Efficient nonlinear algorithm for envelope detection in white light interferometry

Kieran G. Larkin

Department of Physical Optics, The University of Sydney, NSW 2006, Australia

Received October 11, 1994; accepted September 19, 1995; revised manuscript received October 25, 1995

A compact and efficient algorithm for digital envelope detection in white light interferograms is derived from a well-known phase-shifting algorithm. The performance of the new algorithm is compared with that of other schemes currently used. Principal criteria considered are computational efficiency and accuracy in the presence of miscalibration. The new algorithm is shown to be near optimal in terms of computational efficiency and can be represented as a second-order nonlinear filter. In combination with a carefully designed peak detection method the algorithm exhibits exceptionally good performance on simulated interferograms.

Key words: interferometry, white light interferometry, interference microscopy, phase-shifting algorithm, low coherence, nonlinear filter, envelope detection, demodulation.

1. INTRODUCTION

Recently there has been much interest shown in the area variously known as white light interferometry (WLI),[1] coherence radar,[2] coherence probe/scanning,[3,4] correlation microscopy,[5–7] interference microscopy,[3,8,9] and low-coherence interferometry.[10] The main reason for such interest is that the ambiguity present in conventional monochromatic interferometers is not present in WLI. White light interferometers have a virtually unlimited unambiguous range, whereas their monochromatic counterparts are usually limited to not more than half a wavelength (slightly more for systems using high-aperture microscope objectives[11–13]). The close parallel between WLI and confocal (as well as conventional) microscopy was noted in the early literature[3] but has been largely ignored since. Like confocal microscopy, WLI allows surface profiling with high accuracy over a large range, but unlike confocal microscopy WLI allows the entire image field to be captured in one instant without the need for scanning apertures.

Although the objective of WLI can be simply stated—to find the location of peak correlation (or fringe visibility)—a problem arises because of the large three-dimensional sample data sets and the associated computational burden. A typical system[8] collects images containing 256×256 pixels over a series of 64 equispaced sections. If one processed these data using the exact Fourier method[5] to ascertain the peak correlation depth at each pixel, then at least 128×6 multiplications must be evaluated at each pixel, resulting in approximately 5.6×10^7 multiplications. A new algorithm that is many times faster (approximately eight times faster on the above data set) than the Fourier method is developed from a generalized form of the well-known five-step phase-shifting algorithm.[14,15] The application of a spatial carrier phase-shifting algorithm to WLI is, I believe, novel, although a temporal phase-shifting algorithm using three full data sets has been used previously.[2] The original aim of this paper was to investigate the suitability of phase-shifting algorithms for white light fringe analysis, but as the analysis and the simulations progressed, it became clear that one particular algorithm neatly[16] combined the properties of simplicity and robust efficiency. The intent of this paper is to present the rather circuitous development of this new algorithm and to demonstrate its suitability for efficient white light interferogram analysis.

This paper is arranged as follows:

Section 2 outlines the essential details of the structure of white light interferograms.

Section 3 considers the elements of ideal envelope and phase detection.

Section 4 develops an approximation to the Hilbert transform method outlined in Section 3 that corresponds to the phase-shifting algorithm of conventional digital interferometry. Some observations regarding sampling are made.

Section 5 introduces a unique nonlinear algorithm that combines filtering and demodulation in an elegant way.

Section 6 compares envelope detection using the nonlinear algorithm with that of two common phase-shifting algorithms. Performance in the presence of extreme miscalibration errors is evaluated.

Section 7 looks at the interpretation of the calculated envelope and phase. The salient features of peak detection schemes are compared.

Section 8 concludes the paper.

A revealing comparison of several well-known algorithms with the new nonlinear algorithm operating upon simulated, noisy interferograms is included as an appendix.

2. STRUCTURE OF A WHITE LIGHT INTERFEROGRAM

The light intensity, g, measured in a white light interferometer that is spatially incoherent has the following form (see Chim and Kino,[6] for example):

$$g(x, y, z) = a(x, y) + b(x, y)c[z - 2h(x, y)] \times \cos[2\pi w_0 z - \alpha(x, y)]. \quad (1)$$

0740-3232/96/040832-12$06.00

Coordinates x and y correspond to the conventional transverse object or image coordinates, and the coordinate z indicates the axial location or defocus of the object. The quantity $a(x, y)$ is an offset related to the reference and object beam intensities. It is the reflected beam intensity that determines $b(x, y)$. The interferogram envelope function c is related to the spectral profile of the white light, and the spatial frequency of interference fringes in the z direction, w_0, is related to the mean wavelength of the light. A phase change of reflection that is due to the complex reflectance of the surface determines the parameter α. Many papers have assumed that $\alpha = 0$, although this is generally not the case. Arbitrary control of the fringe phase, α, using the geometric phase is now possible and has been recently demonstrated.[17] I shall use the term correlogram[2] to describe the function $g(z)$ when the emphasis is upon the z variation and its characteristic form of fringes within an envelope determined by spectral correlation.

The exact form of the envelope $c = c(z)$ is not critical, although it is usually approximated by a Gaussian function for simplification of the calculations, especially those in the Fourier domain. In interferometer systems without suitably matched reference and object paths $c(z)$ may not be symmetrical because of dispersion. Dispersion has been ignored in the following analysis.

A spatially incoherent source ensures that correlograms can be considered independent of (x, y) location. In a practical system intensity measurements are performed over a uniform array of x, y, and z values. The x and y array values are determined by the pixels of a CCD array. Generally CCD arrays of 256×256 pixels or 512×512 pixels are common, with even larger formats becoming popular. The sequence of values of z at which the intensities are sampled is determined by the sequence of positions of a piezoelectric transducer. The sampling in the x, y, and z directions must satisfy certain constraints. In the transverse directions these constraints can be summarized in terms of the conventional (x, y) image bandwidth. The sampling requirement in the z direction will be covered in more detail in Section 4.

Three-dimensional data sets need significant memory storage capacity. For example, a $256 \times 256 \times 64$ data sets at 1-byte resolution requires 4 Mbytes of memory. Memory cost is rapidly becoming much less significant in digital instrumentation; in fact, the renewed interest in WLI has been partly stimulated by these lower costs.

3. IDEAL ENVELOPE AND PHASE DETECTION

The usual purpose of WLI is to determine the profile [characterized by $h(x, y)$] of a surface too steep for monochromatic interferometry. Also of importance is the phase change on reflection [related to $\alpha(x, y)$], which is determined by the dielectric properties of the surface. Although WLI is often proposed for profiling of rough surfaces, it is important to note that surfaces must be smooth at the scale of the system resolution for both $h(x, y)$ and $\alpha(x, y)$ to be meaningful.

Figure 1 shows a typical correlogram intensity distribution as a function of z. In this particular instance the envelope is Gaussian. Generally it is possible to derive $h(x, y)$ and $\alpha(x, y)$ from a sequence of intensity samples [given by Eq. (1)] over a range of z values. Some insight can be gained if one first considers the (continuous) Fourier transform of Eq. (1) with respect to the z coordinate:

$$G(x, y, w) = \int_{-\infty}^{\infty} g(x, y, z)\exp(-2\pi i w z)\mathrm{d}z\,. \tag{2}$$

Hence

$$\begin{aligned} G(x, y, w) = {} & a(x, y)\delta(w) \\ & + \frac{b(x, y)}{2} C(w)\exp[-4\pi i w h(x, y)] \\ & * [\exp(i\alpha)\delta(w - w_0) + \exp(-i\alpha)\delta(w + w_0)]. \end{aligned} \tag{3}$$

The symbol $*$ indicates one-dimensional convolution.[18] Spatial frequency in the z direction is denoted by w, and $\delta(w)$ is the Dirac delta function. The Fourier transform of $c(z)$ is $C(w)$. Explicit (x, y) variation can be ignored in the following analysis as long as it is remembered that the calculations are performed over an array of points (x, y) in the sampled data.

