay function $\tau(t)$, evaluated at a particular time, t'. Even employing a weighting scheme (see (147) and (148)) which is shown to be optimum (given a particular model of atmospheric path delay), the procedure still relies on an approximation (144), and does not necessarily produce a set of estimates of τ at different points in time which are mutually consistent relative to the expected large-scale structure of the $\tau(t)$ function. A superior estimate can be achieved using the **path estimation** procedure. In path estimation, the domain of solutions consists not of $\tau(t')$, the differential delay function evaluated at a particular point in time, but rather the set of possible **functions** $\tau(t)$ over a **range** of time, $0 < t < T$.

Path estimation involves the search for paths, $\tau(t)$, based upon a figure of merit, Λ. Λ is simply defined (159) as the logarithm of unnormalized *a posteriori* probability. Thus by maximizing the value of Λ, we find solutions which meet the **MAP**, or **Maximum A Posteriory Probability** criterion. There are two aspects to path estimation. The **local problem** involves the maximization of Λ starting from one τ function and leading to the **closest** τ function at which there is a **local** maximum of Λ. The **global problem** addresses the global search for one or more maxima of Λ in order to account for nearly all of the *a posteriori* probability mass described by $k_1 e^{\Lambda}$ where k_1 is an unknown normalization constant.

There are two ingredients to the determination of an *a posteriori* probability density according to Bayes' theorem (158). One ingredient is the **likelihood** function, given by the probability

of the instrument having produced a certain data set **given** a function $\tau(t)$. The **logarithm** of the likelihood (170) is simply denoted λ. The other component contributing to the determination of the *a posteriori* probability density, or its logarithm Λ, is the *a priori* probability of $\tau(t)$. $\tau(t)$ has been characterized as Gaussian colored noise, with a power spectrum following the $-8/3$ power of frequency. The logarithm of its (unnormalized) *a priori* probability is denoted $-\chi^2/2$. Adding this quantity to λ, the logarithm of the likelihood function, yields Λ. The determination (162) of χ^2 is based upon the Fourier decomposition of τ using the cosine transform (53), and the expected energy contained in the components, U_i, of the cosine transform, given by (69).

Convergence toward the local MAP estimate, denoted $\tau^{(\mathrm{MAP})}$, and analysis of the maximization process, is based upon Λ being approximated as a quadratic function in the vicinity of its maximum. Characterizing Λ, the **logarithm** of *a posteriori* probability, as a quadratic function, describes a Gaussian probability density function. Determination of the first and second derivatives of Λ, found in (181)–(183), in principle allows the solution of the maximization problem, through the inversion of a large matrix. However setting the off-diagonal elements of that matrix to zero, and replacement of the diagonal elements with the **expected values** of those numbers, will result in a far simpler procedure (186) which, if performed iteratively, will **converge** toward the MAP solution as shown in Figure 18. Although this procedure was performed on the Fourier coefficients of the function $\tau(t)$, it was shown that the identical result could be achieved using an algorithm in the time-domain (Figure 21) in which a properly designed time-domain filter (208) is employed.

The expected r.m.s. error of this estimation procedure can be determined analytically. The coefficients, A_{ij}, which describe the quadratic expansion of Λ around its peak (172), are assembled into

a matrix A. The covariance matrix applying to the *a posteriori* density of τ, is given by the **inverse** of the matrix A. Some statistics of the elements of A have been determined in (238) – (240). Using these determinations, and a formula derived in Appendix I that applies to the inversion of a matrix of a particular form, the expected values of the diagonal elements of the covariance matrix R, are estimated in (248). Using Parseval's theorem (215), the magnitude of the variances of the frequency components of τ, given by the diagonal elements of R will yields the r.m.s. estimation of error of τ itself. The resultant formula (275) for the estimation error, shows that the r.m.s. error approximately follows the $-5/8$ power of signal-to-noise ratio. Since, when using low-noise detectors, the signal-to-noise ratio (233) follows the **square root** of photon flux, we find that the estimation error will follow the $-5/16$ power of the optical signal level. Simulation results plotted in Figure 22, Figure 23, and Figure 25, clearly verify the performance predicted by theory.

The character of the **total** MAP solution, consisting of a number of possible paths, has been investigated in simulations addressing the global problem. The enumeration of the multitude of paths that would be present over an extended time period, has been simplified by the invention of a topology of **partial paths** connected in between **nodal points**. Thanks to the "almost Markov" nature of the statistical characterization of τ, it is possible to approximate the unnormalized (*a posteriori*) probability of any total path as the **product** of the probabilities of its constituent partial paths. These probabilities, again, are expressed in terms of their logarithms Λ, which are therefore produced as the **sums** of the $\Lambda^{(\mathrm{PATH})}$ corresponding to each constituent partial path. Through a procedure in which the probabilities of disjoint paths connecting to a node from one direction are **summed**, the evaluation of the **marginal** probabilities assigned to a node or a partial path can be evaluated. Using the same principle, the sum of probabilities

of **all** disjoint paths can be determined, providing the normalizing factor by which the unnormalized probabilities can be interpreted.

As viewed at a particular point in time, then, we can form a marginal density which consists of a set of N estimates of τ for that point in time based upon the N partial paths that exist at that point in time. Each of the N possibilities is characterized as a Gaussian probability distribution for τ with a mean, a variance (that is approximately the same for all possibilities), and a confidence (the probability of that Gaussian applying). Given such a characterization of the multiple path solution, it is possible, at least in principle, to use that estimate, albeit ambiguous, to coherently integrate the optical correlation that we have sought to measure. However the estimation of coherence may be less prone to error, if it is based only on the fortuitous time segments in which there is near 100% confidence in a partial path. Even for signal-to-noise ratios as low as 3, simulations have found that periods of high confidence will occur with sufficient frequency for correlation to be integrated in this manner.

Using our algorithm to search for and quantify the multiple path solution of the global path estimation problem, it has been observed that as the signal-to-noise ratio falls below 3 or 4, there is rapid disintegration of the *a posteriori* probability density as can be seen in Figure 32 through Figure 39. One quantification of this phenomenon is possible by identifying the **confidence** of the correct solution. An analysis has been performed which places an upper limit on the expected value of the normalization constant k_1 applicable to the *a posteriori* probability $k_1 e^{\Lambda}$ of multiple solutions. For signal-to-noise ratios **below** a certain level denoted $\mathrm{SNR}^{(\mathrm{CRIT})}$ (approximately equal to 3), the analysis predicts a rapidly decreasing confidence for the correct solution. Comparison of that prediction to simulation results is plotted in Figure 26.

Appendix I:

Determination of the Statistics of a Covariance Matrix

Consider a jointly Gaussian random vector U of dimension M with a mean given by μ. Then the probability density of U is given by:

$$P(\vec{U}) = \frac{\det[A]}{(2\pi)^{M/2}} e^{-\frac{1}{2}(\vec{U}-\vec{\mu})^T A(\vec{U}-\vec{\mu})} \tag{311}$$

where A is the inverse of the covariance matrix R.

As is well known, A is a symmetric positive-definite matrix, as is R. We would like to determine two expectations, given only statistical knowledge concerning A. First we would like to estimate the determinant of A. Secondly, we would like to determine the expected value of the diagonal elements of R, the covariance matrix given by the inverse of A.

To start with, we shall multiply the rows and columns of A by a suitable diagonal matrix to obtain a matrix a whose diagonal elements are unity:

$$a = \begin{Bmatrix} 1 & a_{12} & a_{13} & \cdots \\ a_{21} & 1 & a_{31} & \cdots \\ a_{31} & a_{32} & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{Bmatrix} \tag{312}$$

A is then related to a by:

$$A = KaK \tag{313}$$

where:

$$K = \begin{Bmatrix} \sqrt{A_{11}} & 0 & 0 & \dots \\ 0 & \sqrt{A_{22}} & 0 & \dots \\ 0 & 0 & \sqrt{A_{33}} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{Bmatrix} \tag{314}$$

Then the determinant of A will clearly be related to the determinant of a according to:

$$\det[A] = \det[a] \cdot \prod_{i=1}^{M} A_{ii} \tag{315}$$

If we are able to determine the diagonal elements of the inverse of a, then finding the diagonal elements of R, the inverse of A, is trivial:

$$R_{ii} = A_{ii}^{-1} = \frac{a_{ii}^{-1}}{A_{ii}} \tag{316}$$

Now, let us make the following restrictive assumptions concerning A and a which will make the subsequent analysis possible. Although the diagonal elements of A are random, we will assume that their expectations have been determined and that their variances are relatively small. This is important, since division by A_{ii} in (316) is equivalent to multiplication by the reciprocal of a random variable. The expected value of a reciprocal is the

reciprocal of the expected value only under the condition of a small variance.

