Fourier description of digital phase-measuring interferometry

Klaus Freischlad* and Chris L. Koliopoulos

Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Received May 1, 1989; accepted November 18, 1989

A general description of phase measurement by digital heterodyne techniques is presented in which the heterodyning is explained as a filtering process in the frequency domain. Examples of commonly used algorithms are given. Special emphasis is given to the analysis of systematic errors. Gaussian error propagation is used to derive equations for the random phase errors of common algorithms.

1. INTRODUCTION

In the 1970's Bruning *et al.* introduced to interferometry a phase-detection technique for testing optical components that uses a solid-state detector array to map the optical wave-front phase over a large number of points. In this method, an interference pattern was phase shifted under computer control and discretely sampled repeatedly over different phase values. The digitized intensity values were then correlated with sines and cosines in a digital computer to determine the optical phase variations indicated by the interference pattern. Since then, a number of algorithms have been used to recover the wave-front phase of interference patterns, for which the emphasis is laid on algorithms using only a few samples of the phase-shifted interference patterns: three, four, or five samples 90 or 120 deg apart in overall phase. The measured intensities were derived mostly from two-beam interference patterns and were thus of a simple sinusoidal form. The signal $s(t)$ of a linear photodetector will be proportional to this intensity, that is,

$$s(t) = I[1 + V\cos(2\pi\nu_s t + \Phi)], \tag{1}$$

where V is the fringe contrast, Φ is the phase value to be determined (Φ is a function of position in the interference pattern), and ν_s is the modulation frequency of the signal obtained by varying the path difference between the test and reference beams in a linear fashion.

An often-used procedure[3,4] consists of acquiring four interferograms with a 90-deg constant overall phase shift between consecutive frames. At each pixel location in the interferogram we have the four values

$$s_0 = I[1 + V\cos(\Phi)] \quad \text{for } 2\pi\nu_s t = 0, \tag{2a}$$

$$s_1 = I[1 - V\sin(\Phi)] \quad \text{for } 2\pi\nu_s t = \pi/2, \tag{2b}$$

$$s_2 = I[1 - V\cos(\Phi)] \quad \text{for } 2\pi\nu_s t = \pi, \tag{2c}$$

$$s_3 = I[1 + V\sin(\Phi)] \quad \text{for } 2\pi\nu_s t = 3\pi/2. \tag{2d}$$

The desired phase Φ can be calculated from a combination of these four measurements as

$$\Phi = \arctan\frac{-s_1 + s_3}{s_0 - s_2}. \tag{3}$$

We can rewrite Eq. (3) as

$$\Phi = \arctan\frac{-s\left(\frac{T_s}{4}\right) + s\left(\frac{3T_s}{4}\right)}{s(0) - s\left(\frac{T_s}{2}\right)}, \tag{4}$$

where T_s is the modulation period of the signal, i.e., $T_s = 1/\nu_s$.

Now the measurement data $s_0, \ldots, s_3$ are considered samples of a varying signal $s(t)$ at points nT_s/N, $n = 0, \ldots, 3$, where N is the total number of samples ($N = 4$ in this case).

In practice, these samples can be taken at discrete points of the signal (phase-stepping techniques) or by integrating the time-varying signal into a number of buckets (integrated-bucket technique). In the latter case the signal $s(t)$ is no longer proportional to the intensity but is proportional to the intensity signal convolved with the integration window of the buckets. Equation (4) can still be used; only the modulation of the signal changes.[4]

As will be shown here, systematic errors of the phase measurement that are due to real, imperfect interferometric measurements, such as improper phase shifts, nonlinear photodetector response, and multiple beams producing nonsinusoidal signals as found in Fabry–Perot interferometers, multiple-beam Fizeau interferometers, and grating shearing interferometers, can be analyzed through Fourier methods applied to nonsinusoidal signals and sampling points deviating from nT_s/N.

The development of simple phase algorithms that use only a few samples was based on a method of finding a solution to a set of equations representing the phase-shifted sampled interference patterns. Presented here are analytical techniques that are a consequence of using a Fourier representation of the phase-measuring process to determine the errors in the phase calculation when these simple algorithms are used. Rather than pursuing lengthy numerical simulations,[5] one can use this theory immediately to derive the consequences of each algorithm when the sampled data are corrupted and to indicate which algorithm produces minimal error. An advantage of the Fourier description is the physical insight that allows one to think of the heterodyne process in the commonly known terms of Fourier theory. This Fourier description is not new[1]; however, for the algo-

0740-3232/90/040542-10$02.00

rithms with only a few samples, it has not been used in the literature to its full potential.

The effect of noisy intensity samples on the phase measurement has been described by the spectral methods of communication theory.[2,6] For digital heterodyne procedures with only a few samples, the method of Gaussian error propagation[7] is well suited to determine the random phase errors and will be used in this paper. Thus the analytical methods presented here are in a sense complementary to the previously applied techniques.

2. FOURIER DESCRIPTION OF THE DIGITAL HETERODYNE PROCESS

The basic principle of heterodyne phase measurements consists of mixing a periodic signal, with two sinusoidal signals 90 deg out of phase, from a local oscillator and applying a low-pass filter. Thus the real and imaginary parts of the Fourier component of the signal at the frequency of the local oscillator are determined. The phase of the signal is obtained as the arctangent of the ratio of the imaginary part and the real part.

In general, we write the heterodyne process as correlations of the real signal $s(t)$ with two real functions $f_1(t)$ and $f_2(t)$:

$$c_i = \int_{-\infty}^{\infty} s(t) f_i(t) \mathrm{d}t, \qquad i = 1, 2. \tag{5}$$

These are correlations $c_i(t)$ at $t = 0$. With the central ordinate theorem and the convolution theorem, and using the Hermitian property of the spectra of real functions,[8] we obtain from Eq. (5)

$$c_i = 2\,\mathrm{Re}\left[\int_0^{\infty} S(\nu) F_i^*(\nu) \mathrm{d}\nu\right], \qquad i = 1, 2, \tag{6}$$

where Re stands for the real part, the asterisk denotes the complex conjugate, and a capital letter denotes the Fourier transform of a function. The signal spectrum is first weighted by the spectra of the functions f_i, and then all positive frequency components are integrated. The functions f_i can therefore filter out certain frequency components of the signal; hence we shall call them filter functions.