Equation (3) can be rewritten as

$$\begin{aligned} G(w) = {} & a\delta(w) + \frac{b}{2}\exp[+i(\alpha + 4\pi w_0 h)] \\ & \times C(w - w_0)\exp(-4\pi i w h) \\ & + \frac{b}{2}\exp[-i(\alpha + 4\pi w_0 h)] \\ & \times C(w + w_0)\exp(-4\pi i w h)\,. \end{aligned} \tag{4}$$

In terms of phase and modulus

$$G(w) = |G(w)|\exp[i\phi(w)], \tag{5}$$

with

$$|G(w)| \cong a\delta(w) + \frac{b}{2}|C(w - w_0)| + \frac{b}{2}|C(w + w_0)|, \tag{6}$$

$$\phi(w) \cong \begin{cases} -\alpha - 4\pi(w - w_0)h & w > 0 \\ +\alpha - 4\pi(w + w_0)h & w < 0 \end{cases}. \tag{7}$$

Thus relation (6) shows that $G(w)$ has an impulse at the origin and sidelobes centered at frequencies $w = \pm w_0$. A typical plot of the modulus, $|G(w)|$, is shown in Fig. 2. The approximate equality in the previous equations is achieved when there is minimal overlap of the sidelobes.

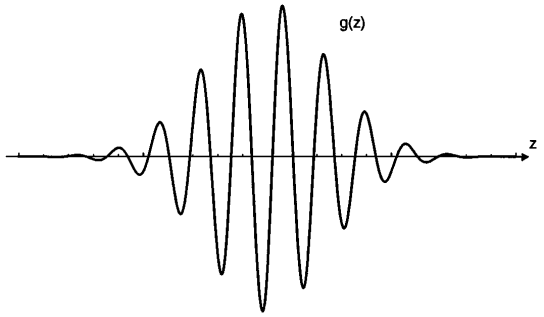

Fig. 1. Typical white light correlogram, $g(z)$. Note the fringe phase offset $\alpha = \pi/4$.

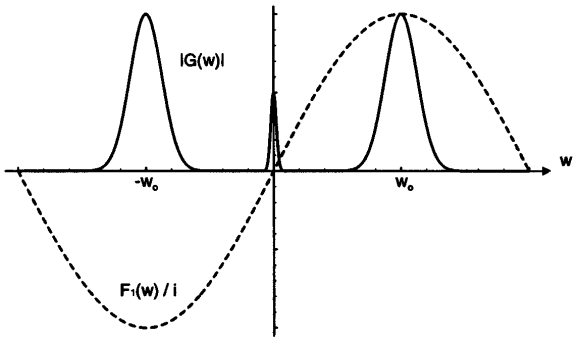

Fig. 2. Modulus of the Fourier-transformed correlogram, $|G(w)|$. The normally large dc term has been reduced in magnitude to fit within the chosen scale. Also shown for comparison is the FT $F_1(w)$ of the finite-difference filter $f_1(z)$ with separation $\Delta = 1/4w_0$.

Generally the impulse is somewhat spread out by noise and a variation of a with z. The two lobes in Fig. 2 are well separated; that is to say, the separation is typically greater than the bandwidth of the lobe. The bandwidth is inversely related to the envelope width in the spatial domain. Careful analysis of the phase function ϕ reveals that the two parameters of interest, h and α, can be simply extracted from the slope and the intercept of the linear portion of the curve. Nevertheless one must take care in performing a linear regression on the phase. Only in the regions where $|G(w)|$ has significant, nonzero values will the phase have meaningful values. A weighted least-squares fit (LSF) to the phase using $|G(w)|^2$ as the weight automatically gives good estimates of h and α. In this context h actually corresponds to the position of the centroid of the envelope squared, as can be shown by a Fourier correspondence theorem.[19] Unfortunately, calculation of h and α using the above method is computationally intensive. The fast Fourier transform (FFT) of a real array alone requires (of order) $N \log_2 N$ floating-point multiplications followed by N multiplications for an optimized LSF over the positive frequency region. A weighted fit requires an additional $2N$ multiplications. Here N is the total number of samples in the z direction (typically $N = 64$). The calculation is then repeated for each element in the x–y array (typically 256×256).

Conceptually, perhaps the easiest way to obtain the phase and the envelope is by use of the transform technique that has been outlined in several papers.[5–7] Briefly, the method entails fast Fourier transformation of the raw data (in the z direction) and then removal of the negative- and zero-frequency components. Finally the transform data are recentered at the midpoint of the sidelobe and then inverse transformed. In fact, the very same technique is better known as the FT method of fringe analysis in interferometry. The signal $s(z)$ that results has the form

$$s(z) = bc(z - h)\exp[i(\alpha - 4\pi w_o h)]. \tag{8}$$

The modulus of $s(z)$ is the envelope that is required, and the argument of $s(z)$ contains the phase offset α. The method outlined above is even more computationally intensive than the LSF method because forward and inverse FFT's are required in addition to the squaring operations required to determine the modulus.

More recently a much improved, computationally efficient method of determining the envelope has been developed.[8] The method relies on a real-space implementation of the previous double FFT method. The crucial point is the introduction of the Hilbert transform (HT) kernel as a digital finite impulse response filter.[20] Computational efficiency is gained by the realization that an approximate HT can give near-perfect results for typical interferograms. A real-space implementation of the HT has also been proposed for two-dimensional interferogram analysis by Zweig and Hufnagel.[21] As I shall show in the following sections of this paper, the real-space implementation can be greatly simplified, so much so that the final algorithm requires only two multiplications per point to produce the envelope squared at each point. In terms of multiplications alone this represents a lower limit upon numerical envelope detection using quadrature functions.

4. APPROXIMATIONS TO HILBERT TRANSFORM ENVELOPE DETECTION

The perfect HT can be considered equivalent to a wideband 90° phase-shift operation,[21] but, as we have already seen, a typical white light interferogram is a bandlimited signal. That is to say, the spatial frequency content is limited to a region centered around a carrier frequency. The wideband property of the perfect HT is not required for such signals. Thus the constraints upon an approximation to the HT can be relaxed—it only must have a 90° phase shift over a limited frequency band. Outside this band the transform can have any phase shift, but the modulus of the response should be low, thus suppressing noise present outside the passband. In the case of discrete sampled data with white noise it is desirable for the transformer to have zero or near-zero response at frequencies below and above the passband. An important practical requirement for a numerical discrete envelope detector is computational efficiency. A discrete implementation standard of efficiency with which to compare any method is the method of Chim and Kino,[8] in which the main computational burden for N samples is due to the $6N$ multiplications and the N square roots required to obtain the envelope. All the above requirements can be met by a pair of quadrature filter functions well known to researchers working in the area known as phase-shifting interferometry. The crucial Fourier properties of these functions have been derived[22] and extended[23] but have not previously been applied to the analysis of white light interferograms. Freischlad and Koliopoulos[22] introduced the two filter functions $f_1(t)$ and $f_2(t)$, which are correlated with a generalized interference pattern $g(x, y, t)$ to produce two quadrature functions, the ratio of which gives the tangent of the phase sought by the technique while the envelope (or the modulation) is given by the root sum of squares. A well-known algorithm that uses five samples[14,15] has the discrete filter functions

$$f_1(t) = 2[\delta(t - \Delta) - \delta(t + \Delta)], \tag{9}$$

$$f_2(t) = -\delta(t - 2\Delta) + 2\delta(t) - \delta(t + 2\Delta). \tag{10}$$

Here Δ is the step between samples. Those familiar with signal processing or digital filtering may recognize

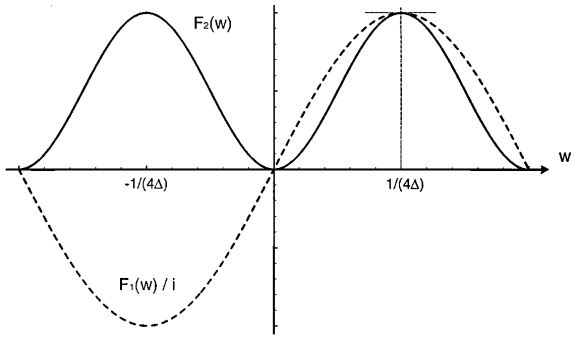

Fig. 3. FT's $F_1(w)$ and $F_2(w)$ of the five-step phase-shifting filters $f_1(z)$ and $f_2(z)$. Note particularly the stationary points at $w = \pm 1/4\Delta$.

these as simple finite impulse response filters. In the case of white light interferograms the temporal parameter t is replaced by a spatial coordinate parameter z. In fact, f_1 and f_2 are symmetrical and antisymmetrical linear phase digital filters, respectively.[24] The function f_1 has the property that its FT is an imaginary odd function, whereas that of f_2 is a real even function. Figure 3 shows F_1 and F_2, the transforms of f_1 and f_2. The choice of the functions f_1 and f_2 is important. There are a multitude of possible choices, with three samples being the minimum number required for such an algorithm. The five-sample algorithm, however, has several important properties not possessed by any other algorithms. The most important of these[23] is related to the matched gradients of the FT's F_1 and F_2 at the fundamental frequency $w = 1/4\Delta$. In Fig. 3 the gradients at the fundamental frequency can be seen to be near zero and slowly varying in this region. A consequence of the zero gradient is that both phase and envelope calculations for the five-step algorithm are somewhat insensitive to sample spacing (or step size) errors; essentially the filter responses change little in this region. Conversely, such an algorithm is insensitive to fringe spacing variation when the sampling is fixed.