Secondly we will wish to restrict the off-diagonal elements of A. The off-diagonal elements of a are necessarily less than unity since a is positive-definite; in fact they will be much less than one. Suppose they are the result of a random process and can be approximately characterized as **zero-mean** Gaussian random variables whose variance takes the special form:

$$Var\{a_{ij}\} = E\{a_{ij}^2\} = g_i g_j \tag{317}$$

where g_i is some function of i. Furthermore, we must insist that the a_{ij} are uncorrelated. This condition on the off-diagonal elements of a is equivalent to the following condition on the off-diagonal elements of A.

$$Var\{a_{ij}\} = \frac{Var\{A_{ij}\}}{A_{ii}A_{jj}} = g_i g_j$$
$$\therefore Var\{A_{ij}\} \approx E\{A_{ii}\}E\{A_{jj}\}g_i g_j \tag{318}$$

So if we can supply the expected values of the diagonal elements of A with relatively small variances, and choose a function g_i such that the off-diagonal elements of A are characterized as uncorrelated zero-mean random variables with variances given by (318), then the following analysis is applicable.

The determinant of the matrix a is the sum of $M!$ terms:

$$\det[a] = \sum_{i_1, i_2, \ldots i_M} \pm a_{1 i_1} a_{2 i_2} \ldots a_{M i_M} \tag{319}$$

where the sum is taken over all sets of indices i_1 through i_m which are permutations of 1, 2, 3 ...M. The + or − is selected depending on whether that is an even or odd permutation, respectively. The expected value of the determinant will be given by the sum of the expected values of these terms. There are three types of terms to consider. The dominant term, due to the product of the diagonal elements, will be +1. Then there are terms which contain only one or more of the a_{ij} squared, present because of the symmetry of the matrix ($a_{ij} = a_{ji}$). Third, there are terms which have one or more a_{ij} in the first degree. These terms have zero expectation value and will be ignored.

Since the a_{ij} are small compared to unity, it makes sense to write the summation in increasing order of the a_{ij}:

$$\begin{aligned} \det[a] \;=\; 1 - & \sum_{\substack{i,\, j \\ No\ Duplicity}} a_{ij}^2 + \sum_{\substack{i,\, j,\, k,\, l \\ No\ Duplicity}} a_{ij}^2 a_{kl}^2 \\ & - \sum_{\substack{i,\, j,\, k,\, l,\, m,\, n \\ NoDuplicity}} a_{ij}^2 a_{kl}^2 a_{mn}^2 + \ldots terms\ in\ a_{ij}^1 \end{aligned} \tag{320}$$

The sums are restricted both so that no two indices are the same, and also so that no term gets duplicated due to interchange of indices. We shall now take the expectation value of each term using (320) Also we will allow each index to now run over the full range of 1 to M. Therefore we must subtract out terms we have thereby introduced which have two (or more) equal indices. We must

also divide by the number of times we have caused the term to be duplicated.

$$
\begin{aligned}
E\{\det[a]\} &= 1 - \frac{1}{2}\left[\sum_{i=1}^{M}\sum_{j=1}^{M} g_i g_j - \sum_{i=1}^{M} g_i^2\right] \\
&+\frac{1}{8}\left[\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} g_i g_j g_k g_l - 6\sum_{i=1}^{M}\sum_{k=1}^{M}\sum_{l=1}^{M} g_i^2 g_k g_l\right] \\
&- \ldots
\end{aligned}
\tag{321}
$$

$$
\begin{aligned}
&= \sum_{N=0}^{M/2} \frac{(-1)^N}{N!2^N}\Big[\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\cdots\sum_{i_{2N}=1}^{M} g_{i_1} g_{i_2}\cdots g_{i_{2N}} \\
&-N(2N-1)\sum_{i_1=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\cdots\sum_{i_{2N}=1}^{M} g_{i_1}^2 g_{i_3} g_{i_4}\cdots g_{i_{2N}}\Big]
\end{aligned}
\tag{322}
$$

Now that each index runs over all values of 1 to M independently, we can use the separability of the terms of these summations to convert the multiple sums into the products of single sums. To simplify, let us assign the following notation to the sums of the g_i and of g_i^2:

$$
\begin{aligned}
J &\triangleq \sum_{i=1}^{M} g_i \\
G &\triangleq \frac{J^2}{2} = \frac{1}{2}\Big[\sum g_i\Big]^2 \\
H &\triangleq \sum_{i=1}^{M} g_i^2
\end{aligned}
\tag{323}
$$

Then we can rewrite (321) as:

$$\begin{aligned} E\{\det[a]\} &= \left[1 - G + \frac{G^2}{2!} - \ldots + (-1)^N \frac{G^N}{N!} + \ldots\right] \\ &\quad + \frac{H}{2}\left[1 - 3G + \ldots + (-1)^{N-1}\frac{2N-1}{(N-1)!}G^{N-1} + \ldots\right] \\ &= \sum_{N=0}^{M/2}(-1)^N\frac{G^N}{N!} + \frac{H}{2}\left[\sum_{N=0}^{\frac{M}{2}-1}(-1)^N\frac{G^N}{N!} - 2G\sum_{N=0}^{\frac{M}{2}-2}(-1)^N\frac{G^N}{N!}\right] \end{aligned} \tag{324}$$

Assuming M is sufficiently large so that the alternating power series have converged to the exponential function of $-G$, we obtain the result:

$$\begin{aligned} E\{\det[a]\} &\approx e^{-G} + \frac{H}{2}e^{-G} - HGe^{-G} \\ &= e^{-\frac{J^2}{2}}\left[1 + \frac{H}{2}(1 - J^2)\right] \end{aligned} \tag{325}$$

Determination of the Diagonal Elements of the Inverse Matrix

Recall that a, the matrix whose determinant we have analyzed, is a column and row scaled version of the matrix A according to (313). We seek the diagonal elements of R, the covariance matrix which is the inverse of A. We shall proceed by first providing estimates of the diagonal elements of the inverse of a. Given a^{-1}, the determination of the diagonal elements of R was given by (316). Because we have stipulated that we will be supplied the expected values of the A_{ii} and that their relative variances will be small, we can write the expected values of the R_{ii} accordingly.

$$E\{R_{ii}\} \approx \frac{E\{a_{ii}^{-1}\}}{E\{A_{ii}^{-1}\}} \tag{326}$$

We shall now proceed to evaluate the diagonal elements of a^{-1} using the above formula for the determinant of such a matrix.

The i-th diagonal term of the inverse of a is given by the formula:

$$[a^{-2}]_{ii} = \frac{\det[a^{(ii)}]}{\det[a]} \tag{327}$$

where $a^{(ii)}$ denotes the matrix formed when the i-th row and the i-th column of a are deleted. That leaves a matrix of the same form as a, with (317) still applying. However the summations J and H defined in (323) have each been reduced by the non-inclusion of the i-th term. Let us only consider the change to the summation J, since the role of H in (325) is of far less consequence.

Call J' the evaluation of J for the matrix $a^{(ii)}$ in which the i-th row and i-th column have been removed. Then clearly:

$$\begin{aligned} J' &= J - g_i \\ \therefore \frac{(J')^2}{2} &= \frac{J^2}{2} - Jg_i + \frac{g_i^2}{2} \end{aligned} \tag{328}$$

We can now use (325) to estimate the determinant of $a^{(ii)}$. We will ignore the g_i^2 term in (328) which is of relatively small consequence, and will likewise ignore the change to H and J^2 which it multiplies in (325).

$$
\begin{aligned}
E\left\{\det[a^{(ii)}]\right\} &\approx e^{-\frac{(Jr)^2}{2}}\left[1+\frac{H}{2}(1-J^2)\right] \\
&\approx e^{-\frac{J^2}{2}}e^{Jg_i}\left[1+\frac{H}{2}(1-J^2)\right] \qquad (329)
\end{aligned}
$$

Now we will find the expected value of the diagonal element of a^{-1} using (327) by crudely dividing the expectation of the numerator by the expected value of the denominator. This is generally an unsound procedure for determining the expected value of a quotient, except in the case of the denominator having a very small variance. We will continue, however, noting that the variability of the numerator and denominator, in this case, largely track each other, since most of the low-order terms in either determinant using (320) would be identical. The following analysis focuses on the change to the determinant due to the elimination of terms included only as a result of elements in row #1 and column #i.

Substituting the result (329) and (325) into (327), we find:

$$
\begin{aligned}
E\left\{[a^{-1}]_{ii}\right\} &\approx \frac{E\{\det[a^{(ii)}]\}}{E\{\det[a]\}} \\
&\approx \frac{e^{-\frac{J^2}{2}}e^{Jg_i}[1+\frac{H}{2}(1-J^2)]}{e^{-\frac{J^2}{2}}[1+\frac{H}{2}(1-J^2)]} \qquad (330) \\
&= e^{Jg_i} \\
&\approx 1+Jg_i
\end{aligned}
$$

(For the purposes to which we will be applying this formula, the exponent in the third line will be small compared to one, justifying the final approximation.) Applying this result to (326) we find for the expected value of the i-th diagonal term of the covariance R (which specifies the **variance** of the i-th element of the vector U):

$$E\{R_{ii}\} \approx \frac{1 + Jg_i}{E\{A_{ii}\}} \tag{331}$$

Appendix II:

Coherent Integration of Optical Correlation (Fringe Visibility) using Inexact Determination of $\tau(t)$

Although separate from the stated subject of this work (estimation of τ), the case for off-line delay-tracking system may seem incomplete without an explanation of the ultimate use for which the off-line estimation of $\tau(t)$ would be employed. The following procedures equally apply to the estimation of $V(\nu)$ given the estimate $\tau(t)$ and the applied **correction** to the delay-line $\tau_c(t)$, whether the delay correction most closely followed the real-time estimate of $\tau(t)$ (**real-time** delay tracking), or is identically zero. In each case we will assume that the record of photon events has been saved, along with the estimate of τ, and knowledge of the expected r.m.s. error of that estimate of τ.