To determine the phase of the input signal with respect to the filter functions, the ratio of the two correlations is defined as

$$r = \frac{2\,\mathrm{Re}\left[\int_0^{\infty} S(\nu) F_1^*(\nu) \mathrm{d}\nu\right]}{2\,\mathrm{Re}\left[\int_0^{\infty} S(\nu) F_2^*(\nu) \mathrm{d}\nu\right]}. \tag{7}$$

If the signal is periodic, it consists of the discrete spectrum

$$S(\nu) = \sum_{n=-\infty}^{\infty} s_n \delta(\nu - n\nu_s), \tag{8}$$

with

$$s_n = |s_n| e^{i\Phi_n},$$

where ν_s is the fundamental frequency of the signal. Equation (7) can be rewritten as

$$r = \frac{c_1}{c_2} = \frac{s_0 F_1(0) + 2\,\mathrm{Re}\left[\sum_{n=1}^{\infty} s_n F_1^*(n\nu_s)\right]}{s_0 F_2(0) + 2\,\mathrm{Re}\left[\sum_{n=1}^{\infty} s_n F_2^*(n\nu_s)\right]}. \tag{9}$$

In order to determine the phase Φ_m of the mth harmonic of the signal, one requires that the following conditions be fulfilled:

$$s_n F_1^*(n\nu_s) = -iA\delta(n, m), \tag{10a}$$

$$s_n F_2^*(n\nu_s) = A\delta(n, m), \tag{10b}$$

where i is the imaginary unit, A is an arbitrary constant, n and m are integers, and $\delta(n, m)$ is the Kronecker delta function. Under these conditions the spectra of the signal and the filter functions overlap only at one common point $m\nu_s$, and all other signal components are filtered out. The Fourier components of f_1 and f_2 at this frequency $m\nu_s$ are 90 deg out of phase (i.e., in quadrature); thus

$$F_1(m\nu_s) = iF_2(m\nu_s). \tag{11}$$

Equation (9) then reduces to

$$r = \frac{2|s_m||F_2(m\nu_s)|\sin(\Phi_m - \Psi)}{2|s_m||F_2(m\nu_s)|\cos(\Phi_m - \Psi)}, \tag{12}$$

with

$$F_2(\nu) = |F_2(\nu)| e^{i\Psi(\nu)},$$

and we have

$$\Phi_m = \arctan(r) + \Psi(m\nu_s). \tag{13}$$

The desired phase is thus equal to the arctangent of the ratio c_1/c_2, independent of the amplitude $|s_m|$ and also independent of the other frequency components of the signal (besides a usually unimportant constant offset Ψ stemming from the definition of the origin of the parameter t). Owing to the periodicity of the functions, the phase can be determined only modulo 2π. The above is the basis of quadrature phase detection of electrical signals.

We want to consider the special case of digital phase-sensitive detection, in which the signal is digitized at N discrete sampling points and the correlations with the filter functions are converted into discrete summations that are performed as numerical calculations in a computer:

$$c_i = \sum_{n=1}^{N} s(t_n) f_i(t_n), \qquad i = 1, 2. \tag{14}$$

The sampling is incorporated into the filter functions by multiplication with the sampling comb $\mathrm{d}(t)$:

$$\mathrm{d}(t) = \sum_{n=-\infty}^{\infty} \delta(t - n\Delta), \tag{15}$$

where Δ is the sampling spacing and the finite number of samples is expressed by multiplication of the filter functions with a square window function. The spectra of these sampled filter functions of finite extent are continuous.

The most common pairs of filter functions consist of the

quadrature sine and cosine functions. In this case, the correlations can be looked at as a simple Fourier transform or equivalently as a least-squares fit of a sine and a cosine to the signal data.[2,9,10] As was mentioned above, for an integrated-bucket technique the detected signal is equal to the convolution of the irradiance signal with the integration window of the buckets. This convolution can also reduce the frequency content of the detected signal if it is not truly represented by simple two-beam interference as represented by Eq. (1).

3. CASE EXAMPLES USING REPORTED PHASE ALGORITHMS

The Fourier approach defined in Section 2 will be used in this section to analyze, without approximation, various algorithms reported in the literature that are based on a minimum number of samples of the phase-shifted interference patterns used to compute the phase. The signal is taken to be of the form of Eq. (1).

The filter functions for these algorithms are a negative sine and a cosine of frequency ν_f, sampled at uniform spacings in different ways at as many as five points (see Fig. 1). For the simple single-frequency sinusoidal signal of Eq. (1), only the bias has to be suppressed; that is,

$$F_1(0) = F_2(0) = 0, \tag{16}$$

or, in other words, the average of the filter functions has to be zero (central ordinate theorem). At least three samples are needed if these two filter functions are to be realized with zero mean. This statement is equivalent to the requirement that at least three measurements are needed to obtain a solution for the three unknowns I, V, and Φ of Eq. (1).

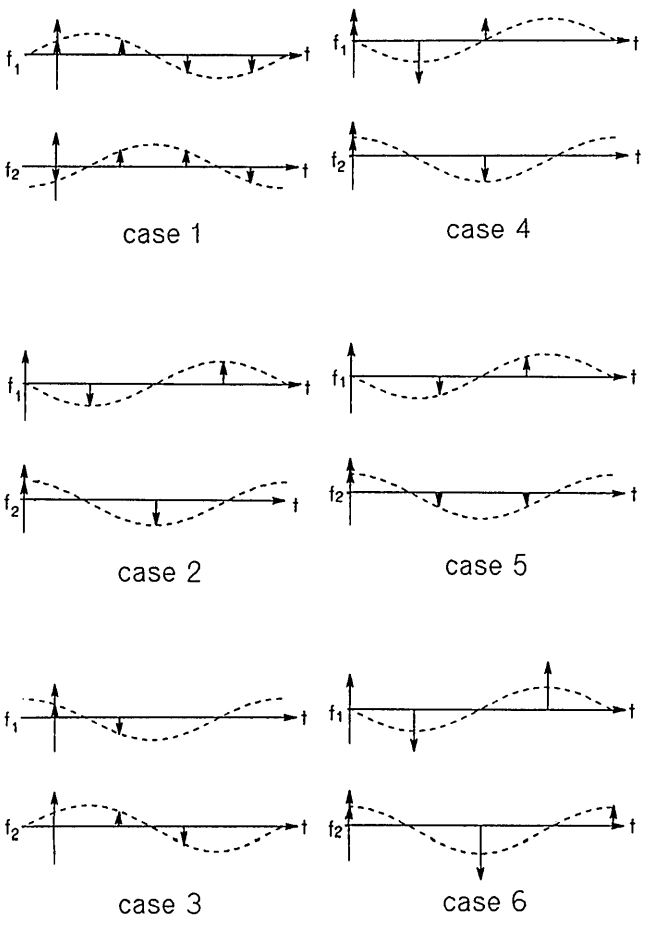


Fig. 1. Filter functions for common heterodyne procedures.

For case 1 the phase equation is given by[11,12]

$$\tan\Phi = \frac{s(0) + s\left(\frac{T_s}{4}\right) - s\left(\frac{T_s}{2}\right) - s\left(\frac{3T_s}{4}\right)}{-s(0) + s\left(\frac{T_s}{4}\right) + s\left(\frac{T_s}{2}\right) - s\left(\frac{3T_s}{4}\right)}; \tag{17}$$

that is, the signal is sampled at four points 90 deg apart. The filter functions for this case are

$$f_1(t) = \delta(t) + \delta\left(t - \frac{T_f}{4}\right) - \delta\left(t - \frac{T_f}{2}\right) - \delta\left(t - \frac{3T_f}{4}\right), \tag{18a}$$

$$f_2(t) = -\delta(t) + \delta\left(t - \frac{T_f}{4}\right) + \delta\left(t - \frac{T_f}{2}\right) - \delta\left(t - \frac{3T_f}{4}\right), \tag{18b}$$

where T_f is the period of the filter functions, $T_f = 1/\nu_f$. Their spectra are

$$F_1(\nu) = 4i\cos\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-\frac{3\pi}{4}i\frac{\nu}{\nu_f}\right), \tag{19a}$$

$$F_2(\nu) = 4\sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-\frac{3\pi}{4}i\frac{\nu}{\nu_f}\right). \tag{19b}$$

Performing the correlations given by Eq. (9), of the signal given by Eq. (1), with the filter functions yields

$$r = \frac{\tan\left(\Phi + \frac{3\pi}{4}\frac{\nu_s}{\nu_f}\right)}{\tan\left(\frac{\pi}{4}\frac{\nu_s}{\nu_f}\right)}, \tag{20}$$

where the spacing of the sampling points with respect to the signal period is now expressed by the ratio of the frequencies ν_s/ν_f.