At this point it is worth considering the sampling requirements for WLI. Most sampling schemes greatly oversample the correlogram data. A typical example, Chim and Kino[8] use approximately 8–10 samples per period. This exceeds the Nyquist criterion by at least a factor of 4. For a typical correlogram this sampling rate gives an unbalanced distribution of information in the Fourier frequency domain. The transform data are centered on the value $1/8\Delta$, which is one-quarter of the Nyquist frequency. A balanced distribution of frequency components requires approximately four samples per fringe and keeps the peak frequency midway between dc and the maximum frequency $1/2\Delta$.

More detailed analyses of the optimum sampling requirements for bandpass signals have been considered elsewhere.[25,26] It is generally agreed that it is the bandwidth of a narrow-band signal that determines the sampling frequency, not the carrier plus bandwidth as recently suggested by Caber.[27] However, the proposition that $4f_c + 2B$ is the minimum sampling frequency necessary to avoid aliasing when the signal is squared (here f_c is the fringe frequency, and B is the envelope bandwidth) fails to account for the fact that the aliasing that results when one is sampling at $4f_c$ occurs at frequency $4f_c - 2B$ and consequently has no effect after low-pass filtering (with a cutoff below $2f_c$).

In spite of this reduced bandpass sampling requirement, a criterion of just four samples per fringe is considered an adequate compromise here because it corresponds to the optimum sampling for a 90° step phase-shifting algorithm. A common misconception is that increased sampling frequency gives a proportionate increase in accuracy, but this is simply not the case. All the following analysis is based on a nominal sampling frequency of four samples per fringe (practically, this will cover the range from approximately three to eight samples per fringe) but is easily modified for other values.

Recently a method has been proposed to sample correlograms at a frequency determined by the bandwidth,[28,29] although such undersampling was previously used in 1991.[10] The technique has been called sub-Nyquist sampling. Instead of sampling the fringe pattern at four samples per carrier fringe the undersampling is by an odd integer factor, so there are $4/(2L + 1)$ samples per fringe, where L is typically 1 or 2. Undersampling of this kind avoids common aliasing problems so long as the envelope bandwidth alone is adequately sampled. An inevitable consequence of undersampling is that the allowable error in the initial prediction of the mean wavelength is inversely proportional to the undersampling factor. Undersampling also increases the relative bandwidth of the bandpass signal and can be expected to degrade the performance of demodulators in the presence of noise. There are some issues regarding the effects of sampling upon the envelope peak detection process that shall be discussed in Section 7.

Another issue worth mentioning is optimality with regard to computational efficiency. It is generally agreed (see Chim and Kino,[8] for example) that mathematical operations such as multiplication, division, and the evaluation of transcendental functions and square roots are much more significant to calculation time than the add and subtract operations. So, for the estimation of the computational burden of a numerical procedure, it is convenient to ignore the add and subtract operations and just count the other operations. Envelope detection algorithms can be compared in efficiency with an idealized numerical scheme that cannot be realized in practice but gives an idea of the limits to efficiency. A perfect envelope detection scheme could consist of the following steps:

1. Read in function values g_n and the perfect quadrature function values $\hat{g}_n$ for $n = 1 \rightarrow N$. (The sequence g_n is assumed to be zero mean.)
2. Calculate envelope values $e_n = \sqrt{g_n{}^2 + \hat{g}_n{}^2}$.

In this idealized scheme evaluation of the envelope function requires $2N$ multiplications and N square-root operations. The sequences g_n and $\hat{g}_n$ are considered known. In practice, $\hat{g}_n$ must be derived from the sequence g_n (which itself must be made zero mean), and more computational steps must be involved. As was mentioned above, the recent computational scheme of Chim and Kino[8] requires $6N$ multiplications and N square roots to yield the envelope, almost a factor of 3 below optimality.

The difficulty of implementing a perfect HT with the use of a real-space filter function is closely related to the well-known problem in communication theory of implementing a wideband 90° phase shifter for single sideband modulation.[30] One way around this is to start with the original function $g(z)$ and produce two new functions $g_a(z)$ and $g_b(z)$. The two new functions are in quadrature to each other, but the phase relationship with $g(z)$ is not constrained. This gives an extra degree of freedom that allows a practical implementation using realistic available electronic components. For example, instead of requiring a 90° phase shifter, one would consider a +45° and a −45° shifter. In terms of numerical filter functions it is in principle trivial to derive a −45° shifter once the +45° shifter is known. Consider a real filter function $f(z)$ and its FT $F(w)$. This can be conventionally represented as

$$f_1(z) = f(z) \Leftrightarrow F(w) = |F(u)|\exp[i\chi(w)]. \tag{11}$$

The symbol ⇔ represents Fourier transformation, and $\chi(w)$ is the FT phase. A filter function with the opposite phase shift is simply

$$f_2(z) = f(-z) \Leftrightarrow F^*(w) = |F(u)|\exp[-i\chi(w)]. \tag{12}$$

In both cases the modulus of the frequency response is the same, $|F|$. The two functions above, f_1 and f_2, are the bases for a whole series of quadrature functions that are linear combinations of f_1 and f_2. For example, a zero phase function f_0 can be defined as

$$f_0(z) = f(z) + f(-z) \Leftrightarrow 2|F|\cos\ \chi\,. \tag{13}$$

In a similar way a 90° phase function can be defined as

$$f_{90}(z) = f(z) - f(-z) \Leftrightarrow 2i|F|\sin\ \chi\,. \tag{14}$$

In this case, although f_0 and f_{90} are exactly 90° in phase, their moduli are no longer necessarily equal at all frequencies. Compare these with f_1 and f_2, which have equal moduli but the phase difference is 2χ, which does not necessarily equal 90° at all frequencies. The function pair f_1 and f_2 or f_0 and f_{90} can be used to generate approximate quadrature pairs of functions from the correlogram $g(z)$ simply by correlation (or by convolution with z-reversed functions):

$$g_1(z) = f_1(z) \otimes g(z)\,, \qquad g_2(z) = f_2(z) \otimes g(z)\,. \tag{15}$$

The symbol ⊗ indicates the correlation operation.[18]

5. CALCULATION OF PHASE AND MODULATION USING PHASE-SHIFTING ALGORITHMS

The following analysis will consider continuous functions and continuous FT's. However, the techniques outlined are applicable to discrete sampled data and discrete FT's. Important differences between the continuous and the sampled case will be noted as necessary.