In order to keep the discussion manageable we will proceed on the basis of three levels of approximation regarding $\tau(t)$. First we will consider estimation of $V(\nu)$ given **exact** knowledge of $\tau^{(ACT)}$, the actual differential atmospheric delay function. We will derive an equation whose solution yields the maximum likelihood estimate of V. However we will see that that solution does not have a closed-form solution, and its basis of operation is hardly transparent.

Then we will form a more "natural" average which appears to proved a **nearly** optimum estimate, and is computationally efficient, That solution will then be extended to the case of employing estimates of $\tau(t)$ which have a Gaussian error structure with a variance specified, for instance, by the analysis pertaining to the local estimation problem. We will also determine a correction to take into account the finite bandwidth of the

spectral channel in which photons for the determination of $V(\nu)$ are detected. Finally, we will consider estimation of $V(\nu)$ using the **poly-Gaussian** estimates of $\tau(t)$ that we found to result at lower signal-to-noise ratios in our study of the global problem.

The Maximum Likelihood Estimate of $V(\nu)$

If some ingredients of the following derivation look like those for the τ estimation problem, that is because they are the same. We will again be forming an estimate to maximize the probability of having received the set of photons that had been detected. However the job, in this case, is easier in two ways. Unlike the τ estimation problem in which the **Maximum A Posteriori Probability** criterion was used for estimation of the **random** process $\tau(t)$, the estimation in this case is of a **non-random** parameter. Thus we do not need to consider the *a posteriori* probability of the unknown, removing the χ^2 component from the determination of the figure of merit (171) driving the estimation.

Secondly, the estimation of $V(\nu_1)$ and $V(\nu_2)$ are independent. This is in contrast to the estimation of τ expressed in terms of its Fourier coefficients U_i. The solution of the M Fourier coefficients would have required the simultaneous solution of M coupled equations, expressed by the dense matrix A in (178). In the estimation of V, it is as if we had instead been presented with a **diagonal** matrix, implying a set of **uncoupled** equations.

To form the maximum likelihood estimate, we will again write the **logarithm** of the likelihood due to photon #k as λ_k in the same form that we had previously used this notation in (167) and (166). However in (166), V is considered to be "known" while τ is an unknown. In the present case, however, we will initially assume

that τ is known exactly, but V will be treated as the unknown, over which, the likelihood is to be maximized. So we can write the logarithm of the likelihood of photon #k, having been detected at time t_k at optical frequency ν_k in an interference channel with photomixing phase angle θ_k as:

$$\begin{aligned}\lambda_k(V') &\stackrel{\wedge}{=} \log \frac{P(\theta_k|V = V')}{P(\theta_k|V = 0)} \\ &= \log(1 + Re\{V' e^{j\theta_k} e^{j2\pi\nu_k(\tau(t_k)-\tau_c(t_k))}\})\end{aligned} \tag{332}$$

Then λ,the logarithm of the **net** likelihood of $V(\nu')$ is, as in (170), given by the sum of the contributing λ_k. However in order to estimate V as evaluated at the optical frequency ν', we shall include **only** photons whose ν is the same as ν'. How large a difference in ν we are willing to tolerate in order to include a photon, is immaterial at this point; it might simply amount to all photons counted in a spectral detection element given a specific hardware configuration. So λ is given by:

$$\lambda(V', \nu') = \sum_{\substack{Photons \\ whose \\ \nu_k = \nu'}} \lambda_k(V') \tag{333}$$

Now the estimation of V' is based on maximization of λ over V'. That maximization may be performed, as usual, by setting the first derivative to zero. In this case, we must be careful, because V' is a **complex** variable, but λ **is not** an analytic function in V'. Thus the derivative of λ with respect to V' is not defined. We must, instead, write $V' = x + jy$, where x and y are real. Then we can write two equations setting equal to zero the derivatives of λ with respect to x and with respect to y.

$$0 = \frac{d\lambda}{dx} = \sum_k \frac{Re\{e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}\}}{1 + Re\{V' e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}\}} \tag{334}$$

$$0 = \frac{d\lambda}{dy} = \sum_k \frac{-Im\{e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}\}}{1 + Re\{V' e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}\}} \tag{335}$$

These simultaneous equations can be combined in the more compact complex equations:

$$0 = \frac{d\lambda}{dx} - j\frac{d\lambda}{dy} = \sum_k \frac{e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}}{1 + Re\{V' e^{j\theta_k} e^{j2\pi\nu'(\tau(t_k)-\tau_c(t_k))}\}} \tag{336}$$

While the solution of this equation correctly yields the maximum likelihood estimate of $V(\nu')$, it does not have a closed-form solution since the unknown appears in unlike denominators of the terms of a large summation. However we will instead proceed with a very different approach in order to produce a closed-form estimate whose genesis is more transparent.

Estimation of $V(\nu)$ using the *Detection Phasor*

Let us define the **detection phasor** $D(\tau, \nu, \theta, t)$ as:

$$D(\tau, \nu, \theta, t) \stackrel{\wedge}{=} e^{j\theta} e^{j2\pi\nu(\tau - \tau_c(t))} \tag{337}$$

The we shall define the **absolute detection phasor** for photon #k, simply denoted D_k with no arguments, as:

$$\begin{aligned} D_k &\stackrel{\wedge}{=} D(\tau^{(\mathrm{ACT})}(t_k), \nu_k^{(\mathrm{ACT})}, \theta_k, t_k) \\ &= e^{j\theta_k} e^{j2\pi\nu_k^{(\mathrm{ACT})}(\tau^{(\mathrm{ACT})}(t_k) - \tau_c(t_k))} \end{aligned} \tag{338}$$

Note that D_k contains the same factors as L_k, the complex likelihood for photon #k, **except** for the non-inclusion of $V(\nu_k)$ (the quantity we presently wish to estimate). Also, unlike the complex likelihood in which $\tau(t_k)$ was considered an unknown, in (338) we are employing for τ the **actual** value of atmospheric delay, as if we had been able to estimate that quantity with no uncertainty (corrections for the estimation error of τ will be dealt with below). $\nu_k^{(\mathrm{ACT})}$ denotes the **actual** optical frequency of that photon, regardless of the precision to which that wavelength was measured.

Then the expected value of the absolute detection phasor can be shown to be proportional to the complex conjugate of the optical correlation $V(\nu_k)$. If there is no real-time delay-tracking being preformed, so that the second complex exponential in (338) is a truly random phase, then the following result will be true. Another sufficient condition for the following result, is use of a balanced discrete beam photomixer with equally separated mixing phases as described by (18) **except** for the case of $N = 2$. Under

either of these conditions the expected value of the absolute detection phasor is:

$$E\{D_k\} = \frac{V^*(\nu_k)}{2} \tag{339}$$

Proof:

Depending on whether we are considering a continuum output or discrete beam photomixer, the expected value of D_k will be expressed as an integral or summation, respectively. We will simply write the **integral** from, with the corresponding summation being obvious.

Let $P_0(\theta)$ denote the probability density of a photon entering the photomixer outputting the photomixer in a position or channel whose photomixing phase is θ, **under the condition of no optical correlation**, $|V| = 0$. In the case of a discrete beam photomixer obeying (16), $P_0(\theta)$ would be written as the sum of N Dirac delta functions, each of strength p_i at $\theta = \theta_i$, for $i = 1, 2, \ldots N$. Then **given** that photon #k was detected at optical frequency ν_k at time t_k, we can write the probability density of that photon being found in the photomixer's output channel at photomixing phase θ in terms of the complex likelihood **evaluated at** $\tau^{(\mathrm{ACT})}(t_k)$, the actual atmospheric delay function at time t_k.

$$\begin{aligned}
&P(\theta) = P_0(\theta)(1 + Re\{L_k(\tau^{(\mathrm{ACT})}(t_k))\}) \\
&= P_0(\theta)(1 + |V|\cos(\arg(V) + \theta + 2\pi\nu_k(\tau^{(\mathrm{ACT})}(t_k) - \tau_c))) \\
&= P_0(\theta)(1 + |V|\cos(\phi + \theta)) \\
&= P_0(\theta)(1 + \frac{|V|}{2}e^{j\phi}e^{j\theta} + \frac{|V|}{2}e^{-j\phi}e^{-j\theta}) \\
&\textit{where} \\
&\phi \stackrel{\triangle}{=} \arg(V) + 2\pi\nu_k(\tau^{(\mathrm{ACT})}(t_k) - \tau_c(t_k))
\end{aligned} \tag{340}$$

The expectation of D_k is required. First, its expectation over all θ is found by integrating over θ.

$$
\begin{aligned}
E\{D_k\} &= \int_0^{2\pi} d\theta\, P(\theta) e^{j\theta} e^{j2\pi\nu_k(\tau^{(\mathrm{ACT})}(t_k)-\tau_c(t_k))} \\
&= e^{-j\arg(V)} \int_0^{2\pi} d\theta\, P(\theta) e^{j\theta} e^{j\phi} \qquad (341)
\end{aligned}
$$

Substitution of (340) for $P(\theta)$ results in the formation of three integrals:

$$
\begin{aligned}
&= e^{-j\arg(V)} \int_0^{2\pi} d\theta\, P_0(\theta)(e^{j\theta}e^{j\phi} + \frac{|V|}{2} e^{2j\theta} e^{2j\phi} + \frac{|V|}{2}) \\
&= e^{j2\pi\nu(\tau^{(\mathrm{ACT})}-\tau_c)} \int d\theta\, P_0(\theta) e^{j\theta} \qquad (342) \\
&\quad + \frac{V}{2} e^{j2\pi\nu(\tau^{(\mathrm{ACT})}-\tau_c)} \int d\theta\, P_0(\theta) e^{j2\theta} \\
&\quad + \frac{V^*}{2} \int d\theta\, P_0(\theta)
\end{aligned}
$$

Note that we have combined phase factors involving $\pm\arg(V)$ with $|V|$ to obtain V and V^* in the second and third integrals. Now the first integral simply amounts to the integral form of the left-hand side of (17), which we have shown must be equal to zero for any lossless photomixer.

