Phase-evaluation methods in whole-field optical measurement techniques

1999 Meas. Sci. Technol. 10 R33

(http://iopscience.iop.org/0957-0233/10/3/005)

Meas. Sci. Technol. **10** (1999) R33–R55. Printed in the UK PII: S0957-0233(99)89880-4

REVIEW ARTICLE

Phase-evaluation methods in whole-field optical measurement techniques

B V Dorrío and J L Fernández

Departamento de Física Aplicada, Universidade de Vigo, Lagoas-Marcosende 9, 36200 Vigo, Spain

Received 30 July 1998, in final form 30 October 1998, accepted for publication 4 November 1998

Abstract. Many optical measurement techniques provide fringe patterns as their results. The decodification processes that employ one or several fringe patterns to automatically retrieve the phase are generally designated as phase-evaluation methods. In this work, an overview of these methods will be schematically presented. Their particular performances will be compared, stressing their main advantages and drawbacks. An important group of these methods employs the well-known phase-shifting algorithms as a tool for calculating the phase. In the general form of these algorithms, the principal value of the optical phase is computed by an inverse trigonometric function whose argument is a combination of phase-shifted intensity values, provided by the modulation of one or several fringe patterns. These algorithms will be also studied in the general context of the phase-evaluation methods.

Keywords: optical measurements, fringe patterns, phase evaluation, phase shifting

1. Introduction

Many whole-field optical measurement techniques provide the measurand data codified as the phase of a periodic intensity profile. When the parallel acquisition capability (a normal attribute of optical methods) is used under these conditions, the obtained image is called a fringe pattern. These patterns, which usually appear as alternating light and dark bands or fringes, can be mathematically formulated by defining a dependence of the intensity i on the spatial coordinates $\boldsymbol{r}$ and the time coordinate t of the image plane such that

$$i(\boldsymbol{r}, t) = i_c(\boldsymbol{r}, t)\{1 + V(\boldsymbol{r}, t) f[\phi(\boldsymbol{r}, t)]\} \qquad (1)$$

where $i_c(\boldsymbol{r}, t)$ is the background intensity, $V(\boldsymbol{r}, t)$ is the visibility or fringe contrast and the periodic nature of the intensity is determined by the unidimensional function f, that takes values in $[-1, 1]$ interval and whose argument is the optical phase $\phi(\boldsymbol{r}, t)$. The function f defines the type of fringe pattern profile, amongst which we could highlight (figure 1) the sinusoidal (typical of double-beam interferometry), the Lorentzian (typical of multiple-beam interferometry), the rectangular (those of binary gratings) and the trapezoidal (associated with the superimposition of different binary gratings, as is usual in moiré methods). At the same time, this periodic dependence permits us to express the intensity values, equation (1), as

$$i(\boldsymbol{r}, t) = i_0(\boldsymbol{r}, t) + \sum_{l=1}^{\infty} a_l(\boldsymbol{r}, t) \cos[l\phi(\boldsymbol{r}, t)] \qquad (2)$$

where $i_0(\boldsymbol{r}, t)$ is the background intensity and $a_l(\boldsymbol{r}, t)$ is the amplitude of the lth harmonic.

From a metrology viewpoint, the optical methods whose output results are in the form of fringe patterns constitute a potent measurement tool, so long as they permit one to obtain information in a bidimensional domain (that is to say, they are whole-field methods rather than point methods which provide only one simultaneous datum); additionally, this is a process which bidimensionally encodes information of a certain magnitude (the measurand), generally defined in a bidimensional or tridimensional space. This encoding process can be seen, in the majority of cases, as a double application which on the one hand assigns a certain fringe-pattern-intensity value to every value of the said magnitude and on the other hand assigns every point of the magnitude space (or line, in cases in which the information is calculated along the wavefront trajectory) to the corresponding fringe-pattern point. The first assignment of corresponding values is usually done through the phase $\phi(\boldsymbol{r}, t)$ via a specific functional relationship for each specific situation. The second relationship is, just like the first one, intimately related to the optical measurement process and is established through

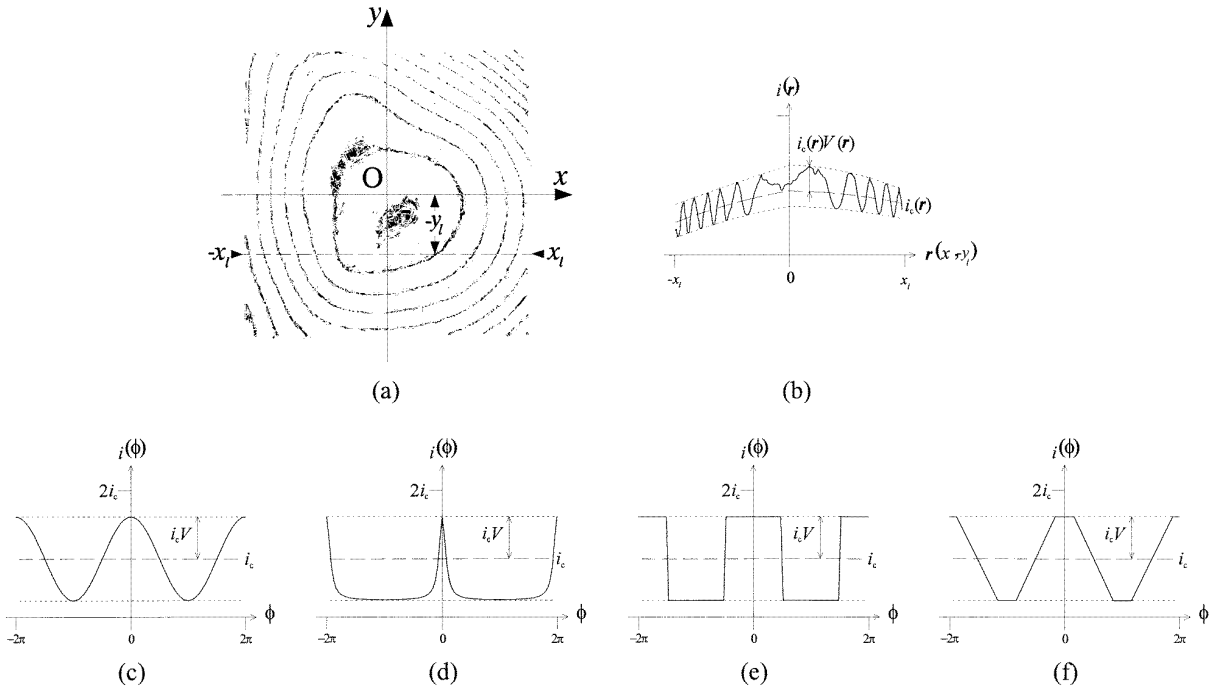


Figure 1. (a) Fringe pattern and (b) intensity distribution for the indicated horizontal line. Distributions of the intensity i with respect to the phase ϕ for patterns $i = i_c\{1 + Vf(\phi)\}$ with (c) sinusoidal, (d) Lorentzian, (e) rectangular and (f) trapezoidal profiles.

the study of image formation when light passes through the various components that comprise the optical system.

Therefore, the fringe patterns provide information wherein, upon taking into account and conveniently cancelling the effects of intensity variations due to $i_c(\boldsymbol{r}, t)$ and $V(\boldsymbol{r}, t)$, the fringes constitute isolines of equal phase. The optical systems are usually designed such that $i_c(\boldsymbol{r}, t)$ and $V(\boldsymbol{r}, t)$ remain practically constant throughout the entire field of view, such that the fringes in the image-system output provide valuable information to the experienced metrologist. However, the technological demands (referring to precision, velocity and measurement-process automation) have resulted in the development of methods which provide a finer interpolation between the extreme intensity values, perform the cancellation of even gross variations of $i_c(\boldsymbol{r}, t)$ and $V(\boldsymbol{r}, t)$, eliminate ambiguities in the optical phase $\phi(\boldsymbol{r}, t)$ (because the phase is generally a multiform function of the intensity) or permit the presentation of the results in a manner understandable to the non-specialist. The entire process which, by analysing one or more fringe patterns, provides a result in a format that is suitable to each case is called fringe-pattern analysis.

The fringe-pattern-analysis process generally consists of the following steps.

(i) Phase evaluation. This consists of obtaining a spatial distribution of the phase $\phi(\boldsymbol{r}, t)$, the so-called phase map, by starting from one or more fringe patterns $i_1, \ldots, i_k, \ldots, i_n$ associated with the same measurand field.

(ii) Phase unwrapping. The previous phase-evaluation stage provides, in most cases, the principal values for the phase $\phi(\boldsymbol{r}, t)$, that is, values indeterminate within a quantity equal to $2\pi p$ rad, where p is an integer number called the fringe order. In such situations, a continuous phase distribution over its definition domain is desired and, therefore, the need for a step that performs this operation arises.

(iii) Elimination of additional terms. The terms possibly introduced within the phase $\phi(\boldsymbol{r}, t)$ which bear no relationship to the measurement magnitude (the simplest case being that of constant terms or those that depend linearly on spatial coordinates), during the formation of the fringe patterns and/or during the evaluation stage, can be removed by an adequate least squares fit, an iterative process or some other method.

(iv) Re-scaling. In most cases, the correct presentation of the results demands that there be a relationship between the values of the measurable magnitude and the coordinates of the space in which it is defined (a mathematical expression, a table, a graph, relationships with grey-scale levels or colour levels, etc.). This stage requires knowledge of the relationships between (a) the phase $\phi(\boldsymbol{r}, t)$ and the measurand magnitude and (b) the bidimensional space of the phase map and the measurand spatial domain. These relationships generally depend on the measurement method used.

The various stages of the fringe-pattern analysis process are shown in figure 2. Once the re-scaling has been done, the measurement process can be considered concluded, although, on certain occasions, the results provided by the previous operations (i), (ii) or (iii) are sufficient for the needs of the experiment. The first one, the phase evaluation, is common and independent of the others and, in general, also independent of the desired measurable magnitude. Therefore, upon completion of the phase-evaluation stage, the obtained principal phase values can be unwrapped and/or their linear terms subtracted and/or re-scaled, depending on the characteristics and needs of each specific case.

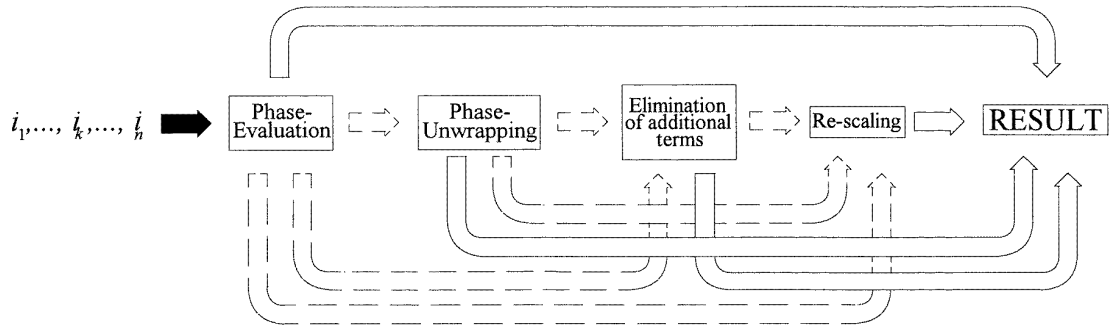


Figure 2. A diagram of the various stages of the fringe-pattern-analysis process. The intermediate stages may or may not be skipped depending on the application.

Phase-evaluation methods (PEMs) have been employed ever since the dawn of optics to measure various physical magnitudes in a variety of scientific and engineering fields. However, their use had been inhibited for a long time since the process had to be carried out manually, wherein, at first, the quantification of the phase was done by locating the extreme points of the fringe patterns either by direct observation or by previously photographing them. These routine procedures were found to be tedious and inefficient because of the great amount of measurements performed manually, together with the fact that the scientist had to further distinguish the real data from noise. During the 1960s, the appearance of solid state TV cameras and computers permitted the acquisition and storage of fringe patterns in an adequate format for their manipulation *a posteriori*. This facilitated the performance of rapid measurements and averaging using specifically developed software, to improve the contrast, reduce noise and to easily locate the extremes of the fringe patterns. These early techniques were rapidly replaced by newer image processing techniques in the mid-1960s which provided the phase maps directly. The improvement in the speed and storage capacity of computers, together with the fall of their prices, heralded the universal use of these techniques and their rapid development. These techniques are presently one of the most potent and versatile metrological tools insofar as applications are concerned.

Owing to the above, the PEMs have become one of the principal world-wide research topics in the field of optical metrology, as can be judged from the number of research groups concentrating their efforts on devising new methods, improving existing ones or using them in new applications. This fact can be confirmed by looking at the increase in the number of papers published on the topic since the end of the 19th century, especially during the 1980s, compared with numbers of papers on other topics in optical metrology (Malacara 1990).

The following sections will describe the present PEMs by indicating their principal properties, advantages and limitations (tables 1 and 2). This synopsis shows that each method possesses features which make it the very best for specific applications, there being no single method that is optimal for all situations. On the other hand, many PEMs use phase-stepping algorithms (PSAs) as a tool for phase calculation, whereby the phase $\phi(\boldsymbol{r}, t)$ is obtained from a series of phase-shifted intensity values, combined to form the argument of an inverse trigonometric function. Owing to their importance, this paper presents an update of PSAs in the general context of PEMs, showing their possible classification, the different strategies available for the phase modulation of the necessary intensity values and, finally, their associated errors.

2. Phase-evaluation methods

Usually the evaluation methods use some kind of phase modulation whereby an additional phase $\alpha(\boldsymbol{r}, t)$ is added to the phase of interest $\phi(\boldsymbol{r}, t)$. In this case, equation (2) can be written as

$$i(\boldsymbol{r}, t, \alpha) = i_0(\boldsymbol{r}, t) + \sum_{l=1}^{\infty} a_l(\boldsymbol{r}, t) \cos\{l[\phi(\boldsymbol{r}, t) + \alpha(\boldsymbol{r}, t)]\}. \tag{3}$$

Because, from the mathematical point of view, it is irrelevant whether the modulation domain is time or space, the additional phase $\alpha(\boldsymbol{r}, t)$ can be expressed (if the modulation domain is unidimensional) in a generic form as

$$\alpha(\boldsymbol{r}, t) = 2\pi\xi_0 u \tag{4}$$

where ξ_0 is a temporal or spatial carrier frequency and u represents temporal or space coordinate, respectively. So, to describe the evaluation process, it is convenient to focus attention on the modulation domain u, having in the general case, for each phase-calculation cell,

$$i(u) = i_0(u) + \sum_{l=1}^{\infty} a_l(u) \cos\{l[\phi(u) + 2\pi\xi_0 u]\} \tag{5}$$

the components of $(\boldsymbol{r}, t)$ that are independent of u being constant in each calculation cell. Also, in many cases, the independence of ϕ, i_0 and/or a_l with respect to u is fulfilled, although they are usually dependent on the cell considered.

A generic evaluation process provides the phase $\phi(u)$ independently of $i_0(u)$ and $a_l(u)$. This process can be better understood if an analysis in the reciprocal space defined by the frequency ξ is carried out. Here the Fourier transform of the intensity is

$$I(\xi) = I_0(\xi) + \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} D_l(\xi - l\xi_0) \tag{6}$$

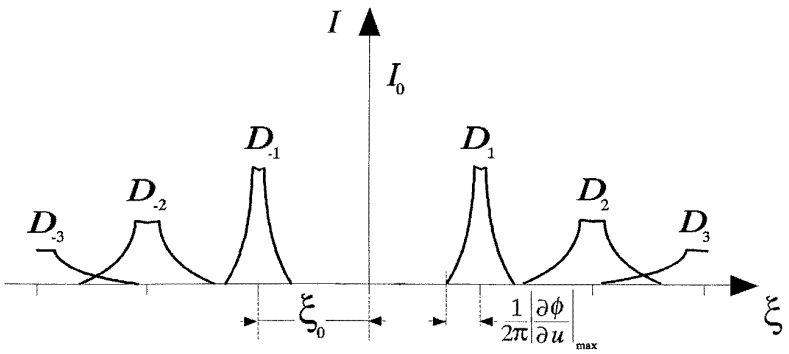


Figure 3. The spectrum (in modulus) of $I(\xi)$, for a constant background intensity.

and $I(\xi_0)$ is the Fourier transform of the background intensity, $D_l(\xi)$ is the Fourier transform of

$$d_l(u) = \tfrac{1}{2}a_l(u)\exp[\mathrm{j}l\phi(u)] \tag{7}$$

and j is the imaginary unit and $a_{-l} = a_l$. Each harmonic l is centred at its carrier frequency $l\xi_0$ and has, considering i_0 and the amplitudes of the lth harmonics to have a constant value, a frequency bandlimit given by

$$l\frac{1}{2\pi}\left|\frac{\partial\phi(u)}{\partial u}\right|_{max} \tag{8}$$

$|\,|_{max}$ being the maximum absolute value of the argument. In order to obtain the phase ϕ, these harmonics must be separated in reciprocal space (figure 3), that is to say, the separability condition must be fulfilled (this condition may be relaxed if previous knowledge of the overlapping harmonics is available). Once one has isolated one of these harmonics $D_l(\xi - l\xi_0)$, the phase ϕ is calculated as

$$l\phi(u) = \mathrm{Arg}\{F^{-1}[D_l(\xi - l\xi_0)]\} - 2\pi l\xi_0 u \tag{9}$$

where F^{-1} means the inverse Fourier transform and the carrier-frequency term can be removed directly if the selected harmonic had previously been shifted to the origin. In this case, the phase is $1/l$ times the argument of the inverse Fourier transform of the order-l harmonic of the pattern. The former model can be extended to a bidimensional modulation domain (u, v), where u and v are the spatial coordinates of the pattern, but a combination of spatial and temporal domains is also possible (Takeda and Kitoh 1992). In this case the reciprocal space coordinates would also be bidimensional, with frequencies (ξ, η).