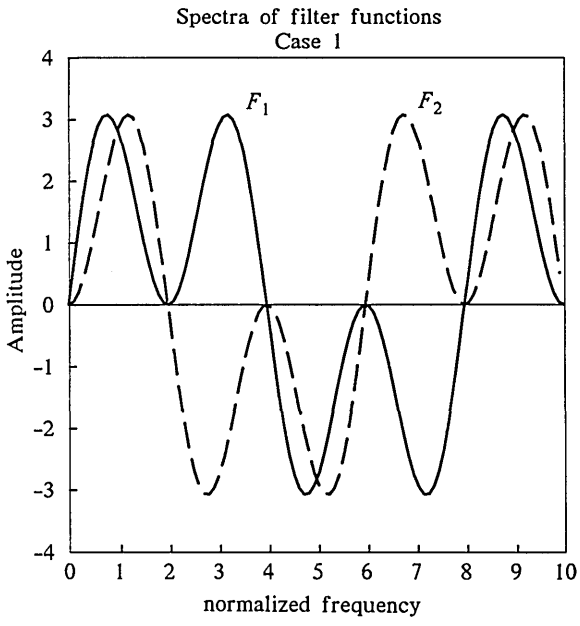
For $\nu_s = \nu_f$ (condition when the sampling is at 90-deg phase intervals) we obtain

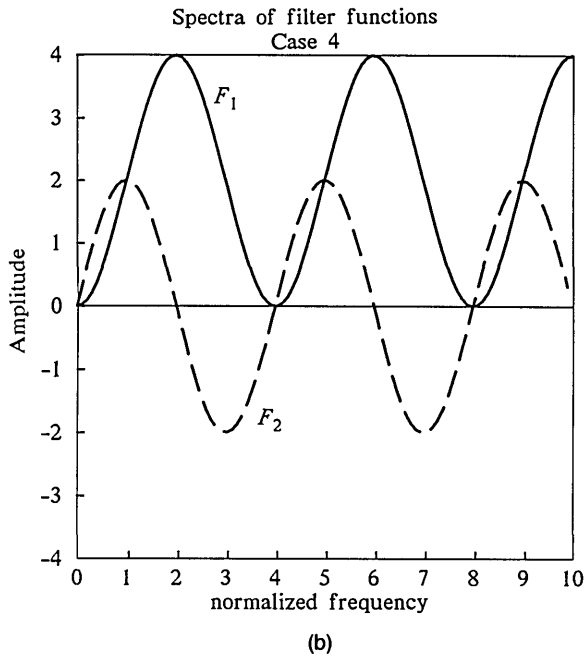
$$r = \tan\left(\Phi + \frac{3\pi}{4}\right). \tag{21}$$

The phase is thus computed by the arctangent of the ratio of the two correlations.

The spectra of the filter functions of this case are shown in Fig. 2(a) (without constant or common phase factors). We see that at $\nu_s = \nu_f$ both filter functions have the same amplitude. Since the functions have a constant phase shift of 90 deg between them, the requirements of Eq. (11) are fulfilled at this frequency. This is not the case at other frequencies, where the amplitudes of the two filter functions are different. However, at $\nu_s/\nu_f = 5, 9, \ldots$ (corresponding to phase steps of 450 deg, 810 deg, ...) the phase is calculated correctly, whereas at $\nu_s/\nu_f = 3, 7, \ldots$ (corresponding to phase steps of 270 deg, 630 deg, ...) negative phase values are obtained. For $\nu_s/\nu_f = 2, 4, \ldots$ (corresponding to phase steps of 180 deg, 360 deg, ...) both filter functions are zero, and hence no meaningful results can be obtained.

Case 2, as derived in Section 1 as a simple phase algorithm,[3,4] is similar to case 1. The phase equation is given by





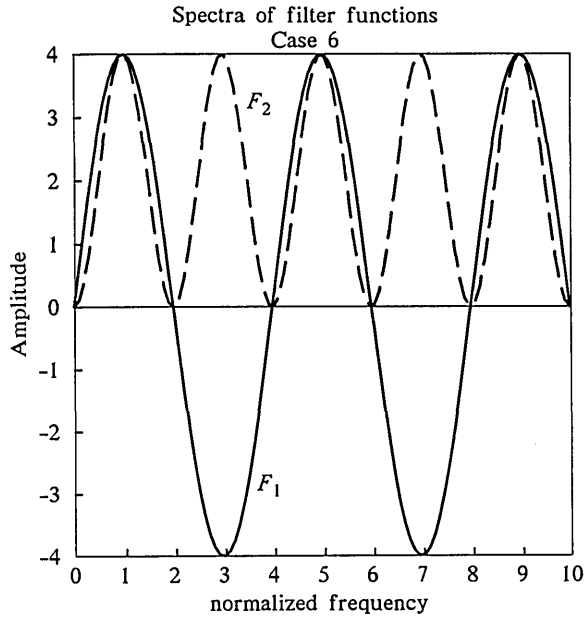


Fig. 2. Spectra of filter functions as a function of the normalized frequency ν_s/ν_f: (a) Case 1, (b) case 4, (c) case 6.

$$\tan\Phi = \frac{-s\left(\frac{T_s}{4}\right) + s\left(\frac{3T_s}{4}\right)}{s(0) - s\left(\frac{T_s}{2}\right)}. \tag{22}$$

Again we have four samples taken at 90-deg intervals, but now they are shifted with respect to the sine and cosine compared with case 1. Now the spectra of the filter functions are

$$F_1(\nu) = -2i \sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-i\pi\frac{\nu}{\nu_f}\right), \tag{23a}$$

$$F_2(\nu) = 2i \sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-i\frac{\pi}{2}\frac{\nu}{\nu_f}\right), \tag{23b}$$

and we have

$$r = \frac{\sin(\Phi + \pi\epsilon)}{\cos\left(\Phi + \frac{\pi\epsilon}{2}\right)}, \tag{24}$$

where ϵ is a detuning parameter defined as

$$\epsilon = \frac{\nu_s - \nu_f}{\nu_f}. \tag{25}$$

For $\nu_s = \nu_f$, Eq. (24) becomes

$$r = \tan(\Phi). \tag{26}$$

Case 3 uses only three samples at 90-deg steps[13]; the sine and the cosine are sampled only over half a period. The phase equation is now given by

$$\tan\Phi = \frac{s(0) - s\left(\frac{T_s}{4}\right)}{s\left(\frac{T_s}{4}\right) - s\left(\frac{T_s}{2}\right)}. \tag{27}$$