The second integral is a little more tricky. It can be thought of as the Fourier series transform of the periodic function P_0 evaluated at the frequency of 2 cycles (per 2π radians). **If** we have chosen a balanced photomixer with equally spaced phases as in (18), then P_0, consisting of N equally spaced delta functions on the unit circle, will only contain frequency components of N and its harmonics. Therefore **unless** $N = 2$, the second integral will be equal to zero for a photomixer satisfying (18).

Even if the integral itself cannot be shown to be zero, we should still look at the leading factor multiplying the integral. We must still evaluate the expectation of the product **over** $\tau^{(\mathrm{ACT})} - \tau_c$, the uncorrected component of delay error. If we can be certain that τ is **not being closely tracked** in real-time hardware, then it can be

presumed that this phase factor would be totally random, thus uniformly distributed over the unit circle. The expectation of an expression containing such a quantity (and none other correlated with it) is clearly zero. So we have identified two individually sufficient conditions for the second term to vanish.

Finally, the **third** integral in (342), being the integral of a probability density only, over the entire domain of that density, is clearly unity. In the discrete case, this integral is identical to the left hand side of (15), which we have shown to be unity for any lossless photomixer. So, assuming at least one of the conditions has been met to insure that the second term will vanish, the determination of (342) reduces to the leading factor multiplying the third integral, thus proving (339).

Having shown the absolute detection phasor to be proportional to the complex conjugate of the optical correlation, it is a simple matter to form an estimate of $V(\nu')$ using the D_k for photons whose $\nu_k = \nu'$. If, during the course of an observation we can identify K photons at the desired optical frequency ν', the estimate of $V(\nu')$ would clearly be:

$$\hat{V} = \frac{2}{K} \sum_k D_k^* \tag{343}$$

Unfortunately determination of the **absolute** detection phasor requires, among other things, **exact** knowledge of the underlying atmospheric delay, $\tau^{(\mathrm{ACT})}(t)$. So let us instead look at the estimation of V using inexact **estimates** of τ that may have been obtained using one or another procedure.

The Corrupted Detection Phasor

Using the concept of the absolute detection phasor as defined in (338), and the estimate (343) of $V(\nu')$ based on it, we shall now attempt to formulate a similar estimate using a comparable quantity, but one which is actually **observable**. We will do this by retaining the result (339) we found for the expected value of the absolute detection phasor. Then, analyzing the **corrupted detection phasor**, we will calculate an applicable correction factor.

Let $\hat{\tau}$ represent our best **estimate** of τ. We shall assume, as we found in the discussion of MAP path estimation of τ, that the *a posteriori* density of τ around its mean is Gaussian. Let the r.m.s. error of the estimate be denoted σ_τ. Using the MPA path estimation procedure, for instance, σ_τ would be given by (275). The we shall form (one version of) the **corrupted detection phasor** by applying (337) as we did for the **absolute** detection phasor, but this time using the **estimated** value of τ, rather than the **actual** value of τ, which we would have no way of knowing.

$$\begin{aligned} \tilde{D}_k &\stackrel{\wedge}{=} D(\hat{\tau}(t_k), \nu_k^{(\mathrm{ACT})}, \theta_k, t_k) \\ &= e^{j\theta_k} e^{j2\pi\nu_k(\hat{\tau}(t_k) - \tau^{(\mathrm{ACT})}(t_k))} \end{aligned} \tag{344}$$

It is a simple matter to express this corrupted detection phasor in terms of the absolute detection phasor.

$$\tilde{D}_k = D_k e^{j2\pi\nu_k(\hat{\tau}(t_k) - \tau^{(\mathrm{ACT})}(t_k))} \tag{345}$$

We would like to determine the expected value of the corrupted detection phasor. The expectation of the product in (345) is given by the product of the expected values since the two quantities are uncorrelated. Note that the difference inside of the parenthesis

of the exponent of the phase factor of (345) is simply the **estimation error** of τ, having zero mean and an r.m.s. magnitude of σ_τ. We can then use a standard result to determine the expected value of the phase factor.

$$\begin{aligned} E\{\tilde{D}_k\} &= E\{D_k\}E\{e^{j2\pi\nu_k(\hat{\tau}(t_k)-\tau^{(\mathrm{ACT})}(t_k))}\} \\ &= \frac{V^*}{2}e^{-2\pi^2\nu_k^2\sigma_\tau^2} \end{aligned} \tag{346}$$

Again, we are assuming that at least one of the two conditions enabling the determination of (339) has been met. Then, assuming that during an observation there has been no substantial change in the statistics of the atmosphere or in the equipment, so that σ_τ has remained constant, we can form a **corrected** estimate of V as in (343), using, this time, the **corrupted** detection phasor.

$$\hat{V} = \frac{2}{K}e^{2\pi^2\nu^2\sigma_\tau^2}\sum_k \tilde{D}_k^* \tag{347}$$

Now let us look at a second way in which a naive measurement of the detection phasor could introduce an inaccuracy in the estimation of V. We have assumed that the photomixer outputs have been spectrally dispersed, and that only photons whose optical frequency $\nu_k^{(\mathrm{ACT})} = \nu'$ are used in the summations (343) and (347), for the estimation of $V(\nu')$. Of course, for any specified ν' we will **never** find a photon of that **exact** wavelength, but we will detect many photons in the spectrometer **channel** which includes ν'. Let us assume that the spectrometer channel is **centered** at ν' and is of spectral width $\Delta\nu$. Then we will use all photons detected in that spectral **channel** as if each had an optical frequency of exactly ν'. Since they are actually spread about ν', there would be an inaccuracy introduced into the estimation of V using (343).

Let us assume that the spectral width of the channel is relatively small, and that the received light has a continuum spectrum which does not fluctuate significantly over the spectral width of the detection channel. Then we can calculate the expected value of this **new** corrupted detection phasor, in which we will focus on the effect of the finite bandwidth of the spectral channel, and ignore the perturbation, already calculated, due to inexact estimation of τ.

Let us consider all photons whose **detected** $\nu_k = \nu'$, whereas the **actual** optical frequency of each such photon is denoted $\nu_k^{(\mathrm{ACT})}$. For these selected photons, $\nu_k^{(\mathrm{ACT})}$ is uniformly distributed from $\nu' - \Delta\nu/2$ to $\nu' + \Delta\nu/2$. The **corrupted** detection phasor, in this case, is what would be found by applying (337), but inserting for the optical frequency in (337), the **center** optical frequency of the spectral channel, ν'. Then we can express this corrupted detection phasor in terms of the absolute detection phasor:

$$\begin{aligned} \tilde{D}_k &= D(\tau^{(\mathrm{ACT})}, \nu', \theta_k, t_k) \\ &= e^{j\theta_k} e^{j2\pi\nu'(\tau^{(\mathrm{ACT})}(t_k) - \tau_c(t_k))} \\ &= D_k e^{j2\pi(\nu' - \nu_k^{(\mathrm{ACT})})(\tau^{(\mathrm{ACT})}(t_k) - \tau_c(t_k))} \end{aligned} \tag{348}$$

The expectation of a phasor of zero mean phase uniformly distributed over an interval, is, as usual, given by a **sinc** function. Again assuming that one of the two conditions enabling the validity of (339) has been met, we can determine the expected value of this corrupted detection phasor.

$$
\begin{aligned}
E\{\tilde{D}_k\} &= E\{D_k\}E\{e^{j2\pi(\nu'-\nu_k^{(\mathrm{ACT})})(\tau^{(\mathrm{ACT})}(t_k)-\tau_c(t_k))}\} \\
&= E\left\{\frac{V^*(\nu_k^{(\mathrm{ACT})})}{2}\right\}\mathrm{sinc}\left((\tau^{(\mathrm{ACT})}(t_k)-\tau_c(t_k))\Delta\nu\right) \qquad (349) \\
&\approx \frac{V^*(\nu')}{2}\,\mathrm{sinc}\left((\hat{\tau}(t_k)-\tau_c(t_k))\Delta\nu\right)
\end{aligned}
$$

In the final step, the average of V over the optical frequency **interval** has been approximated by its value at the center of that interval. Note that this is totally consistent with a **slope** in the function V; a discrepancy would only occur in the case of V having a large **curvature** over the interval. We have also replaced the **actual** τ inside the argument of the sinc function, with the **estimated** τ. Since the error in τ is symmetric about zero, the expected value of the sinc of a slightly random argument would be virtually identical to its value at the mean value of that random quantity.