Although the process of phase extraction can be explained in a quite general manner in the above-mentioned reciprocal space framework, the combination of different ways of performing phase modulation, sampling of intensity values and demodulation had led to a great variety of PEMs, with their individual algorithms for phase calculation and with their particular needs and performances. Most of these methods use patterns with a sinusoidal profile, although a good number of them can be employed for other fringe profiles after appropriate adjustments. Several classifications of PEMs have been reported so far, attending to different relevant attributes. A first classification of these methods (Kreis 1987) is by distinguishing between local and global methods. A local method is one that calculates the optical phase $\phi(\boldsymbol{r}, t)$ at a point $(\boldsymbol{r}, t)$ using successive intensity values at that point or simultaneous values in a group of points surrounding it, whereas a global method is one for which one needs to know the intensity at all of the pattern points to find the phase $\phi(\boldsymbol{r}, t)$ at each point. An important difference between these two approaches is that, in local methods, values of $i_0(\boldsymbol{r}, t)$, $a_l(\boldsymbol{r}, t)$ and $\phi(\boldsymbol{r}, t)$ are assumed constant in each phase-calculation cell throughout the evaluation process, a three-sample process at each cell being sufficient to recover the phase $\phi(\boldsymbol{r}, t)$; on the other hand, in global methods the cell comprises the whole pattern, the phase distribution $\phi(\boldsymbol{r}, t)$ is variable within it and its calculation relies on the possibility of cancelling out the effects of the variations of $i_0(\boldsymbol{r}, t)$ and $a_l(\boldsymbol{r}, t)$ within the pattern by isolating their contribution in some transformation plane (to the best of our knowledge, the only well reported global method is based on the Fourier transform, i.e. the Fourier transform method). Another type of classification (Kujawinska and Wójciak 1991a) distinguishes between temporal and spatial methods, wherein, in the first case, the data necessary for the evaluation are obtained at different time intervals, whereas in the second case, all of the data are obtained simultaneously. Finally (Creath 1993, Kujawinska 1993b), the PEMs can be classified as electronic or analytical, wherein an electronic method is one that uses a detector to measure phase increments in a temporal domain at each point, whereas an analytical method is one in which the phase $\phi(\boldsymbol{r}, t)$ is obtained by acquisition of intensity values while the phase $\phi(\boldsymbol{r}, t)$ is being modulated temporally or spatially.

Although practically all of the PEMs can be introduced under any of the three above-mentioned classifications, in the present study we have preferentially performed a new systematization of the same, taking into account the need or lack of a need for the introduction of a spatial carrier as a substantial element of the method. The reason for this classification scheme is that it stresses the important characteristic of the methods with a spatial carrier that they use, for the phase modulation, the same domain as that for the image representation (i.e. the image pixels), in contrast to other methods in which each pixel is interrogated at a given instant independently of the other pixels (which leads to the need to acquire several intensity values at different time intervals). So, in methods with a spatial carrier limitations arise in the phase-measurement range (more exactly in the

maximum phase gradient) and also in the admissible spatial variations of $i_0(\boldsymbol{r}, t)$ and $a_l(\boldsymbol{r}, t)$. These limits are related to the avoidance of aliasing effects and can be expressed as a fraction of the value of the carrier frequency, limited ultimately by the image-sampling density (i.e. the number of image pixels). In the counterpart, the methods with a spatial carrier need only one fringe pattern to perform the phase evaluation and are very adequate for the study of dynamic events. According to this, these methods are considered real-time ones. Following the former criterion, the evaluation methods can be classified as

(i) methods without a spatial carrier such as the temporal phase-shifting method (Greivenkamp and Bruning 1992) (TPSM) and synchronous detection (Bruning *et al* 1974) (SD), the spatial phase-shifting method (Kwon 1984) (SPSM), the sinusoidal phase-modulating method (Sasaki and Okazaki 1986a) (SPMM), phase-locked interferometry (Johnson *et al* 1979) (PLI) and the heterodyne method (Massie *et al* 1979) (HM); and

(ii) methods with a spatial carrier such as the Fourier transform method (Takeda and Mutoh 1983) (FTM), the spatial carrier phase-shifting method (Williams *et al* 1991) (SCPSM) and the sinusoidal fitting method (Macy 1983) (SFM), spatial synchronous detection (Womack 1984) (SSD), logical moiré (Asundi and Yung 1991a) (LM) and the multiplicative analogical moiré method (Dorrío *et al* 1995, Schwider *et al* 1986, 1987) (MAMM).

2.1. Methods without a spatial carrier

2.1.1. The temporal phase-shifting method. This method requires that the phase be stepped or increased linearly with time, usually in a uniform manner within the whole field by using a modulator (Creath 1988, Greivenkamp and Bruning 1992). During this process, a series of patterns with a certain phase increment $\alpha(\boldsymbol{r}, t)$ between them is obtained. The temporal bandwidth requisite for each of the obtained phase-shifted patterns is sufficiently low for a TV camera to capture. In this way, once these phase-shifted patterns have been combined using a PSA, we can obtain the phase $\phi(\boldsymbol{r}, t)$ values modulo 2π for all of the points simultaneously, such that the phase $\phi(\boldsymbol{r}, t)$ calculated at each point is independent of the phase values at the remaining points. This method can provide very high accuracy if the ambient conditions are well controlled. On the other hand, due to the need to obtain different separate patterns in time, the TPSM cannot be applied to dynamic processes or applied under adverse conditions, because it is necessary for the TPSM that the background intensity, the visibility and, especially, the phase $\phi(\boldsymbol{r}, t)$ be stationary magnitudes.

Synchronous detection (Bruning *et al* 1974) was the predecessor to the actual TPSM, in which the phase $\phi(\boldsymbol{r}, t)$ is obtained by phase modulation of the initial pattern. This process is carried out by introducing a phase term $\alpha(\boldsymbol{r}, t)$ that exhibits a linear dependence on time and later correlating the obtained phase-modulated pattern to one sinusoidal and another cosinusoidal pattern of the same frequency. Both products are averaged over various oscillation periods to obtain two magnitudes proportional to the sine and cosine of the phase $\phi(\boldsymbol{r}, t)$. Finally, the two magnitudes can be combined to obtain the phase $\phi(\boldsymbol{r}, t)$ modulo 2π as the arctangent of its quotient.

2.1.2. Spatial phase-shifting method. The spatial phase-shifting method (SPSM) has been designed to conserve some advantages of the TPSM and to avoid problems arising due to the acquisition of different patterns at different times. This method combines a set of phase-shifted patterns located at different points in space, acquired at the same instant (Kujawinska 1993a). The separation in space of the patterns can be achieved by using rotational polarizing components (Symthe and Moore 1984), diffraction gratings (Kwon 1984) or computer-generated diffractive optical elements (Barrientos *et al* 1997). In this case, errors due to environmental instabilities are avoided due to the simultaneous acquisition of the patterns. However, other types of errors appear because the patterns are obtained using different TV cameras or in different parts of the same camera, which introduces variation of $i_0(\boldsymbol{r}, t)$ and $a_l(\boldsymbol{r}, t)$ inside each phase-calculation cell. Therefore a real-time evaluation method is obtained at the cost of accuracy (Kwon *et al* 1987).

2.1.3. The sinusoidal phase-modulating method. In this method, the phase is modulated sinusoidally in a range of frequencies from tens to hundreds of hertz. This modulation is easier to perform than the one described above and can be carried out, for example, by modulating the diode-laser injection current which acts as an illuminating source (Sasaki *et al* 1990a) or by vibrating a surface of the optical set-up using piezoelectric transducers (Sasaki *et al* 1990b, Wang *et al* 1994). All time-varying fringe-pattern points are detected simultaneously by a video camera which integrates the intensity of the pattern during a quarter of the period of the introduced sinusoidal signal. Four resultant patterns are obtained in a phase-modulation period and are combined such that the phase $\phi(\boldsymbol{r}, t)$ modulo 2π is calculated as the arctangent of their linear combination quotient, by adequately selecting the sinusoidal modulation parameters (Sasaki *et al* 1987, Suzuki *et al* 1995). Other algorithms use fewer patterns at the cost of a longer processing time, because, in this case, it is necessary to obtain the temporal Fourier transforms of the patterns in order to operate with them later (Sasaki and Okazaki 1986a).

This method, whose drawback is a relatively complex calculation compared with those of the rest of the methods, provides a high accuracy (Sasaki *et al* 1987). It is very insensitive to external perturbations such as temperature fluctuations and mechanical vibrations of low frequency compared with the modulation frequency (Sasaki *et al* 1990a).

2.1.4. Phase-locked interferometry. In this method, the phase is modulated by the introduction of an additional phase $\alpha(\boldsymbol{r}, t)$ that varies linearly with time, plus a sinusoidal modulation of a small amplitude in the kilohertz range, either by displacing an optical surface using piezoelectric transducers (Johnson *et al* 1979, Matthews *et al* 1986) or by controlling the diode-laser current that acts as an illumination source (Suzuki *et al* 1989). This double modulation produces

a time-varying intensity pattern, with terms dependent on the form of Bessel function products of the order l multiplied by harmonic functions each with their phase proportional to the same order. *A posteriori* filtering out of the higher order terms permits one to isolate a pattern proportional to the sine of the sum of the original phase $\phi(\boldsymbol{r}, t)$ plus the additional reference phase $\alpha(\boldsymbol{r}, t)$. The phase of this signal, recorded with a point detector, is used to produce a feedback signal which modifies the reference phase $\alpha(\boldsymbol{r}, t)$ in order to cancel out the detected signal. When the null value is obtained, the sum of the original phase $\phi(\boldsymbol{r}, t)$ and the reference phase $\alpha(\boldsymbol{r}, t)$ is zero modulo 2π. In this way, one can obtain the original phase $\phi(\boldsymbol{r}, t)$ modulo 2π since the reference phase $\alpha(\boldsymbol{r}, t)$ is known. Because the phase changes with the detector movement, the feedback signal is used to vary the reference phase $\alpha(\boldsymbol{r}, t)$, thus maintaining the null-intensity condition. When this exceeds the value 2π, the original phase $\phi(\boldsymbol{r}, t)$ value is increased or decreased by the same quantity and the reference value is maintained at zero. This supplies an automatic count of the fringe order which makes *a posteriori* phase-unwrapping unnecessary.

The main limitation of the method is that its accuracy is limited for mechanical reasons (Schwider 1990), because of the high sensitivity to phase drift during the slow point-by-point measurement process. This value can be improved by using a CCD camera as the detector, together with the direct modulation of the diode laser (Suzuki *et al* 1989). Additionally, the region to be measured must be previously determined to avoid taking measurements outside it, which could break the continuity in the reference phase $\alpha(\boldsymbol{r}, t)$ feedback and thus the measurement of the whole pattern (Schwider 1990).

2.1.5. The heterodyne method. In the heterodyne method (HM), the intensity measurements are replaced by phase-increment measurements in the temporal domain, for each point in the pattern. To this end, one introduces a temporal carrier into the phase, ranging from hundreds of kilohertz (Kreis 1993, 1996, Massie *et al* 1979) to tens of megahertz (Lesne *et al* 1987). In interferometric techniques, this difference is obtained via a frequency shift in one or both wavefronts, by using, for example, Bragg cells (Massie 1980), rotating polarizing components (Rosenbluth and Bobroff 1990) and the transversal (Toyooka *et al* 1984) or rotational (Barnes 1987) movement of a grating. The overlap of these wavefronts gives rise to a series of intensity terms fluctuating in time, with frequencies in the optical range, except one term corresponding to the frequency difference. The employed photodetector should have a bandwidth matched to this beat frequency. The original phase $\phi(\boldsymbol{r}, t)$ is precisely the beat phase and can be obtained by comparing the value obtained from a mobile detector at each point with that obtained from a fixed detector at a certain reference point. The whole phase field is obtained by mechanically displacing the mobile detector along the pattern, no additional unwrapping process being necessary.

In this case, the obtained phase difference has a very high theoretical limit for accuracy because this process is performed independently of the visibility and of the background intensity (Dändliker and Thalmann 1985, Reid 1986). Also, it is unnecessary to perform *a posteriori* phase-unwrapping. However, there is a need for a stable environment, because a considerable time is taken up in mechanically moving the point detector along the whole pattern (Kreis *et al* 1993). This drawback, coupled with the fact that the measurement is carried out point by point, has resulted in the fact that this method is practically nonexistent in industry and therefore is relegated to being a precise laboratory evaluation method.

2.2. Methods with a spatial carrier

2.2.1. The Fourier transform method. This method uses the rapid Fourier transform techniques in its unidimensional (Takeda *et al* 1982, Takeda and Mutoh 1983) and bidimensional forms (Bone *et al* 1986, Nugent 1985), the former using less processing time and the latter appreciably reducing noise. In both cases, upon Fourier transformation of the fringe patterns, which are real, a spectrum is obtained with its real part even and its imaginary part odd. This provides a symmetrical tri-modal function with respect to the origin. The terms that are symmetric with respect to the continuous term contain the same phase information. If the phase $\phi(\boldsymbol{r}, t)$ variation is slow, band-pass filtering in the spatial frequency domain isolates one of these terms. This filtering process is performed easily if the spatial carriers have been selected to conveniently separate the different terms in reciprocal space, although it can also be done without the spatial carrier in certain specific cases (Kreis 1986). The isolated term is transferred to the origin to remove the carrier and is again Fourier transformed to obtain a complex function with real and imaginary non-zero parts. The phase $\phi(\boldsymbol{r}, t)$ modulo 2π is obtained as the arctangent of this imaginary term divided by the corresponding real term. In this case there is a global ambiguity in the phase polarity, that is, all of the points have a correct or an incorrect sign. The determination of the correct phase sign in those cases, wherever necessary, can be done by comparing the results with those obtained with the same pattern in quadrature (Kreis 1986), in which case the method loses its nature as a real-time process. This method provides a high accuracy (Kujawinska and Wójciak 1991c).

The principal limitations of the method are related to its global nature (Kreis 1987) which requires the previous definition of the phase domain to avoid spurious points in its calculations. Thus human intervention to supervise the filtering operation is generally needed, making this method difficult to automate (Kreis *et al* 1993, Perry and McKelvie 1993). Likewise, the carrier frequency must be an integer number of pixels and of constant value. In some cases, these limitations can be avoided (Liu and Ronney 1997).

2.2.2. The spatial carrier phase-shifting method. In this method, a phase shift of known value is introduced between successive points of a unique sinusoidal fringe pattern by inserting a spatial carrier (Józwicki *et al* 1992, Kujawinska and Wójciak 1991b, Takeda 1990, Takeda and Kitoh 1992, Williams *et al* 1991). If the spatial phase $\phi(\boldsymbol{r}, t)$ pattern variation is slow, the phase increase $\alpha(\boldsymbol{r}, t)$ between consecutive points will also be nearly constant ($\pi/2$ rad is usually selected). The phase $\phi(\boldsymbol{r}, t)$ modulo 2π at

Table 1. Principal characteristics of PEMs without a spatial carrier.

	Method				
	TPSM, SD	SPSM	SPMM	PLI	HM
Number of patterns	$\geqslant 3$	$\geqslant 3$	4	Continuous detection	Continuous detection
Real-time operation	No	Yes	No	No	No
Accuracy	Very high	Low	High	Low	Very high
Influence of $\frac{\text{static}}{\text{dynamic}}$ noise	$\frac{\text{Low}}{\text{High}}$	$\frac{\text{High}}{\text{Low}}$	$\frac{\text{Low}}{\text{Low}}$	$\frac{\text{High}}{\text{High}}$	$\frac{\text{Low}}{\text{High}}$
Experimental requisites	High	Low	High	High	Very high
Processing complexity	Medium	Medium	High	Medium	Medium
Operator interaction	Not possible	Not possible	Not possible	Not possible	Not possible
Process duration	Short	Short	Short	Long	Long
Commercial systems	Yes	No	No	No	Yes
Measuring range	High	Medium	High	High	High
Cost	Medium	High	Medium	Medium	Medium

each point is obtained by using the usual PSAs, for which the intensity values that appear in the arctangent function argument correspond in this case to neighbouring points in the unique pattern. Later, if need be, the introduced carrier can be subtracted. On the other hand, in order to increase the measurement range of the method and to permit measurement with a greater phase variation, the process can be carried out in two dimensions, by introducing two spatial carriers in perpendicular directions. In this case a lower accuracy than that of the one-dimensional SCPSM is achieved (Pirga and Kujawinska 1995).