The idea of using spatial carrier phase-shifting algorithms for white light envelope detection has not, to my knowledge, appeared in the literature. Typically such algorithms are used in interferometry to detect the phase of a wave front over a two-dimensional array of points when only one interferogram is available. Initially it is simpler to consider temporal phase shifting for the derivation of the new algorithm. In the case of the five-step algorithm five interferograms are captured by a digital imaging system. At each pixel location in the image one combines all five intensity values to extract the phase $\phi(x, y)$ at each pixel. The algorithm can be simply defined as

$$\phi(x, y) = \tan^{-1}\left\{\frac{2[I_2(x, y) - I_4(x, y)]}{-I_1(x, y) + 2I_3(x, y) - I_5(x, y)}\right\}, \tag{16}$$

where I_1 to I_5 are the interferogram intensities. The modulation at each point, $M(x, y)$, can also be calculated easily from

$$M(x, y) = 1/4[4(I_2 - I_4)^2 + (-I_1 + 2I_3 - I_5)^2]^{1/2}. \tag{17}$$

Both these formulas are exact when the phase shift between interferograms is 90°. For phase shifts in the region near 90° the errors in both ϕ and M are small because the first-order error terms cancel.[15] This error-compensating property of the five-step algorithm has made it popular in many digital interferometer systems. Other error-compensating algorithms exist, the most well known being that of Carré.[31] The interesting point about the Carré algorithm is that it compensates exactly for step errors and requires only four interferograms, whereas the five-step algorithm compensates only partially (second-order residuals are present). The problem with the Carré algorithm is that it is more complex and hence more time consuming to compute. A little known fact is that an exact compensating form of the five-step algorithm exists and is somewhat less complex than the Carré algorithm. The essential background to this exact compensating five-step algorithm is present in the paper of Hariharan *et al.*,[15] but it is not shown explicitly. It can be easily shown that the following phase and modulation expressions are exact:

$$\phi = \tan^{-1}\left\{\frac{[4(I_2 - I_4)^2 - (I_1 - I_5)^2]^{1/2}}{-I_1 + 2I_3 - I_5}\right\}, \tag{18}$$

$$M = \frac{(I_2 - I_4)^2\{[4(I_2 - I_4)^2 - (I_1 - I_5)^2] + (-I_1 + 2I_3 - I_5)^2\}^{1/2}}{4(I_2 - I_4)^2 - (I_1 - I_5)^2}. \tag{19}$$

The phase step between interferograms here is ψ, which can have any value not equal to an integral multiple of π (this condition also applies to most phase-shifting algorithms including the Carré algorithm). The denominator of Eq. (19) can also be expressed in a form that avoids problematic zero-by-zero divisions whenever $\sin\ \phi = 0$.

A serendipitous simplification of Eq. (19) leads to

$$4M^2\ \sin^4\ \psi = (I_2 - I_4)^2 - (I_1 - I_3)(I_3 - I_5)\,, \tag{20}$$

$$M^2 \propto (I_2 - I_4)^2 - (I_1 - I_3)(I_3 - I_5)\,. \tag{20a}$$

For values of ψ near 90° (and odd multiples thereof) the sine factor is near unity, and so M can be calculated with just two multiplications and one square-root operation; precisely the number of operators required for the ideal

detection scheme. The division operation [Eq. (19)] and its associated computational problems are neatly avoided. The right-hand side of relation (20a) defines what is now referred to as the five-sample-adaptive (FSA) nonlinear algorithm. Both the Carré algorithm and the conventional five-sample algorithm require an additional multiplication to calculate M. Remarkably the optimum sampling (Nyquist and sub-Nyquist) is predetermined by the maxima of $\sin^4 \psi$, which is entirely in accord with the bandpass signal sampling discussed earlier in this section.

The temporal phase-shifting analysis immediately preceding, strictly speaking, applies only where the modulation and the offset remain constant between interferograms. A technique known as spatial carrier phase detection[32,33] applies the same algorithms to a single interferogram. The intensity values I_1 to I_5 now represent adjacent pixels instead of separate interferograms. If the modulation and the offset are assumed to vary slowly across the interferogram, then the algorithms are approximately correct. Generally, it is necessary to introduce a large number of tilt (or carrier) fringes into the interferogram. This is because the maximum phase variation detectable is proportional to the mean phase variation, which in turn is related to the total number of fringes. The application of spatial carrier phase detection algorithms to white light correlograms initially appears counterintuitive because spatial phase techniques normally assume slowly varying offset and modulation, whereas white light correlograms, generally, have rapidly varying modulation along the z direction.

6. APPLICATION OF (SPATIAL CARRIER) PHASE-SHIFT ALGORITHMS TO ENVELOPE DETECTION

Perhaps the easiest way to investigate envelope detection with the use of phase-shift algorithms is by computer simulation.[34] Three representative algorithms are compared in this section. It should be emphasized that from here on the values $I_1 \rightarrow I_5$ represent adjacent sample values in the z direction. The Carré-derived envelope algorithm has been omitted because it requires three multiplications to be evaluated (also, preliminary simulations indicate poor performance). Two of the (five-step) envelope algorithms have been mentioned in the previous section. The third is based on the simplest three-step algorithm utilizing 90° phase steps.[35] The modulation factor in this case is

$$M(x, y) = 1/2[(I_3 - I_2)^2 + (I_2 - I_1)^2]^{1/2} \qquad (21)$$

and is correct only for exact 90° steps. The above algorithm has been selected because it requires only two squaring operations and one square root (both operations are possible with the use of fast lookup table computation).

A number of other three-, four-, or five-step algorithms[36] could have been chosen with and without error-compensating properties. Initial testing has shown that conventional phase-shifting algorithms have inferior performance to the algorithm defined in relation (20a). The three algorithms chosen here illustrate the performance of no error compensation [Eq. (21)], partial (first-order) error compensation [Eq. (17)], and exact error compensation [relation (20a)]. For convenience the algorithms shall be denoted (1), (2), and (3), respectively. Each algorithm has been applied in turn to a simulated white light correlogram without noise. The correlogram is shown in Fig. 1. The calculated envelopes are shown in Fig. 4 for the case in which the carrier, or rather the mean fringe spacing, is known precisely and the algorithm step size and sampling step are set at 90° or, equivalently, one quarter of the mean period. Figures 1 and 4 show continuous functions, but it can be easily shown that the discretely sampled case involves samples that occur at points located on the continuous curves.

Often the exact value of the carrier frequency is unknown before a measurement is made. The approximate value can be readily calculated from a knowledge of the spectral profile of the illumination and the numerical aperture of the imaging system (typically a microscope objective). The exact value depends on other system parameters and the spectral reflectance of the sample being measured. Therefore it cannot be known exactly before a measurement is made. Calibration of

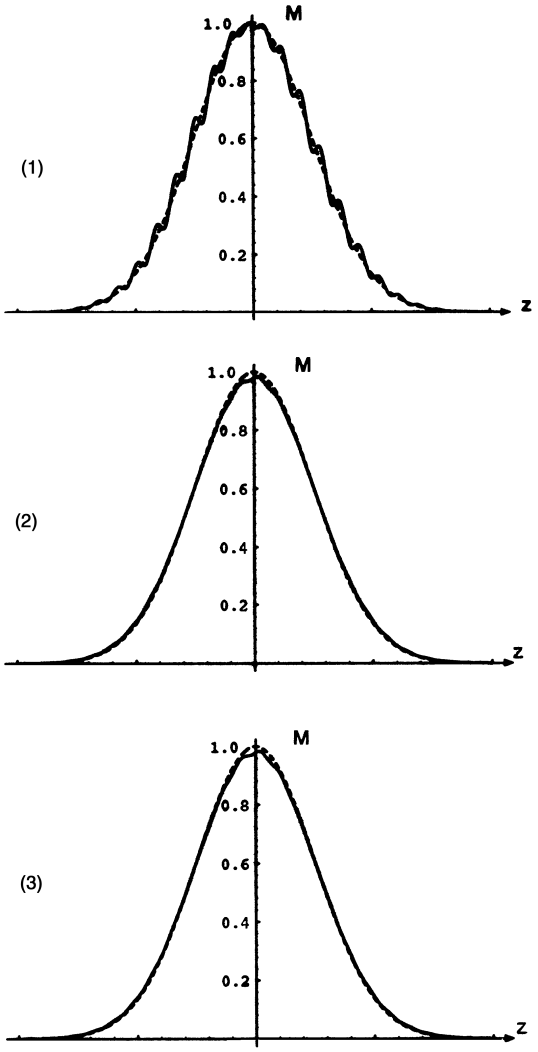

Fig. 4. Envelope detection for all three algorithms with use of a 90° step size. In this particular case the curves produced by algorithms (2) and (3) are the same. Algorithm (1) has noticeable fringe structure.