Again, using this determination for the expectation of the **corrupted** detection phasor, we can, as in (374), write the estimate for $V(\nu')$ as a **corrected** average over the detection phasors of all photons detected in the spectral channel centered about ν'. However in **this case**, the correction for different photons will be unalike, thus we must apply the correction photon-by-photon[25].

[25]To form the estimate with the absolutely smallest estimation error, we would really want to **weight** the contribution of the photons differently depending on the value of the sinc function associated with each one. That weighting will be included in the next level of analysis. For the present discussion, however, we will note that practically speaking, in a spectrally dispersed detection system with good spectral resolution, the sinc function will be near unity for all photons, and such a weighting will add little to the quality of the estimate. Non-inclusion of optimum weighting does **not** contribute any **bias** to the estimation; only the expected estimation **error** is affected.

Including the previous correction due to the estimation error of τ, we would obtain for the estimate of V:

$$\hat{V} = \frac{2}{K} e^{2\pi^2 \nu^2 \sigma_\tau^2} \sum_k \frac{\tilde{D}_k^*}{\text{sinc}\left((\hat{\tau}(t_k) - \tau_c(t_k))\Delta\nu\right)} \tag{350}$$

Again, this summation is taken over the K photons found in the spectral channel centered about ν'.

The degradation of the expected magnitude of the corrupted detection phasor caused by the detection of the interfered spectra in a limited number of discrete channels, has implications regarding the optimum mode of operation. Given absolutely noiseless detectors, the effect of the degradation due to the sinc function in (349) can be arbitrarily reduced by increasing the spectral resolution of the spectrally-dispersed detection hardware, thus decreasing $\Delta\nu$. However using detectors with a finite dark count, there will be a limit beyond which further dividing the light among detectors reduces the signal-to-noise ratio obtained when observing dim objects.

An alternative means of mitigating the effect of the sinc function in (349) is to reduce the **uncorrected** component of delay error, $\Delta\tau$, by adjusting τ_c, the hardware **correction** delay, to follow, as closely as possible, $\tau^{(\text{ACT})}(t)$ in real-time. While the emphasis has been on the **off-line** estimation of τ, we have noted that, even for signals too weak to successfully track closely in real-time, there is still the possibility of performing **medium-coarse** real-time delay-tracking in order to reduce the peak uncorrected delay excursions from, perhaps, a **hundred** wavelengths, to, perhaps, only **several** wavelengths.

One may ask why even bother reducing $\Delta\tau = \tau - \tau_c$ in real-time, when the degradation predicted by (349) can be perfectly compensated in the estimation of V using (350). There are two good reasons to reduce $\Delta\tau$ if possible. First, even though we will, following the off-line estimation of τ, quantify the degradation predicted by (349) and correct for it, it is always preferable to require a **small** correction rather than a **large** one! Secondly, the degradation of detected interference amplitude due to the sinc function in (349) affects not only the measurement of V, but **also** the detection of interference for the sake of estimating τ itself (using **any** estimation procedure). We did not include any correction for this effect in the τ estimation algorithm discussed because it does not **bias** the estimators, and, for $\Delta\tau\Delta\nu$ not approaching unity, a modification to the estimator's operating parameters would hardly be required. However it can be shown that the **effective** signal-to-noise ratio determining the performance of the estimation of $\tau(t)$ (for instance, as used in (256)) will be reduced by the sinc function shown in (349).

Estimation of V using the Poly-Gaussian Estimate of τ

We have seen that the measurable quantity that we have called the **corrupted detection phasor** can be used to form an estimate of $V(\nu')$ by taking into account the degree to which inexactness in the estimation of τ, and the combining of a **range** of optical wavelengths into a single spectral channel, have attenuated the measurement of V using the measured detection phasor. If we had considered the use of the delay-dispersed photomixer with detector elements covering a range of interference phase, we similarly could have estimated an additional correction factor.

Now, however, we would like to consider estimation of V using the results of the **global MAP solution**, as depicted, for instance, in Figure 41. Integration of V using photons around $t = 0$ or $t = 25$ is straight forward, since the estimate of τ at these times is characterized as a mean estimate with an expected r.m.s. error, the ingredients determining the corrupted detection phasor according to (344), from which we have derived an unbiased estimate of V in (350). However we would like to know how to employ the photons occurring, for instance, during the time period $t = 10$–20, during which the estimate of τ consists of the **union** of two distinct solutions, one of which has a confidence of 82%, the other 18%.

Let $\hat{\tau}_1$, $\hat{\tau}_2$,... denote the mean estimates of $\tau(t_k)$, with *a posteriori* probabilities of p_1, p_2,... according to the results of the global MAP solution. In this case we will define the **corrupted** detection phasor for photon #k, for estimation $V(\nu')$, as the weighted average of the detection phasor for photon #k evaluated at ν' at each mean estimate of τ, weighted according to the confidence of each solution. We will also further attenuate terms in the average according to the degradation that they would already suffer due to $\Delta\nu$, given each mean estimate of $\tau(t_k)$.[26]

$$\tilde{D}_k \stackrel{\wedge}{=} \sum_i p_i sinc((\hat{\tau}_i - \tau_c)\Delta\nu) D(\hat{\tau}_i, \nu', \theta_k, t_k) \tag{351}$$

We can analyze this corrupted detection phasor as follows:

[26]Note that this is the weighting that we neglected to perform earlier, and for which we noted there would be only a small degradation in the efficiency of the estimate of V, and no bias introduced. In the present derivation, however, failure to include this weighting would eventually lead to a less symmetric result which we can easily avoid by simply introducing this factor at this point, without further explanation.

$$\tilde{D}_k = D(\tau, \nu_k^{(\mathrm{ACT})}) \sum_i \Big[p_i \,\mathrm{sinc}((\hat{\tau}_i - \tau_c)\Delta\nu) \cdot e^{j2\pi\nu'(\hat{\tau}_i - \tau)} e^{j2\pi(\tau - \tau_c)(\nu' - \nu_k^{(\mathrm{ACT})})} \Big] \tag{352}$$

τ in this expression is at this point simply an arbitrary variable. However let us suppose that it represents our estimate of $\tau^{(\mathrm{ACT})}$. Now let us determine the expected value of (352). In this case, the new feature we have included is the multiplicity of estimates of $\tau(t_k)$. We will thus treat τ as being characterized by the poly-Gaussian *a posteriori* density derived from the results of the multiple path MAP solutions. Thus we will average over each such solution weighted by the confidence of that solution.

$$\begin{aligned} E\{\tilde{D}_k\} &= \sum_j p_j E\{\tilde{D}_k | \tau \sim N(\tilde{\tau}_j, \sigma_\tau^2)\} \\ &= \sum_j p_j E\{D(\tau, \nu_k^{(\mathrm{ACT})}, \theta_k, t_k) | \tau \sim N(\tilde{\tau}_j, \sigma_\tau^2)\} \cdot \\ &\quad \sum_i p_i \,\mathrm{sinc}((\hat{\tau}_i - \tau_c)\Delta\nu) e^{j2\pi\nu'(\hat{\tau}_i - \hat{\tau}_j)} e^{-2\pi^2\nu^2\sigma_\tau^2} \,\mathrm{sinc}((\hat{\tau}_j - \tau_c)\Delta\nu) \end{aligned} \tag{353}$$

Taking the expectation of the product of the terms in (351) has recovered the degradation factors due to σ_τ and $\Delta\nu$ that we have previously calculated (the similar appearing **weighting** factor containing $\Delta\nu$ has also been propagated into this expression). Now careful examination will reveal that the detection phasor inside the conditional expectation of the lower expression of (353), is none other than the **absolute detection phasor** (338). That is because it is evaluated **conditionally** on τ being the actual τ function that generated the photon, the specified probability of that τ then being irrelevant. Therefore we have already determined a value for this expectation. Continuing:

$$
\begin{aligned}
&= \frac{V^*}{2} e^{-2\pi^2 \nu'^2 \sigma_\tau^2} \sum_j p_j \sum_i [p_i sinc((\hat{\tau}_i - \tau_c)\Delta\nu) \cdot \\
&\quad e^{j2\pi\nu'(\hat{\tau}_i - \hat{\tau}_j)} sinc((\hat{\tau}_j - \tau_c)\Delta\nu)] \qquad (354) \\
&= \frac{V^*}{2} S_k
\end{aligned}
$$

S_k is the **strength** of photon #k toward the estimation of $V(\nu_k)$. Observing that the summations in (354) are **separable**, and that the separate sums are complex conjugates of each other, we can write S_k as:

$$
S_k \stackrel{\wedge}{=} e^{-2\pi^2 \nu'^2 \sigma_\tau^2} \left| \sum_i p_i sinc((\hat{\tau}_i - \tau_c)\Delta\nu) e^{j2\pi\nu_k \hat{\tau}_i} \right|^2 \qquad (355)
$$

Having found the expected value of this corrupted detection phasor, it is possible to apply the proper normalization to any average involving such quantities to determine $V(\nu')$. However we may choose to weight each detection phasor in order to form a more reliable average. Let us apply a weight of w_k to photon #k and form the weighted average called Q, from which the formation of the estimate of V is trivial.

$$Q \triangleq \sum_k w_k \tilde{D}_k^*$$

$$E\{Q\} = \frac{V(\nu')}{2} \sum_k w_k S_k \tag{356}$$

$$\therefore \hat{V}(\nu') = 2\frac{\sum_k w_k \tilde{D}_k^*}{\sum_k w_k S_k}$$

It appears that a constant weighting in (356) would form the estimate with the smallest r.m.s. error. That is because, as analysis would reveal, the expected value of the **signal** in each corrupted detection phasor is proportional to S_k. However the expected r.m.s. noise in each corrupted detection phasor is proportional to $\sqrt{S_k}$. This, it turns out, already implements the optimum weighting, mathematically speaking.