The spectra of the filter functions for this case are

$$F_1(\nu) = 2i \sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp\left(-i\frac{\pi}{4}\frac{\nu}{\nu_f}\right), \tag{28a}$$

$$F_2(\nu) = 2i \sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp\left(-i\frac{3\pi}{4}\frac{\nu}{\nu_f}\right), \tag{28b}$$

and

$$r = \frac{\sin\left(\Phi + \frac{\pi}{4}\frac{\nu_s}{\nu_f}\right)}{\cos\left(\Phi + \frac{\pi}{4}\frac{\nu_s}{\nu_f} + \frac{\pi\epsilon}{2}\right)}, \tag{29}$$

which for $\nu_s = \nu_f$ becomes

$$r = \tan\left(\Phi + \frac{\pi}{4}\right). \tag{30}$$

Case 4 also uses only three samples 90 deg apart.[4] For $f_1(t)$ special care had to be taken to suppress the signal bias. The phase equation is given by

$$\tan\Phi = \frac{s(0) - 2s\left(\frac{T_s}{4}\right) + s\left(\frac{T_s}{2}\right)}{s(0) - s\left(\frac{T_s}{2}\right)}. \tag{31}$$

The spectra of $f_1(t)$ and $f_2(t)$ illustrated in Fig. 2(b) are now

$$F_1(\nu) = 4 \sin^2\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp\left(-i\frac{\pi}{2}\frac{\nu}{\nu_f}\right), \tag{32a}$$

$$F_2(\nu) = 4i \sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\cos\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp\left(-i\frac{\pi}{2}\frac{\nu}{\nu_f}\right), \tag{32b}$$

leading to

$$r = \tan\left(\frac{\pi}{4}\frac{\nu_s}{\nu_f}\right)\tan\left(\Phi + \frac{\pi\epsilon}{2}\right), \tag{33}$$

and for $\nu_s = \nu_f$

$$r = \tan(\Phi). \tag{34}$$

In case 5 the sine and the cosine are sampled at three points 120 deg apart.[4] The phase equation is given by

$$\tan\Phi = \frac{\frac{\sqrt{3}}{2}\left[-s\left(\frac{T_s}{3}\right) + s\left(\frac{2T_s}{3}\right)\right]}{s(0) - \frac{1}{2}\left[s\left(\frac{T_s}{3}\right) + s\left(\frac{2T_s}{3}\right)\right]}. \tag{35}$$

The spectra of the filter functions are

$$F_1(\nu) = -i\sqrt{3} \sin\left(\frac{\pi}{3}\frac{\nu}{\nu_f}\right)\exp\left(-i\pi\frac{\nu}{\nu_f}\right), \tag{36a}$$

$$F_2(\nu) = 1 - \cos\left(\frac{\pi}{3}\frac{\nu}{\nu_f}\right)\exp\left(-i\pi\frac{\nu}{\nu_f}\right). \tag{36b}$$

The correlations then lead to

$$r = \frac{\sqrt{3}\sin\left(\frac{\pi}{3}\frac{\nu_s}{\nu_f}\right)\sin(\Phi + \pi\epsilon)}{\cos(\Phi) + \cos\left(\frac{\pi}{3}\frac{\nu_s}{\nu_f}\right)\cos(\Phi + \pi\epsilon)}, \tag{37}$$

which reduces, for $\nu_s = \nu_f$, to

$$r = \tan(\Phi). \tag{38}$$

Finally, in case 6 five samples are used at a separation of 90 deg.[14,15] As in case 4, the samples at the edges of the sampling interval had to be weighted by 0.5 in order to suppress the bias term of the signal. The phase equation is given by

$$\tan\Phi = \frac{-2s\left(\frac{T_s}{4}\right) + 2s\left(\frac{3T_s}{4}\right)}{s(0) - 2s\left(\frac{T_s}{2}\right) + s(T_s)}. \tag{39}$$

For the spectra of the filter functions we obtain now

$$F_1(\nu) = -4i \sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-i\pi\frac{\nu}{\nu_f}\right), \tag{40a}$$

$$F_2(\nu) = -2\left[1 - \cos\left(\pi\frac{\nu}{\nu_f}\right)\right]\exp\left(-i\pi\frac{\nu}{\nu_f}\right), \tag{40b}$$

and thus

$$r = \frac{2\sin\left(\frac{\pi}{2}\frac{\nu_s}{\nu_f}\right)}{1 - \cos\left(\pi\frac{\nu_s}{\nu_f}\right)}\tan\left(\Phi - \pi\frac{\nu_s}{\nu_f}\right). \tag{41}$$

For $\nu_s = \nu_f$, Eq. (41) becomes

$$r = \tan(\Phi). \tag{42}$$

Figure 2(c) shows the spectra of the filter functions of case 6. We see again that at $\nu_s = \nu_f$ both filter functions have the same amplitude. Since the functions have a constant phase shift of 90 deg between them, the requirements of Eqs. (10) are fulfilled at this frequency. Similar statements can be derived for the other phase-measurement algorithms illustrated in Fig. 1.

With the filtering interpretation in the frequency domain presented in this paper, more insight in the heterodyne process can be gained than by looking at the phase-measurement problem as the solution of a set of equations for the unknown phase. The effects of the finite number of samples can be analytically determined quickly, rather than laboriously by attempting computer simulations. This approach is also advantageous for the analysis of systematic errors of the phase-measurement processes, as will be shown next.

4. SYSTEMATIC ERRORS

Detuning Errors

Detuning, where $\nu_s \neq \nu_f$, is the most common source of error in the implementation of the phase algorithms described above. In the space domain, this means that the sampling of the signal is uniform but does not occur at a spacing of the proper fraction of the signal period. Detuning is referred to

as a linear phase-shift calibration error, in which, for example, the phase shift between samples might be 85 deg rather than 90 deg. For the sinusoidal signal of Eq. (1), we determined in Section 2 the results of the heterodyne Eq. (9) for the six discussed algorithms [see Eqs. (20), (24), (29), (33), (37), and (41)] from which the phase error can easily be determined as a function of Φ and ν_s. The effects of averaging techniques such as those introduced by Schwider[14] can also be analyzed.

Looking at Cases 1, 4, and 6, we see that a detuning results in a factor of δ in the ratio of the correlations that represents the phase to be determined:

$$r = \delta \tan(\Phi), \tag{43}$$

where the unimportant constant phase shift due to the definition of the origin of t is dropped. For the interesting condition of the sampling of the filter functions' frequency close to the signal frequency, δ is nearly unity, and the measurement error $\Delta\Phi$, given by

$$\Delta\Phi = \arctan(r) - \Phi, \tag{44}$$

can be approximated by

$$\Delta\Phi \cong \frac{\delta - 1}{2}\sin(2\Phi). \tag{45}$$

The phase error $\Delta\Phi$ is modulated with twice the original phase and has a maximum peak-to-valley value of $\delta - 1$. Figure 3 shows $\delta - 1$ for cases 1, 4, and 6. All curves go though zero at $\nu_s = \nu_f$, but the algorithm presented in case 6 has a minimum at that point, leading to much smaller errors when a detuning occurs than for the algorithm presented in cases 1 and 4. An intuitive illustration of this can also be seen by inspection of the spectra of the filter functions for cases 1, 4, and 6 (Fig. 2). In order for δ to be close to one (so that the phase error $\Delta\Phi$ is near zero), $F_1(\nu)$ and $F_2(\nu)$ have to be similar at values other than for $\nu_s = \nu_f$. For case 6, $F_1(\nu)$ is close to $F_2(\nu)$ over a wide range, whereas for case 1 and 4, $F_1(\nu)$ and $F_2(\nu)$ cross at $\nu_s = \nu_f$.