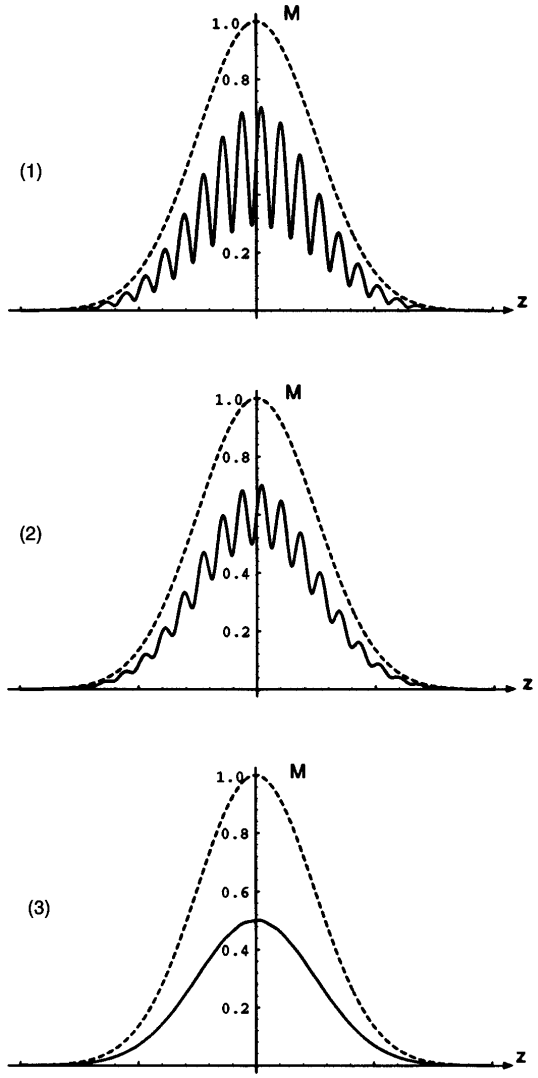

Fig. 5. Envelope detection for all three algorithms with use of a 45° step size. Both algorithms (1) and (2) have significant fringe structure visible. Algorithm (3) performs exceptionally well and has a 50% reduction as predicted.

the system for each sample can be a rather tedious addition to a measurement procedure. A preferable method requires a self-calibrating (or error-compensating) algorithm that works effectively over a range of carrier frequencies. To show the effects of carrier frequency variation, I have calculated the envelopes for two extreme values of the frequency. In Fig. 5 the frequency is 0.5 times its nominal value, and hence the samples are now 45° apart. Figure 6 shows the envelopes calculated when the frequency is 1.5 times its nominal value and the samples are thus 135° apart. In a system with white light in the range 400 nm to 700 nm the extreme frequency variation that is due to spectral effects alone is in the range 0.72–1.28 times nominal. Such extremes can be achieved only if the reflected light is narrow band at either 400 nm or 700 nm, in which case the envelope becomes broad and the FSA algorithm can again be expected to perform well. The dotted curves in Figs. 4–6 represent the ideal envelope.

A detailed analysis of the effect of noise in the correlogram will be the subject a subsequent paper. The performance of the FSA algorithm versus that of three other well-known white light profiling algorithms is covered in Appendix A; nevertheless some general observations can be made. The linear filtering properties of all three algorithms[22,37] as well as the perfect HT method can be determined. To summarize: (1) is a high-pass filter, (2) and (3) are bandpass filters centered on the nominal carrier frequency, and the HT is a wideband (all-pass) filter. In the presence of zero-mean Gaussian white noise, algorithms (2) and (3) suppress spectral components of noise outside the signal bandwidth and can therefore be expected to perform well compared with both (1), which boosts high-frequency noise, and the HT method, which neither suppresses nor boosts noise.

7. INTERPRETATION OF CALCULATED CORRELOGRAM ENVELOPES

In the previous section three algorithms have been used to estimate the envelope of the white light correlogram. In all cases some fringe structure propagates through into

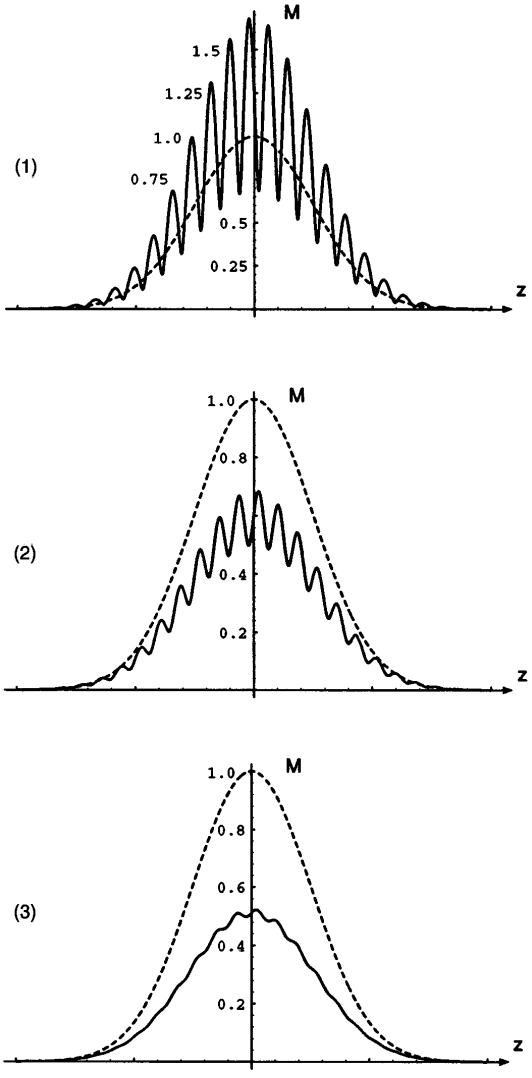

Fig. 6. Envelope detection for all three algorithms with use of a 135° step size. Again both algorithms (1) and (2) have significant fringe structure visible, whereas algorithm (3) shows only a trace of the second-harmonic fringe structure.

the calculated envelopes. This problem does not occur in the FT method and occurs only to a minuscule degree in the real-space HT technique. Of the algorithms, (3) has by far the smallest residual of fringe structure over the full range of sampling intervals from 45° to 135°. This factor is important in the process of finding the envelope peak, which is the crucial parameter. In Section 2 the height of the sample surface at any point, $h(x, y)$, was directly linked to the ideal envelope peak position $z_p(x, y)$:

$$z_p(x, y) = h(x, y). \tag{22}$$

Inevitably, the three algorithms tested only approximate the desired envelope. Applying a simple point-to-point peak detection process[38] to the calculated envelope can give significant errors with respect to the ideal peak position and requires significantly more than four samples per fringe to work correctly. A better way to find the peak is through use of the overall shape of the envelope around the approximate peak position. Simple curve fitting using three points has been proposed in an alternative approach to the envelope detection process.[27,39] In the region of the peak the calculated envelope can be expected to be well approximated by a Gaussian function, $\exp(-\beta z^2)$, where β characterizes the ideal envelope. A better estimate of the peak position can thus be obtained from a LSF to this function. For example, the nearest-neighbor and next-nearest-neighbor samples can be used for a five-point, symmetrical LSF to the function's exponent, $[\beta_0 - \beta(z - z_p)^2]$, which is a quadratic in z. The use of a symmetrical LSF greatly simplifies the calculation.[40,41] The full process is simply implemented if one takes the logarithm of the calculated envelope values and computes the peak position from an explicit solution of the symmetrical LSF.

The peak detection process can just as easily be applied to the envelope squared, as this produces only a factor of 2 in the envelope exponent. Thus the overall computation can be reduced by N square-root operations. The increased computational burden of the five-point LSF is only 19 operations (5 logarithms, 12 multiplications, and 2 divisions). Initial analysis indicates that the dominant error in the calculated envelope occurs at the second harmonic of the carrier frequency, which is typical of a second-order nonlinearity. As a result the conventional LSF gives a significant error in the peak prediction. However, it is possible to define a weighted five-point LSF that is insensitive to second-harmonic errors and thus gives much improved peak prediction. Five is the minimum number of points required to satisfy both LSF and harmonic criteria.[42] Equation (23) defines the five-point, frequency-selective, LSF peak predictor, where the symbol L_n represents the logarithm of the envelope value I_n and the distance z_p is measured from the third sample:

$$z_p = 0.4\Delta\left(\frac{L_1 + 3L_2 + 0L_3 - 3L_4 - L_5}{L_1 + 0L_2 - 2L_3 + 0L_4 + L_5}\right). \tag{23}$$

Figure 7 shows the result of applying the aforementioned peak detection process to the algorithm (3) envelope shown in Fig. 4. Again a continuous function analysis has been performed, but the result for a sampled function is just one point on the curves shown. In this particular instance the error of the proposed procedure is less than 1/20 sample for calculations with initial estimates of peak location within two samples of the actual value. The unweighted LSF is 1 order of magnitude less accurate. Careful scrutiny of Fig. 7 reveals a small bias in the predicted value; a bias related to the nonzero phase change (in this case $\alpha = \pi/4$).