This technique is easier to implement than the majority of the methods that use PSAs as long as there is no need for a mechanism to modulate the wavefront phases or for synchronization of the pattern-acquisition system with the phase modulation process. This method provides an order of accuracy analogous to that of the TPSM.

Its principal restriction consists of the requirement for appropriately introducing the spatial carrier into the pattern, this being a difficult process to automate. Besides, this method is very sensitive to static noise in the pattern and needs a high-spatial-resolution detector with a uniform sensitivity over all points (Creath and Schmit 1996).

On the other hand, the sinusoidal fitting method (SFM) (Macy 1983, Mertz 1983, Ransom and Kokal 1986) is analogous to the above-explained SCPSM, for an additional phase shift introduced between successive points of $2\pi/3$ rad. Likewise, this method is a simplified variant of the FTM. The spatial carrier is selected such that the pattern period is of three points. If the phase $\phi(\boldsymbol{r}, t)$ fluctuations are minor, the pattern can be adjusted to a Fourier series formed by their first two harmonics. These terms, which correspond to the sine and cosine of the phase $\phi(\boldsymbol{r}, t)$ at a point, are obtained from the pattern-intensity values at that point and at the points before and after it. Finally, the phase $\phi(\boldsymbol{r}, t)$ modulo 2π is obtained as the arctangent of the quotient of these terms minus the introduced carrier phase $\alpha(\boldsymbol{r}, t)$. In this case, the calculation times with respect to the FTM are reduced by a factor of five (Sirkis *et al* 1992).

2.2.3. Spatial synchronous detection. This method is similar to SD with the difference that in this case, a spatial carrier is introduced into the pattern. The decodification process (which provides the original phase $\phi(\boldsymbol{r}, t)$) is carried out by multiplying, in the spatial domain, this phase-modulated carrier with a carrier of the same frequency, without a continuous term, (i) in phase and (ii) in quadrature. This process generates two patterns with one fundamental frequency term and the other with a double carrier frequency. By applying a low-pass filter to both, one can obtain two patterns proportional to the cosine and the sine of the phase $\phi(\boldsymbol{r}, t)$, which can be combined in one PSA (Ichioka and Inuiya 1972, Womack 1984, Freischlad and Koliopoulos 1990).

This method provides a moderate accuracy, the main inconvenience being the impossibility of adequately separating the different spectral terms if a big phase $\phi(\boldsymbol{r}, t)$ variation exists. In such a case, these terms overlap and their complete separation using a filter is impossible (Womack 1984).

The demodulation process can be carried out by multiplying the high-frequency pattern by a periodic integer-valued pattern instead of a sinusoidal one (Gao *et al* 1997). In such a case, the phase $\phi(\boldsymbol{r}, t)$ is calculated in a similar manner with a similar accuracy but in half the time.

2.2.4. Logical moiré. In this method, the pattern, into which a spatial carrier has been introduced, is combined with a computer-generated binary reference grating (Asundi 1993), assigning zero and unit values to the various points to obtain a one-bit binary carrier. The carrier period can be easily varied and its minimum value is two pixels. The fringe pattern with a carrier is also converted into a binary format by a process whereby a threshold is defined in order to assign its extreme values. Using the AND and XOR logic operators, a multiplicative moiré pattern is obtained (Asundi and Yung 1991a) by the overlapping of the computer generated binary grating and the binary pattern. This moiré pattern can be phase modulated by shifting the digital binary grating by an integer number of points. The various moiré patterns can be combined using the usual PSAs to obtain the original phase $\phi(\boldsymbol{r}, t)$ (Asundi and Yung 1991b). In this case, the minimum binary grating shift is one pixel. However, in order to avoid result fluctuations, the binary grating shift must be equal to its period divided by the number of patterns necessary for each PSA.

The entire process is carried out directly in the computer. Knowledge of the carrier frequency is not a critical factor because the binary-grating period is generated after the pattern has been acquired and can be modified in a simple manner.

On the other hand, this method requires a greater processing time than do the methods mentioned earlier because (i) the pattern must be converted into a binary format; (ii) its average period must be obtained with a view to defining the binary-grating period, to obtain a low-frequency moiré pattern; and (iii) this moiré pattern must be isolated from the rest of the terms which originate from the superposition.

2.2.5. The multiplicative analogical moiré method. This method has principally been developed to evaluate fringe patterns with non-sinusoidal profiles (Dorrío *et al* 1995). It is based on the combination of the moiré effect and the PSAs (Józwicki *et al* 1992, Malacara *et al* 1998, Schwider *et al* 1986, 1987). The introduction of a high-frequency spatial carrier into the pattern to be evaluated provides a phase-modulated carrier, whose demodulation is carried out by overlapping it with a transmission grating located outside the optical module which supplies the fringe pattern. In this case (Dorrío *et al* 1995), an output intensity distribution equal to the product of the phase-modulated carrier intensity and the corresponding transmission function is observed. If a coherent source is used, a diffusing screen is usually placed just behind the grating or, alternatively, a low-spatial-coherence illumination source may be used, in order to re-image this multiplicative moiré pattern in the detector. The lowest frequency pattern (i.e. the moiré image) is usually selected using a low-pass-filtering process. The moiré image phase contains the original phase $\phi(\boldsymbol{r}, t)$, a linear phase term and an additional phase term $\alpha(\boldsymbol{r}, t)$ proportional to the in-plane displacement of the grating. By acquiring a series of moiré images with various additional phase terms, one can apply a standard PSA.

This method contemplates the possibility that the fringe profile is of non-sinusoidal nature. Moreover, the modulator can be located outside the optical module which provides the fringe patterns, thus improving the measurement-process stability. Additionally, (i) the modulator dimensions can be reduced considerably with respect to the wavefronts and (ii) it is independent of the illumination-source wavelength and the image-forming system. Likewise, it is not necessary to know the frequency of the spatial carrier. It does not have to be resolved by the acquisition system (which improves the measuring range) and can be analogically cancelled out using infinite moiré fringes. On the other hand, the method shows sensitivity to environmental perturbations, similar to TPSM, and is not a real-time method. All of the above makes the MAMM an appropriate alternative when (Dorrío *et al* 1997, 1998) (i) the cost is an important factor, (ii) the wavefronts have big dimensions, (iii) there is the need to add an evaluation module to a closed optical measurement system or (iv) high spatial resolution and measurement range are not relevant factors.

3. Phase-unwrapping algorithms

Owing to the multiple value of the inverse trigonometric function employed to calculate the phase $\phi(\boldsymbol{r}, t)$ in the majority of the PEMs (all except HM and PLI), these usually provide the phase $\phi(\boldsymbol{r}, t)$ value modulo π. However, it is possible to identify the quadrant where the phase $\phi(\boldsymbol{r}, t)$ is located if the signs of the corresponding argument are taken into account. These PEMs therefore provide phase $\phi(\boldsymbol{r}, t)$ values in the range $[0, 2\pi)$ rad, the so-called principal values. In this way, the obtained results exhibit a discontinuity every time the phase $\phi(\boldsymbol{r}, t)$ increases or decreases by 2π rad with respect to a certain arbitrary phase origin. These discontinuities need to be resolved, for there is usually a need to obtain continuous values of the phase $\phi(\boldsymbol{r}, t)$. This process is called phase unwrapping and it is a trivial process, as long as the patterns are continuous, sampled with high enough density and present a good signal-to-noise ratio. In such a case, the process is carried out by adding (or subtracting) 2π rad every time a discontinuity is detected in the principal phase $\phi(\boldsymbol{r}, t)$ values, with a view to obtaining the unwrapped phase $\phi^U(\boldsymbol{r}, t)$ (figure 4) (Macy 1983, Takeda *et al* 1982). However, in many situations, the fringe patterns are of a low quality and the unwrapping process becomes dramatically complicated.

Many different algorithms have been designed for phase unwrapping which carry out the process by defining criteria for the correct identification of the discontinuities in the corresponding principal values, assuming the existence of a unique one-to-one correspondence between the measurement magnitude and the phase map (Judge and Bryanston-Cross 1994, Malacara *et al* 1998, Robinson 1993, Takeda 1996). These algorithms can be classified (Charette and Hunter 1996, Ettl and Creath 1996) into (i) path-independent or global algorithms, those that bring about the unwrapping process in one step, identifying, isolating and excluding those principal phase $\phi(\boldsymbol{r}, t)$ value zones which could lead to an error-prone identification of the discontinuities, in which case the unwrapping process is performed via any arbitrary path; and (ii) path-dependent or local algorithms, those that provide a continuous point-by-point phase $\phi(\boldsymbol{r}, t)$ value, via a determined path. All of these algorithms are complex to process but compensate by providing an efficient treatment of the problematic points during the unwrapping process, that is, the intensity values which (i) present a noise comparable to the fringe amplitude, (ii) have a low modulation, (iii) exhibit an abrupt phase change due to a discontinuity of the measurand, thus nonrelated with the use of an inverse trigonometric function or (iv) present a very low sampling density. The various algorithms can be distinguished by (i) their processing times, (ii) their sensitivities to error propagation during the unwrapping process, (iii) their initial hypotheses about the principal phase $\phi(\boldsymbol{r}, t)$ value distribution inconsistencies and (iv) their robustness against the problematic points mentioned earlier. However, none of them can account for all of the possible effects that might arise, each of them being dedicated to partially resolving the problem and in many a case needing additional information. Therefore, at present there is no totally automated phase-unwrapping process, although this is one of the research fields within fringe-pattern analysis which receives a lot of research attention.

Table 2. Principal characteristics of PEMs with a spatial carrier.

	Method				
	FTM	SCPSM, SFM	SSD	LM	MAMM
Number of patterns	1	1	1	1	$\geqslant 3$
Real-time operation	Yes	Yes	Yes	Yes	No
Accuracy	High	High	Medium	Medium	High
Influence of $\frac{\text{static}}{\text{dynamic}}$ noise	$\frac{\text{High}}{\text{Low}}$	$\frac{\text{High}}{\text{Low}}$	$\frac{\text{High}}{\text{Low}}$	$\frac{\text{High}}{\text{Low}}$	$\frac{\text{Low}}{\text{High}}$
Experimental requisites	Low	Low	Low	Low	High
Processing complexity	High	Medium	Low	High	Medium
Operator interaction	Possible	Possible	Possible	Possible	Impossible
Process duration	Long	Short	Short	Long	Short
Commercial systems	Yes	Yes	Yes	No	No
Range	Low	Low	Low	Low	Medium
Cost	Medium	Medium	Medium	Medium	Low

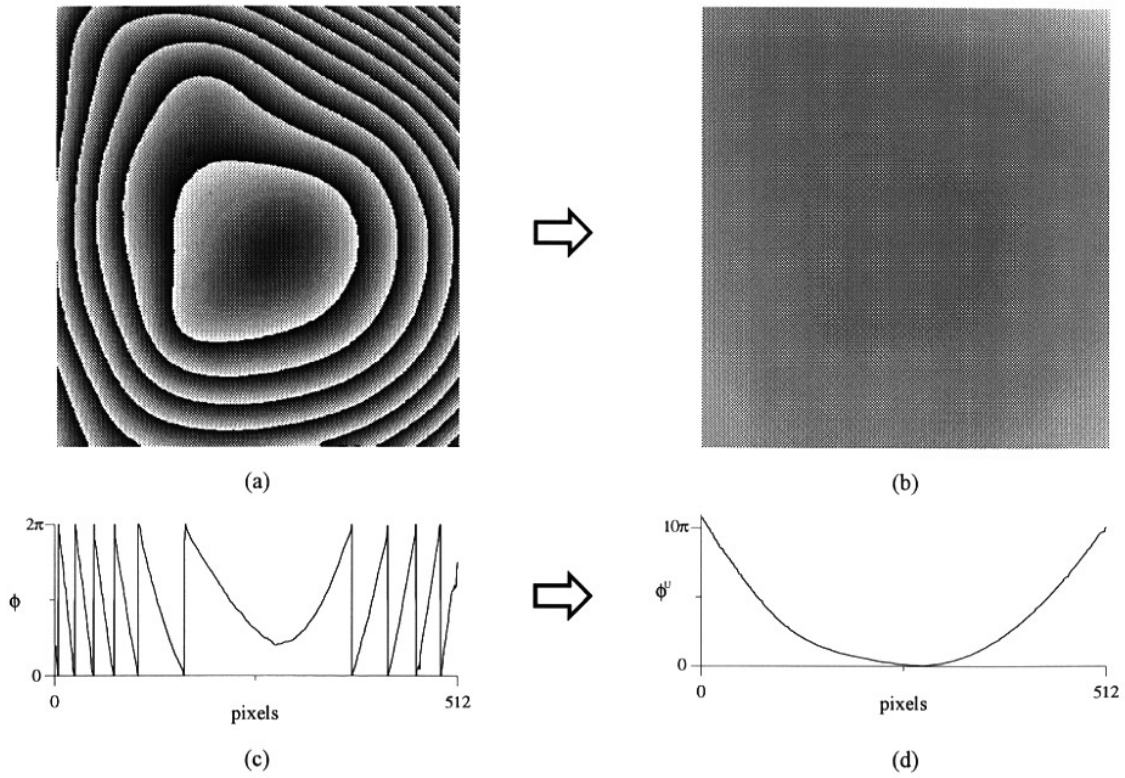


Figure 4. A bidimensional representation on a grey scale of (a) principal phase ϕ values and (b) the unwrapped phase ϕ^U of a fringe pattern. In (c) and (d) an example of the unwrapping of the central horizontal image line is shown.

4. Phase-stepping algorithms

Many PEMs employ phase-stepping algorithms (PSAs) as a tool for obtaining the phase $\phi(\boldsymbol{r}, t)$: SCPSM, SSD, LM, MAMM, SFM, TPSM, SPSM and SD. Some of them generally use all PSAs whereas others use just a few of the PSAs. The PSAs, which date back to the 1960s (Carré 1966), permit calculation of a phase with its polarity within a short time interval, with a low computational complexity (even in the presence of boundaries or discontinuities) and without the need for researcher intervention, after the intensity values have been obtained. The above reasons justify their use in the great majority of the present commercial optical measurement devices.

The PSAs generally assume that the fringe pattern to be used presents a sinusoidal profile. In such a case, equation (3) would be written as

$$i(\boldsymbol{r}, t, \alpha) = i_0(\boldsymbol{r}, t)\{1 + V(\boldsymbol{r}, t)\cos[\phi(\boldsymbol{r}, t) + \alpha(\boldsymbol{r}, t)]\} \quad (10)$$

where the additional phase $\alpha(\boldsymbol{r}, t)$ is the so-called phase shift. In such cases, the PSAs calculate the phase $\phi(\boldsymbol{r}, t)$ by combining n intensity values $i_k(\boldsymbol{r}, t)$ proceeding from one or more fringe patterns in which the phase shift takes different values $\alpha_k(\boldsymbol{r}, t)$. This phase $\alpha_k(\boldsymbol{r}, t)$ may vary within the different intensity values $i_k(\boldsymbol{r}, t)$ in three alternative ways: (a) for point pertaining (i) to several analogically obtained fringe patterns at different times (MAMM, TPSM and SD),

either discretely or in a continuous form; or (ii) to several digitally obtained fringe patterns, which arise from a single pattern, at different processing times (SSD and LM) (in both cases the intensity values can be expressed as $i_k(\boldsymbol{r}, t_k)$); (b) for a single pattern between different points of the same fringe pattern (SCPSM and SFM), in which case the intensity values are expressed as $i_k(\boldsymbol{r}, t)$; (c) or for the same point of different fringe patterns acquired at the same instant of time (SPSM), for which the intensity values to be employed will be denoted $i_k(\boldsymbol{r}^k, t)$. In any case, the intensity values that are combined in a PSA are phase shifted and can be generically represented by i_k, the temporal and spatial dependencies of which would vary according to the evaluation method considered (figure 5). These dependencies are explicitly avoided from now on, in order to adequately abbreviate the notations of the respective expressions. Therefore, the phase-shifted intensity values can be written according to equation (10) as

$$i_k = i_0 + i_0 V \cos(\phi + \alpha_k) \qquad k = 1, 2, \ldots, n \tag{11}$$

The main advantages of the PSAs are as follows.

(i) Low sensitivities to noise stationary in the domain in which α varies. For example, if $\alpha = \alpha(t)$, the PSA is insensitive to non-uniformities in the background intensity i_0, in the visibility V and in the local detector gain.

(ii) They can be used with low-contrast patterns.