For cases 2, 3, and 5, detuning leads not only to different amplitudes of the numerator and the denominator in the ratio of the correlations but also to different phase offsets in the arguments of the sine and cosine, so that the measurement error of the phase calculation has to be determined separately for each case.

A procedure to determine the sampling spacing, or equivalently the signal frequency, was given first by Rowley and Hamon[11] and later by Carre.[12] Using our Fourier frequency-domain approach, we see that the signal frequency can be determined by correlating the signal with two filter functions with frequencies ν_f and $2\nu_f$, which are both odd with respect to a common center of symmetry. The unknown phase of the signal drops out because we correlate both times with an odd function. If the signal is even, i.e., $\Phi = n\pi$, n an integer, the ratio r becomes 0/0 and cannot be determined.

For the case of four samples, the fundamental and second-harmonic odd filter functions that will be used for the frequency estimate are given by

$$f_1(t) = \delta(t) + \delta(t - \Delta) - \delta(t - 2\Delta) - \delta(t - 3\Delta), \tag{46a}$$

$$f_2(t) = \delta(t) - \delta(t - \Delta) + \delta(t - 2\Delta) - \delta(t - 3\Delta), \tag{46b}$$

with

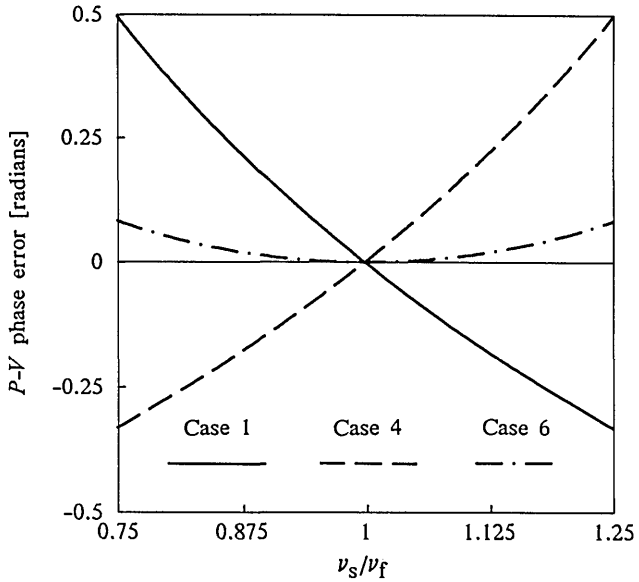


Fig. 3. P–V phase error ($\delta - 1$) as a function of the normalized frequency ν_s/ν_f.

$$\Delta = \frac{1}{4\nu_f} = \frac{T_f}{4}.$$

In the frequency domain we have

$$F_1(\nu) = 4i\cos\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\sin\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-\frac{3\pi i}{4}\frac{\nu}{\nu_f}\right), \tag{47a}$$

$$F_2(\nu) = 4i\sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\cos\left(\frac{\pi}{2}\frac{\nu}{\nu_f}\right)\exp\left(-\frac{3\pi i}{4}\frac{\nu}{\nu_f}\right). \tag{47b}$$

Using those functions as filter functions, we obtain as the ratio of the correlations from Eq. (9)

$$r = \frac{2}{1 - \tan^2\left(\frac{\pi}{4}\frac{\nu_s}{\nu_f}\right)}, \tag{48}$$

which can be solved for $\nu_s = \nu_f$. The solution for $\tan(\pi\nu_s/4\nu_f)$ can be used directly in Eq. (20) or (33), compensating for the error that is due to the detuning and for linear calibration phase-shift error. The five-bucket method given by Cheng and Wyant[16] also resorts to a correlation with two odd filter functions, one having twice the frequency of the other. In order to be able to generate these two odd functions at the sample points, we need at least four samples. This mirrors the fact that now we have to solve the four unknowns, I, V, Φ, and ν_s, and therefore need four equations. An extension of this method to more than four samples can be found in Ref. 17.

Phase Measurement of Harmonic Signals

We saw that the correlation Eq. (9) does not yield the correct phase if more than one selected frequency component of the signal is filtered and if at this frequency the two filter functions are not in quadrature.

Multiple detected frequency components typically arise from nonsinusoidal signals from multiple-beam interferometers.[25] For example, in a grating lateral-shear interferome-

ter the different diffraction orders of the gratings are superposed in the detector plane.[18] If at any point more than two orders overlap, the overall irradiance will consist of the sum of the interference patterns of all pairs of existing diffraction orders. By using a Ronchi ruling as the diffraction grating, the Fourier components of the signal of the lateral shear interferometer can be written as

$$s_n = \begin{cases} \Gamma_n'(g_0 g_{-n}^* + g_0^* g_n)\exp[-2\pi i W'(x,y)nd/\lambda] & n \text{ odd} \\ \Gamma_n \sum_{m=-\infty, m0, n}^{\infty} g_m g_{m-n}^* \exp(2\pi i\{W(x-md,y) & \\ \quad - W[x-(m-n)d, y]\}/\lambda) & n \text{ even} \end{cases}, \tag{49}$$

where the Γ_n and Γ_n' are coherence factors, W is the wave front to be determined and W' is its derivative, d is the shear distance between adjacent orders, and λ is the wavelength of the light. The g_n are the spectral components of the Ronchi grating, given by

$$g_n = 1/2 \sum_{n=-\infty}^{\infty} \frac{2}{n\pi} \sin(n\pi/2). \tag{50}$$

When the phase of s_1 is measured, the signal at the fundamental frequency yields

$$\Phi = -\frac{2\pi}{\lambda} W'(x,y)d. \tag{51}$$

The heterodyne procedure has to suppress all the other frequency components[18] to prevent errors due to the higher orders interferring at other phase relationships. This total suppression is not possible with the digital phase algorithms described here. Look at case 1 [Fig. 2(a)]: the spectra of the filter functions are such that we can filter out all even orders if $\nu_s = \nu_f$ but not the odd orders. The odd-order spectra weighted by $1/n$ that are due to the Ronchi ruling will contribute to the phase measurement. We now can easily see with this frequency-domain description of the phase-measurement process that it is better to use an integrated-bucket technique than a phase-stepping technique in connection with the Ronchi ruling because the integration of the buckets results in another weighting function, reducing the higher orders by a factor of $b_n < 1/n$ (see Fig. 4). To determine the error due to the higher orders, we obtain from Eq. (9) (including only the first and third orders)

$$r \cong \frac{\sin(\Phi) - \delta \sin(3\Phi)}{\cos(\Phi) + \delta \cos(3\Phi)}, \tag{52}$$

where δ is given by $b_n = (\Gamma_3')/(3b_3\Gamma_1')$ for case 1, leading to a phase error of

$$\Delta\Phi \cong \delta \sin(4\Phi). \tag{53}$$

If the phase varies from 0 to 2π, the phase error goes through four cycles with a maximum phase error of δ.