Once the peak position of the envelope has been estimated, it is a straightforward task to find the phase at that position. The fully compensating five-step algorithm phase given in Eq. (18) can also be shown to have much smaller errors (which are due to miscalibration) than those of the two other algorithms. The phase at the estimated peak must be interpolated from actual calculations of the phase at sample positions on either side of the peak. The expected form of the phase near the peak is linear with respect to z. Hence a two-point linear interpolation will give a good estimate of the phase at the peak, in other words, an estimate of $\alpha(x, y)$. More points could be used for a linear least-squares estimate. However, the main difficulty in estimating the phase is the occurrence of phase discontinuities that are due to the modulo 2π restriction of the arctangent function. A phase discontinuity in the region of the peak renders the interpolation useless. To avoid such discontinuities, one must subtract a linear phase component from the calculated phase and reevaluate modulo 2π:

$$\phi_1 = \mathrm{mod}_{2\pi} \times \left(\tan^{-1}\left\{\frac{[4(I_2 - I_4)^2 - (I_1 - I_5)^2]^{1/2}}{-I_1 + 2I_3 - I_5}\right\} - 4\pi w_0 z\right). \tag{24}$$

The interpolation scheme can then be applied and the mean phase term added afterward to give α_p'. A rough estimate of w_0 is sufficient because errors cancel completely:

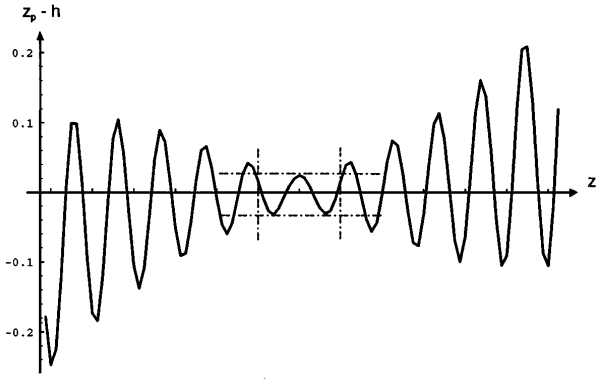

Fig. 7. Error in the predicted peak position with the weighted LSF defined by Eq. (23). Note that, for an initial estimate of peak position within two samples of the actual value (delimited by the vertical dotted lines), the peak prediction has only 1/20 sample error. The small bias in the predicted peak is related to the nonzero phase change on reflection. The range of z is just half that shown in Figs. 1 and 4–6. Clearly the error is related to the second harmonic of the original fringe. A second application of the peak detection process ensures that the estimate is within one-half sample, and errors decrease proportionately.

$$\alpha_p' = \phi_1 + 4\pi w_0 z\,. \tag{25}$$

This process works well in the region of the envelope peak for values of α not equal to $\pm\pi/2$ or $\pm\pi$. The number of additional significant computational steps for a two-point interpolation is 10. So for a data set with N values of z the total number of operations required for an estimate of $h(x, y)$ and $\alpha(x, y)$ is $2N + 29$. This method compares favorably with the HT method of Chim and Kino,[8] which requires $6N$ nontrivial multiplications just to produce the envelope squared. Typically $N = 64$, which means that the phase-shift algorithm method is nearly two and a half times faster than the HT kernel method. Essentially this speed gain is due to the remarkable adaptive properties of the fully compensating five-step algorithm of relation (20a) when applied to envelope detection. Also, by limiting the accurate estimates of h and α to small regions near the estimated peak of the envelope, one avoids much global calculation.

The comparisons in speed are valid only if the methods compared have similar accuracy. Certainly the procedure consisting of algorithm (3) followed by a weighted least-squares peak prediction given by Eq. (23) has an error of less than a small fraction of one sample interval. There appears to be no published work that considers the performance of any WLI peak prediction schemes in the presence of noise and other degradations. A preliminary analysis of this kind is included as Appendix A to indicate the veracity of the general principle discussed here. All methods that estimate the envelope can also utilize some form of LSF peak prediction and so, presumably, are capable of subsample resolution. In the past not all methods have used simple curve fitting to such advantage and therefore have a crude resolution limited to half a sample at best.

A recent paper by Caber[27] developed a communication theory approach to the interferogram envelope detection (see also Liu *et al.*[43]). The well-known demodulation process of a bandpass filter followed by a square-law nonlinearity followed, in turn, by a low-pass filter is implemented as a sequence of digital filters. Although details are not given, the known computational efficiency of two digital infinite impulse response high-pass or low-pass filters[20]in series with a squarer is lower than that of the compensating five-step algorithm outlined in the preceding sections. The availability of digital signal processing boards with special digital filter hardware may counterbalance the lower efficiency in practice. A minimum of just eight frames of data need to be stored at any moment in this scheme, compared with ten frames needed for the FSA algorithm proposed here. An accuracy of 1/25 sample spacing is claimed for this method.[39] The accuracy of this technique has not been tested in the appended simulation because details of the infinite impulse response filters used have not been disclosed in the open literature.

Some final remarks about the potential accuracy of the sub-Nyquist sampling method of de Groot[13] follow. The method relies upon a best fit to the phase gradient calculated from the FT of the sampled data. The phase gradient at the carrier frequency is easily shown to be proportional to the first moment (or centroid) of the envelope by a well-known Fourier correspondence theorem. Similarly the weighted LSF to the phase gradient (weighted by the magnitude squared) is proportional to the centroid of the square of the envelope.[19] Several authors have studied the effects of sampling upon centroid estimation,[44–46] essentially concluding that sampling must satisfy the bandpass sampling requirements mentioned in Section 4. The main point, however, is that phase gradient estimation (equivalently FT centroid estimation) is quite distinct from peak detection. In statistics the centroid is known to be susceptible to noise, that is to say, it is not a robust estimator. The effect of noise upon the centroid increases with the interval over which the centroid is evaluated. In contrast the peak prediction schemes outlined earlier depend only upon values of a distribution near the peak. A balanced assessment of the two techniques must compare accuracy versus computational complexity. Both the centroid of the envelope and the centroid of the squared envelope have been calculated for simulated data sets, and the results are presented in Appendix A.

A Fourier description of the mechanism defined by relation (20a) shows some similarities to the Caber method. The essential difference is that the Caber method uses two conventional infinite impulse response filters to remove low frequency and the second-harmonic components produced by the square-law nonlinearity, whereas the FSA nonlinear filter bandpass filters the signal and then shifts it to dc (i.e., demodulates) in one operation. From the point of view of classification the new procedure can be seen as a (nonlinear) second-order polynomial (Volterra series) digital filter[47] followed by least-squares peak prediction. The filter can be defined in general terms as a finite-difference operation followed by a nonlinear difference operation. There are similarities to the quadrature receiver (see, for example, Whalen[48]), except that the sine and cosine modulation terms are derived from the signal itself instead of an external source. Yet another classification known as the bilinear (quadratic with memory) transformation[49] covers such nonlinearities and offers a tractable analysis of noise propagation. The nonlinear filter can be explicitly defined by f_b, where

$$f_a(z) = g(z + \Delta) - g(z - \Delta) \qquad \text{finite difference}\,, \tag{26}$$

$$f_b(z) = f_a{}^2(z) - f_a(z - \Delta) f_a(z + \Delta) \qquad \text{nonlinear difference}\,. \tag{27}$$

Such a definition is amenable to Fourier analysis, and the following relations can be demonstrated:

$$F_a(w) = 2i\,\sin(2\pi w\Delta) G(w) \qquad \text{bandpass filter}\,, \tag{28}$$

$$F_b(w) = F_a(w) * F_a(w) - [\exp(-2\pi i w\Delta) F_a(w)] * [\exp(2\pi i w\Delta) F_a(w)]\,. \tag{29}$$

The last equation represents zero- and second-harmonic generation (from autoconvolution), with out-of-phase components canceling at the second harmonic.