(iii) The accuracy of the results is finally limited by the fringe-pattern signal-to-noise ratio (quantification errors, etc.). The systematic error effect can be minimized to a level such that only random noise limits the measurement accuracy.

(iv) They can be completely automated.

(v) The phase sign $\phi(\boldsymbol{r}, t)$ is determined for each point.

(vi) The spatial resolution is high, because the number of measurement points coincides with the number of detector elements used to acquire the intensity values.

(vii) Upon calculating the phase ϕ over a fixed matrix of points (the image digitizer detector matrix), a high geometric fidelity and a uniform spatial sampling are ensured even at the borders or in the presence of discontinuities. In this case, the aberrations of the image-forming system produce a distortion which has a typical value of less than 5%.

(viii) The computer calculating power at present available, including that of personal computers, permits the phase ϕ calculation operations to take place within a short time (even less than 1 s) for a whole image.

In like manner, the principal limitation of these algorithms is the fact that the phase ϕ and its variations between successive points are known as modulo 2π. This means that the maximum spatial fringe frequency measurable without undersampling is equal to half the detector sampling frequency, that is, its Nyquist frequency. However, this limit can be extended if (i) the sensitivity is reduced via a synthetic wavelength, using two patterns of the same measurand obtained with different sensitivities (Creath *et al* 1985, Wyant 1971); (ii) previous information about the wavefront and its first derivatives are known (Greivenkamp 1987); (iii) the wavefront form is approximately known such that a lower frequency fringe pattern is obtained as a correlation of the original pattern and another reference one generated with this information (Servín *et al* 1994); or (iv) the different wavefront zones are analysed sequentially according to the progressive introduction of different grades of defocusing (Malacara *et al* 1995).

4.1. Classification of the PSAs

The PSAs can be classified and compared according to diverse criteria (Creath 1988, 1993, Hibino *et al* 1997, Greivenkamp and Bruning 1992, Malacara *et al* 1998). The taxonomy which is proposed here divides them into two big groups. The biggest group is made of what we shall call generic PSAs obtained by a systematic process and that is why new algorithms of this type continuously appear in the literature. The rest shall be accommodated among the so-called specific PSAs, as long as the original hypothesis defining the generic PSAs is not verified and the phase ϕ is not the directly obtained result.

4.1.1. Generic PSAs. These algorithms generally use known additional phase values α_k, as a result of which, in the expression of the intensity values i_k, equation (11), there are only three unknowns, namely i_0, V and ϕ. Therefore, three values of intensities i_k are needed in order to obtain the corresponding analytical solution, for example (Frantz *et al* 1979, Joenathan 1994, Wyant *et al* 1984)

$$^{3,\pi/2}\phi = \arctan\left(\frac{2i_2 - i_1 - i_3}{i_3 - i_1}\right) \tag{12a}$$

$$^{3,\pi/2}\phi = \arctan\left(\frac{i_3 - i_2}{i_1 - i_2}\right) - \frac{\pi}{4} \tag{12b}$$

$$^{3,2\pi/3}\phi = \arctan\left(\frac{\sqrt{3}i_1 - \sqrt{3}i_3}{2i_2 - i_1 - i_3}\right) + \frac{2\pi}{3} \tag{12c}$$

with $\alpha_k = (k-1)\pi/2$ rad. Although in some specific cases for $n > 3$ one can develop PSAs in an intuitive manner, like for example, the popular Wyant PSA (Wyant 1975)

$$^{4,\pi/2}\phi = \arctan\left(\frac{i_4 - i_2}{i_1 - i_3}\right) \tag{13}$$

where $\alpha_k = (k-1)\pi/2$ rad, the quickest systematic form for finding a solution to this type of problem is by using a least squares fit of the intensity values to a sinusoidal function (Greivenkamp 1984, Lai and Yatagai 1991, Morgan 1982). The general complexity derived from the obtained expressions is considerably reduced if the relative phase shift α_r (i.e. the phase increment between consecutive intensity values) given by

$$\alpha_r = \alpha_{k+1} - \alpha_k \tag{14}$$

is constant for any k values in the interval $n > k \geqslant 1$, in which case the value of the additional phase α_k is an integer multiple of the relative phase shift α_r (see equations (12) and (13)) and can be generally expressed as

$$\alpha_k = (k-1)\alpha_r. \tag{15}$$

Unless stated otherwise, from now on it is assumed that equation (15) is fulfilled, in which case the phase ϕ can be calculated as

$$^{n,\alpha_r}\phi = \arctan\left(\frac{N}{D}\right) + {}^{n,\alpha_r}\phi_0 \tag{16}$$

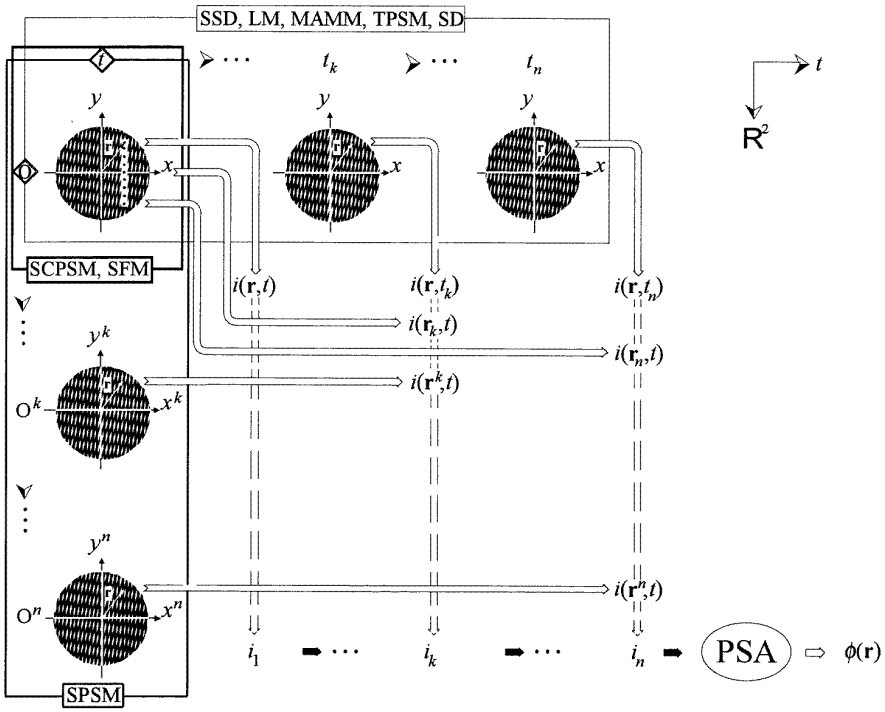


Figure 5. Various approaches for obtaining phase-shifted intensity values i_k: in the processing temporal domain with a spatial carrier (SSD, LM), in the temporal domain with a spatial carrier (MAMM), in the temporal domain without a spatial carrier (TPSM and SD), in the spatial domain with a spatial carrier (SCPSM and SFM) and in the spatial domain without a spatial carrier (SPSM).

where N and D are linear combinations of n ($n \geqslant 3$) phase-shifted intensity values i_k and defined as

$$N = \sum_{k=1}^{n} \mu_k i_k = C \sin\phi \tag{17a}$$

$$D = \sum_{k=1}^{n} \nu_k i_k = C \cos\phi \tag{17b}$$

μ_k and ν_k being real numbers, called PSA sampling amplitudes (Freischlad and Koliopoulos 1990). These sampling amplitudes satisfy the following two conditions, obtained by substituting equation (11) into equations (17a) and (17b):

$$\sum_{k=1}^{n} \mu_k \sin\alpha_k = \sum_{k=1}^{n} \nu_k \cos\alpha_k = 1 \tag{18a}$$

$$\sum_{k=1}^{n} \mu_k \cos\alpha_k = \sum_{k=1}^{n} \nu_k \sin\alpha_k = 0 \tag{18b}$$

The term ${}^{n,\alpha_r}\phi_0$ in equation (16) is an additional constant value which can be annulled by appropriately selecting the initial phase value α_1, in which case equation (15) can be written as $\alpha_k = (k - s)\alpha_r$, where s is a certain real number. Likewise, C is equal to the product of the background intensity i_0 and the visibility V and can be easily expressed using expressions (17a) and (17b) as

$$C = \left[\left(\sum_{k=1}^{n} \mu_k i_k \right)^2 + \left(\sum_{k=1}^{n} \nu_k i_k \right)^2 \right]^{1/2}. \tag{19}$$

The low-modulation points (those in which C takes an almost null value, the signal-to-noise ratio reaches low levels and, therefore, expression (16) provides an unreliable result) are identified by analysing the value of C. This permits the interpolation of its values or allows one simply to avoid them in subsequent analysis processes.

The simplicity of the formulation of generic PSAs is made evident by the fact that any of these algorithms, or any other which can be expressed using equation (16), can be constructed in a simple manner if the phase-shifted intensity values i_k are represented as vectors in a bidimensional intensity space (Brophy 1990). Each of these vectors located on radii of a circumference is angularly separated from the consecutive one by a quantity equal to the relative phase shift α_r (this angle is measured clockwise if its value is positive and anti-clockwise if its value is negative). In cases in which there exists an initial constant phase, the first vector is angularly displaced from the abscissa by a quantity equal to ${}^{n,\alpha_r}\phi_0$, clockwise if its value is negative and anti-clockwise if it is positive. An analysis of the symmetry around the coordinate axes of this intensity space shows us that a vector on the abscissa has just the one $\cos\phi$ dependency and a vector on the ordinate axis has just the one $\sin\phi$ dependency. Thus, these vectors are distributed such that a linear combination of them, N, is found on the ordinate and the other, D, is found on the abscissa, so that the sums of the vector coefficients in each of the linear combinations are zero, to cancel the i_0 effect; that is,

$$\sum_{k=1}^{n} \mu_k = \sum_{k=1}^{n} \upsilon_k = 0 \tag{20}$$

while the quotient for both combinations should annul the dependency on $i_0 V$, and can be appropriately used later as the argument of the arctangent function in equation (16). Figure 6 graphically shows this simple tool for construction of various PSAs (Schmit and Creath 1995) with six i_k intensity values.

On the other hand, the PSAs designed by a least squares fit do not always have an optimal behaviour against certain errors. This has resulted in the design of alternative techniques. The design of PSAs robust against random errors has been tackled using the maximum-likelihood estimation theory (Rogala and Barrett 1997) and similar results to the above have been obtained, whereas the optimization of PSA against certain systematic errors has been based on strategies which combine pre-existing generic PSAs (Joenathan 1994, Schmit and Creath 1995, Schwider *et al* 1983) or undertake the task of the design of a filter in reciprocal space (Freischlad and Koliopoulos 1990, Hibino *et al* 1995, Larkin and Oreb 1992, Zhang *et al* 1998). Likewise, there exist other PSAs which are obtained by imposing additional fitting conditions related to algorithm 'robustness' against certain systematic errors (Hariharan *et al* 1987). Thus they can be distinguished as follows.

4.1.1.1. Average PSAs. On the one hand, this averaging can be done directly with calculated phase values with different PSAs, for example (Creath 1986)

$$ {}^{4,\pi/2}\phi = \frac{1}{2}\left[\arctan\left(\frac{i_3 - i_2}{i_1 - i_2}\right) + \arctan\left(\frac{i_4 - i_3}{i_2 - i_3}\right)\right] - \frac{\pi}{4} \tag{21} $$

obtaining the latter as the average of two series of calculated phases according to equation (12*b*), or by appropriate weighting of PSAs to improve their immunity to certain systematic errors, like for example (Joenathan 1994)

$$ {}^{5,\pi/2}\phi = \frac{1}{2}\left[\arctan\left(\frac{i_4 - i_2}{i_1 - i_3}\right) + \arctan\left(\frac{i_4 - i_2}{i_5 - i_3}\right)\right] + \frac{15}{100}\left[\arctan\left(\frac{i_5 - i_2}{i_1 - i_4}\right) - \frac{\pi}{4}\right]. \tag{22} $$

On the other hand, one can obtain an average PSA by following a procedure (Schwider *et al* 1983, 1993) which consists of putting forward a new PSA with two series of n intensity values, phase shifted by $\pi/2$ rad and by respectively averaging the terms N and D, equations (17*a*) and (17*b*), of two precursor PSAs. The results are taken as the argument of the corresponding arctangent function. The PSAs so obtained are a lot less sensitive to calibration errors than are their precursors. Generally, for the application of these types of PSAs one needs a greater number of intensity values ($2n$) but if one starts with one PSA with a relative phase shift α_r equal to $\pi/2$ rad, there is a need for just one additional intensity value ($n + 1$) in order to get both series (Surrel 1993). The first algorithm of this type, called the Schwider–Hariharan PSA (Hariharan *et al* 1987, Schwider *et al* 1983), had its origin in the Wyant PSA, equation (13), and is expressed as

$$ {}^{5,\pi/2}\phi = \arctan\left(\frac{2i_2 - 2i_4}{2i_3 - i_5 - i_1}\right). \tag{23} $$

Repeated use of this technique to obtain an entire family of average PSAs from generic PSAs, as well as the average of multiple series of intensity values, like for example

$$ {}^{6,\pi/2}\phi = \arctan\left(\frac{-i_1 - 3i_2 + 4i_3 + 4i_4 - 3i_5 - i_6}{i_1 - 3i_2 - 4i_3 + 4i_4 + 3i_5 - i_6}\right) + \frac{\pi}{4} \tag{24} $$

has recently been proposed (Schmit and Creath 1995,1996).

4.1.1.2. Reciprocal-space-adjusted PSAs. The capacity of a generic PSA to compensate for certain systematic and random errors can be analysed by its Fourier representation, identifying the application of a PSA with a filtering process in the reciprocal space and by analysing the PSA according to its frequency representation (Freischlad and Koliopoulos 1990, Hibino *et al* 1995, 1997, Malacara *et al* 1998, Onodera and Ishii 1996, Schmit and Creath 1995, Zhang *et al* 1998). Therefore, the phase-shifted intensity values i_k are considered to be sampled values of the intensity $i(\alpha)$, equation (10), at points α_k. The sampling functions of the numerator N and the denominator D of a generic PSA, equations (17*a*) and (17*b*), are

$$ f_N(\alpha) = \sum_{k=1}^{n} \mu_k \delta(\alpha - \alpha_k) \tag{25a} $$

$$ f_D(\alpha) = \sum_{k=1}^{n} \nu_k \delta(\alpha - \alpha_k) \tag{25b} $$

where $\delta(\alpha)$ is the Dirac delta function, α acts as a phase-shift parameter with a modulation frequency ξ_0, according to equation (4), and in this case the phase ϕ can be expressed as (Hibino *et al* 1995)

$$ \phi = \arctan\left(\frac{\int_{-\infty}^{\infty} i(\alpha) f_N(\alpha)\,\mathrm{d}\alpha}{\int_{-\infty}^{\infty} i(\alpha) f_D(\alpha)\,\mathrm{d}\alpha}\right) \tag{26} $$

or, according to Parseval's identity, in the form

$$ \phi = \arctan\left(\frac{\int_{-\infty}^{\infty} I(\Omega) F_N(\Omega)\,\mathrm{d}\Omega}{\int_{-\infty}^{\infty} I(\Omega) F_D(\Omega)\,\mathrm{d}\Omega}\right) \tag{27} $$

where Ω is the angular frequency, with a fundamental value $\Omega_f = 2\pi\xi_0$, and $I(\Omega)$ is the Fourier transform of $i(\alpha)$. The respective N and D filtering functions are obtained by calculating the Fourier transforms of $f_N(\alpha)$ and $f_D(\alpha)$:

$$ F_N(\Omega) = \sum_{k=1}^{n} \mu_k \exp(-\mathrm{j}\alpha_k\Omega) \tag{28a} $$

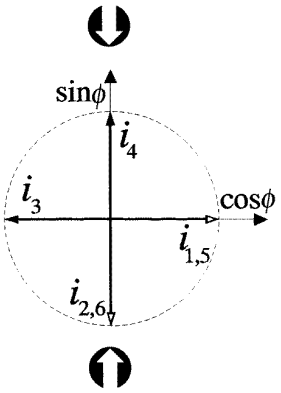
$$ F_D(\Omega) = \sum_{k=1}^{n} \nu_k \exp(-\mathrm{j}\alpha_k\Omega). \tag{28b} $$

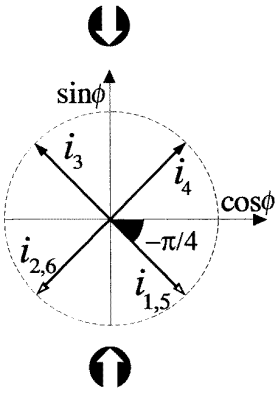
These functions determine the sensitivity of each PSA to the various systematic errors, such that, in order to obtain the correct phase ϕ, $F_N(\Omega)$ and $F_D(\Omega)$ must essentially satisfy the following conditions (Freischlad and Koliopoulos 1990, Malacara-Doblado *et al* 1997):

$$ F_N(0) = F_D(0) \tag{29a} $$

$$ F_N(\Omega_f) + \mathrm{j}F_D(\Omega_f) = 0. \tag{29b} $$

$$ {}^{6,\pi/2}\phi = \tan^{-1}\left\{\frac{-3i_2 + 4i_4 - i_6}{i_1 - 4i_3 + 3i_5}\right\} \qquad {}^{6,\pi/2}\phi = \tan^{-1}\left\{\frac{-i_1 - 3i_2 + 4i_3 + 4i_4 - 3i_5 - i_6}{i_1 - 3i_2 - 4i_3 + 4i_4 + 3i_5 - i_6}\right\} - \frac{\pi}{4} $$





$$ {}^{6,\pi/2}\phi = \tan^{-1}\left\{\frac{i_1 - 5i_2 - 2i_3 + 10i_4 - 3i_5 - i_6}{i_1 + 3i_2 - 10i_3 + 2i_4 + 5i_5 - i_6}\right\} \qquad {}^{6,\pi/2}\phi = \tan^{-1}\left\{\frac{-4i_2 + 4i_3 + 4i_4 - 4i_5}{i_1 - i_2 - 6i_3 + 6i_4 + i_5 - i_6}\right\} - \frac{\pi}{4} $$

Figure 6. A graphical representation of various PSAs in the intensity space.