Multiple detected frequency components within the signal can also be due to detector nonlinearities. If the sinusoidal irradiance distribution of a two-beam interferometer is detected by a device with a nonlinear responsivity curve, higher harmonics will be present in the signal. Knowing the spectra of the filter functions, we can immediately tell which harmonics and therefore which nonlinearities will affect the result of the phase measurement and to what extent. The

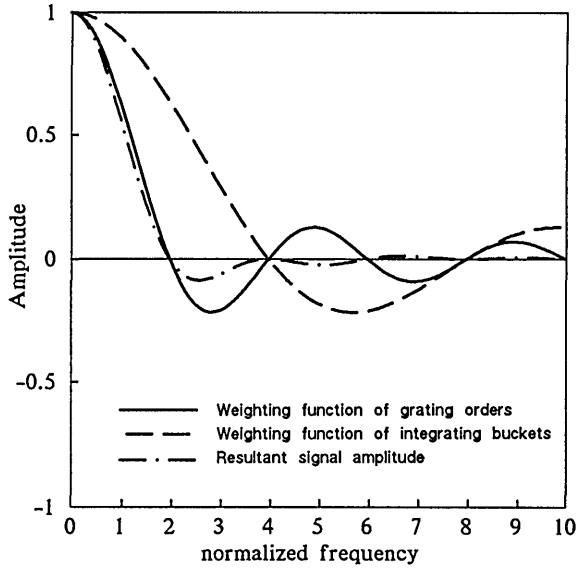


Fig. 4. Suppression of higher harmonics by the integrating-bucket technique.

simplest case is a quadratic response of the detector, generating an additional second harmonic of the signal that can be suppressed if the filter functions are zero at the frequency $\nu = 2\nu_f$, which is true for the phase algorithms in cases 1, 2, and 6. For these algorithms, a quadratic detector response does not introduce a phase-measurement error as long as $\nu_s = \nu_f$. Larger detector nonlinearities may introduce third harmonics, which can also produce errors in the phase measurement as described for the above example with the grating lateral-shear interferometer, depending on the phase algorithm used.

Another difficulty in some phase-measuring interferometers is nonlinearities of the phase shifter by which the phase difference between the test beam and the object beam is varied.[19] For this condition, the signal is not strictly periodic, and its spectrum is continuous. Equation (7) must be used to describe the heterodyne process, in which the integrals are extended over the product of two continuous spectra, leading to an error of the phase measurement.

In another interpretation, the nonlinearity is ascribed to the filter functions alone, and the signal is left unchanged. The filter functions are transformed in a nonlinear fashion, causing the signal to be sampled at unevenly spaced intervals. Using case 1 and a quadratic nonlinearity as an example, we obtain for the filter functions (recentered to drop the phase bias of $3\pi/4$)

$$f_1(t) = \delta\left(t + \frac{3\pi}{4} - 9\alpha^2\right) + \delta\left(t + \frac{\pi}{4} - \alpha^2\right) - \delta\left(t - \frac{\pi}{4} - \alpha^2\right) - \delta\left(t - \frac{3\pi}{4} - 9\alpha^2\right), \tag{54a}$$

$$f_1(t) = -\delta\left(t + \frac{3\pi}{4} - 9\alpha^2\right) + \delta\left(t + \frac{\pi}{4} - \alpha^2\right) + \delta\left(t - \frac{\pi}{4} - \alpha^2\right) - \delta\left(t - \frac{3\pi}{4} - 9\alpha^2\right), \tag{54b}$$

where the parameter α describes the quadratic component of the phase shift. The phase shift is shown in Fig. 5(a). It

leads to a shift of the sampling points of the filter functions as shown in Fig. 5(b) owing to the difference between the linear and quadratic shifts at the sampling points. The spectra of the filter functions are now

$$F_1(\nu) = 2i\left[\sin\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp(-2\pi i\alpha^2\nu) + \sin\left(\frac{3\pi}{4}\frac{\nu}{\nu_f}\right)\exp(-2\pi i 9\alpha^2\nu)\right], \quad (55a)$$

$$F_2(\nu) = 2\left[\cos\left(\frac{\pi}{4}\frac{\nu}{\nu_f}\right)\exp(-2\pi i\alpha^2\nu) - \cos\left(\frac{3\pi}{4}\frac{\nu}{\nu_f}\right)\exp(-2\pi i 9\alpha^2\nu)\right]. \quad (55b)$$

Equation (9) then gives an analytical solution for the phase calculation as a function of phase, detuning, and quadratic component.

From Eqs. (55) we see that for $\nu_s = \nu_f$ Eq. (11) is fulfilled. Therefore the correct phase is determined (with a constant phase offset) if the signal is of the sinusoidal form of Eq. (1).

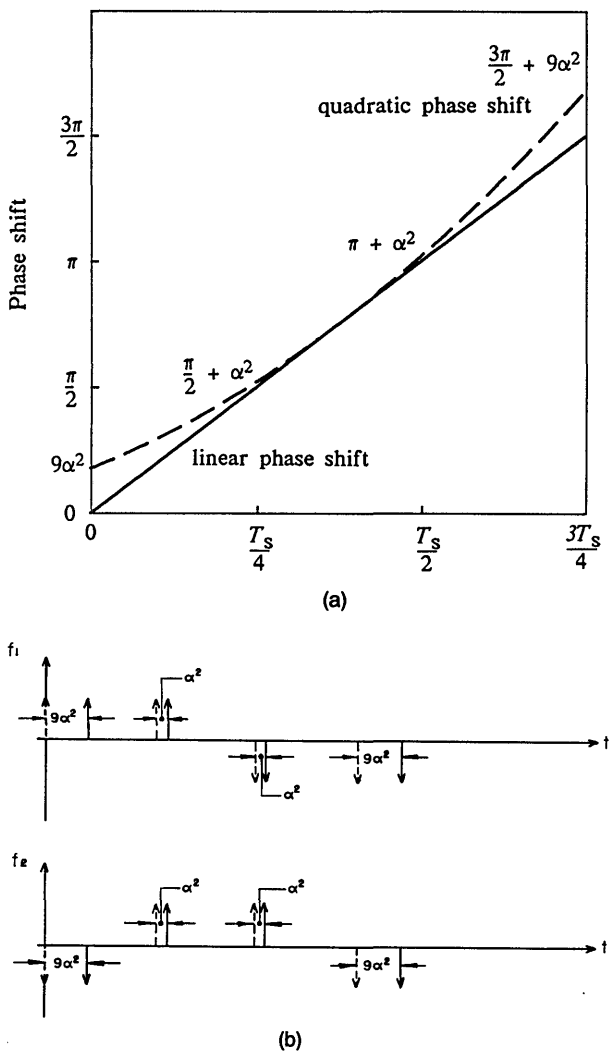


Fig. 5. Influence of a quadratic phase shift: (a) quadratic phase shift, (b) sampling points.