8. CONCLUSION

A simple but highly effective method for envelope detection in white light correlograms has been introduced and demonstrated. The speed that is due to increased com-

putational efficiency is between two and three times that of the real-space HT technique. Combined with a new procedure for peak prediction using a weighted LSF algorithm that removes the residual second-harmonic error in the envelope, the fully compensating five-sample envelope detection algorithm is remarkably simple yet is effective over a wide range of carrier frequencies. The simulations presented used a fixed (but wide) bandwidth and show that repeatable subsample accuracy can be attained for envelope peak detection. In fact, the FSA algorithm, when combined with recursive peak detection, gives exceptionally good results compared with a number of conventional (and relative inefficient) WLI profiling algorithms. In terms of multiplication operations the new algorithm has been shown to be near the ideal limit of two multiplications per sample, suggesting that any further speed improvements from other methods can only be marginal. In situations in which the bandwidth is small enough the proposed algorithm can be combined with sub-Nyquist sampling to improve efficacy further.

The method is not limited to WLI and is applicable to any bandpass signals where either the envelope or the phase, or both, need to be detected. Optical measurement techniques such as confocal interferometry and spatial carrier phase-shifting interferometry could benefit from such a method.

Note added in final revision: A number of algorithms closely related to the FSA nonlinear algorithm in this paper have appeared recently in a number of papers related to amplitude and frequency demodulation of speech signals. The derivation of these algorithms uses a construct known as the energy operator,[50] which is defined for continuous, oscillating signals. Discrete approximation of the energy operator leads to a number of algorithms called discrete energy separation algorithms,[51–53] one of which is very much like Eq. (27). The derivation of the FSA nonlinear algorithm that I have presented here (from phase-shifting algorithms) is more general in that the separation of samples is initially assumed arbitrary [Eqs. (18) and (19)] rather than infinitesimal, as it is for the energy operator. The energy operator approach also assumes that signals have the dc, or background, component removed in a preprocessing operation. Notably the adaptive properties of the algorithm are particularly striking when applied to speech demodulation. Interestingly the idea of using these algorithms with undersampled (or, more correctly, bandpass-sampled) signals has been overlooked until recently.[54]

APPENDIX A: NUMERICAL SIMULATION OF ALGORITHMS APPLIED TO NOISY WHITE LIGHT INTERFEROGRAMS

This appendix contains a summary of results obtained by application of a number of well-known algorithms to simulated white light correlograms. The simulated data are available from the author as tagged image format files containing 512×64 pixels with 1-byte resolution. Sufficient information is provided for interested researchers to recreate simulated data with similar statistical properties. A full analysis of error propagation in WLI is in preparation.

1. Simulated Data

The fundamental parameters of the simulated data closely resemble the experimental data shown in a number of papers by Chim and Kino.[5–8] Essentially the correlogram is sampled at 64 locations in depth (z). The sampling occurs at a sample spacing of one-eighth the mean wavelength and three-eights the mean wavelength in the undersampling case, corresponding to step sizes of 90° and 270°, respectively. The envelope chosen corresponds approximately to a spectral range from 400 nm to 700 nm. In order that sufficient data exist for useful statistical inferences to be made, there are 512 independent measurements of the correlogram. To eliminate some systematic errors, the correlogram shifts z position progressively over the full 512 range. The total shift is one sample period over the full 512 range. Mathematically the image files used can be defined as

$$g(x, z) = \mathrm{INT}[128 + 100 \exp(-z_s^2/\sigma^2)\cos(4\pi z_s/\lambda_m) + n(x, z)]. \quad \text{(A1)}$$

The INT(·) function outputs the nearest integer to the argument input. The z sample locations are defined by $z_s = z - 32\Delta - x/512$. The sample spacing is defined by $\Delta = \lambda_m/8$ or $\Delta = 3\lambda_m/8$ in the undersampling case considered. The noise added to the interferogram is $n(x, y)$. The coordinates x and z take on only the following integer values: $x = l\Delta$ and $z = m\Delta$, where $0 \le l < 512$ and $0 \le m < 64$. For the selected spectrum $\sigma = 3.85\Delta$.

The noise $n(x, y)$ is zero-mean Gaussian-distributed random noise with a standard deviation (or rms) value specified in the range 0%–8% of the modulation value. The modulation is chosen to be 100 in this case. Note that even in the case of zero noise the quantization introduces some systematic (or correlated) noise. The actual noise characteristics for WLI are rather complex, being a combination of such factors as vibration, photon noise, and quantization, to name just a few. A full analysis requires a multidimensional statistical procedure. Zero-mean Gaussian noise has been chosen as a simple and well-defined starting point.

All the algorithms tested were set up to predict the z peak position at all 512 values of x. The ideal results lie upon a straight line in the x–z plane. The distribution of actual values around the best-fit (least-squares) line is computed in each case, and the standard deviation of the error is tabulated in Tables 1 and 2.

Four algorithms were tested:

(1) The FSA nonlinear envelope demodulator [relation (20a)] in conjunction with the specialized five-point peak detector of Eq. (23). The peak detector is applied twice if the first estimate is more than half a sample from the raw data peak. This iterative technique removes some systematic errors shown in Fig. 7.

(2) The Fourier–Hilbert transform method[5–7] is used to generate the envelope, and a simple three-point peak detector is used.

(3) The envelope is predicted by the exact Fourier method as for algorithm (2). The centroid of the envelope is then calculated. This is numerically identical to the instantaneous phase derivative method of de Groot.

(4) The square of the envelope is predicted by the exact Fourier method as for algorithm (2). The centroid of the

Table 1. rms Peak Location Error for Various Algorithms with Four Samples per Period

	Position Error in Sample Spacings			
rms Noise (%)[a]	FSA Algorithm	Fourier–Hilbert Algorithm	Envelope Centroid	Square of Envelope Centroid
0	0.010	0.003	0.000	0.000
1	0.034	0.056	0.175	0.027
2	0.064	0.112	0.322	0.061
4	0.126	0.233	0.553	0.160
8	0.248	0.479	0.856	0.476

[a]Expressed as a percentage of the modulation.

Table 2. rms Peak Location Error for Various Algorithms with 4/3 Samples per Period (3 Times Undersampling)

	Position Error in Sample Spacings			
rms Noise (%)[a]	FSA Algorithm	Fourier–Hilbert Algorithm	Envelope Centroid	Square of Envelope Centroid
0	0.055	0.166	0.305	0.170
1	0.058	0.166	0.430	0.175
2	0.065	0.168	0.625	0.204
4	0.094	0.176	0.909	0.396
8	0.172	0.206	1.190	1.046

[a]Expressed as a percentage of the modulation.

square of the envelope is then calculated. This is numerically identical to the weighted least-squares phase-derivative method of de Groot.

2. Results

The results for ideal sampling and 3 times undersampling are shown in Tables 1 and 2, respectively. It is interesting to note that the FSA algorithm gives the best results in the undersampling case and that for full sampling the FSA algorithm is superior for noise levels above 2%.

Note that, unfortunately, the method of Caber could not be simulated because the digital filter parameters needed to define the algorithm have not been disclosed.

ACKNOWLEDGMENTS

The development of this new algorithm was inspired by an impromptu remark by James C. Wyant about the need for a simple algorithm to analyze white light interferograms (February 1993, Eighth Australian Optical Society Conference, Sydney). I thank Colin Sheppard for the many enlightening discussions on short coherence length interferometry that stimulated this work. Gordon Kino kindly provided advice on the significance of computational (and general) aspects of WLI, and this further stimulated my interest. John Quartel kindly provided a suite of C programs utilized in the production and the analysis of simulated white light interferograms.

The author's e-mail address is k.larkin@physics.usyd.edu.au.