The gradients of the filtering functions $F_N(\Omega)$ and $F_D(\Omega)$, at the fundamental frequency Ω_f, describe the sensitivity of a PSA to errors in the relative phase shift α_r, whereas the values of these functions and their gradients at higher order frequencies, $\Omega = l\Omega_f$, determine their sensitivities to higher order harmonics l in the profile, equation (2) and to detuning in their detection.

Likewise, any PSA can be represented, according to equation (28), as a function of its Fourier transform vectors (Malacara-Doblado *et al* 1997):

$$ \vec{F}_N = \left(\sum_{k=1}^{n} \mu_k \cos(\alpha_k \Omega_f), \sum_{k=1}^{n} \mu_k \sin(\alpha_k \Omega_f)\right) \qquad (30a) $$

$$ \vec{F}_D = \left(\sum_{k=1}^{n} \upsilon_k \cos(\alpha_k \Omega_f), \sum_{k=1}^{n} \upsilon_k \sin(\alpha_k \Omega_f)\right) \qquad (30b) $$

in which case the orthogonal and equal amplitude conditions of the sampling functions, equations (29), are equivalent to the orthogonality and equal modulus conditions of these sampling vectors in reciprocal space (Malacara-Doblado *et al* 1997).

The so-called symmetrical PSAs, which are able to simultaneously reduce the effects of the phase-calibration error and the non-sinusoidal profile of the pattern, can be obtained for example by the above analysis strategy in reciprocal space. They are also called $n+1$ PSAs because they use n intensity values with regularly distributed phase shifts plus an additional intensity shifted by $\alpha_{n+1} = \alpha_1 + 2\pi$. The Schwider–Hariharan PSA, equation (23), can be analysed from this perspective and also alternatively considered as an $n + 1$ PSA (figure 7). Within the last few years, a technique to systematically design this type of PSA with an arbitrary number of intensity values ($n + 1 \geqslant 4$) has been proposed (Larkin and Oreb 1992), for example

$$ {}^{7,\pi/3}\phi = \arctan\left(\frac{\sqrt{3}(i_2 + i_3 - i_5 - i_6) + (i_7 - i_1)/\sqrt{3}}{-i_1 - i_2 + i_3 + 2i_4 + i_5 - i_6 - i_7}\right) \qquad (31) $$

where $\alpha_k = (k - 4)\pi/3$ rad.

In like manner, by applying additional conditions to the intensity values, one can obtain 'robust' PSAs insensitive to certain systematic errors, like for example (Zhao and Surrel 1995)

$$ {}^{6,\pi/2}\phi = \arctan\left(\frac{2i_4 + 2i_5 - 2i_2 - 2i_3}{i_1 + i_2 + i_5 + i_6 - 2(i_3 + i_4)}\right) - \frac{\pi}{4} \qquad (32) $$

that improves their immunity against second-order harmonics in the intensity profile and that given by Hibino *et al* (1997):

$$ \begin{aligned} {}^{9,\pi/2}\phi = \arctan[&(i_1 - i_9 - i_2 + i_8 - 7i_3 + 7i_7 - 9i_4 - 9i_6) \\ &\times (2i_1 + 2i_9 + 8i_2 + 8i_8 + 8i_3 + 8i_7 - 8i_4 - 8i_6 \\ &- 20i_5)^{-1}] \end{aligned} \qquad (33) $$

which is insensitive to a quadratic and non-uniform relative phase shift.

4.1.2. Specific PSAs. Specific PSAs are those which cannot be obtained by a least squares fit or by any equivalent procedure. Therefore, their expressions do not generally amount to that given by equation (16). Amongst these, we can distinguish the following.

4.1.2.1. Compensated relative phase shift PSAs. The characteristics of these PSAs are that they use an unknown relative phase shift α_r, having an identical value for each of the intensity values used. They can be applied although α_r is not uniform in space or there exist great spatial calibration errors. The first published PSA (Carré 1966), known as

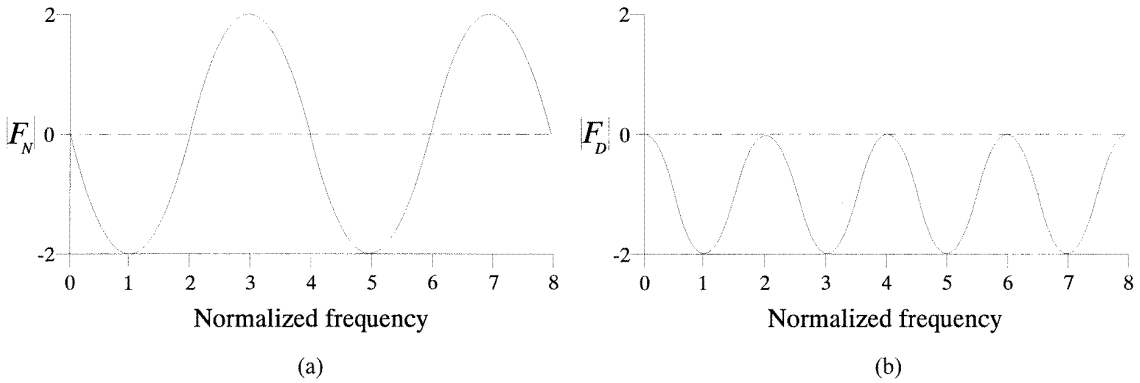


Figure 7. A graphical representation in reciprocal space of the Schwider–Hariharan PSA. The amplitudes of its filtering functions are associated with the normalized frequency, Ω/Ω_f.

the Carré PSA, belongs to this class and permits phase ϕ calculation by combining four intensity values using the expression

$$^{4,\alpha_r}\phi = \arctan[\{[(i_1 - i_4) + (i_2 - i_3)] \times [3(i_2 - i_3) - (i_1 - i_4)]\}^{1/2}[(i_2 + i_3) - (i_1 + i_4)]^{-1}] \quad (34)$$

with $\alpha_4 = (k - 5/2)\alpha_r$ rad. In this case, it is not necessary to explicitly calculate the value of α_r, in contrast to other specific PSAs with an arbitrary relative phase shift α_r (Colonna de Lega and Jacquot 1996, Farrel and Player 1992, Lai and Yatagai 1991, Kong and Kim 1995, Kreis 1987, 1993, 1996). This PSA is nowadays widely used, above all in those cases in which a precision calibration of the relative phase shift α_r is not possible. On the other hand, the main inconveniences are the fact that it produces inaccuracies in the phase ϕ when both the numerator and the denominator simultaneously exhibit values close to zero (which happens when the phase ϕ is near to a multiple of π rad) and when the root argument takes negative values. The Carré PSA also exhibits a certain sensitivity to nonlinear miscalibration of the phase shift. The latter problem can be partially eliminated by combining five intensity patterns (Larkin 1996):

$$^{5,\alpha_r}\phi = \arctan\left(\frac{[4(i_2 - i_4)^2 - (i_1 - i_5)^2]^{1/2}}{-i_1 + 2i_3 - i_5}\right). \quad (35)$$

One can finally obtain a certain insensitivity to external vibrations by calculating the phase as (Sotoilov and Dragostinov 1997)

$$^{5,\alpha_r}\phi = \arctan[2(i_2 - i_4)((2i_3 - i_1 - i_5) \times \{1 - [(i_1 - i_5)/2(i_2 - i_4)]^2\}^{1/2})^{-1}] \quad (36)$$

where $\alpha_k = (k - 3)\alpha_r$ rad.

4.1.2.2. Bönsch PSA. This PSA (Bönsch and Böhme 1989) was designed to evaluate the coded phase in fringe patterns provided by multiple wavefront interference. The results provided by the generic PSAs in this case give an error that becomes relevant in the presence of harmonics a_l in the pattern profile with an order higher than the first one, equation (2), or the number of intensity values used in the PSA is low. Both problems are simultaneously solved by combining only four intensity values, with a relative phase shift α_r equal to $\pi/2$ rad:

$$^{4,\pi/2}\phi = \arctan\left[\left(\frac{(i_1 - i_4)(i_2 - i_3)}{(i_2 - i_1)(i_3 - i_4)}\right)^{1/2}\right]. \quad (37)$$

This PSA shares the same limitations as those of the Carré PSA, and besides the effects on the phase ϕ produced by most of the systematic errors (Bönsch and Böhme 1989, Nicolaus 1993) have practically not been studied. Both reasons at present limit the general use of these PSAs for the purpose for which they were designed.

4.1.2.3. Variable-background-intensity PSAs. Almost all of the PSAs assume that the background intensity i_0 and the pattern visibility V are independent of the phase variations α, equation (10). However, in some situations of great practical interest, the mechanism used to introduce the phase shift induces a variation of the background intensity i_0. During the last few years, a few specific PSAs that permit the phase ϕ to be calculated with precision under these conditions, as long as one knows how the phase shift α_k affects i_0, have been developed.

Therefore, if there is a linear relationship (Ishii and Onodera 1995, Onodera and Ishii 1996) between i_0 and α_k, the phase ϕ is calculated as

$$^{6,\pi/2}\phi = \arctan\left(\frac{3i_1 - 5i_2 + 5i_5 - 3i_6}{i_1 + 3i_2 - 4i_3 - 4i_4 + 3i_5 + i_6}\right) \quad (38)$$

or as (Surrel 1997b)

$$^{5,\pi/2}\phi = \arctan\left(\frac{-i_1 + 2i_2 - 2i_4 + i_5}{-2i_2 + 4i_3 - 2i_4}\right) \quad (39)$$

but if i_0 depends sinusoidally (Moore *et al* 1994) on the value of α_k

$$^{4,\pi/2}\phi = \arctan\left(\frac{i_2^2 - i_4^2}{i_3^2 - i_1^2}\right). \quad (40)$$

4.1.2.4. The 2(+1) pattern PSA. This algorithm (Colucci and Wizinowich 1992, Wizinowich 1990) allows a small acquisition time, which permits phase evaluation in the presence of vibrations or perturbations that would otherwise

cause appreciable intensity variations in the pattern. It uses two fringe patterns, phase shifted by $\pi/2$ rad, which are acquired in rapid sequence using a video camera. A third pattern obtained as an average of these two patterns displaced by π rad provides the background intensity i_0. The phase is obtained as

$$^{2(+1),\pi/2}\phi = \arctan\left(\frac{i_2 - i_0}{i_1 - i_0}\right). \tag{41}$$

In certain cases (Ng 1996), this third pattern which permits the calculation of the background intensity i_0 can be achieved by an appropriate sinusoidal phase modulation. On the other hand, the background intensity i_0 can be eliminated by high-pass filtering (Kerr *et al* 1990) or by evaluating the difference between the two patterns (Joenathan and Khorana 1992). Thus, there is need for the use of just two patterns in the last two cases.

4.1.2.5. Phase-difference-evaluation PSAs. These are PSAs for directly calculating the phase difference $(\phi' - \phi)$ between two sets of intensity values for fringe patterns i'_k and i_k, without explicitly evaluating their respective phases. Amongst these, one can highlight the PSA with 2+2 intensity values (Owner-Petersen 1991, Fachini *et al* 1993):

$$^{2+2,\pi/2}[\phi' - \phi] = 2\arctan\left(\frac{i'_1 - i_2}{i'_2 - i_1}\right) \tag{42}$$

with $\alpha'_k = (k-1)\pi/2$ and $\alpha_k = (k-2)\pi/2$; the PSA with $3+3$ intensity values (Vikram *et al* 1993):

$$\begin{aligned}^{3+3,\alpha_r}[\phi' - \phi] = \arccos[&((i'_3 - i'_2)[(i'_3 - i'_1) + (i'_2 - i'_1)] \\ &+(i_1 - i_3)[(i_3 - i_2) + (i_1 - i_2)]) \\ &\times((i'_3 - i'_2)[(i'_3 - i'_2) + (i'_1 - i'_2)] \\ &+(i_1 - i_3)[(i_3 - i_1) + (i_2 - i_1)])^{-1}]\end{aligned} \tag{43}$$

for $\alpha_k = (k-2)\alpha_r$ and the PSA with $4+4$ intensity values (Pryputniewicz 1992, Stetson 1990):

$$\begin{aligned}^{4+4,\pi/2}[\phi' - \phi] = \arctan[&([i_1 - i_3 + i'_2 - i'_4]^2 \\ &+[i_2 - i_4 - i'_1 + i'_3]^2 - [i_1 - i_3 - i'_2 + i'_4]^2 \\ &-[i_2 - i_4 + i'_1 - i'_3]^2) \times ([i_1 - i_3 + i'_1 - i'_3]^2 \\ &+[i_2 - i_4 + i'_2 - i'_4]^2 - [i_1 - i_3 - i'_1 + i'_3]^2 \\ &-[i_2 - i_4 + i'_2 + i'_4]^2)^{-1}].\end{aligned} \tag{44}$$

4.1.2.6. Phase-gradient evaluation PSAs. These PSAs directly calculate the phase–gradient $\nabla\phi$ component in a fixed direction (generally horizontal or vertical) either by numerically deriving the fringe patterns during the evaluation process (Singh and Sirkis 1994) or by using a finite-difference approximation (Facchini and Zanetta 1995, Kaufmann and Galizzi 1997) of the intensity gradient values:

$$^{4,\pi/2}\nabla\phi = \frac{1}{\Delta}\frac{[i_4^\Delta - i_2^\Delta][i_1 - i_3] - [i_4 - i_2][i_1^\Delta - i_3^\Delta]}{[i_4 - i_2]^2 + [i_1 - i_3]^2} \tag{45}$$

where Δ is the distance between adjacent points along the direction for which $\Delta\phi$ is calculated and the values i_k^Δ are those corresponding to the point $r+\Delta$. When phase gradients are obtained, there is obviously no need for unwrapping.

Table 3. Modulation techniques and their relationships with types of PEMs.

Analogue			Digital
Temporal	Spatial		
Continuous or discrete	Continuous	Discrete	LM
TPSM, SD MAMM	SCPSM SFM	SPSM	SSD

4.1.2.7. The max–min exploration PSA. This type of PSA (Vikhagen 1990, Wang and Grant 1995) employs a great number of intensity values ($n \geqslant 10$), which are generally acquired sequentially. The relative phase shift α_r does not have to be constant, known or uniform in space. The PSA essentially consists of determining the maximum i_{max} and minimum i_{min} intensity values at each point as the phase α changes and simultaneously calculating the background intensity i_0 and the visibility V. Thus, the phase ϕ can be calculated using the expression

$$^{n,\alpha_r}\phi = \arccos\left(\frac{2i - [i_{max} + i_{min}]}{i_{max} - i_{min}}\right). \tag{46}$$

4.2. Modulation techniques

We can classify the various techniques for phase modulation of the intensity values i_k by placing them into two big groups (table 3), analogue (temporal or spatial, continuous or discrete) and digital. A description of them follows, detailing the principal strategies which are used in each case to undertake their appropriate implementation.

4.2.1. Temporal analogue modulation. Temporal analogue modulation (Bruning *et al* 1974, Dorrío *et al* 1995, Greivenkamp and Bruning 1992) (used in TPSM, SD and MAMM) can be done either continuously or discretely. In the first case, an integration of the intensity values is obtained, which provokes a reduction of the visibility to which most PSAs are insensitive (Greivenkamp 1984). In this type of modulation, an analogue mechanism is used for deliberately introducing an additional variable phase into the pattern (the phase modulator). Also, this process can be induced by the slow changes and the instability of the measurement magnitude (Colonna de Lega and Jacquot 1996, Davies and Bryanston-Cross 1996). There are several techniques for carrying out this process, amongst which the modulators can be primarily grouped as mechanical or non-mechanical, according to whether or not there is the need for a linear (or angular) displacement of any of the modulator components for their operation. During this process, either one (or more) of the wavefronts which provide the fringe pattern becomes phase-shifted or the fringe pattern is directly modulated. The principal phase modulators are explained in detail below.