Practically speaking, however, we may wish to apply a weighting function in order to eliminate or attenuate the contribution of photons which, for any reason, we don't trust. For instance, it may be found that, at points in multiple path solutions where the *a posteriori* probability is spread among several paths, the **reported** *a posteriori* probabilities of those solutions are in error. If that error has a systematic character (such as consistently exaggerating the difference in *a posteriori* probability between likely and unlikely solutions), then a **bias** could be introduced into the estimation of V using the **reported** *a posteriori* probabilities. A possible remedy would be to assign zero weight to photons occurring at times at which the reported *a posteriori* probability of the dominant path does not, say, exceed 90%. Again, whatever weighting rule is chosen, there will be no bias introduced into the estimation of V using (356); the only

deleterious effect will be a certain increase in the r.m.s. error of that estimation.

Estimator Bias Caused by the Dual Use of Photons

We previously saw that the corrupted detection phasor in which we used not the actual atmospheric delay function, but rather an **estimate** of $\tau(t)$, responded to $V(\nu)$ with an attenuation factor quantified in (346). There is one more way in which the use of an estimate of τ can affect the accuracy of the estimation of V using the corrupted detection phasor. That occurs when the estimate of V is based on the corrupted detection phasors of photons **which were also used in the estimation of** τ.

Consider the estimation of $\tau(t)$ based upon the maximization of Λ. The maximization of Λ given by (171) involves the maximization of (170) or (168), and thus the maximization of the real part of the photons' complex likelihoods given by (166). This will produce an unbiased estimate of τ. However the specific errors in the resultant estimate of τ will generally be so as to increase the real part of the complex likelihoods. If that estimate of τ is then **also** used to form the corrupted detection phasor for the **same photons**, then the estimate of $V(\nu)$ will be biased in the direction of the initially assumed value of $V(\nu)$ used in (166).

This is simply a result of the randomness of photon detections being inseparable from the information contained in those photons For instance, the upper left and upper right of Figure 19 in each depict MAP estimates of τ using different sets of 50 randomly generated photons. Around $t = 15$, the upper left estimate of τ is lower than the actual τ function, whereas the upper right estimate of τ is greater than the actual τ. If the estimate of τ in the upper left of Figure 19 were used to form the corrupted detection phasor for the photons used in the upper right figure (or any other

realization of photons based on the same underlying $\tau^{(ACT)}$ function) then the effect of the estimation error seen in the upper left figure would simply be to attenuate the magnitude of the corrupted detection phasor according to (346). Based on our knowledge of the r.m.s. error of the estimation procedure, this effect can be corrected using (347). **However**, the effect of the **specific** estimation error seen in the upper left figure, **if applied to the same photons used in that estimation**, would be to increase the resultant estimate of V.

The expected resulting bias could be quantified. However a more direct strategy for eliminating this nuisance effect is to compute the corrupted detection phasor for photon $\#k$ using an estimate of τ in which photon $\#k$ has not been used. Rather than performing a complete re-estimation of τ for each photon to be used in the estimation of V, it would be more practical to, perhaps, eliminate a certain **portion** of the set of photons, re-estimate τ using the reduced set of photons, and use that estimate of τ to compute the detection phasors for only the photons that had been ignored in that estimation of τ. Performing this procedure several times on different portions of the entire set of photons, could then produce corrupted detection phasors for all photons in which this bias has been prevented. Simulations using this method have produced estimates of V in which any such bias appears to have been avoided.

The Expected Error in the Estimation of $V(\nu)$

Let us briefly examine the expected r.m.s. estimation error of $V(\nu')$ using estimates based upon averaging the detection phasors of K photons whose $\nu_k = \nu'$. For simplicity, let us look at the error of the estimate (343) based on the averaging of the (unobservable)

absolute detection phasor (338). To simplify the problem, let us assume that $|V| \ll 1$. This yields the "worst-case" estimation error (but not **much** worse than for large $|V|$).

In order to employ the following mathematics, we must deal not with D_k and $V(\nu)$, but rather with the **real** and **imaginary** parts of these complex quantities separately. Let us denote the real part of D_k as d_k, and denote the real part of $V(\nu')$ as v. Then clearly, from (343), the estimation of v from d_k would be given by:

$$\hat{V} = \frac{2}{K} \sum_k d_k \tag{357}$$

where again the summation includes K photons whose ν_k is (approximately) the same as ν'. We have previously calculated the expected value of D_k in (339) which justified the estimator (343) (or (357)). Now let us look at the **variance** of d_k.

We specified that $|V|$ would be regarded as being small compared to unity, for the sake of this analysis. In the case that it is exactly **zero**, then clearly the net phase factors in (338) would be totally random, with the expected value of the phasor therefore being zero (as it should be according to (339) for $V = 0$). d_k, the **real part** of D_k, would clearly be given by $\cos(\phi)$ where ϕ is a totally random phase. Then the **mean squared** value of d_k would be the mean squared value of the cosine of a random angle, or $\frac{1}{2}$. That is also the **variance** of d_k, given that its mean value is zero. The variance of the summation in (357) is given by the sum of the variances, or $K/2$. Then the r.m.s. estimation error is determined as:

$$\tilde{v} \triangleq \sqrt{E\{\hat{V}^2\}} = \sqrt{\frac{4}{K^2} Var\left\{\sum_k d_k\right\}} = \sqrt{\frac{2}{K}} \tag{358}$$

It is obvious that the estimate of the **imaginary** part of $V(\nu')$ using (343) would be subject to an error of the same magnitude.

To present a numerical example, suppose that we wish to estimate $V(\nu')$ with an expected error of no more than .01 in both the real and imaginary components. That would require, according to (358), a total of 20000 photons to be included in the average toward V for the optical frequency ν'. If the spectrally-dispersed detection system included 100 wavelength channels, one of which was used for the estimation of each of 100 points of V, then a total of approximately 2 million photons would be required to achieve this accuracy in each wavelength channel. The faintest objects for which delay-tracking is possible, provide on the order of 1000 detected photons per second in the interferometer. Thus in this very worst case, an observation time of about half an hour would be required to achieve the specified accuracy in the estimation of V.

More typically, the estimation of V in 100 wavelength channels, **each** having an individual estimation error of .01, would probably not be required. Typical functions of $V(\nu)$ (in which the specification of ν implies not only the result of imaging in different colors, but more importantly, the correlation corresponding to different **spatial frequency** components of the object) are relatively smooth, so that a curve of V specified at 100 points with a **larger** estimation error **at each point**, could reasonably be **smoothed** to extract the underlying curve with an accuracy well in excess of the accuracy of each data point on the curve. In any case, it can be seen that required observation

periods using delay-tracking and coherent integration of optical correlation, will generally be on the order of minutes, not hours.

Glossary

Symbols:

α Parameter defining the magnitude of the *a posteriori* variances of the components of the cosine transform of τ, according to (69), page 62.

β An intermediate parameter which is proportional to the square of the signal-to-noise ratio and to T. Given by (244), page 133. In the case of determinations using the **exact logarithm**, β is slightly modified: see (272), page 147.

κ The dimensionless parameter (see page 29) expressing the range of delay presented in the fringe pattern of the delay-dispersed photomixer (Figure 5). For a two-slit interferometer, κ is defined as the ratio of slit separation to slit width, **as measured at the plane of the beam's waists** (where the beams are afocal).

θ Phase by which the signal from input channel 1 is delayed relative to the signal from input channel 2 in the production of a particular output channel of a photomixer. θ_i refers to the phase for output channel i. See (16), page 22.

λ An ingredient in the expression for Λ in (171). Defined as a constant plus the logarithm of the likelihood of $\tau(t)$ given a set of received data. See (170), page 102. Equal to the sum of the λ_k (each given by (167)) for each photon.

Λ Defined as a constant plus the logarithm of the *a posteriori* probability density of $\tau(t)$ given a set of received data. See (159), page 99.

$\Lambda^{(\mathrm{PATH})}$ The value of Λ determined for a **partial path**, based on the limited period of time over which that path is defined. These values, for each path, are the ingredients on which the calculation of the Λ^{+}, Λ^{-}, $\Lambda^{(\mathrm{NODE})}$, $\Lambda^{((full))}$, and all marginal probabilities in the global solution, are based. Refer to the discussion starting on page 186.

ν Optical frequency. Note that frequency references to directly observable time-scales (such as the responses of electronic or digital filters) are instead given by ω.

ν_k Optical frequency of photon #k.

ν_0 The "nominal" optical frequency, defined as the root-mean-squared optical frequency averaged over all photons. See (231), page 129.