We also see that this is true for all even harmonics of the signal. In general, if the two filter functions can be represented as even or odd pairs of δ functions centered about a common center of symmetry, then for nonlinear phase shifts symmetrical about this center of symmetry the spectra of the filter functions will be of a form similar to Eqs. (55), and the above statements hold.

From using this Fourier frequency-domain analysis of digital phase-measuring interferometry, some general statements can be made concerning the phase-computation process and its associated measurement errors.

(1) If the signal is purely sinusoidal, i.e., of a single frequency, any set of zero mean periodic filter functions of arbitrary functional form can be used to compute the phase without error as long as the filter functions have spectral components in quadrature at the signal's frequency.

(2) The phase of any periodic irradiance function can be determined by using filter functions of pure sine and cosine. This corresponds to using $N = \infty$ samples.

(3) The larger the number of samples of the irradiance function, the better the filtering at the fundamental frequency and the less the phase-measurement error caused by higher harmonics. Most of the phase algorithms presented here use filter functions with equal weights of the samples for computational ease. Filter functions with weighted samples can also be used that will reduce the higher-order spectral content of the filter functions, thus improving the phase-measurement accuracy when the irradiance also consists of higher spectral terms.

5. RANDOM ERRORS

Until now we have considered only systematic errors. However, every intensity measurement will suffer from noise, causing random errors in the calculated phase. For the digital heterodyne techniques discussed here, the straightforward method of Gaussian error propagation presented here is feasible as long as the number of samples does not become too large. This method uses the fact that the variance of any linear combination y of statistically independent variables x_n,

$$y = \sum_{n=1}^{N} a_n x_n, \quad (56)$$

is given by

$$\sigma_y^2 = \sum_{n=1}^{N} a_n^2 \sigma_n^2, \quad (57)$$

where σ_n^2 is the variance of the nth variable x_n. This result is valid independently of the probability distributions of the individual variables.[20] For any functional dependence

$$y = f(x_1, \ldots, x_N), \quad (58)$$

we can approximate the variance of y by linearizing f, using only the first order of a Taylor expansion[7]:

$$\sigma_y^2 \cong \sum_{n=1}^{N}\left[\frac{\partial f(\langle x_1\rangle, \ldots, \langle x_N\rangle)}{\partial x_n}\right]^2 \sigma_n^2, \quad (59)$$

where $\langle x \rangle$ denotes the mean of x. Therefore we can determine the variance of any calculated quantity as long as the linearization is a good approximation.

In the heterodyne procedures the calculated phase is a function of the intensity measurements. Since they are taken sequentially in time, the measurement errors that are due to detection are uncorrelated, and we obtain for the cases of Fig. 1, for $\nu_s = \nu_f$, as variance of the phase,

$$\sigma_\Phi^2 \cong \frac{1}{(2|S(\nu_s)||F_2(\nu_s)|)^4} \sum_{n=1}^{N} [f_1(t_n)c_2 - f_2(t_n)c_1]^2 \sigma_n^2, \quad (60)$$

from which Table 1 is calculated for the sinusoidal signal of Eq. (1). A comparison with computer simulations shows that relation (60) holds for signal-to-noise ratios as low as 2. It is valid for all practical applications, for which the signal-to-noise ratio is usually much higher.

We have to distinguish two cases. Background noise, detector noise, and quantization noise[4] are independent from the signal (additive noise), whereas shot noise, which is due to the quantized nature of light, has a variance equal to the mean signal for laser light as well as for broadband thermal light.[21] In the latter case, the intensity signal is replaced by the photon-count signal, and I in Table 1 is the bias photon count per sample.

From Table 1 we see that σ_Φ^2 is proportional to the variance of the noise and inversely proportional to the power $(IV)^2$ in the signal component filtered by the phase-measurement process. Only a shift of the sampling points of the signal occurs between case 1 and case 2. The same holds for case 3 and case 4. Therefore these pairs of cases are pairs of equal error propagation. Table 1 also shows that for both types of noise only heterodyne procedures 1 and 2 yield variances of the phase error independent from the phase value itself. That is, the phase error is additive. This is also true for case 5 with additive noise. The other cases show sinusoidal variations with the phase value.

Table 1. Variance of Phase Error σ_Φ^2 of Common Heterodyne Procedures[a]

Case	Additive Noise	Shot Noise
1	$\frac{1}{2}\frac{\sigma_I^2}{(IV)^2}$	$\frac{1}{2}\frac{I}{(IV)^2}$
2	$\frac{1}{2}\frac{\sigma_I^2}{(IV)^2}$	$\frac{1}{2}\frac{I}{(IV)^2}$
3	$\frac{\sigma_I^2}{(IV)^2}[1 + \frac{1}{2}\cos(2\Phi)]$	$\frac{1}{(IV)^2}\left\{1 + \frac{1}{2}\cos(2\Phi) - \frac{V}{2}[\sin(\Phi) + \sin(3\Phi)]\right\}$
4	$\frac{\sigma_I^2}{(IV)^2}[1 + \frac{1}{2}\cos(2\Phi)]$	$\frac{1}{(IV)^2}\left\{1 + \frac{1}{2}\cos(2\Phi) - \frac{V}{2}[\sin(\Phi) + \sin(3\Phi)]\right\}$
5	$\frac{2}{3}\frac{\sigma_I^2}{(IV)^2}$	$\frac{I}{(IV)^2}\left[1 - \frac{V}{2}\cos(3\Phi)\right]$
6	$\frac{1}{16}\frac{\sigma_I^2}{(IV)^2}[7 + \cos(2\Phi)]$	$\frac{I}{32(IV)^2}[14 + \cos(\Phi) + 2\cos(2\Phi) - \cos(3\Phi)]$

[a] Variance of additive noise, σ_I^2; variance of shot noise, I.

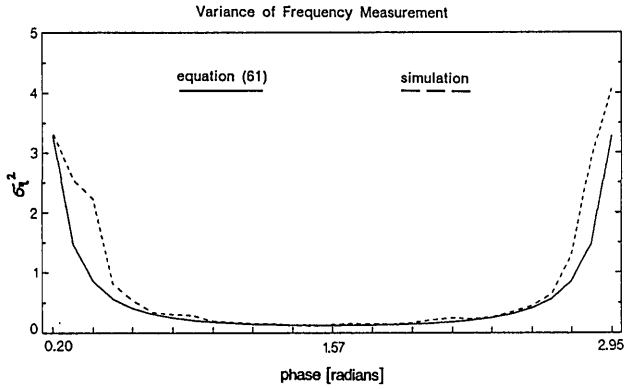


Fig. 6. Normalized variance of the frequency measurement with additive noise.

We can also apply the general relation (59) to the frequency measurement given by Eq. (48) to obtain, in the presence of additive noise of variance σ_I^2, for a measurement of ν_s,

$$\sigma_\eta^2 \cong \frac{\sigma_I^2}{64I^2V^2} \frac{5 - 12\sin^2(\eta) + 8\sin^4(\eta)}{\cos^2(\eta)\sin^4(\eta)} \frac{1}{\sin^2(\Phi)}, \quad (61)$$

with $\eta = \pi\nu_s/4\nu_f$.