Amongst mechanical modulators one can highlight the following variants.

(i) A modulator that is made up of a set of piezoelectric transducers (PZTs) which define the position of an optical

surface (Ai and Wyant 1987, Boebel *et al* 1991, Cheng and Wyant 1985, Kong and Kim 1995, Wyant 1982). The application of an electrical voltage to the PZT induces a contraction or expansion, which produces the movement of the surface and, therefore, results in a relative displacement between the wavefronts, giving rise to the subsequent phase shift in the fringe pattern. These simple modulators, which are incorporated into most of the commercial metrological systems which use TPSM and SD, have a series of limitations due to their hysteresis and nonlinearity. They are additionally found to be difficult to use with big and heavy optical elements.

(ii) The modulator made up of an optical fibre, through which one of the wavefronts travels, and a PZT cylinder. In its most typical configuration, the fibre is wound onto the external surface of the PZT. The application of a voltage induces a radial expansion of the cylinder, which varies the fibre dimensions and its refractive index, bringing about the phase shift (Davies and Kingsley 1974, Martini 1987). In fact, one can apply any physical phenomenon which would cause a change in the optical path of the fibre, like for example, a temperature variation (Schwider 1990). The main drawback of this mechanical modulator lies in the appearance of a certain change in the polarization, together with nonlinearity between the electric voltage applied and the depth of modulation (Davies and Kingsley 1974, Martini 1987, Mercer and Beheim 1991).

(iii) A plano-parallel plate. This is introduced into the path of one of the wavefronts, with a certain tilt with respect to its propagation direction. By varying the angle with respect to the normal incidence of the wavefront, one can obtain a series of phase shifts. The plate must be of a high quality in order to obtain a uniform value for the additional phase (Brown 1993, Kaufmann and Jacquot 1990, Poon *et al* 1993).

(iv) A wedge compensator located in a similar configuration to that in the above plate method, whereby the additional phase is introduced into a wavefront by moving one of the compensator parts perpendicularly to the incidence direction. Just like in the plate method, the compensator must be of a high quality in order to obtain a uniform value for the additional phase (Owner-Petersen 1991, Schwider 1990).

(v) A diffraction grating (linear or radial, of reflection or transmission type) situated (i) perpendicular to the direction of one or more wavefronts, which are diffracted by the grating (Schwider *et al* 1986, 1987), or (ii) superimposed upon the fringe pattern directly (Dorrío *et al* 1995, 1997, 1998). In both cases, the displacement of the grating in its own plane provides a phase shift independently of the illumination-source wavelength and proportional to the diffraction order and/or to the grating spatial frequency. In order not to significantly modify the measurement magnitude, there is a need for a high-quality grating which presents spatial variations of a lower magnitude than its period (Dorrío *et al* 1998). If the diffraction effects are relevant, it becomes necessary to avoid overlapping of the various diffracted wavefronts and to provide the same efficiency for all diffraction orders (Schwider *et al* 1986, 1987).

(vi) Polarizing components that are situated perpendicular to the direction of the wavefront propagation. Apart from using the linear displacements mentioned above, one can also use rotational movements applied to polarization-changing optical systems (like $\lambda/2$ and $\lambda/4$ plates and polarizers) when the wavefronts present mutually perpendicular polarizations. This procedure is followed in order to bring about a relative phase shift of a multiple of π rad, for one complete rotation of the polarizing components. In this case, the various additional phase values are obtained by precisely varying the azimuthal angles of certain components (Kadono *et al* 1987, Kothiyal and Delisle 1985, Jin *et al* 1994, Salbut and Patorsky 1990).

Other modulators have been designed as alternatives to mechanical modulators (which have, due to inertia, a limited bandwidth in the kilohertz and in the tens of kilohertz ranges). These dispense with the mechanical movement of any component to obtain the additional phase; they are the so-called non-mechanical modulators (with a bandwidth that can be extended to the megahertz and even gigahertz ranges). Amongst the latter, we can highlight the following.

(i) The direct modulation of a laser diode, when it is used as an illumination source. This can be obtained by varying the injection current or the active region temperature of the laser diode. This produces a change in the wavelength, resulting in an additional phase α that is proportional to the given variation of the wavelength and the optical path difference (Chen *et al* 1988, Hariharan 1989a, 1989b, Ishii 1991, Kato *et al* 1993, Onodera and Ishii 1996, Tatsuno and Tsunoda 1987). Since the reference phase depends on the optical path difference, the latter must be carefully controlled and this technique can only be applied to those cases in which it does not have a null value.

(ii) The liquid crystal cells with an induced birefringence controlled with a low electrical voltage. This process presents a high design flexibility without the hysteresis phenomenon like in the PZT systems. The main limitation is the low modulation range in the obtained phase, together with the possibility of the undesirable appearance of an intensity modulation or a non-uniform spatial relative phase shift (Hibino *et al* 1997, Kadono *et al* 1991, 1994, Konforti *et al* 1988).

Finally, other non-mechanical modulators that can be used to produce phase shifts are pressure chambers (Schwider 1990), electro-optical modulators (Georges and Lemaire 1995) and acousto-optical modulators (Massie and Nelson 1978, Colucci and Wizinowich 1992).

4.2.2. Spatial analogue modulation. These techniques can be further divided into (i) continuous ones (Macy 1983, Williams *et al* 1991), in which the phase shift between intensity values is obtained by introducing a spatial carrier into just one fringe pattern (this is commonly used in the SCPSM and SFM); and (ii) discrete ones (which are used in the SPSM), whereby n phase-shifted intensity patterns are simultaneously produced and are separated in space using a diffraction grating (Kwon 1984) or polarizing components (Smythe and Moore 1984).

4.2.3. Digital modulation. In this case, the intensity values used in the PSAs, which arise from a single pattern, are obtained by carrying out the modulation process either (i) electronically (this is used in SSD) (Womack 1984), or (ii) by computer (this is employed in LM) (Asundi 1993, Asundi and Yung 1991a, 1991b).

4.3. Associated errors

Various error sources affect measurements performed by PEMs using PSAs. Some of them are of a general type, due to the very nature of these PSAs and, therefore, common to all evaluation methods that use them. Others are of specific types, which originate from the manner in which the PSAs are used in the various evaluation methods. These error sources modify the accuracy and reproducibility of the measurements, so it is important to identify and quantify their influence on the results, in order to minimize their effect on the phase ϕ.

There are many published scientific works on the errors associated with PSAs, carried out using various strategies. A first approach to the study of the influence of the various error sources consisted of a computer simulation (Asundi and Chan 1994, Creath 1988, 1992, Creath and Schmit 1996, Groot and Deck 1996, Joenathan 1994, Nicolaus 1993, Pirga and Kujawinska 1995, Schmit and Creath 1995, Wizinowich 1990). These numerical calculations are done by comparing the output phase maps from two sets of intensity values obtained with the same PSA. One of the sets acts as a reference, whereas a known perturbation has been introduced into the other. This simple quantitative method provides numerical information on the effect produced by a determined error source in a certain PSA, even permitting the selection of an appropriate PSA from amongst the analysed group of PSAs.

Alternatively, one can use methods which provide analytical results. The phase error for a generic PSA has been calculated using a linear approximation in the Taylor expansion of the phase (Brophy 1990, Dorrío *et al* 1996, Groot 1995a, Hibino *et al* 1997, Kinnstaetter *et al* 1988, Wingerden *et al* 1991); the Fourier representation of PSAs has been used (Freischlad and Koliopoulos 1990, Larkin and Oreb 1992, Malacara *et al* 1998, Malacara-Hernández and Malacara-Doblado 1997, Ohyama *et al* 1988, Onodera and Ishii 1996) to identify and determine the errors in phase ϕ calculation; the PSAs have been analysed using a data-sampling window associated with the corresponding sampling amplitudes (Groot 1995c, Schmit and Creath 1996); the characteristic polynomial associated with the PSAs has been studied (Surrel 1996, 1997b); and the statistical properties of the PSAs have also been analysed (Ho and Kahn 1995, Rathjen 1995, Zhao 1997b).

The main errors (table 4) affecting the PSAs and their characteristics are detailed below. Although the nature of the errors is shared by all of the evaluation methods that use PSAs, their effect on the phase ϕ can differ from one case to another. No PSA is immune to all types of errors. Usually, they either reduce or cancel out certain errors, which sometimes increases their sensitivity to other types of errors. Likewise, it is observed that some PSAs are more sensitive than others to certain types of errors, while certain other errors affect all of them equally. Therefore, the most appropriate PSA for each case depends upon the peculiarities of a specific application.

4.3.1. Random errors. These errors are also called environmental or stochastic errors. They are common to practically all of the PEMs and affect both the accuracy and the reproducibility of the measurements. Their effects can be studied and reduced by designing specific strategies (Rathjen 1995, Wingerden *et al* 1991).

Amongst the different sources of random errors, we can highlight the following.

(i) Turbulent and laminar air flows. The effects of the former can be reduced by protecting the optical paths, whereas those arising from the latter can be reduced by using an optical path perpendicular to the flow.

(ii) Thermal drifts and mechanical relaxation. In order to reduce their effect on the measurements, one must provide thermal insulation and a sufficiently long waiting time such that the system is able to stabilize. The latter can be obtained by using materials with low thermal expansion coefficients and high thermal conductivities.

(iii) Vibrations. One of the most common problems is the appearance of an additional phase term proportional to the mechanical vibration of the optical set-up components. If the vibration frequency is higher than the acquisition frequency (usually tens of hertz), a reduction in the modulation of the intensity values (Kinnstaetter *et al* 1988) is produced. If the vibration is sinusoidal and its frequency is low relative to the acquisition frequency, its effect can be reduced by using certain PSAs (Groot and Deck 1996). In this case, there appears to be an error in the phase $E(\phi)$ which has a sinusoidal dependency on 2ϕ; that is, the error frequency is twice that of the fringes (Creath 1992, Deck and Groot 1998). The general procedures used to mitigate this effect consist of using (i) rigid components with high internal damping and isolated from the external perturbations via a passive or even active suspension (Stahl 1989); (ii) common path optical techniques (Mallick 1992); (iii) PSAs with a high number of intensity values (Groot 1995a, 1995c); (iv) PSAs which average the phases obtained with quadrature patterns (Deck and Groot 1998, Schwider *et al* 1983) and (v) specific PSAs (Sotoilov and Dragostinov 1997), equation (36), or (2 + 1) pattern PSA (Colucci and Wizinowich 1992, Joenathan and Khorana 1992, Kerr *et al* 1990, Ng 1996, Wizinowich 1990), equation (41).

The three above-mentioned random errors greatly affect methods like the TPSM, SD and MAMM and are conveniently avoided in the rest of the methods that employ PSAs (SCPSM, SSD, LM, SFM and SPSM), so long as, in this case, the acquisition of intensity values can be carried out within a very short period of time. These errors usually present an average null distribution or are stationary, so their effect can be reduced firstly by averaging various independent measurements. This averaging can be done for the obtained phase ϕ or over the intensity values i_k. If each intensity value i_k is acquired sufficiently fast to freeze the noise (at least one order of magnitude faster than the noise variation), then

Table 4. Errors associated with PSAs.

Random	Systematic
Turbulent and laminar air flow	Optical system aberrations
Thermal drift and mechanical relaxation	Parasitic fringes
Vibrations	Calibration errors
Speckle	Non-sinusoidal nature
Electronic noise	Photodetection
	Quantification

the result, after evaluation, represents the phase ϕ plus an additional spatially random variation term. By averaging many of these ϕ phases, the additional random terms cancel out, leaving behind just the phase ϕ. On the other hand, when the frequency of the noise terms is high relative to the total acquisition-process frequency, it is more efficient to average the intensity values to reduce their effect on the phase ϕ (Bruning *et al* 1974, Ellingsrud and Rosvold 1992, Ovryn and Haacke 1993, Stahl 1989).

The following random errors (iv) and (v) affect all PSAs equally. Their effect can be diminished by designing the system adequately and by signal treatment. The phase error $E(\phi)$ is characterized by the sampling amplitudes (μ_k and ν_k) and the relative phase shift α_r of the PSA employed (Brophy 1990, Freischlad and Koliopoulos 1990, Hibino 1997, Surrel 1997a).

(iv) Electronic noise. This originates in the photodetection and during amplification of the detected signal. It is due, amongst other causes, to the thermal agitation of the charge carriers and to the quantum nature of the luminance energy. It can be reduced by proper averaging techniques (Ellingsrud and Rosvold 1992).

(v) Optical noise ('speckle'). The high-frequency component of the speckle noise can be reduced by diverse methods: (i) digital spatial filtering of the intensity values i_k by replacing them at each point by the average, the weighted average or the median of the measured intensities at that point and its neighbours (Reid 1986); (ii) coherent spatial filtering of the intensity values i_k; (iii) averaging among the intensity values i_k (Ellingsrud and Rosvold 1992); (iv) the use of a rotational diffuser which produces a variable speckle with time, which results are to be averaged during the integration period of the acquisition system (Chang *et al* 1985, Lowenthal and Joyeux 1971); and (v) the use of specific PSAs (Moore *et al* 1994).

4.3.2. Systematic errors. The systematic errors, also called the deterministic errors, have been studied exhaustively because of their appreciable influence on the accuracy of measurements. The effects of almost all systematic errors can be conveniently reduced or even cancelled out by selecting appropriate PSAs. The errors are of the following types.

(i) Aberrations of the optical system which provides the fringe patterns or of the phase modulator used. Their effect on the phase ϕ can be cancelled out by using an adequate calibration process (Greivenkamp and Bruning 1992, Józwicki 1992).

(ii) Parasitic fringes. These can be provoked into existence when one uses an illumination source with a high temporal coherence, which produces interference of the additional wavefronts (Bruning *et al* 1974). In certain cases, the error $E(\phi)$ that appears has a phase dependency of the same frequency, that the pattern, if the additional wavefront phase is constant over the pupil (Ai and Wyant 1988). Their effect can be minimized using a generic PSA which combines the usual phase-shifted patterns with other additional ones, wherein a phase shift of π rad is introduced into one of the wavefronts (Schwider *et al* 1983) or the wavefront under test has been blocked (Ai and Wyant 1988).

(iii) Calibration errors of the additional phase α_k. These are probably the most problematic errors from which PSAs can suffer; nevertheless, they are intimately related to their very nature. That is why their study has been a principal topic within the field of evaluation using PSAs and their effect on the phase ϕ is being widely studied and analysed. This type of error shows up when there happens to be a discrepancy between the real and nominal values of the additional phase at a given point of the pattern.

On the one hand, this discrepancy can be spatially uniform for the set of intensity values, either of a linear type or of a higher order with respect to α_k. This fact in the first case provokes a phase error $E(\phi)$ at the calculation point of frequency twice the pattern frequency (Cheng and Wyant 1985, Creath 1993, Creath and Schmit 1996, Pirga and Kujawinska 1995, Wingerden *et al* 1991). The mentioned dependency of the error on the phase ϕ shows us a simple way to compensate for it by using the arithmetic average of the set of intensity values globally shifted by $\pi/2$ rad. This is because the errors both in these two cases present the same period but are phase shifted by π rad. One can reduce this error by averaging. In this way, one is able to obtain the phase ϕ with a certain level of insensitivity to linear errors in the additional phase α_k, by using simple PSAs such as those given in equation (21). The last averaging becomes more effective if the combination of these intensity values is realized in the arctangent argument, by adding numerators and denominators of the averaged PSAs, like in the case of the Schwider–Hariharan PSA, equation (23). However, the results provided by the above can be bettered (i) by using PSAs that use a greater number of intensity values (Groot 1995c); (ii) by using weighted linear combinations of the PSAs, like in equation (22); (iii) by generalizing the averaging process in the arctangent argument, similarly to that taking place in the Schwider–Hariharan PSA (Hariharan *et al* 1987, Schwider *et al* 1983), such as for example is given by equation (24) (Schmit and Creath 1995, Zhang *et al* 1998); (iv) or by means of average PSAs with a bell-shaped data-sampling window (Schmit and Creath 1996). In these cases, the linear effects are cancelled out and the higher order effects are considerably reduced at the cost of a relatively greater

calculational complexity. Alternatively, one can use PSAs that use a low number of intensity values, for example those in equations (12), and later the phase error can be cancelled out via iterative corrections (Kinnstaetter *et al* 1988) or calculated by polynomial fitting of the obtained phase (Schwider 1989).