$\bar{\nu^2}$ The mean-squared optical frequency averaged over all photons. See (230), page 129.

$\overline{\nu^2 V^2}$ The expected value of $\nu^2|V(\nu)|^2$ averaged over all detected photons according to $F(\nu)$. See (228), page 128.

$\overline{\nu^4 V^2}$ The expected value of $\nu^4|V(\nu)|^2$ averaged over all detected photons according to $F(\nu)$. See (228), page 128.

σ_i^2 The variance of U_i, the i-th frequency component of the cosine transform of $\tau(t)$, according to its *a priori* characterization. Since the i-th frequency component occurs at the (radian) frequency $\omega = nT/i$, and the frequency spacing between i and $i+1$ is also a function of T, the variance σ_i^2 is dependent upon T, the time period over which the cosine transform is performed. See (53), (54), and (69), page 59.

$\tau(t)$ The difference in the optical delay between the light received by two arms of an interferometer due to atmospheric effects and/or other random process, but **excluding** the geometrical delay due to the position of the object in the sky and the location of the telescopes on earth. Since this is the quantity that we wish to estimate, $\tau(t)$ more often refers not to a definite function, but rather represents a **test function**, that is an **independent variable** (of high dimension), from which a figure of merit (Λ in the case of path estimation) is obtained.

$\tau^{(\mathrm{ACT})}$ The **actual** underlying differential delay function that the delay-tracker is attempting to estimate.

$\tau^{(\mathrm{MAP})}$ The estimate of τ based upon the maximization of Λ or the *a posteriori* probability function. Whether this represents the **global** maximization of Λ or only a local maximum, is determined by the context of its use.

$\tau_0(t)$ The **one-way** atmospheric optical delay affecting the transmission of starlight received at a given location on earth. τ, the **differential** atmospheric delay, is defined as the **difference** between τ_0 at two different points on earth.

τ_c The (optional) **correction** delay introduced by the delay line (in addition to τ_g, the geometric delay) in response to the estimation of τ. See Figure 1.

τ_g The **geometric** delay affecting the received light received by telescope 2 relative to telescope 1. Since this delay is deterministic, it is **not** considered as part of the estimation problem. See Figure 1.

$\Delta\tau$ The **uncorrected** component of $\tau(t)$ when real-time control of the delay line is used to partially compensate for atmospheric delay fluctuations. $\Delta\tau = \tau - \tau_c$. See (3), page 15. Not to be confused with $\Delta\tau$, below.

$\Delta\tau$ Used following page 123 to denote the difference between an arbitrary function $\tau(t)$, and $\tau^{(\mathrm{MAP})}$, the function which maximizes Λ. Not to be confused with $\Delta\tau$, above.

$\tilde{\tau}$ The expected r.m.s. error in the estimation of τ using one or another procedure.

$\tilde{\Phi}$ The expected r.m.s. **phase** error in the estimation of the atmospheric delay, given by $2\pi\nu_0\tilde{\tau}$, using one or another procedure.

χ^2 Defined as $-\frac{1}{2}$ of the logarithm of the *a priori* probability density for $\tau(t)$ over the interval $0 < t < T$, based upon the expansion of τ using the cosine transform truncated to M terms (see (161), (162), page 100). In the path estimation procedure, therefore, χ^2 acts as a **penalty** term in the figure of merit Λ (see (171), page 103), so that smoother solutions are favored. As suggested by its symbol, our χ^2 has an *a priori* probability density which is χ^2 with M degrees of freedom.

ω Radian frequency. Applies to **directly observable** time scales (Hertz, Kilo-Hertz). Note that **optical frequency** is instead denoted by ν.

ω_0 The -6 dB cutoff frequency of the filter $H(\omega)$ (see below) used in the time-domain algorithm. See (210), (211), page 122.

A The matrix contained in the exponential of the expression for the *a posteriori* probability of $\tau(t)$ expressed in U space (see (173), (172), page 104). A is the inverse of R, the covariance matrix describing the *a posteriori* statistics of the vector U.

b The supplementary gain parameter affecting the stability and rate of convergence of the time-domain algorithm shown in Figure 21. See (203), page 119.

C See **Confidence**, below.

C.I. See **Confidence Index**, below.

D See **Structure Function**, below.

D_k The **absolute detection phasor** (338) (page 215) includes all of the ingredients of the **complex likelihood** function (see L_k below) **except** for the optical correlation, $V(\nu_k)$. Since the real part of the complex likelihood is proportional to the probability of a photon detection, the expected value of the absolute detection phasor in which V has been **excluded**, is proportional to V^*. Based upon that determination, averaging of the absolute detection phasor (338), or one of its relatives (344), (348), (351), yields an estimate of $V(\nu)$ using (343), (347), (350) or (356).

e_1 The first (and more important) **enhancement** factors that enters into the formula for expected estimation error only when using the **exact logarithm** (and not the LLA approximation) to perform estimation of τ. e_1 is always between 1 and 2. It only approaches 2 for an unresolved object ($|V| = 1$) with

a low rate of dark counts. When using the LLA approximation, the formulae require setting $e_1 = 1$. See (265), page 143. Plotted vs. constant $|V|$ in Figure 24.

e_2 The second "enhancement factor" involved in the formula for expected estimation error when using the **exact logarithm**. Compared to e_1, e_2 has a very minor effect, and actually acts to **worsen** the expected estimation error. Although for $|V| = 1$ it approaches infinity, the expected estimation error does not suffer according to the prediction of e_2, and thus its inclusion in the formula for estimation error is questionable at best. See (271), page 146. Plotted vs. constant $|V|$ in Figure 24.

e_3 When using the **exact logarithm** only, the expected value of λ_k at $\tau^{(\mathrm{ACT})}$ is equal to $|V|^2/4$ **times** e_3 (see (300)). At low $|V|$, e_3 is 1, but can reach a value as high as 1.23, when $|V| = 1$.

$F(\nu)$ The photon spectral density. Normalized so that its integral over ν is unity. Defined so that the probability of detection (in one or another detector channel) a photon of optical frequency between ν and $\nu + \Delta\nu$ in a time period Δt is $(\Delta t)I_0(\Delta\nu)F(\nu)$.

G An intermediate result used in Appendix I. See (323), page 206.

g_i A function of the discrete frequency index i, which, if correctly chosen, will allow the expression of the variance of the off-diagonal elements of A or a to be of the form of (318) (see page 204). If g_i can be found to satisfy (318), then the conclusions of Appendix I are applicable.

G_τ A parameter expressing the magnitude of the variance of the random atmospheric delay field τ_0. G_τ contains the same information as r_0, the Fried parameter, but unlike r_0, G_τ is **not** a function of wavelength (optical frequency). G_τ includes no information regarding the **dynamics** of τ_0. See (42), (43), page 54.

H An intermediate result used in Appendix I. See (323), page 206.

$H(\omega)$ The frequency response of the **smoothing filter** which is used in the time-domain algorithm for maximizing Λ, shown in Figure 21. See (208), page 121.

I_0 The total photon detection rate, in counts per second. See page 128.

J An intermediate result used in Appendix I. Defined as the sum of the g_i (see above) over all i. See (323), page 206.

K The total number of detected photons being considered in the time interval of length T.

k_2 See (257), page 138. A correction factor based upon a combination of source statistics. Defined to be on the order of unity. Exactly equal to 1 in the narrowband case, or when $|V|$ is not a function of ν.

k_4 See (257), page 138. A correction factor based upon a combination of source statistics. Defined to be on the order of unity. Exactly equal to 1 in the narrowband case.

L_k The **complex likelihood** function for photon #k. An analytic function of τ, the **complex likelihood** is defined so that 1

plus its **real part** will yield the likelihood function itself. Using the LLA approximation, the log-likelihood λ_k is exactly given by the real part of L_k. See (80), page 70.

L_Σ The **net complex likelihood** function, formed as the (possibly weighted) sum of the complex likelihoods of a number of photons. See (93), page 75; (129), page 82. Using the LLA approximation, the net log-likelihood λ is exactly given by the real part of L_Σ. See (92), page 74.

p_i The portion of power form either photomixer input channel routed to output channel i. See (14), page 21.

$p(t)$ The series of **photon impulses** which is used in the time domain algorithm for maximizing Λ, shown in Figure 21. See (199), page 117.

P_i The Fourier cosine transform of $p(t)$, above. See (195), page 116.

R The covariance matrix of the estimated vector U, the cosine transform of the estimated $\tau(t)$ function. Parseval's theorem asserts that the **trace** of R yields (with proper normalization) the mean squared error of τ.

r_o The **Fried Parameter**. According to the widely accepted model introduced by D. L. Fried [1], the structure function of atmospheric turbulence varies as the $5/3$ power of baseline separation. r_0 specifies the separation at which the mean squared phase error due to atmospheric turbulence is 6.88 radians^2. r_0 varies with wavelength according to $\lambda^{6/5}$, as can be seen from (43), page 54.

SNR The **intrinsic** signal-to-noise ratio (233) is defined on the basis of the number of photons received in the standard time period T_0 as evaluated at the nominal optical frequency ν_0. Since quantum noise is given by the **square root** of photon flux, the signal-to-noise **ratio** increases according to the square root of photon flux, in the case of photon noise being dominant. See page 129.