REFERENCES AND NOTES

1. P. A. Flourney, R. W. McClure, and G. Wyntjes, "White-light interferometric thickness gauge," Appl. Opt. **11,** 1907–1915 (1972).
2. T. Dresel, G. Häusler, and H. Venzke, "Three dimensional sensing of rough surfaces by coherence radar," Appl. Opt. **31,** 919–925 (1992).
3. M. Davidson, K. Kaufman, I. Mazor, and F. Cohen, "An application of interference microscopy to integrated circuit inspection and metrology," in *Integrated Circuit Metrology, Inspection, and Process Control*, K. M. Monahan, ed., Proc. SPIE **775,** 233–247 (1987).
4. B. S. Lee and T. C. Strand, "Profilometry with a coherence scanning microscope," Appl. Opt. **29,** 3784–3788 (1990).
5. S. S. C. Chim and G. S. Kino, "Correlation microscope," Opt. Lett. **15,** 579–581 (1990).
6. S. S. C. Chim and G. S. Kino, "Phase measurements using the Mirau correlation microscope," Appl. Opt. **30,** 2197–2201 (1991).
7. G. S. Kino and S. S. C. Chim, "Mirau correlation microscope," Appl. Opt. **29,** 3775–3783 (1990).
8. S. S. C. Chim and G. S. Kino, "Three-dimensional image realization in interference microscopy," Appl. Opt. **31,** 2550–2553 (1992).
9. M. Davidson, "Method and apparatus for using a two beam interference microscope for inspection of integrated circuits and the like," U.S. patent 4,818,110 (April 4, 1989).
10. B. L. Danielson and C. Y. Boisrobert, "Absolute optical ranging using low coherence interferometry," Appl. Opt. **30,** 2975–2979 (1991).
11. K. Creath, "Calibration of numerical aperture effects in interferometric microscope objectives," Appl. Opt. **28,** 3333–3338 (1989).
12. G. Schulz and K.-E. Elssner, "Errors in phase-measurement interferometry with high numerical apertures," Appl. Opt. **30,** 4500–4506 (1991).
13. C. J. R. Sheppard and K. G. Larkin, "Effect of numerical aperture on interference fringe spacing," Appl. Opt. **34,** 4731–4734 (1995).
14. J. Schwider, R. Burow, K.-E. Elssner, J. Grzanna, R. Spolaczyk, and K. Merkel, "Digital wave-front measuring interferometry: some systematic errors sources," Appl. Opt. **22,** 3421–3432 (1983).
15. P. Hariharan, B. F. Oreb, and T. Eiju, "Digital phase-shifting interferometer: a simple error-compensating phase calculation algorithm," Appl. Opt. **26,** 2504–2506 (1987).
16. Neat: cleverly effective in character or execution. *The Macquarie Dictionary*, 2nd ed. (The Macquarie Library, Macquarie University, Sydney, Australia 1991).
17. P. Hariharan, K. G. Larkin, and M. Roy, "The geometric phase: interferometric observations with white light," J. Mod. Opt. **41,** 663–667 (1994).
18. R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978).
19. S. C. Pohlig, "Signal duration and the Fourier transform," Proc. IEEE **68,** 629–630 (1980).
20. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1975).
21. D. A. Zweig and R. E. Hufnagel, "A Hilbert transform algorithm for fringe-pattern analysis," in *Advanced Optical Manufacturing and Testing*, L. R. Baker, P. B. Reid, and G. M. Sanger, eds., Proc. SPIE **1333,** 295–302 (1990).
22. K. Freischlad and C. L. Koliopoulos, "Fourier description of digital phase-measuring interferometry," J. Opt. Soc. Am. A **7,** 542–551 (1990).
23. K. G. Larkin and B. F. Oreb, "Design and assessment of symmetrical phase-shifting algorithms," J. Opt. Soc. Am. A **9,** 1740–1748 (1992).
24. R. E. Bogner and A. G. Constantinides, *Introduction to Digital Filtering* (Wiley, New York, 1975).
25. O. Brigham, *The Fast Fourier Transform*, 2nd ed. (Prentice-Hall, Englewood Cliffs, N.J., 1988).
26. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (Wiley, New York, 1978).

27. P. J. Caber, "Interferometric profiler for rough surfaces," Appl. Opt. **32,** 3438–3441 (1993).
28. P. de Groot and L. Deck, "Three-dimensional imaging by sub-Nyquist sampling of white-light interferograms," Opt. Lett. **18,** 1462–1464 (1993).
29. Zygo product brochure, "New view 100: 3D imaging surface structure analyzer," Zygo Corporation, Middlefield, Conn. (1993).
30. F. G. Stremler, *Introduction to Communication Systems* (Addison-Wesley, Reading, Mass., 1982).
31. P. Carré, "Installation et utilisation du comparateur photoélectrique et interferentiel du Bureau International des Poids et Mesures," Metrologia **2,** 13–23 (1966).
32. K. H. Womack, "Interferometric phase measurement using spatial synchronous detection," Opt. Eng. **23,** 391–395 (1984).
33. M. Kujawinska, "Spatial phase measurement methods," in *Interferogram Analysis: Digital Fringe Pattern Measurement Techniques*, D. W. Robinson and G. T. Reid, eds. (Institute of Physics, Bristol, UK, 1993).
34. Analytically, Eq. (20) is an approximate solution of the linear demodulation problem $g(z) = a + c(z)\cos(2\pi u_0 z + \alpha)$ using five equally spaced measurements to solve for c and having only second-order errors [related to the derivatives $(\partial c/\partial z)^2$ and $\partial^2 c/\partial z^2$]. The other algorithms have various first-order errors.
35. B. Bushan, J. C. Wyant, and C. L. Koliopoulos, "Measurement of surface topography of magnetic tapes by Mirau interferometry," Appl. Opt. **24,** 1489–1497 (1985).
36. There are a number of review papers and book chapters that consider the ever increasing range of phase-shifting algorithms. One of the more recent is K. Creath, "Temporal phase-measurement methods," in *Interferogram Analysis: Digital Processing Techniques for Fringe Pattern Measurement*, D. W. Robinson and G. T. Reid (Institute of Physics, Bristol, UK, 1993).
37. The nonlinear filter defined by relation (20a) can be represented by a bandpass filter (equivalent to a finite-difference filter) followed by a quadratic Volterra series filter. The bandpass filter has the same form as that of F_1 shown in Figs. 2 and 3 when the sampling is the nominal four per period. Interestingly the Hilbert envelope detector can be expressed as a quadratic filter with many terms, but strangely truncation to the first few terms does not give relation (20a).
38. P. Sandoz and G. Tribillon, "Profilometry by zero-order interference fringe identification," J. Mod. Opt. **40,** 1691–1700 (1993).
39. D. K. Cohen, P. J. Caber, and C. P. Brophy, "Rough surface profiler and method," U.S. patent 5,133,601 (July 28, 1992).
40. Closely related to Savitsky–Golay, or digital smoothing polynomial, filters. A LSF over a symmetrical domain causes many off-diagonal elements in the matrix equation to be zero, and hence matrix inversion is trivial. The resulting kernels have even or odd symmetry. See the following reference for more information.
41. W. H. Press, S. A. Teulolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge U. Press, Cambridge, 1992).
42. Typical peak detection algorithms involve a combination of samples in both numerator and denominator of a quotient resembling Eq. (23). Linear combinations are equivalent to correlation or convolution with a kernel function. In the Fourier domain the transformed signal is multiplied by the kernel transform. If the kernel is suitably chosen yet still satisfies the LSF criteria, then zeros (of the transform) can occur at certain frequencies and these frequencies are thus removed from the signal. Essentially filtering and peak detection have been combined into one. Equation (23) resembles the filtered signal quotient analyzed in detail in Refs. 22 and 23.
43. H.-H. Liu, P.-H. Cheng, and J. Wang, "Spatially coherent white-light interferometer based on a point fluorescent source," Opt. Lett. **18,** 678–680 (1993).
44. N. Bareket, "Undersampling errors in measuring the moments of images aberrated by turbulence," Appl. Opt. **18,** 3064–3069 (1979).
45. R. P. Loce and R. E. Jodoin, "Sampling theorem for geometric moment determination and its application to a laser beam position detector," Appl. Opt. **29,** 3835–3843 (1990).
46. B. F. Alexander and K. C. Ng, "Elimination of systematic error in subpixel accuracy centroid estimation," Opt. Eng. **30,** 1320–1331 (1991).
47. I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications* (Kluwer, Boston, 1990).
48. A. D. Whalen, *Detection of Signals in Noise* (Academic, New York, 1971), p. 200.
49. B. E. A. Saleh, "Optical bilinear transformations: general properties," Opt. Acta **26,** 777–799 (1979).
50. J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, N.M., 1990 (IEEE, New York, 1990), pp. 381–384.
51. P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," IEEE Trans. Signal Process. **41,** 3024–3051 (1993).
52. P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On amplitude and frequency demodulation using energy operators," IEEE Trans. Signal Process. **41,** 1532–1550 (1993).
53. A. C. Bovik and P. Maragos, "Conditions for positivity of an energy operator," IEEE Trans. Signal Process. **42,** 469–471 (1994).
54. K. G. Larkin, "Efficient demodulator for bandpass sampled AM signals," Electron. Lett. **32**(2) (1996).