On the other hand, one can find non-uniform spatial variations of the additional phase (like those due to high-numerical-aperture surfaces in cases in which one employs a PZT system (Creath and Hariharan 1994, Schulz and Elssner 1991) or to the use of liquid crystal cells as modulators (Hibino *et al* 1997). The α_k variation from point to point can become high, producing an error with a complicated dependency on the phase ϕ. This error can be (i) cancelled out by using compensated relative-phase-shift PSAs, equations (34)–(36); (ii) minimized (Creath and Hariharan 1994) by using the Schwider–Hariharan PSA, equation (23), or just like for most of the other systematic error types, the error can be minimized by using PSAs with a high number of intensity values (Groot 1995b); or (iii) cancelled out using equation (33), if the nonlinearity is quadratic (Hibino *et al* 1997).

(iv) Non-sinusoidal profile patterns. As indicated earlier, most of the PSAs assume that the intensity values have a sinusoidal profile, equation (10). When these algorithms are used directly to analyse patterns which do not have such a profile, equation (2), there appears an error with a complicated dependency on the phase ϕ (Hariharan 1987). This depends on the PSA employed and the amplitudes of the various higher order harmonics a_l. Their effect can be (i) cancelled out in certain cases by using specific PSAs like the Bönsch PSA, equation (37), or (ii) minimized using generic PSAs with a great number of intensity values (Asundi and Chan 1994). In this way, for an intensity profile with various harmonic components, equation (2), the effects of order-l harmonics are reduced if symmetric PSAs of $l + 2$ intensity values shifted by $2\pi/(l+2)$ rad are used (Hariharan 1987, Hibino *et al* 1995), such as those in equations (31) and (32). When an analytical expression for the error is available, the results obtained with generic PSAs can be corrected by subtracting from the calculated phase ϕ the theoretical error values provided by the corresponding analytical expression (Dorrío *et al* 1996).

(v) Photodetection errors. The PSAs assume that the photodetection is linear; i.e., there is a linear relationship between the intensity incident on each detector element and its output signal. When the photodetection is not linear, there appears an error in the measured phase analogous to that produced by the non-sinusoidal fringe profile, as long as the nonlinearity of the photodetection gives rise to the appearance of intensity terms with harmonic dependencies on the phase, typically of second or third order (Creath 1988, 1992), in which case the phase error is proportional to a sine function of four times the fringe frequency (Wingerden *et al* 1991). In such a specific situation, the Carré PSA, equation (34) becomes totally inefficient as a means to reduce the effect of this error on the phase, whereas the Wyant PSA, equation (13), is insensitive to nonlinearity of the second order and slightly sensitive to third-order nonlinearity. On the other hand, the Schwider–Hariharan PSA is insensitive to both nonlinearities (Kinnstaetter *et al* 1988, Schmit and Creath 1995). In like manner, if the intensity values are acquired using different cameras or in different parts of a single TV camera (like in SCPSM, SFM and SPSM), variations in i_0 and V at each phase calculation cell are found to appear (Pirga and Kujawinska 1995), which introduce an error $E(\phi)$.

(vi) Quantification errors. The quantification error is induced due to the sampling of the intensity values i_k before manipulating them by computer. The latter process is usually carried out using an analogue-to-digital converter (A/D), wherein a continuous-valued signal is converted into a digital signal of discrete values (grey levels). The usual converters use 8 bits, which means that there exist $2^8 = 256$ grey levels. However, in practice, in low-visibility zones, the number of effective grey levels diminishes and the error $E(\phi)$ increases. The most effective way to reduce this error is to fit the TV camera output to the A/D converter, such that its whole dynamic range is used. However, one can also reduce the error by using PSAs with a high number of phase-shifted intensity values. In this way, if the relative phase shift α_r between the n intensity values i_k is constant, the error in phase $E(\phi)$ due to the use of Q quantification levels, expressed in the form of the standard deviation of the phase ϕ, is inversely proportional to the square root of the product of n and Q (Brophy 1990, Greivenkamp and Bruning 1992). A more general treatment (Zhao 1997a, Zhao and Surrel 1997) shows that by using eight or more bits, the error in phase ϕ is depreciable if the intensity values cover the whole quantification range. In this case, the magnitude of the error primarily depends on the presence of noise in the pattern. The standard deviation of the error in the phase is reduced due to the interaction between the quantification and the noise in inverse relationship with n (Zhao 1997b).

5. Conclusions

The field of phase-evaluation methods (PEMs) has experienced a dramatic development during the present decade. Basically, the process of phase extraction can be explained in a quite general way in the reciprocal space framework, but the combination of different techniques of modulating, sampling the intensity values and demodulating has led to a great variety of particular methods. The approximation to the real applications is indeed made harder due to the various domains in which phase modulation may take place, the different instrumental needs and possibilities associated with the implementation of each method and the wide range of types of error to be managed. Several classifications of the PEMs have already been reported in the literature, attending to different relevant attributes. In this paper a new criterion has been chosen, i.e., the use of a spatial carrier, to make an updated classification of the PEMs. This criterion stresses the fact that, when the phase modulation takes place in the same domain of variation as the magnitude that one desires to measure (i.e. the phase ϕ), then there arises a limit for the maximum phase gradient allowed, related with the separability condition in reciprocal space and to the Nyquist sampling frequency. So, whereas the PEMs with a spatial carrier are adequate for real-time acquisition

(all use just a single pattern except for MAMM), the PEMs without a spatial carrier provide a greater measurement range.

On the other hand, we have analysed the phase-stepping algorithms (PSAs) in the general context of the PEMs which make use of them. Amongst the numerous topics covered in this review, we would like to highlight the generic formulation of the PSAs, their classification, the various modulation techniques and the various types of errors which can affect the PSAs, the last point being justified by the fact that their high potential accuracy can only be attained if the degrading processes involved are conveniently taken in account.

Insofar as the errors associated with the PSAs are concerned, there is no single PSA which would be insensitive to all possible sources of systematic errors; moreover, the reduction of one of them by using a certain PSA implies the increase of sensitivity to another or the generation of other error sources. The PSAs that exhibit a better behaviour when confronted with a whole set of systematic errors are those that use a higher number of intensity values, but these in certain cases require exhaustive control of the stochastic conditions in order to avoid an excessive influence of random errors. Therefore, a compromise solution at the time of selecting a PSA is best, since the use of a low number of intensity values reduces the influence of random errors as well as the storage space and processing time, whereas using a higher number reduces the effects of the systematic errors.

Acknowledgments

This work was funded by the Xunta de Galicia (XUGA 32101B97).

References

Ai Ch and Wyant J C 1987 Effect of piezoelectric transducer non-linearity on phase shift interferometry *Appl. Opt.* **26** 1112–16

——1988 Effect of spurious reflection on phase shift interferometry *Appl. Opt.* **27** 3039–45

Asundi A 1993 Moiré methods using computer-generated gratings *Opt. Engng.* **32** 107–16

Asundi A and Chan C S 1994 Phase shifting applied to non-sinusoidal intensity – an error simulation *Opt. Lasers Engng.* **21** 3–30

Asundi A and Yung K H 1991a Logical moiré and its applications *Exp. Mech.* **1** 236–42

——1991b Phase-shifting and logical moiré *J. Opt. Soc. Am.* A **8** 1591–600

Barnes T H 1987 Heterodyne Fizeau interferometer for testing flat surfaces *Appl. Opt.* **26** 2804–9

Barrientos B, Moore A J, Pérez-López C, Wang L L and Tschudi T 1997 Transient deformation measurement with ESPI using a diffractive optical element for spatial phase-stepping *Automatic Processing of Fringe Patterns* ed W Jüptner and W Osten (Berlin: Akademie) pp 371–5

Boebel D, Packross B and Tiziani H J 1991 Phase shifting in an oblique incidence interferometer *Opt. Engng.* **30** 1910–14

Bone D J, Bachor H A and Sandeman J 1986 Fringe-pattern analysis using a 2-D Fourier transform *Appl. Opt.* **25** 1653–60

Bönsch G and Böhme H 1989 Phase-determination of Fizeau interferences by phase-shifting interferometry *Optik* **82** 161–4

Brophy C P 1990 Effect of intensity error correlation on the computed phase of phase-shifting interferometry *J. Opt. Soc. Am.* A **7** 537–41

Brown G M 1993 Fringe analysis for automotive applications *Opt. Lasers Engng.* **19** 203–20

Bruning J H, Herriot D R, Gallager J E, Rosenfeld D P, White A D and Brangaccio D J 1974 Digital wavefront measuring interferometer for testing optical surfaces and lenses *Appl. Opt.* **13** 2693–703

Carré P 1966 Installation et utilisation du comparateur photoélectrique et interferentiel du Bureau International des Poids et Mesures *Metrologia* **2** 13–23

Chang M, Hu Ch-P, Lam P and Wyant J C 1985 High precision deformation measurement by digital phase shifting holographic interferometry *Appl. Opt.* **24** 3780–3

Charette P G and Hunter I W 1996 Robust phase-unwrapping method for phase images with high noise *Appl. Opt.* **35** 3506–13

Cheng Y Y and Wyant J C 1985 Phase shifter calibration in phase-shifting interferometry *Appl. Opt.* **24** 3049–52

Chen J, Ishii Y and Murata K 1988 Heterodyne interferometry with a frequency-modulated laser diode *Appl. Opt.* **27** 124–8

Colonna de Lega X and Jacquot P 1996 Deformation measurement with object-induced dynamic phase-shifting *Appl. Opt.* **35** 5115–21

Colucci D and Wizinowich P L 1992 Millisecond phase acquisition at video rates *Appl. Opt.* **31** 5919–25

Creath K 1988 Phase-measurement interferometry techniques *Progress in Optics* vol 26, ed E Wolf (Amsterdam: Elsevier) pp 349–93

——1986 Comparison of phase-measurement algorithms *Proc. SPIE* **680** 19–28

——1992 Error sources in phase-measuring interferometry *Proc. SPIE* **1720** 428–35

——1993 Temporal phase measurement methods *Interferogram Analysis* ed D W Robinson and G T Reid (Bristol: Institute of Physics) pp 94–140

Creath K, Cheng Y-Y and Wyant J C 1985 Contouring aspheric surfaces using two-wavelength phase-shifting interferometry *Opt. Acta* **32** 1455–64

Creath K and Hariharan P 1994 Phase-shifting error in interferometric tests with high-numerical-aperture reference surfaces *Appl. Opt.* **33** 24–5

Creath K and Schmit J 1996 N-point spatial phase-measurement techniques for non-destructive testing *Opt. Lasers Engng.* **24** 365–79

Dändliker R and Thalmann R 1985 Heterodyne and quasi-heterodyne holographic interferometry *Opt. Engng.* **24** 824–31

Davies R N and Bryanston-Cross P J 1996 Application of automatic phase stepping to the measurement of heat transfer *Simulation and Experiment in Laser Metrology* ed Z Füzessy *et al* (Berlin: Akademic) pp 295–9

Davies D E N and Kingsley S 1974 Method of phase modulating in optical fibers: application to optical-telemetry systems *Electron. Lett.* **10** 21–2

Deck L L and Groot P J 1998 Punctuated quadrature phase-shifting interferometry *Opt. Lett.* **23** 19–21

Dorrío B V, Blanco-García J, López C, Doval A F, Soto R, Fernández J L, Pérez-Amor M 1996 Phase error calculation in a Fizeau interferometer by Fourier expansion of the intensity profile *Appl. Opt.* **35** 61–4

Dorrío B V, Doval A F, López C, Soto R, Blanco-García J, Fernández J L and Pérez-Amor M 1995 Fizeau phase-measuring interferometry using the moiré effect *Appl. Opt.* **34** 3639–43

Dorrío B V, López C, Alén J M, Bugarín J, Fernández J L, Doval A F, Blanco-García J, Pérez-Amor M and Fernández J L 1998 Multiplicative moiré two-beam phase-stepping and Fourier transform methods for the evaluation of multiple-beam Fizeau patterns: a comparison *Appl. Opt.* **37** 1945–52

Dorrío B V, López C, Doval A F, Alén J M, Bugarín J, Fernández A, Blanco-García J, Fernández J L and Pérez-Amor M 1997 Measurement range analysis in moiré evaluation Fizeau interferometry *Appl. Opt.* **36** 3635–44

Ellingsrud S and Rosvold G O 1992 Analysis of data-based TV-holography system used to measure small vibration amplitudes *J. Opt. Soc. Am.* A **9** 237–51
Ettl P and Creath K 1996 Comparison of phase-unwrapping algorithms using gradient first failure *Appl. Opt.* **35** 5108–14
Facchini M, Alberecht D and Zanetta P 1993 Phase detection algorithm in ESPI fringes with speckle noise reduction by using preliminary time averaging *Proc. FASIC Fringe '93* ed W Jüptner and W Osten (Bermen: Akademie) pp 45–50
Facchini M and Zanetta P 1995 Derivatives of displacement obtained by direct manipulation of phase-shifted interferograms *Appl. Opt.* **34** 7202–6
Farrell C T and Player M A 1992 Phase step measurement and variable step algorithms in phase-shifting interferometry *Meas. Sci. Technol.* **3** 953–8
Frantz L M, Sawchuk A A and Ohe W 1979 Optical phase measurement in real time *Appl. Opt.* **18** 3301–6
Freischlad K and Koliopoulos Ch L 1990 Fourier description of digital phase-measuring interferometry *J. Opt. Soc. Am.* A **7** 542–51
Gao Z, Zhou S and Hu Y 1997 High-speed fringe analysis by using stair-shaped virtual grating demodulation *Opt. Lasers Engng.* **28** 411–22
Georges M P and Lemaire Ph C 1995 Phase-shifting holographic interferometry that uses bismuth silicon oxide crystals *Appl. Opt.* **34** 7497–506
Greivenkamp J E 1984 Generalized data reduction for heterodyne interferometry *Opt. Engng.* **23** 350–2
——1987 Sub-Nyquist interferometry *Appl. Opt.* **26** 5245–58
Greivenkamp J E and Bruning J H 1992 Phase shifting interferometry *Optical Shop Testing* ed D Malacara (New York: Wiley) pp 501–98
Groot P 1995a Vibration in phase-shifting interferometry *J. Opt. Soc. Am.* A **12** 354–65
——1995b Phase-shift calibration errors in interferometers with spherical Fizeau cavities *Appl. Opt.* **34** 2856–63
——1995c Derivation of algorithms for phase-shifting interferometry using the concept of a data-sampling window *Appl. Opt.* **34** 4723–30
Groot P and Deck L 1996 Numerical simulations of vibration in phase-shifting interferometry *Appl. Opt.* **35** 2172–8
Hariharan P 1987 Digital phase-stepping interferometry: effects of multiple reflected beams *Appl. Opt.* **26** 2506–7
——1989a Phase stepping interferometry with laser diodes: effect of changes in laser power with output wavelength *Appl. Opt.* **28** 27–9
——1989b Phase stepping interferometry with laser diodes 2: effects of laser wavelength modulation *Appl. Opt.* **28** 1749–50
Hariharan P, Oreb B F and Eiju T 1987 Digital phase-stepping interferometry: a simple error-compensating phase calculation algorithm *Appl. Opt.* **26** 2504–5
Hibino J M 1997 Susceptibility of systematic error-compensating algorithms to random noise in phase-shifting interferometry *Appl. Opt.* **36** 2084–93
Hibino K, Oreb B F, Farrant D I and Larkin K G 1995 Phase shifting for nonsinusoidal waveforms with phase-shift errors *J. Opt. Soc. Am.* A **12** 761–8
——1997 Phase-shifting algorithms for nonlinear and spatially nonuniform phase shifts *J. Opt. Soc. Am.* A **14** 918–30
Ho K-P and Kahn JM 1995 Exact probability-density function for phase-measurement interferometry *J. Opt. Soc. Am.* **12** 1984–9
Ichioka Y and Inuiya M 1972 Direct phase detecting system *Appl. Opt.* **11** 1507–14
Ishii Y 1991 Recent developments in laser-diode interferometry *Opt. Lasers Engng.* **14** 293–309
Ishii Y and Onodera R 1995 Laser-diode phase-shifting interferometer insensitive to changes in laser power *Proc. SPIE* **2544** 173–6
Jin G, Bao N and Chung P S 1994 Applications of a novel phase-shift method using a computer-controlled polarization mechanism *Opt. Engng.* **33** 2733–7
Joenathan C 1994 Phase-measuring interferometry: new methods and error analysis *Appl. Opt.* **33** 4147–55
Joenathan C and Khorana B M 1992 Phase measurement by differentiating interferometric fringes *J. Mod. Opt.* **39** 2075–87
Johnson G W, Leiner D C and Moore D T 1979 Phase-locked interferometry *Opt. Engng.* **18** 46–52
Józwicki R, Kujawinska M and Salbut M 1992 New *contra* old wavefront measurement concepts for interferometric optical testing *Opt. Engng.* **31** 422-33
Judge T R and Bryanston-Cross P J 1994 A review of phase unwrapping techniques in fringe analysis *Opt. Lasers Engng.* **21** 199–239
Kadono H, Ogusu M and Toyooka S 1994 Phase shifting common path interferometer using a liquid-crystal phase modulator *Opt. Commun.* **110** 391–400
Kadono H, Takai N and Asakura T 1987 New common-path phase shifting interferometer using a polarization technique *Appl. Opt.* **26** 898–904
Kadono H, Toyooka S and Iwasaki Y 1991 Speckle-shearing interferometry using a liquid-crystal cell as a phase modulator *J. Opt. Soc. Am.* A **8** 2001–8
Kato J-I, Yamaguchi I and Ping Q 1993 Automatic deformation analysis by a TV speckle interferometer using a laser diode *Appl. Opt.* **32** 77–88
Kaufmann G H and Galizzi G E 1997 Evaluation of a method to determine interferometric phase derivatives *Opt. Lasers Engng.* **27** 451–65
Kaufmann G H and Jacquot P 1990 Phase shifting of whole field speckle photography fringes *Appl. Opt.* **29** 3570–2
Kerr D, Mendoza Santoyo F and Tyrer J R 1990 Extraction of phase data from electronic speckle pattern interferometric fringes using a single-phase-step method: a novel approach *J. Opt. Soc. Am.* **7** 820–6
Kinnstaetter K, Lohmann A W, Schwider J and Streibl N 1988 Accuracy of phase shifting interferometry *Appl. Opt.* **27** 5082–9
Konforti N, Marom E and Wu S T 1988 Phase-only modulation with twisted pneumatic liquid-crystal spatial light modulators *Opt. Lett.* **3** 251–3
Kong I B and Kim S W 1995 Portable inspection of precision surfaces by phase-shifting interferometry with automatic suppression of phase-shift errors *Opt. Engng.* **34** 1400–4
Kothiyal M P and Delisle C 1985 Shearing interferometer for phase-shifting interferometry with polarization phase shifter *Appl. Opt.* **24** 4439–42
Kreis T 1986 Digital holographic interference-phase measurement using Fourier-transform method *J. Opt. Soc. Am.* A **3** 847–55
——1987 Quantitative evaluation of interference patterns *Proc. SPIE* **863** 68–77
——1993 Computer aided evaluation of fringe patterns *Opt. Lasers Engng.* **19** 221–40
——1996 Quantitative evaluation of the interference phase *Holographic Interferometry* (Berlin: Akademic) pp 55–64
Kreis T, Geldmacher J and Jüptner W 1993 A comparison of interference phase determination methods with respect to the achievable accuracy *Proc. FASIC Fringe '93* ed W Jüptner and W Osten (Bremen: Akademie) pp 51–9
Kujawinska M 1993a Spatial phase measurement methods *Interferogram Analysis* ed D W Robinson and G T Reid (Bristol: Institute of Physics) pp 141–93
——1993b The architecture of a multipurpose fringe pattern analysis system *Opt. Lasers Engng.* **19** 261–3
Kujawinska M and Wójciak J 1991a Spatial phase-shifting techniques of fringe pattern analysis in photomechanics *Proc. SPIE* **1554B** 503–13
——1991b Spatial-carrier phase shifting technique of fringe pattern analysis *Proc. SPIE* **1508** 61–7
——1991c High accuracy Fourier transform fringe pattern analysis *Opt. Lasers Engng.* **14** 325-39