Numerical simulations have shown relation (61) to be a good approximation over a limited range where σ_η^2 is small (Fig. 6); that is,

$$1 \le \frac{\nu_s}{\nu_f} \le 1.75 \quad (62a)$$

and

$$\frac{\pi}{4} \le \Phi \le \frac{3\pi}{4}. \quad (62b)$$

There is a strong dependence on the phase as well as on the ratio ν_s/ν_f. Since the frequency measurement uses the odd part of the signal, the variance goes toward infinity for even signals, i.e., $\Phi = n\pi$, with n an integer. Also, as ν_s/ν_f goes to zero, which means small sampling intervals as compared with the period of the signal, the numerator and the denominator of r in Eq. (48) become small, and again σ_η^2 diverges. For $\nu_s \cong \nu_f$, relation (61) reduces to

$$\sigma_\eta^2 \cong \frac{\sigma_I^2}{8I^2V^2} \frac{1}{\sin^2(\Phi)}. \quad (63)$$

Figure 6 shows this Φ dependence of σ_η^2. If this method is used to calibrate the phase shift by a separate measurement of the signal frequency, phase points near 0 and π should be excluded owing to the large increase in the variance.

If Eq. (20) from case 1 is used to determine the phase and if the signal frequency is calculated from the same set of four samples,[11] we obtain for the variance of the phase in the presence of additive noise

$$\sigma_\Phi{}^2 = \frac{\sigma_I{}^2}{I^2V^2\sin^4(2\eta)} \times\left[\cos^2(\eta) + \frac{5 - 12\sin^2(\eta) + 8\sin^4(\eta)}{4\sin^2(\eta)}\cos^2(\Phi)\right]. \quad (64)$$

Here the approximation is good over a range of

$$1 \le \frac{\nu_s}{\nu_f} \le 1.5. \quad (65)$$

The divergences at $\Phi = n\pi$ are no longer present as with the frequency determination. From Eq. (64) the optimum value of η that minimizes the average variance $\sigma_\Phi{}^2$ for uniformly distributed Φ is found to be 1.222. This represents sampling steps equal to 0.3055 signal periods, or 110 deg, in excellent agreement with Carre.[12]

For $\nu_s \cong \nu_f$ we obtain for this case a variance

$$\sigma_\Phi{}^2 = \frac{\sigma_I{}^2}{2I^2V^2}[1 + \cos^2(\Phi)]; \quad (66)$$

that is, the phase error is not simply additive but varies with phase.

6. CONCLUSION

The Fourier description of the digital heterodyne process as a filtering operation has proved to be useful for the analysis of systematic errors. All we need to know are the spectra of the signal and the filter functions of the chosen algorithm. Since the parameter t of the equations does not need to represent time, the general formalism is also suited for the analysis of spatial synchronous detection techniques.[22–24]

We could quantify the phase error that is due to multiple-beam interference in a grating shearing interferometer or a multiple-beam Fizeau interferometer[25] and decide whether phase stepping or integrated buckets leads to smaller errors. Also, Rowley's algorithm for the measurement of the sampling interval could be interpreted on a physical basis as a filtering operation, leading to the generalization of the process to more than four samples.

We also showed that Gaussian error propagation is well suited for the analysis of random errors of the phase measurement by heterodyning. It is a versatile tool: it can be applied not only to the standard phase measurements but also to the determination of the sampling steps or to more complicated phase algorithms, such as the averaging techniques by Schwider.[14] The range of validity of the approximate results is sufficient for any practical interferometry data. Thus the Fourier description of the digital phase-measurement process together with Gaussian error propagation is a valuable tool for the design and analysis of digital phase-measuring interferometric systems.

* Present address, Carl Zeiss, Postfach 1369/1380, D-7082 Oberkochen, Federal Republic of Germany.

REFERENCES

1. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surfaces and lenses," Appl. Opt. **13**, 2693–2703 (1974).
2. J. H. Bruning, "Fringe scanning interferometers," in *Optical Shop Testing*, D. Malacara, ed. (Wiley, New York, 1978).
3. J. C. Wyant, "Use of an ac heterodyne lateral shear interferometer with real time wavefront correction systems," Appl. Opt. **14**, 2622–2626 (1975).
4. C. L. Koliopoulos, "Interferometric optical phase measurement techniques," Ph.D. dissertation (University of Arizona, Tucson, Arizona, 1981).
5. K. Creath, "Comparison of phase measuring algorithms," in *Surface Characterization and Testing*, K. Creath, ed., Soc. Photo-Opt. Instrum. Eng. **680**, 19–28 (1986).
6. C. W. Helstrom, *Statistical Theory of Signal Detection* (Pergamon, Oxford, 1968).
7. E. Hardtwig, *Fehler- und Ausgleichsrechnung* (Bibliographisches Institut, Mannheim, 1968).
8. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (Wiley, New York, 1978).
9. C. J. Morgan, "Least-squares estimation in phase-measurement interferometry," Opt. Lett. **7**, 368–370 (1982).
10. J. E. Greivenkamp, "Generalized data reduction for heterodyne interferometry," Opt. Eng. **23**, 350–351 (1984).
11. W. R. C. Rowley and J. Hamon, "Quelques mésures de dyssymétrie de profils spectraux," R. Opt. Theor. Instrum. **42**, 519–523 (1963).
12. P. Carre, "Installation et utilisation du comparateur photoelectrique et interferentiel du Bureau International des Poids et Mésures," Metrologia **2**, 13–16 (1966).
13. J. B. Hayes, "Linear methods of computer controlled optical figuring," Ph.D. dissertation (University of Arizona, Tucson, Arizona, 1984).
14. J. Schwider, R. Burow, K.-E. Elssner, J. Grzanna, R. Spolaszyk, and K. Merkel, "Digital wave-front measuring interferometry: some systematic error sources," Appl. Opt. **22**, 3421–3432 (1983).
15. P. Hariharan, "Digital phase shifting interferometry: a simple error-compensating phase calculation algorithm," Appl. Opt. **26**, 2504–2506 (1987).
16. Y.-Y. Cheng and J. C. Wyant, "Phase shifter calibration in phase-shifting interferometry," Appl. Opt. **24**, 3049–3052 (1985).
17. K. Freischlad, "Wavefront sensing by heterodyne shearing interferometry," Ph.D. dissertation (University of Arizona, Tucson, Arizona, 1986).
18. C. L. Koliopoulos, "Radial grating lateral shear heterodyne interferometer," Appl. Opt. **19**, 1523–1528 (1980).
19. C. Ai and J. C. Wyant, "Effect of piezoelectric transducer nonlinearity on phase-shift interferometry," Appl. Opt. **26**, 1112–1116 (1987).
20. B. R. Frieden, *Probability, Statistical Optics, and Data Testing* (Springer-Verlag, Berlin, 1983).
21. J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
22. Y. Ichioka and M. Inuiya, "Direct phase detecting system," Appl. Opt. **11**, 1507–1514 (1972).
23. L. Mertz, "Real-time fringe-pattern analysis," Appl. Opt. **22**, 1535–1539 (1983).
24. K. H. Womack, "Frequency domain description of interferogram analysis," Opt. Eng. **23**, 396–400 (1984).
25. P. Hariharan, "Digital phase stepping interferometry: effects of multiply reflected beams," Appl. Opt. **26**, 2506–2507 (1987).