$\mathrm{SNR}^{(b)}$ The signal-to-noise ratio especially defined for the case of estimation using **time bins**. See (125), page 81. Because this figure is proportional to T which may be **chosen**, it is not simply a characteristic of the physical circumstances, but is specific to the binning estimation procedure. However it is **analogous** to the intrinsic signal-to-noise ratio (see above) which **is** determined only on the basis of the statistics of the received photons.

$\mathrm{SNR}^{(\mathrm{CRIT})}$ The **critical** level of intrinsic signal-to-noise ratio, below which, the confidence of the correct solution is predicted to deteriorate. See (303), page 160.

T The time period over which the Fourier transform of $\tau(t)$ is performed (see (53), page 59), and/or over which photons are considered in the estimation problem. See page 99; also (75), page 66.

T_τ The single parameter controlling the **dynamics** of $\tau(t)$, the differential atmospheric delay, according to the simplified model proposed in (50). Combines the information on G_τ (or r_0) and V_0. See (51), page 56.

T_0 Simply defined as r_0/V_0, and thus a function of optical wavelength. Often referred to as the "atmospheric coherence time" or the "speckle lifetime." T_0 is best viewed as a time

constant expressing the time period over which the phase of interference subject to the random process $\tau(t)$ is expected to vary by an r.m.s. phase angle of 3.7 radians. Note that different authors have defined similar time periods which differ by a constant factor. See (52) and footnote 9 on page 57.

V "Fringe Visibility" or **normalized cross-correlation** between two light sources. A function of optical frequency ν. In most contexts, $V(\nu)$ refers to the cross-correlation of A_1 and A_2, the light that would be observed prior to entering the earth's atmosphere, and not subject to atmospheric or geometrical delays, **but reduced in magnitude** by dark counts, photomixer mode impurity, or other instrumental effects.

$\overline{V^2}$ Shorthand for the expectation of the squared magnitude of V (fringe visibility) averaged over all detected photons. See (225), page 128.

V_0 The (hypothetical) wind velocity which, acting upon $\tau_0(x, y)$, the random field of atmospheric delay **as a function of position**, would account for the **temporal** characteristics of τ_0 evaluated at a fixed position. See page 55.

x_1,x_2 The field amplitudes of the light entering the photomixer after having passed through the delay line. See (10), (2), page 15.

X_1,X_2 The field amplitudes of the light entering the telescopes. See Figure 1.

Terms:

Baseline	The vector separating the two telescopes receiving the light to be interfered.
Complex Likelihood	The **complex likelihood** is an analytic function of τ. It is defined so that 1 plus its **real part** will yield the likelihood function itself. It has the special property that the derivative of likelihood can be obtained directly from its **imaginary part** scaled by ν. See (80), page 70. See L_k, L_Σ above.
Confidence	Suppose some estimation procedure produces an estimate, a claimed maximum estimation error, and a figure of merit called Q. The **confidence** of the estimate is the *a priori* probability that an estimate using this procedure which yielded a figure of merit equal to Q, would in fact have an error no larger than claimed. In the case of the global path estimation problem for $\tau(t)$ in which the *a posteriori* probability consists of a number of possibilities, each of which is characterized as a mean path with a Gaussian covariance structure, then the **confidence** of any individual possibility is identical to the **total** *a posteriori* probability contained in that solution.
Confidence Index	As T grows, the confidence of the correct solution is expected to decay exponentially. The **confidence index** is a measure of the rate of this decay. See (277), page 152.

Exact Logarithm

Used to indicate derivations in which the LLA approximation (see below) has **not been employed** in the determination of λ and similar quantities. See, for instance, (168), page 102.

Fried Parameter

See r_0, above.

Likelihood

If x is an underlying process to be estimated and y is an observation on the basis of which x is to be estimated, then $P(y|x)$ is the **instrumental response** function. However **given** an observation y, then we call $P(y|x)$ **viewed as a function of** x, the **likelihood** of x given y. **Multiplying** the likelihood by the *a priori* density of x, yields the (un-normalized) *a posteriori* probability density of x given y, $k_1 P(x|y)$ (see (78), page 68). The **likelihood** function itself, however, is blind to the *a priori* characterization of x.

Linear-Log Approximation (LLA)

See (84), page 71. The approximation: $\log(1 + x) \approx x$, which is accurate for $|x| \ll 1$. See Figure 12. Used throughout this work to evaluate the logarithm of probabilities in a simplified but reasonably accurate form. Equations in which the LLA approximation is employed are punctuated with "(LLA)." Equations **not** using the LLA approximation are punctuated with "(EL)" for **exact logarithm**. Using the **exact logarithm** enhances the estimation performance by a modest amount only when $|V|$ is close to unity.

MAP

Maximum *a posteriori* **probability**. The MAP estimate refers to the estimation of an unknown quantity from observed data, based upon the maximization of the *a posteriori* probability density for the unknown quantity given the observation. The latter is determined by Bayes' Theorem (78) given the *a priori* probability of the unknown, the probabilistic model connecting the observable data to the unknown, and the observation itself.

Photomixer

A passive linear optical network with two (or more) input modes and at least as many output modes. A **balanced lossless 2-input** photomixer can be described by (16), page 22. One with output modes of equal power and equally separated phase is given by (18).

Scattering Matrix

See (10), page 20.

Signal-to-Noise Ratio (intrinsic)

See SNR, above.

Structure Function

The **structure function** $D_x(\Delta t)$ characterizes a stochastic process $x(t)$, where $x(t)$ must be **incrementally stationary**. $D_x(\Delta t)$ is defined as the **mean squared difference** between $x(t)$ and $x(t + \Delta t)$. Note that D_x contains the same information embraced by the autocorrelation function R_{xx}, when the latter exists (see (49), page 56.

White Light Fringe

Also **Central Fringe**. In an interferometer processing white (wideband) light, at the point where the net differential delay is zero, all wavelengths of light interfere **in phase** and reinforce the peak at that special point. As the bandwidth is decreased, then the prominence of the white light fringe decreases compared to the sidelobes. In a delay-dispersed photomixer, the **position** of the white light fringe is directly proportional to the temporal delay affecting the correlation between the two optical fields, τ.

Bibliography

[1] Fried D.L., Statistics of a Geometric Representation of Wavefront Distortion, **J. Opt Soc Am.** #55, 1965, page 1427

[2] A. Labeyrie, Interference Fringes Obtained on Vega with Two Optical Telescopes, **Astrophysical Journal** #196, March 1,1975, page L71

[3] D. Morancais, P. Nisenson, Phase Reconstruction in Optical Interferometry, **Optics Communications**, v. 67 #1, June 1,1988, page 39

[4] D. Buscher, Low Light Level Limits to Tracking Atmospheric Fringe Wander, Summary submitted to conference: **Quantum Limited Imaging and Information Processing**, Cap cod, Massachusetts June 12–13, 1989, Cosponsored by Optical Society of America, Air Force Office of Scientific Research

[5] M. M. Colavita, M. Shao, Atmospheric Phase Measurements with the Mk II and Mk II Interferometers at Mt. Wilson, **Proceedings of the Joint Workshop on High-Resolution Imaging from the Ground Using Interferometric Techniques,** Oracle Arizona, Jan. 12–15, 1987, Cosponsored by European Southern Observatory, National Optical Astronomy Observatories, page 205

[6] P. Nisenson, W. Traub, Magnitude Limit of the Group Delay Fringe Tracking Method for Long Baseline Interferometry, ibid., page 129

[7] S. Ridgeway, J. Mariotti, **Method for Multispectral Infrared Interferometry**, ibid, page 93

[8] W. A. Traub, M. G. Lacasse, N. P. Carleton, Spectral Dispersion and Fringe Detection in IOTA, Proc. SPIE v. 1237, **Amplitude and Intensity Spatial Interferometry**, 1990, page 145

[9] M. Colavita, Design Considerations for Very Long Baseline Fringe-Tracking Interferometers, ibid., page 80

[10] F. Peter Shloerb, Image Reconstruction Techniques for the Infrared-Optical Telescope Array, ibid., page 154

[11] R. D. Reasenberg, Kalman-Filter Fringe Tracking in an Optical Interferometer, ibid., page 172

[12] H. M. Dyck, et. al., Recent Progress on the Infrared Michelson Array, ibid., page 31

[13] M. Shao, M. M. Colavita, Status of the Mark III Interferometer, **Proceedings of the Joint Workhshop on High-Resolution Imaging from the Ground Using Interferometric Techniques**, Oracle Arizona, Jan. 12–15, 1987, Cosponsored by European Southern Observatory, Nation Optical Astronomy Observatories, page 115

[14] P. Lawson, A. Booth, The Fringe Tracking Servo in SUSI, **Very High Angular Resolution Imaging, Proceedings of the 158th Symposium of the International Astronomical Union,** Women's College, University of Sydney, Australia, Jan. 11–15, 1993, Ed. J. G. Robertson, W. J. Tango, page 184

[15] I. Tallon-Bosc et. al., Progress on GI2T, ibid., page 155

PhD. Thesis

Estimation and Tracking of Atmospheric Delay Noise in a Long-Baseline Optical Stellar Interferometer and Determination of the Expected Estimation Error

Jeffrey Meisner