Kwon O Y 1984 Multichannel phase-shifted interferometer *Opt. Lett.* **9** 59–61
Kwon O Y, Shough D M and Willians R A 1987 Stroboscopic phase-shifting interferometry *Opt. Lett.* **12** 855–7
Lai G and Yatagai T 1991 Generalized phase-shifting interferometry *J. Opt. Soc. Am.* A **8** 822–7
Larkin K G 1996 Efficient nonlinear algorithm for envelope detection in white light interferometry *J. Opt. Soc. Am.* A **13** 832–4
Larkin K G and Oreb B F 1992 Design and assessment of symmetrical phase-shifting algorithms *J. Opt. Soc. Am.* A **9** 1740–8
Lesne J L, Le Brun A, Roger D and Dieulesaint E 1987 High bandwidth laser heterodyne interferometer to measure transient mechanical displacements *Proc. SPIE* **863** 13–22
Liu J B and Ronney P D 1997 Modified Fourier transform method for interferogram fringe pattern analysis *Appl. Opt.* **36** 6231–41
Lowenthal S and Joyeux D 1971 Speckle removal by a slowly moving diffuser associated with a motionless diffuser *J. Opt. Soc. Am.* **61** 847–51
Macy W W 1983 Two-dimensional fringe-pattern analysis *Appl. Opt.* **22** 3898–901
Malacara D 1990 Updated optical testing bibliography *Appl. Opt.* **29** 1384–7
Malacara D, Servín M and Malacara Z 1998 *Interferogram Analysis for Optical Testing* (New York: Dekker)
Malacara D, Servín M, Morales A and Malacara Z 1995 Aspherical wavefront testing with several defocusing steps *Proc. SPIE* **2576** 190–2
Malacara-Doblado D, Servín M, Malacara-Hernández D 1997 Graphical vector description of sampling weights in phase-detecting algorithms *Opt. Engng.* **36** 2086–91
Malacara-Hernández D and Malacara-Doblado D 1997 Error analysis of phase detection algorithms *Automatic Processing of Fringe Patterns* ed W Jüptner and W Osten (Berlin: Akademie) pp 45–51
Mallick S 1992 Common-path interferometers *Optical Shop Testing* ed D Malacara (New York: Wiley) pp 95–122
Martini G 1987 Analysis of a single-mode optical fibre piezoceramic phase modulator *Opt. Quant. Electron.* **19** 179–90
Massie N A 1980 Real-time digital heterodyne interferometry: a system *Appl. Opt.* **19** 154–60
Massie N A and Nelson R D 1978 Beam quality of acousto-optic phase shifters *Opt. Lett.* **3** 46-7
Massie N A, Nelson R D and Holly S 1979 High-performance real-time heterodyne interferometry *Appl. Opt.* **18** 1797–803
Matthews H J, Hamilton D K and Sheppard J R 1986 Surface profiling by phase-locked interferometry *Appl. Opt.* **25** 2372–4
Mercer C R and Beheim G 1991 Fiber optic phase stepping system for interferometry *Appl. Opt.* **30** 729–34
Mertz L 1983 Real-time fringe-pattern analysis *Appl. Opt.* **22** 1535–9
Morgan C J 1982 Least-squares estimation in phase-measurement interferometry *Opt. Lett.* **7** 368–70
Moore A J, Tyrer J R and Mendoza-Santoyo F 1994 Phase extraction from electronic speckle pattern interferometry addition fringes *Appl. Opt.* **33** 7312–20
Ng T G 1996 The one-step phase-shifting technique for wave-front interferometry *J. Mod. Opt.* **43** 2129–38
Nicolaus R A 1993 Precise method of determining systematic errors in phase-shifting interferometry on Fizeau interferences *Appl. Opt.* **32** 6380–6
Nugent K A 1985 Interferogram analysis using an accurate fully automatic algorithm *Appl. Opt.* **24** 3101–5
Ohyama N, Kinoshita S, Cornejo-Rodríguez A, Honda T and Tsujiuchi J 1988 Accuracy of phase determination with unequal reference phase shift *J. Opt. Soc. Am.* A **5** 2019–25
Onodera R and Ishii Y 1996 Phase-extraction analysis of laser-diode phase-shifting interferometry that is insensitive to changes in laser power *J. Opt. Soc. Am.* **13** 139–46
Ovryn B and Haacke E M 1993 Temporal averaging of phase measurements in the presence of spurious phase drift: application to phase-stepped real-time holographic interferometry *Appl. Opt.* **32** 1087–94
Owner-Petersen M 1991 Digital speckle pattern shearing interferometry: limitations and prospects *Appl. Opt.* **30** 2730–8
Perry K E and McKelvie J 1993 A comparison of phase shifting and Fourier methods in the analysis of discontinuous fringe patterns *Opt. Lasers Engng.* **19** 269–84
Pirga M and Kujawinska M 1995 Errors in two-directional spatial-carrier phase shifting method *Proc. SPIE* **2544** 112–21
Pirga M and Kujawinska M 1995 Two directional spatial-carrier phase-shifting method for analysis of crossed and closed fringe patterns *Opt. Engng.* **34** 2459–66
Poon C Y, Kujawinska M and Ruiz C 1993 Automated fringe pattern analysis for moiré interferometry *Exp. Mech.* **33** 234–41
Pryputniewicz R J 1992 Electro-optic holography *Proc. SPIE* **CR46** 148–74
Ransom P L and Kokal J V 1986 Interferometric analysis by a modified sinusoid fitting technique *Appl. Opt.* **25** 4199–204
Rathjen C 1995 Statistical properties of phase-shift algorithms *J. Opt. Soc. Am.* A **12** 1997–2008
Reid G T 1986 Automatic fringe pattern analysis: a review *Opt. Lasers Engng.* **7** 37–68
Robinson D W 1993 Phase unwrapping methods *Interferogram Analysis* ed D W Robinson and G T Reid (Bristol: Institute of Physics) pp 194–229
Rogala E W and Barrett H H 1997 Phase-shifting interferometry and maximum-likelihood estimation theory *Appl. Opt.* **36** 8871–6
Rosenbluth A E and Bobroff N 1990 Optical sources of non-linearity in heterodyne interferometers *Prec. Eng.* **12** 7-11
Salbut L and Patorsky K 1990 Polarization phase shifting method for moiré interferometry and flatness testing *Appl. Opt.* **29** 1471–3
Sasaki O, Okamura T and Nakamura T 1990b Sinusoidal phase modulating Fizeau interferometry *Appl. Opt.* **29** 512–15
Sasaki O and Okazaki H 1986a Sinusoidal phase modulating interferometry for surface profile measurement *Appl. Opt.* **25** 3137–40
——1986b Analysis of measurement accuracy in sinusoidal phase modulating interferometry *Appl. Opt.* **25** 3152–8
Sasaki O, Okazaki H and Sakai M 1987 Sinusoidal phase modulating interferometer using the integrating-bucket method *Appl. Opt.* **26** 1089–93
Sasaki O, Takahashi K and Suzuki T 1990a Sinusoidal phase modulating laser interferometer with a feedback control system to eliminate external disturbance *Opt. Engng.* **29** 1511–15
Schmit J and Creath K 1995 Extended averaging technique for derivation of error-compensating algorithms in phase-shifting interferometry *Appl. Opt.* **34** 3610–19
——1996 Window function influence on phase error in phase-shifting algorithms *Appl. Opt.* **35** 5642–9
Schulz G and Elssner K-E 1991 Errors in phase-measurement interferometry with numerical apertures *Appl. Opt.* **30** 4500–6
Schwider J 1989 Phase shifting interferometry: reference phase error reduction *Appl. Opt.* **28** 3889–92
——1990 Advanced evaluation techniques in interferometry *Progress in Optics* vol 28, ed E Wolf (Amsterdam: Elsevier) pp 271-359
Schwider J, Burow R, Elssner K-E, Grzanna J and Spolaczyk R 1986 Semiconductor wafer and technical flat planeness testing interferometer *Appl. Opt.* **25** 1117–21
——1987 Semiconductor wafer and technical flat planeness testing interferometer *Measurement* **5** 98–101

Schwider J, Burow R, Elssner K-E, Grzanna J, Spolaczyk R and Merkel K 1983 Digital wave-front measuring interferometry: some systematic error sources *Appl. Opt.* **22** 3421–32
Schwider J, Falkenstörfer O, Schreiber H, Zöller A and Sreibl N 1993 New compensating four-phase algorithm for phase-shifting interferometry *Opt. Engng.* **32** 1883–5
Servín M, Malacara D, Malacara Z and Vlad V 1994 Sub-Nyquist null aspheric testing using a computer stored compensator *Appl. Opt.* **33** 4103–8
Singh H and Sirkis J S 1994 Direct extraction of phase gradients from Fourier-transform and phase-step fringe patterns *Appl. Opt.* **33** 5016–20
Sirkis J S, Chen Y-M, Singh H and Cheng A Y 1992 Computerized optical fringe pattern analysis in photomechanics: a review *Opt. Engng.* **31** 304–14
Smythe R and Moore R 1984 Instantaneous phase measuring interferometers *Opt. Engng.* **23** 361–4
Sotoilov G and Dragostinov T 1997 Phase-stepping interferometry: five-frame algorithm with an arbitrary step *Opt. Lasers Engng.* **28** 61–9
Stahl H P 1989 Testing large optics: high-speed phase-measuring interferometry *Photonics Spectra.* **12** 105–12
Stetson K A 1990 Theory and applications of electronic holography *Proc. SEM Conf. Hologram Interferometry and Speckle Metrology* pp 294–300
Surrel Y 1993 Phase-stepping: a new self-calibrating algorithm *Appl. Opt.* **32** 3598–600
——1996 Design of algorithms for phase measurements by the use of phase stepping *Appl. Opt.* **35** 51–60
——1997a Additive noise effect in digital phase detection *Appl. Opt.* **36** 271–6
——1997b Design of phase-detection algorithms insensitive to bias modulation *Appl. Opt.* **36** 805–7
Suzuki T, Hasegawa J, Sasaki O and Maruyama T 1995 Sinusoidal phase-modulating laser diode interferometer using a self-pumped phase-conjugate mirror *Proc. SPIE* **2576** 299–307
Suzuki T, Sasaki O and Maruyama T 1989 Phase locked laser diode interferometry for surface profile measurement *Appl. Opt.* **28** 4407–10
Takeda M 1990 Spatial-carrier fringe-pattern analysis and its applications to precision interferometry and profilometry: an overview *Indust. Metrol.* **1** 79–99
——1996 Recent progress in phase unwrapping techniques *Proc. SPIE* **2782** 334–43
Takeda M, Ina H and Kobayashi S 1982 Fourier-transform method of fringe-pattern analysis computer-based topography and interferometry *J. Opt. Soc. Am.* **72** 156–60
Takeda M and Kitoh M 1992 Spatiotemporal frequency multiplex heterodyne interferometry *J. Opt. Soc. Am.* A **9** 1607–14
Takeda M and Mutoh K 1983 Fourier transform profilometry for the automatic measurement of 3-D object shapes *Appl. Opt.* **22** 3977–82
Tatsuno K and Tsunoda Y 1987 Diode laser direct modulation heterodyne interferometer *Appl. Opt.* **26** 37–40
Toyooka S, Tanahashi T and Tominaga M 1984 Simple heterodyne interferometry using a holographic common-path interferometer *Appl. Opt.* **23** 1460–3
Vikhagen E 1990 TV holography: spatial resolution and signal resolution in deformation analysis *Appl. Opt.* **29** 137–44
Vikram C S, Witherow W K and Trolinger J D 1993 Algorithm for phase-difference measurement in phase-shifting interferometry *Appl. Opt.* **32** 6250–2
Wang J and Grant I 1995 Electronic speckle interferometry, phase-mapping, and nondestructive testing techniques applied to real-time, thermal loading *Appl. Opt.* **34** 3620–7
Wang X, Sasaki O, Takebayashi Y, Suzuki T and Maruyama T 1994 Sinusoidal phase-modulating Fizeau interferometer using a self-pumped phase conjugator for surface profile measurements *Opt. Engng.* **33** 2670–4
Williams D C, Nassar N S, Banyard J E and Virdee M S 1991 Digital phase-step interferometry: a simplified approach *Opt. Lasers Technol.* **23** 147–50
Wingerden J, Frankena H J and Smorenburg C 1991 Linear approximation for measurement error in phase shifting interferometry *Appl. Opt.* **30** 2718–29
Wizinowich P L 1990 Phase shifting interferometry in the presence of vibration: a new algorithm and system *Appl. Opt.* **29** 3271–9
Womack K H 1984 Interferometric phase measurement using spatial synchronous detection *Opt. Engng.* **23** 391–5
Wyant J C 1971 Testing aspherics using two-wavelength holography *Appl. Opt.* **10** 2113–18
——1975 Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems *Appl. Opt.* **14** 2622–6
——1982 Interferometric optical metrology: basic principles and new systems *Laser Focus* **5** 65–71
Wyant J C, Koliopoulos C L, Bhushan B and George O E 1984 An optical profilometer for surface characterization of magnetic media *Trans. ASLE* **27** 101–13
Zhao B 1997a A statistical method for fringe intensity-correlated error in phase-shifting measurement: the effect of quantization error on the *N*-bucket algorithm *Meas. Sci. Technol.* **8** 147–53
——1997b Effect of intensity-correlated error due to quantization and noise on phase-shifting method *Opt. Lasers Engng.* **28** 199–211
Zhao B and Surrel Y 1995 Phase shifting: six-sample self-calibrating algorithm insensitive to the second harmonic in the fringe signal *Opt. Engng.* **34** 2821–2
——1997 Effect of quantization error on the computed phase of phase-shifting measurements *Appl. Opt.* **36** 2070–5
Zhang H, Lalor M J and Burton D R 1998 A comparative study of error-compensating phase-shifting algorithms *Applied Optics and Optoelectronics 1998* ed K T V Grattan (Bristol: Institute of Physics) pp 33–8